Diplomarbeit aus Technischer Mathematik, TU-Graz

Modellierung und Vergleich Spirometrischer Parameter von Nichtrauchern und Rauchern

Bischof Herbert
September 1997

Vorgelegt der Technisch-Naturwissenschaftlichen Fakultät
an der Technischen Universität Graz

Begutachter und Betreuer: Prof. Dr. Ernst Stadlober

Institut für Statistik der Technischen Universität Graz

Ich versichere, diese Arbeit selbständig verfaßt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfsmittel bedient zu haben.

Ich möchte mich bei Herrn Prof. Dr. Ulrich Dieter für die großzügige Bereitstellung der Institutseinrichtungen, die für die Durchführung dieser Arbeit notwendig waren, bedanken.

Besonders möchte ich mich bei Herrn Prof. Dr. Ernst Stadlober für die intensive fachliche Betreuung meiner Diplomarbeit bedanken.

Danken möchte ich auch Herrn Dr. Herwig Friedl für die Unterstützung bezüglich Statistik-Software und sonstiger zahlreicher nützlicher Tips im Fachgebiet und im Umgang mit dem Textsystem.

Besonderer Dank gilt jedoch meinen Eltern für ihre jahrelange Unterstützung meiner Studientätigkeit.

ZUSAMMENFASSUNG

Die zu analysierenden Spirometriedaten wurden in den Jahren 1989 bis 1993 in der Steiermark erhoben. Dabei wurden an den insgesamt 54.123 Probanden folgende Lungenfunktionsparameter (LFP) gemessen: **FVC**, $\mathbf{FEV}_1$, **PEF**, $\mathbf{MEF}_{75}$, $\mathbf{MEF}_{50}$, $\mathbf{MEF}_{25}$. In dieser Arbeit werden nur die Daten jener Probanden verwendet, die zum Zeitpunkt der Untersuchung zumindest 19 Jahre alt waren (16.389 Frauen und 20.679 Männer).

Durch die Erklärung der medizinischen Hintergründe der Spirometrie (Kapitel 1) werden die Zusammenhänge zwischen den LFP eines Probanden und den daraus abgeleiteten Krankheitsbildern dargelegt. Um die gemessenen LFP eines Individuums im Vergleich zu einer 'Normalpopulation' richtig einschätzen zu können, sind diese mit Normwerten zu vergleichen, welche aus Referenzgleichungen von Nichtrauchern ermittelt werden.

Besonders wichtig ist in diesem Zusammenhang, daß die Referenzgleichungen anhand eines möglichst homogenen Datenmaterials erstellt werden. Die Selektion der Daten (Kapitel 2) erfolgt demnach nach Kriterien, welche Größe, Gewicht, Lungenkrankheiten und mögliche Ausreißer berücksichtigen. Die auf diese Weise selektierten Daten (7.182 Frauen und 10.383 Männer) werden aufgrund der eigenen Angaben der Probanden in eine Nichtraucher, sowie in fünf Rauchergruppen unterteilt.

Alle Analysen werden jeweils für alle sechs Frauen- und Männergruppen erstellt.

Die Homogenität bzgl. Alter, Größe und Gewicht wird durch Boxplots, Histogramme sowie Mittelwert- und Standardabweichungsbetrachtungen beurteilt (Kapitel 4). Im Anschluß daran ergibt eine Kovarianzanalyse erste Hinweise auf Unterschiede zwischen den Niemals- und den Rauchergruppen bzgl. der LFP.

Für alle 12 Gruppen werden die LFP in Abhängigkeit von Größen- und Alterstermen unter Anwendung der multiplen linearen Regression dargestellt (Kapitel 6-10). Durch diese Modelle erhält man Informationen über die Art der geschlechtsspezifischen Unterschiede in den Parametern. Die Beurteilung der Güte der Modelle erfolgt mittels Medianplots und ausführlicher Analyse der stand. Residuen.

Letztendlich werden die LFP der Nichtraucher und der schweren Raucher miteinander verglichen (Kapitel 12). Der negative Einfluß des Rauchens ist besonders an den Modellkurven der Parameter FEV_1 und MEF_{50} zu erkennen. Auch sind die nach Größe und Gewicht adjustierten Mittelwerte der schweren Raucher(innen) durchwegs niedriger als jene der Niemalsraucher(innen). Dieser Unterschied zwischen Nichtrauchern und schweren Rauchern bei Männern (Frauen) läßt sich statistisch ab der Altersklasse [19-24) ([34-39)) bei allen Parametern nachweisen.

Die statistischen Auswertungen erfolgten mit den statistischen Softwarepaketen SAS 6.11, SPSS 6.1.3 und S-PLUS 3.3.

Inhaltsverzeichnis

Kapitel 1

Medizinische Grundlagen

1.1 Die Lunge

Die Lunge hat die Aufgabe, dem Blut Sauerstoff zuzuführen und das Kohlendioxid aus dem Blut auszuscheiden. Den Sauerstoff, den wir zur Aufrechterhaltung des Lebens der einzelnen Körperteile benötigen, beziehen wir aus der Umgebungsluft. Die Luft, die wir einatmen, setzt sich wie folgt zusammen:

- 20,96% Sauerstoff(O_2)
- 78% Stickstoff(N_2)
- 0,04% Kohlendioxid(CO_2)
- 1% Edelgase

Die Ausatemluft hat eine andere Zusammensetzung:

- ca. 16% Sauerstoff(O_2)
- 78% Stickstoff(N_2)
- ca. 5% Kohlendioxid(CO_2)
- 1% Edelgase

Aus der Gegenüberstellung Einatem/Ausatemluft ist zu ersehen, daß ca. 4-5% vom eingeatmeten Sauerstoff nicht wieder ausgeatmet werden, der Gehalt an Kohlendioxid aber um diesen Anteil steigt. Innerhalb des Körpers hat ein chemischer Prozeß stattgefunden, bei dem Sauerstoff verbraucht und Kohlendioxid erzeugt wurde. Dieser Prozeß findet in der Lunge, in den Alveolen statt. Für weitere Details siehe Dräger [2].

1.1.1 Anatomischer Aufbau der Lunge

Die Lunge besteht im wesentlichen aus Luftröhre, Bronchien, Bronchiolen und Alveolen (siehe Abbildung 1.1).

Die Luftröhre

Die Luftröhre schließt sich an den Kehlkopf an. Sie ist ein etwa 12 cm langes, aus Knorpelringen bestehendes Rohr, das sich etwa in der Höhe des sechsten Brustwirbels in zwei Röhren aufteilt. Diese Röhren, die man linken und rechten Hauptbronchus nennt, gehen in die Bronchialbäume über.

Die Bronchien

Bronchien sind die Äste der Bronchialbäume.

Die Bronchiolen

Bronchiolen sind die feineren Verzweigungen in diesen Bäumen.

Die Alveolen

Die Alveolen oder Lungenbläschen sind die "Blätter" der Bäume. Bis hierher strömt die Luft von außen über die oberen und unteren Luftwege ein, es erfolgt der Gasaustausch, und die Luft wird wieder über denselben Weg nach außen abgegeben.

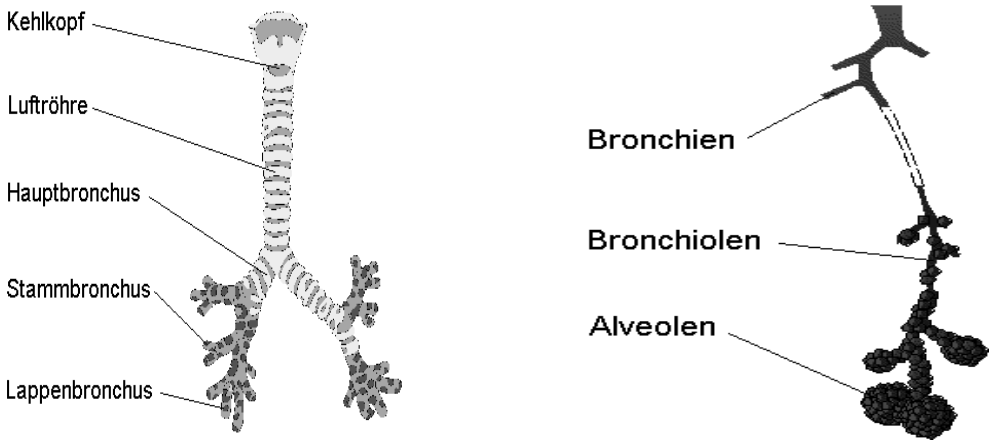


Abbildung 1.1: Anatomischer Aufbau der Lunge

Die Alveolen bestehen aus zarten feuchten elastischen Doppelwänden. Jedes einzelne ist von einem Netz feinster Blutgefäße (Kapillaren) umsponnen. Sie haben ausgeatmet einen Durchmesser von 0,1 ... 0,2 mm, eingeatmet 0,3 ... 0,5 mm. Die Anzahl der Alveolen wird auf etwa 300 Millionen geschätzt. Ihre Gesamtoberfläche, die als Gasaustauschfläche zur Verfügung steht, beträgt etwa 70 bis 100 m^2. Die Wände der Alveolen sind so dünn, daß eine Diffusion von Gasen möglich ist. In ihnen erfolgt der Gasaustausch.

1.1.2 Morphologische Entwicklung der Lunge

Die Lunge entwickelt sich nach Art einer Drüse aus der Ventralwand des Rumpfdarms. Die Entwicklung der Lunge beginnt in der 4. Embryonalwoche. Bereits am Anfang erfolgt die Aufteilung der Lunge in zwei Primäräste, welche die Grundlage für die spätere Zweiteilung der Lunge in einen rechten und linken Lungenflügel bilden. Bis zur 16. Woche werden etwa 17 Generationen von Bronchien und Bronchiolen gebildet. Nach der Geburt dauert es noch 2 Monate, bis sich die Alveolen voll entwickelt und alle Elemente eines reifen Azinus vorhanden sind. Bis zum Alter von 7 Jahren entwickeln sich neue Generationen von respiratorischen Bronchiolen und Alveolen durch Umwandlung. Nach dem 8. Lebensjahr wächst die Lunge nur noch durch Vergrößerung der vorhandenen Einheiten. Insgesamt steigert sich das Lungenvolumen von der Geburt bis zum Erwachsenenalter um das 22fache (entnommen aus Bachmann et al. [1]).

1.1.3 Erkrankungen der Lunge

In den USA waren im Jahre 1991 laut Angaben des National Jewish Medical and Research Center [12] über 100.000 Todesfälle direkt auf Erkrankungen der Lunge zurückzuführen. Damit lagen sie an 4. Stelle der häufigsten Todesursachen. Während die Raten der anderen führenden Todesursachen sinken, ist die der Lungenerkrankungen steigend. Aus diesem Grund sollte die Messung der Lungenfunktionswerte bei jeder routinemäßigen ärztlichen Untersuchung ebenso selbstverständlich sein, wie die Messung des Blutdrucks. Auch ist die Messung der wichtigsten Lungenfunktionswerte ohne großen Aufwand innnerhalb von ein paar Minuten durchzuführen.

Bei den Erkrankungen der Lunge unterscheidet man restriktive und obstruktive.

restriktive Erkrankungen

Restriktive Erkrankungen gehen meist einher mit einer Reduktion der Totalen Lungenkapazität (TLC), das ist jenes Luftvolumen, welches sich nach max. Inspiration in der Lunge befindet. Zu den restriktiven Erkrankungen zählen z.B. respiratorische Muskelschwäche, Brustkorbdeformitäten und Fibrosis.

obstruktive Erkrankungen

Eine Beeinträchtigung der Ausatmung kennzeichnet obstruktive Lungenerkrankungen. Dazu zählen allgemein bekannte Erkrankungen wie Asthma, Bronchitis und Emphysema.

1.2 Spirometrie

Die Spirometrie mißt das Volumen (wieviel) und die Flußrate (wie schnell) Luft ein- bzw. ausgeatmet wird. Mit einem sogenannten Spirometer werden verschiedene Lungenfunktionsparameter gemessen. Die Probanden werden gebeten, tief einzuatmen und möglichst schnell, durch das mit dem Spirometer verbundene Mundstück, wieder auszuatmen. Dieser Vorgang wird normalerweise dreimal wiederholt und die höchsten dabei erzielten Werte werden aufgezeichnet. Bei der Messung der Parameter ist es wichtig die Vorschriften genau einzuhalten, um die gemessenen Werte der Probanden untereinander und mit Referenzwerten vergleichen zu können. Die gemessenen Werte variieren dabei unter anderem zwischen den Geräten von verschiedenen Herstellern.

Die Lungenfunktionsparameter werden unterteilt in Volums- und Flußvolumsparameter. Die Volumsparameter geben wieder, wieviel ein Mensch max. oder in einer bestimmten Zeit ein- bzw. ausatmen kann. Die Flußvolumsparameter messen, wie schnell die eingeatmete Luft wieder ausgeatmet wird. Besonders zur Beurteilung der Flußvolumsparameter ist es wichtig nicht nur die absoluten Werte, sondern auch deren Verlauf, die Flußvolumskurve zu analysieren.

1.2.1 Volumsparameter

Zu den Volumsparamtern zählen die forcierte Vitalkapazität (FVC) und das forcierte expiratorische Volumen einer Sekunde (FEV_1). Ausgehend von einer maximalen Inspiration wird jenes Volumen gemessen das innerhalb einer Sekunde maximal ausgeatmet werden kann. Dieses Volumen entspricht dem FEV_1-Wert. Ebenfalls ausgehend von einer maximalen Inspiration wird jenes Volumen gemessen, welches bei einer forcierten Ausatmung maximal ausgeatmet werden kann. Dieses Volumen entspricht dem FVC-Wert. d.h., daß beide Parameter gleichtzeitig gemessen werden können (siehe Abbildung 1.2).

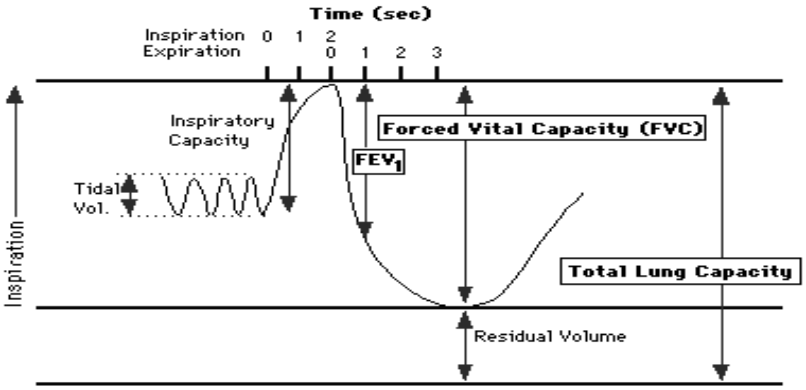


Abbildung 1.2: FVC und FEV_1 eines gesunden Probanden

Wie in der Abbildung 1.2 ersichtlich kann nicht die gesamte sich in der Lunge befindliche

Luft ausgeatmet werden. Das Residual-Volumen (RV) verbleibt in der Lunge. Bei einer Erkrankung der Lunge z.B. Emphysema kann dieses RV, aufgrund eingeschlossener Luft, erhöht sein und gleichzeitig verringert sich die FVC (siehe Abildung 1.3).

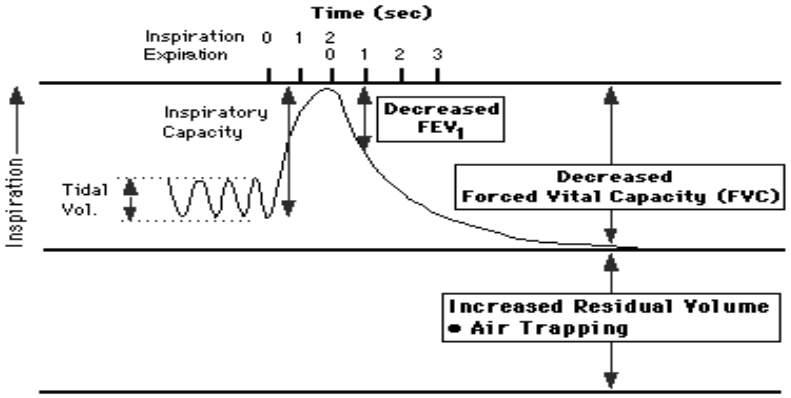


Abbildung 1.3: FVC und FEV_1 eines kranken Probanden

Ein Ziel dieser Arbeit ist es Referenzgleichungen anzugeben, die auf einem Kollektiv gesunder Probanden beruhen. Mit diesen Gleichungen ist es nun möglich über das Geschlecht, das Alter und die Größe Referenzwerte für die Volums- und Flußvolumsparameter zu berechnen. Vergleicht man nun diese vorhergesagten Werte mit den tatsächlich gemessenen, so kann bei einer wesentlichen Unterschreitung der vorhergesagten Werte auf eine Erkrankung der Lunge geschlossen werden.

Für FVC und FEV_1 liegt diese Grenze bei etwa 80% vom vorhergesagten Wert d.h. sind die gemessenen Werte eines Probanden um mehr als 20% geringer, als die vorhergesagten, so spricht man von einer Beeinträchtigung der Atmung. Eine weitere Festlegung der Grenzen kann über die Normalverteilung erfolgen. Unter der Annahme, daß die Werte der Probanden in jeder Altersklasse normalverteilt sind, können die Grenzwerte wie folgt festgelegt werden. Man subtrahiert vom vorhergesagten Mittelwert das 1,64-fache der Standardabweichung (Modellstreuung) aus der Regressionsanalyse. Mit dieser Festlegung liegen dann 95% aller Meßergebnisse der Referenzprobanden über diesem Grenzwert. Um die Referenzwerte (Grenzwerte) festzulegen wird eine wohldefinierte Menge von Niemalsrauchern verwendet. Ist die Annahme der Normalverteilung nicht gerechtfertigt, so können auch die 5%-Perzentile verwendet werden.

Eine weitere wichtige Größe ist der Quotient von FEV_1/FVC in Prozent. Wobei hier Werte über 70% als normal angesehen werden. Dieser Wert ist vor allem zur Erkennung von obstruktiven Erkrankungen wichtig. In der Tabelle 1.4 und Tabelle 1.5 ist dargelegt, welche Auswirkungen obstruktive, restriktive bzw. wenn beide Arten gleichzeitig Auftreten, auf die oben erwähnten Volumsparameter haben.

	OBSTRUCTIVE	RESTRICTIVE	MIXED
FEV1	↓	↓ or N	↓
FEVC	↓ or N	↓	↓
FEV1/FVC	↓	N or ↑	↓

*N = Normal

Abbildung 1.4: Klassifikation von Lungenerkrankungen durch Spirometrie I

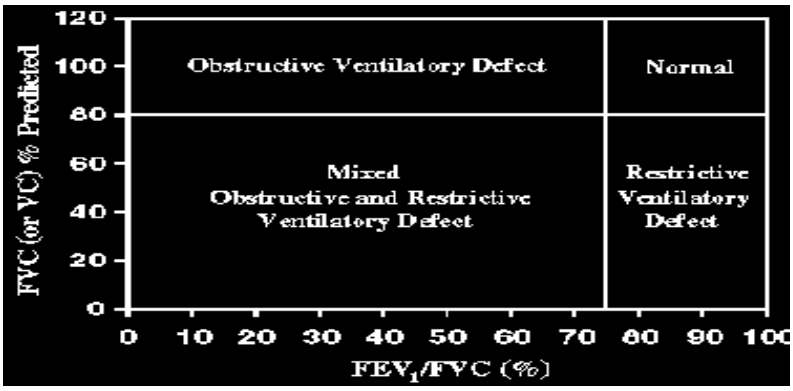


Abbildung 1.5: Klassifikation von Lungenerkrankungen durch Spirometrie II

In der Abbildung 1.6 wird noch einmal deutlich gemacht, wie obstruktive und restriktive Erkrankungen zu unterschiedlichen Meßergebnissen führen, im Vergleich zur Messung von FVC und FEV_1 bei einem gesunden Probanden.

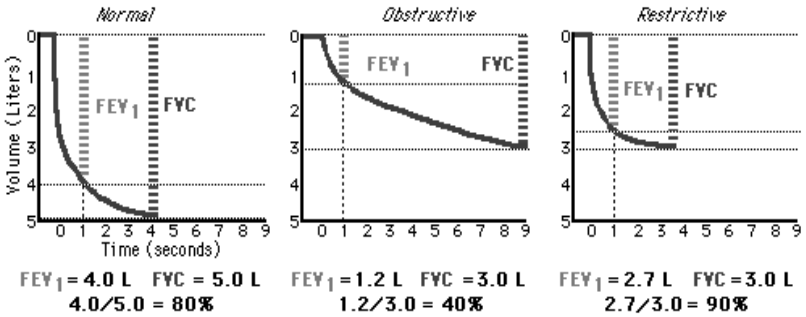


Abbildung 1.6: Klassifikation von Lungenerkrankungen durch Spirometrie III

1.2.2 Flußvolumsparameter

Die Flußvolumsparameter werden nach max. Inspiration während einer forcierten Ausatmung gemessen. Bereits kurz nach Beginn der Ausatmung wird der max. Fluß (in Liter/Sekunde) erreicht (siehe Abbildung 1.7). Diesen Wert bezeichnet man als PEF. In weiterer Folge werden drei Flußwerte gemessen, nachdem 25%, 50% und 75% der FVC ausgeatmet worden sind. Diese drei Werte werden als MEF_{75}, MEF_{50} und MEF_{25} bezeichnet.

Wie in der Abbildung 1.7 ersichtlich haben restriktive und obstruktive Erkrankungen der Lunge drastische Auswirkungen auf das Aussehen der Fluß-Volumskurve. So ist z.B besonders für Asthmatiker die tägliche Messung des PEF-Wertes ein wichtiger Indikator für den Zustand ihrer Erkrankung. Ausgehend von einem persönlichen Bestwert, gibt die Höhe des Momentanwertes Auskunft über einen möglichen zu erwartenden asthmatischen Anfall.

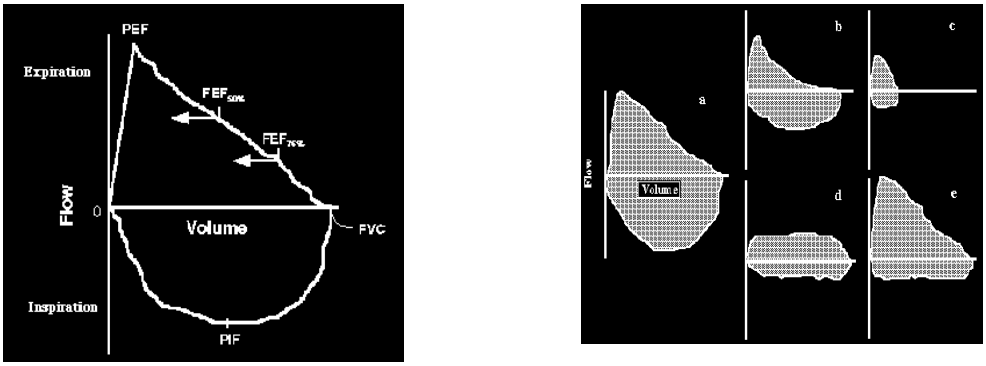


Abbildung 1.7: Flußvolumskurven

A=Atmungsvolumen
B=maximale Inspiration (max l)
C=Spitzenfluß (PEF)
D=maximale Exspiration
E=Einsekundenkapazität
V=Fluá
VK=Vitalkapazität
FVK=forcierte Vitalkapazität
RV=Residualvolumen
TLK=totale Lungenkapazität

a) normal
b) obstruktive Erkrankung
c) restriktive Erkrankung
d) dauerhafte Beeinträchtigung der Luftwege
e) variable Beeinträchtigung der oberen Luftwege

Wie in der Abbildung 1.7 ersichtlich, ist nicht nur die Höhe der vier gemessenen Flußvolumsparameter wichtig zur Diagnostik, sondern in besonderer Weise auch der Verlauf der Flußvolumskurve. Für weiter Informationen zur Spirometrie siehe [9], [13] und [14].

Kapitel 2

Datenmaterial

2.1 Datengewinnung

In den Jahren 1989 bis 1993 wurden mit einer mobilen Einheit, dem Pneumobil, in der gesamten Steiermark 54.123 Personen auf ihre Lungenfunktionswerte hin untersucht. Davon waren 24.965 weiblichen und 29.158 männlichen Geschlechts. Im Rahmen dieser Erhebungen wurden Personen im Alter von 5 - 94 Jahren untersucht. Die Probanden entstammten verschiedenen Berufsgruppen mit unterschiedlichen körperlicher Belastungen.

Zusätzlich zu den Lungenfunktionsparametern wurden auch noch andere Merkmale der Probanden erhoben:

Variable: **Alter** [Jahre], **Groesse** [Meter], **Gewicht** [kg], **Geschlecht** [w/m] sowie

Berufsgruppe: unterteilt in leichte und schwere Arbeit, stressbelastende Arbeit: je nach Eigeneinschätzung und in eine Risikogruppe: das sind Probanden, die bei ihrer Arbeit Stäube, Dämpfe oder organische Materialen inhalieren.

Wohlbefinden: die Atmung betreffend; nach eigener Einschätzung (Ja/Nein)

Raucher: hier wurde eine Unterteilung in 5 Gruppen vorgenommen in Niemalsraucher, Passivraucher, Exraucher, gelegentliche Raucher und Raucher. Bei den Rauchern wurde zusätzlich die Art (Pfeife, Zigarette, Zigarre) und die Anzahl (z.B.: eine Packung Zigaretten/Tag) erhoben.

Anamnese: die Probanden wurden gebeten anzugeben, ob sie gesund sind oder an einer im folgenden angeführten Erschwernis oder Krankheit leiden:

1. keine Krankheit
2. Husten oder Auswurf
3. Allergie oder Asthma
4. chron. Bronchitis oder Emphysem

5. Lungen- oder Rippfellentzündung
6. TBC oder Lungenresektion

2.1.1 Lungenfunktionsparameter und Meßmethode

Es wurden folgende Lungenfunktionsparameter gemessen:

FVC: **F**orced **V**ital **C**apacity [l]
(ist jenes Luftvolumen in Litern, das der Proband nach maximaler Inspiration maximal exspirieren kann)

FEV_1: **F**orced **E**xspiratory **V**olume at one second [l]
(ist jenes Luftvolumen in Litern, das der Proband nach maximaler Inspiration innerhalb einer Sekunde ausatmen kann)

FEV_1%FVC: ist der Quotient FEV_1/FVC in Prozent [%]

PEF: **P**eak **E**xspiratory **F**low [l/s]
(ist der größte, während einer forcierten Exspiration erreichbare Fluß in Litern pro Sekunde, beginnend aus der maximalen Inspirationsstellung)

MEF: **M**aximal **E**xspiratory **F**low [l/s]
(ist der maximale Fluß in Litern pro Sekunde, gemessen bei einem definierten Lungenvolumen)

MEF_{75}: Wert nach Ausatmung von 25% des FVC

MEF_{50}: Wert nach Ausatmung von 50% des FVC

MEF_{25}: Wert nach Ausatmung von 75% des FVC

Als Meßgerät wurde ein Pneumatograph der Medical Graphics Corporation, USA verwendet. Die morgendliche Inbetriebnahme erfolgte bei einer Raumtemperatur von mindestens 15 Grad. Die Kalibrierung der Geräte wurde morgens vor Arbeitsbeginn, nach ca. einer Arbeitsstunde, bzw. nach Bedarf durchgeführt.

Der Proband wurde vor der Messung über den Meßablauf informiert und zu einer aktiven Teilnahme motiviert. Vor allem die Flußvolumsparameter (FVP) PEF und MEF_{75} hängen sehr von der forcierten Exspiration der Testperson ab. Aufrecht stehend, ohne beengende Kleidung und ohne Nasenklemme, wurde die zu untersuchende Person aufgefordert, tief einzuatmen und anschließend kräftig und möglichst lange in den Schlauch zu blasen. Nach mindestens zwei Versuchen wurde ein zufriedenstellender Kurvenverlauf direkt am PC abgespeichert. Der Meßvorgang wurde von einem Arzt und einer medizinisch-technischen Assistentin durchgeführt und kontrolliert.

2.2 Datenselektion

Aufgrund des Zieles dieser Auswertungen, eventuell vorhandene Unterschiede zwischen Raucher- und Nichtrauchergruppen aufzuzeigen, wurden in diese Studie nur Probanden ab einem Alter von 19 Jahren aufgenommen.

Im folgenden wird die Selektion und Aufteilung der Daten bei Frauen und Männern in jeweils sechs Untergruppen durchgeführt. Die Aufteilung erfolgt nach den eigenen Angaben der Probanden. Anhand dieser Angaben werden folgende Untergruppen gebildet:

Niemalsraucher: nicht mehr als 100 Zigaretten während des gesamten Lebens.

Passivraucher: jene, in deren häuslicher und/oder beruflicher Umgebung regelmäßig geraucht wird.

Ex- und gelegentliche Raucher: Abstinenz seit mindestens 12 Monaten oder Raucher, die nicht täglich und weniger als 5 Zigaretten in der Woche rauchen.

Raucher_leicht: jene, die 1 - 10 Stück pro Tag rauchen.

Raucher_mittel: jene, die 11 - 20 Stück pro Tag rauchen.

Raucher_schwer: jene, die mehr als 20 Stück pro Tag rauchen.

2.2.1 Datenselektion der weiblichen und männlichen Population

Die Selektion der Daten erfolgt bei Frauen und Männern nach dem gleichen Schema in zehn Schritten. In der rechten Spalte der Abbildungen am Ende dieses Abschnitts sind jene Probanden protokolliert, welche nicht in die Analyse aufgenommen werden, d.h. welche das im jeweiligen Schritt gestellte Ausschlußkriterium erfüllen. In der linken Spalte befinden sich die Probanden, welche weitergeführt werden. Ab dem zweiten Schritt sind die Probanden noch zusätzlich in Untergruppen aufgeschlüsselt. Probanden, welche sich aufgrund fehlender Daten keiner Gruppe zuordnen lassen, werden bereits im ersten Schritt von weiteren Untersuchungen ausgeschlossen.

Im 4. Schritt wird die Selektion auf Basis des Relativgewichtes vorgenommen. Das Relativgewicht berechnet sich nach folgender Formel: $Rel.Gew. = \frac{Gewicht[kg]}{Gre[cm]-100}$

Die praktische Selektion der Daten erfolgt nach folgenden Schritten:

1. Daten mit falschen Größen-, Alters-, oder Gewichtsangaben sowie Meßfehlern und nicht möglicher Gruppenzuteilung
2. Probanden jünger als 19 Jahre
3. Frauen mit einer Körpergröße von weniger als 1,45 m und Männer kleiner als 1,55 m

4. Frauen mit einem Relativgewicht von unter 0,8 und über 1,3448 sowie Männer mit einem Relativgewicht von unter 0,8548 und über 1,282. Die Grenzen wurden dabei so festgelegt, daß sowohl bei den Frauen als auch bei den Männern jeweils 5% der niedrigsten und höchsten Werte ausselektiert werden, z.B. 5% der Frauen haben ein Relativgewicht von weniger als 0,8.

5. jenen Probanden, die sich als lungenkrank bezeichnen, d.h., daß sie bei der Erhebung ihrer Anamnese eine der angeführten Krankheiten angegeben haben.

6. Probanden, die bei der Frage (Wohlbefinden), wie sie sich bei der Atmung fühlen, mit "Nein" geantwortet haben.

7. Frauen mit einem PEF-Wert größer als 13,8 Liter und Männer mit einem PEF-Wert größer als 20 Liter.

8. Probanden bei denen der PEF-Wert größer als 3,5×FVC ist.

9. $0{,}99 \times \mathrm{PEF} \leq \mathrm{MEF}_{75}$; diese Bedingung dient zur Überprüfung, ob ein Plateau in der Flußvolumskurve vorliegt; dies deutet auf eine obstruktive Beschränkung der Atmung hin

10. $0{,}99 \times \mathrm{MEF}_{75} \leq \mathrm{MEF}_{50}$; dies kann als Identifizierung eines Knicks in der Flußvolumskurve interpretiert werden, wenn man von einem starken Abfall zwischen dem PEF und dem MEF_{75} ausgeht, dem ein geringer Unterschied zwischen MEF_{75} und MEF_{50} folgt.

Frauen: Ausgehend von 24.965 untersuchten Probanden, bleiben nach Prüfung aller Kriterien noch **7.182 Probandinnen** (=28,77%) übrig. Die Aufteilung der 7.182 Probandinnen in sechs Untergruppen sieht folgendermaßen aus:

Niemalsraucherinnen:	4124 (57,42%)	Probandinnen
Passivraucherinnen:	411 (5,72%)	Probandinnen
Ex-gel. Raucherinnen:	1006 (14,01%)	Probandinnen
leichte Raucherinnen:	723 (10,07%)	Probandinnen
mittlere Raucherinnen:	760 (10,58%)	Probandinnen
schwere Raucherinnen:	158 (2,20%)	Probandinnen

Hier ist schon ein mögliches Problem für spätere Auswertungen zu erkennen. Die Anzahl der schweren Raucherinnen ist mit 158 im Vergleich zu anderen Gruppen, insbesondere jener der Niemalsraucher mit 4124 Probanden sehr niedrig. In Kapitel Nr. 12, in dem spezielle Vergleiche zwischen Niemals- und schweren Raucherinnen durchgeführt werden, wird deshalb die Gruppe der schweren Raucherinnen um jene mittleren Raucherinnen, welche mindestens 15 Zigaretten pro Tag rauchen, vergrößert.

Männer: Ausgehend von 29.158 Probanden bleiben nach vollzogener Selektionierung noch genau **10.383 Probanden** (=35,61%) übrig. Die zuletzt angeführte Aufteilung in sechs Gruppen hat folgendes Aussehen:

Niemalsraucher:	3459 (33,31%)	Probanden
Passivraucher:	382 (3,68%)	Probanden
Ex-gel. Raucher:	3072 (29,59%)	Probanden
leichte Raucher:	813 (7,83%)	Probanden
mittlere Raucher:	1620 (15,60%)	Probanden
schwere Raucher:	1037 (9,99%)	Probanden

Von Vorteil ist hier, daßdie Gruppe der schweren Raucher noch mit 1.037 Probanden besetzt ist.

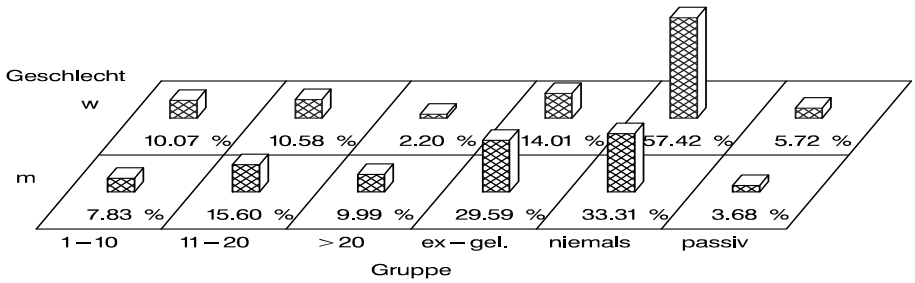
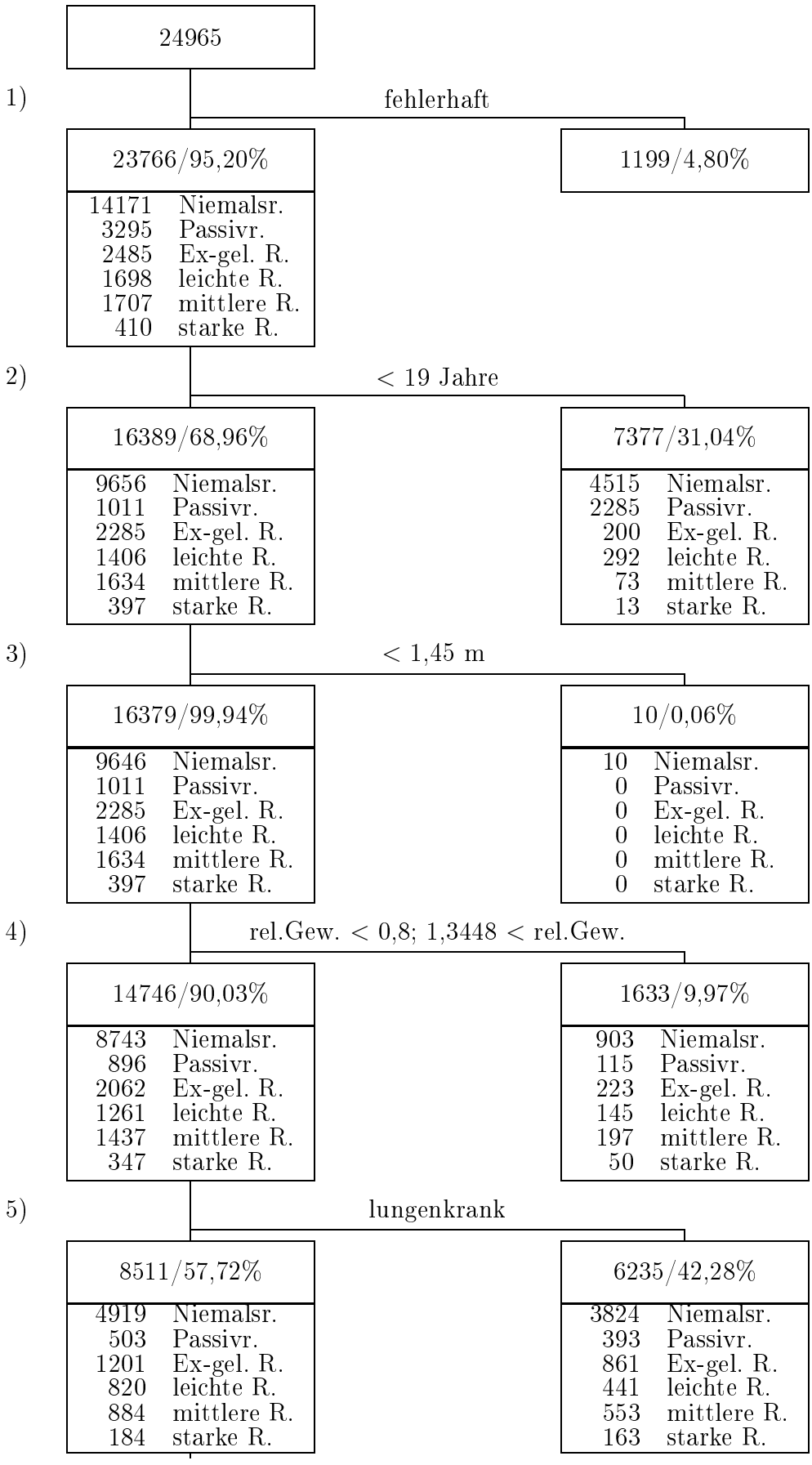


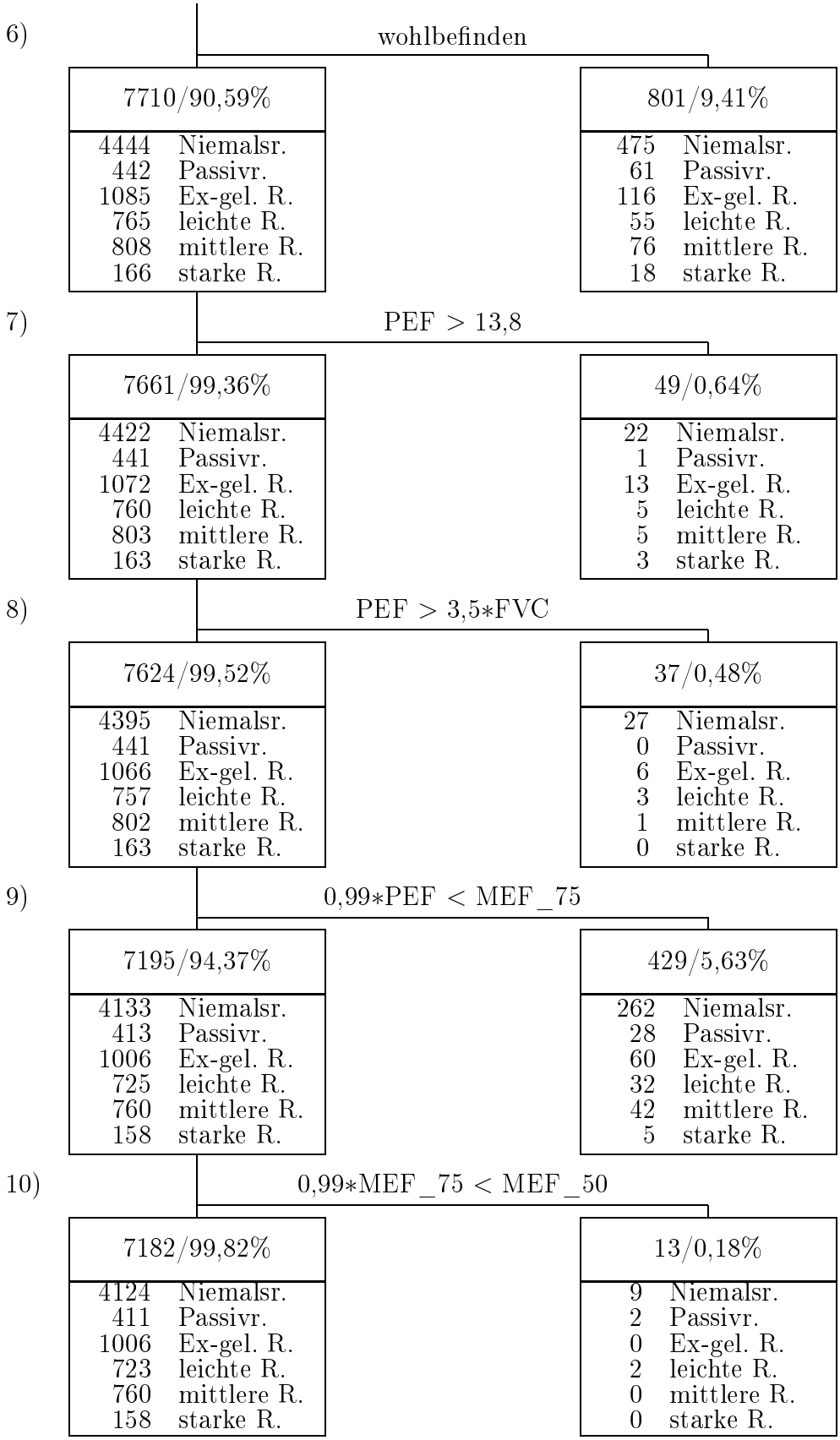
Abbildung 2.1: BlockHistogramm

Abbildung 2.1 ermöglicht einen graphischen Vergleich der prozentuellen Anteile der einzelnen Gruppen. Die Prozentwerte sind jeweils getrennt für Frauen und Männer berechnet. Bei den Frauen ist deutlich der überproportionale Anteil der Niemalsraucherinnen und der geringe Anteil der schweren Raucherinnen zu sehen. Hingegen ist die Gruppenaufteilung bei den Männern etwas ausgewogener. Vergleicht man die Geschlechter miteinander, so sieht man, daß vor allem der Anteil der schweren und ex-gel. Raucherinnen geringer ist als der Anteil der entsprechenden Gruppen bei den Männern.

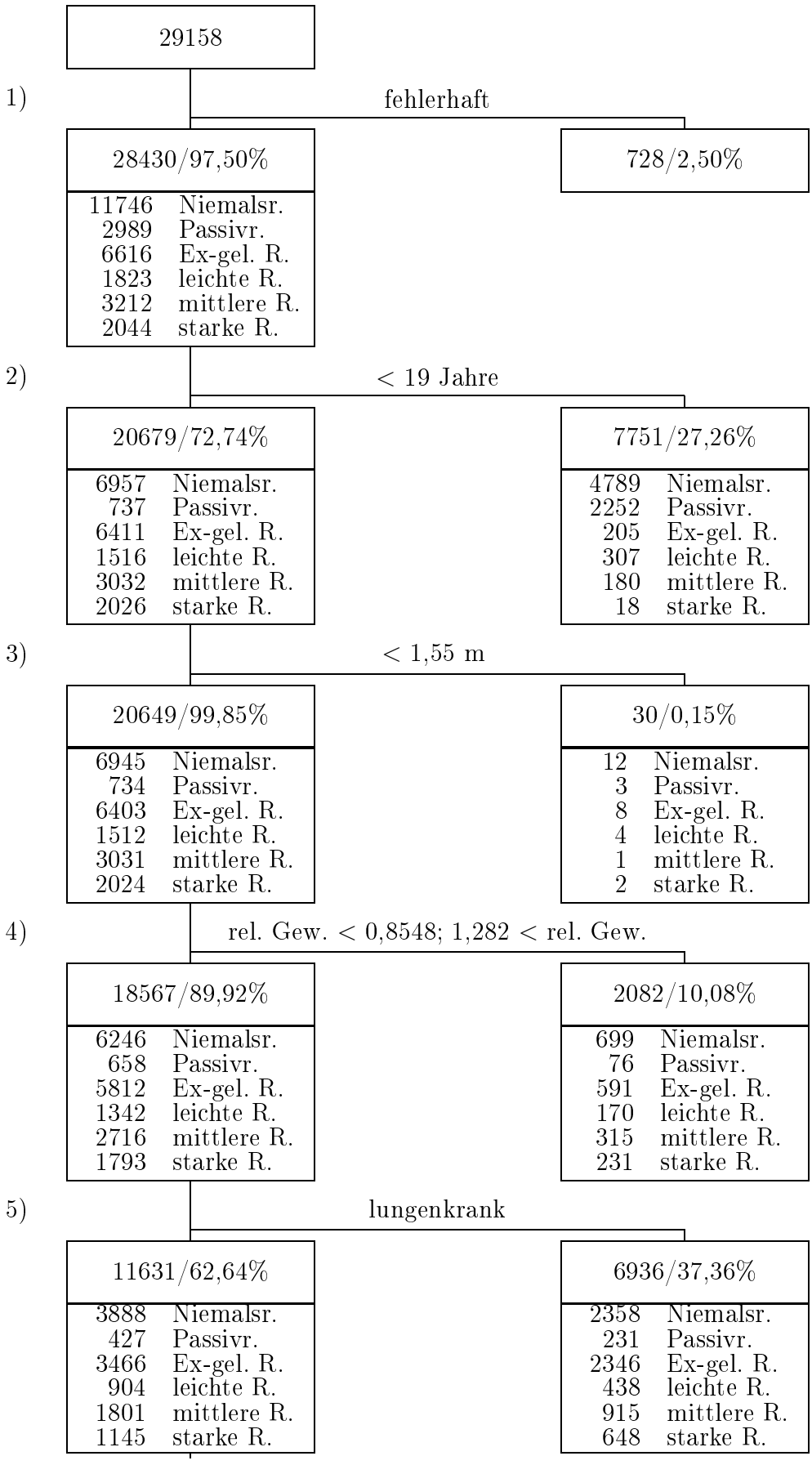
Auf den folgenden vier Seiten ist die Datenselektion bei den Frauen und bei den Männern graphisch in Form eines Flußdiagramms wiedergegeben. Dabei wird zusätzlich die prozentuelle Aufteilung der Probanden angegeben, bezogen auf die Anzahl der in vorherigen Schritt weitergeführten Probanden. Auffallend ist, daß bei den Frauen im Schritt 5 ein höherer Anteil von Lungenkranken (42,28%) auftritt als bei den Männern (37,36%).

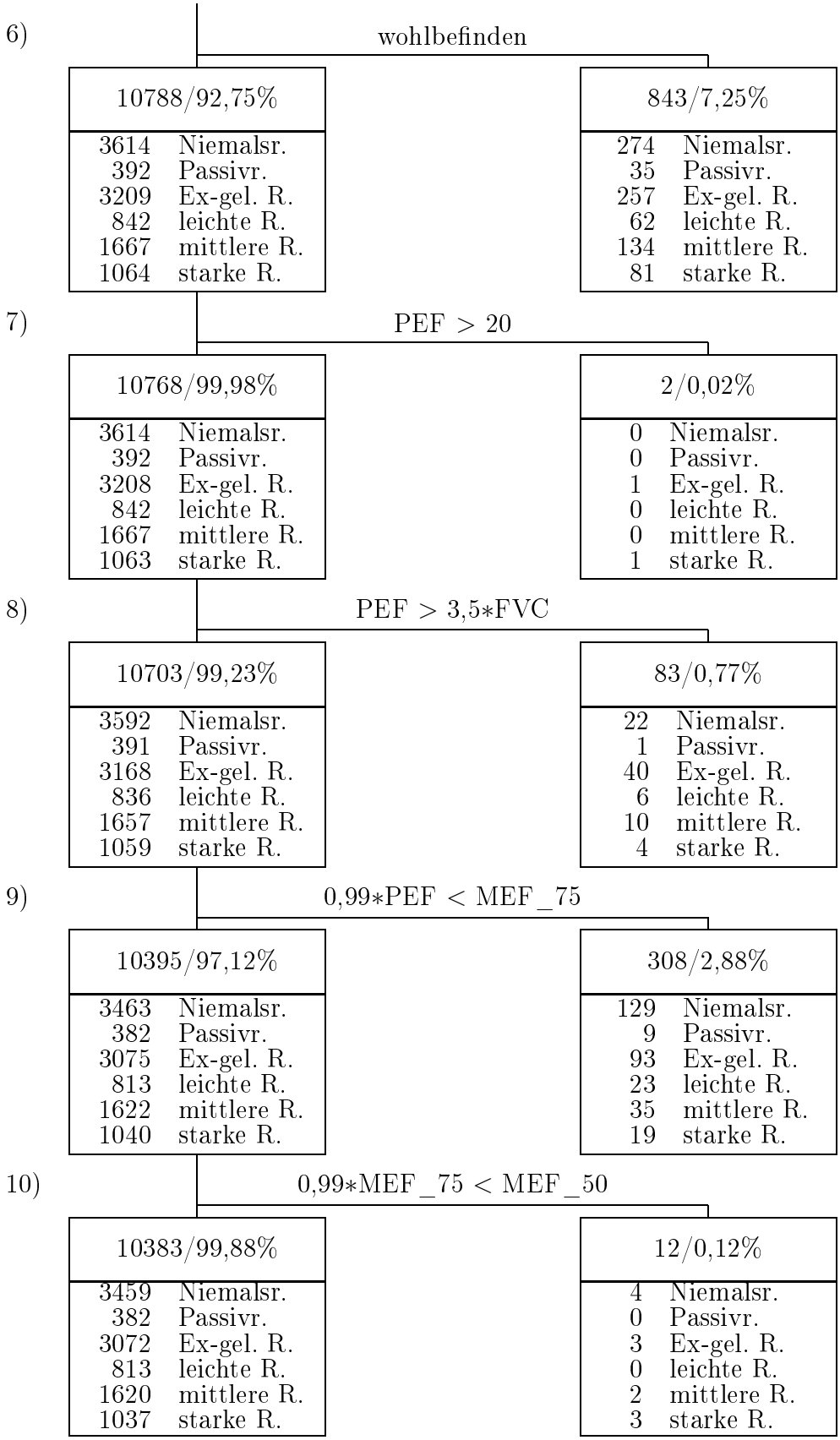
Datenselektion bei den Frauen




6)
wohlbefinden
7710/90,59%
4444 Niemalsr.
442 Passivr.
1085 Ex-gel. R.
765 leichte R.
808 mittlere R.
166 starke R.
801/9,41%
475 Niemalsr.
61 Passivr.
116 Ex-gel. R.
55 leichte R.
76 mittlere R.
18 starke R.
7)
PEF > 13,8
7661/99,36%
4422 Niemalsr.
441 Passivr.
1072 Ex-gel. R.
760 leichte R.
803 mittlere R.
163 starke R.
49/0,64%
22 Niemalsr.
1 Passivr.
13 Ex-gel. R.
5 leichte R.
5 mittlere R.
3 starke R.
8)
PEF > 3,5*FVC
7624/99,52%
4395 Niemalsr.
441 Passivr.
1066 Ex-gel. R.
757 leichte R.
802 mittlere R.
163 starke R.
37/0,48%
27 Niemalsr.
0 Passivr.
6 Ex-gel. R.
3 leichte R.
1 mittlere R.
0 starke R.
9)
0,99*PEF < MEF_75
7195/94,37%
4133 Niemalsr.
413 Passivr.
1006 Ex-gel. R.
725 leichte R.
760 mittlere R.
158 starke R.
429/5,63%
262 Niemalsr.
28 Passivr.
60 Ex-gel. R.
32 leichte R.
42 mittlere R.
5 starke R.
10)
0,99*MEF_75 < MEF_50
7182/99,82%
4124 Niemalsr.
411 Passivr.
1006 Ex-gel. R.
723 leichte R.
760 mittlere R.
158 starke R.
13/0,18%
9 Niemalsr.
2 Passivr.
0 Ex-gel. R.
2 leichte R.
0 mittlere R.
0 starke R.

Datenselektion bei den Männern




6)
wohlbefinden
10788/92,75%
3614 Niemalsr.
392 Passivr.
3209 Ex-gel. R.
842 leichte R.
1667 mittlere R.
1064 starke R.
843/7,25%
274 Niemalsr.
35 Passivr.
257 Ex-gel. R.
62 leichte R.
134 mittlere R.
81 starke R.
7)
PEF > 20
10768/99,98%
3614 Niemalsr.
392 Passivr.
3208 Ex-gel. R.
842 leichte R.
1667 mittlere R.
1063 starke R.
2/0,02%
0 Niemalsr.
0 Passivr.
1 Ex-gel. R.
0 leichte R.
0 mittlere R.
1 starke R.
8)
PEF > 3,5*FVC
10703/99,23%
3592 Niemalsr.
391 Passivr.
3168 Ex-gel. R.
836 leichte R.
1657 mittlere R.
1059 starke R.
83/0,77%
22 Niemalsr.
1 Passivr.
40 Ex-gel. R.
6 leichte R.
10 mittlere R.
4 starke R.
9)
0,99*PEF < MEF_75
10395/97,12%
3463 Niemalsr.
382 Passivr.
3075 Ex-gel. R.
813 leichte R.
1622 mittlere R.
1040 starke R.
308/2,88%
129 Niemalsr.
9 Passivr.
93 Ex-gel. R.
23 leichte R.
35 mittlere R.
19 starke R.
10)
0,99*MEF_75 < MEF_50
10383/99,88%
3459 Niemalsr.
382 Passivr.
3072 Ex-gel. R.
813 leichte R.
1620 mittlere R.
1037 starke R.
12/0,12%
4 Niemalsr.
0 Passivr.
3 Ex-gel. R.
0 leichte R.
2 mittlere R.
3 starke R.

Kapitel 3

Statistische Grundlagen I

3.1 Lagemaße von Häufigkeitsverteilungen

Lagemaße sind Kennzahlen, die eine Beobachtungsreihe $X_i \overset{iid}{\sim} F, i = 1, \ldots, n$, mit $E(X_i) = \mu, Var(X_i) = \sigma^2$, charakterisieren. Folgende Maßzahlen werden verwendet:

⊙ Das arithmetische Mittel: $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
ist ein sinnvolles Lagemaß vor allem bei metrisch skalierten Merkmalen. Es stellt einen erwartungstreuen Schätzer von μ, dem Erwartungswert (X_i) dar. Liegen jedoch Ausreißer bezüglich des Verteilungsmodells F vor, so können diese den Schätzer $\overline{X}$ stark beeinflußen, und somit kann ein falscher Eindruck über die Lage des Zentrums der Verteilung entstehen.

⊙ Der empirische Median:
Sind die Beobachtungsvariablen der Größe nach geordnet, also $X_{(1)} \leq \cdots \leq X_{(n)}$, so kann der empirische Median folgendermaßen angegeben werden:

$$\tilde{X}_{0,5} = \begin{cases} \frac{1}{2}\left(X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})}\right) & \text{falls} \quad n \quad \text{gerade} \\ X_{([\frac{n}{2}]+1)} & \text{falls} \quad n \quad \text{ungerade} \end{cases}$$

Das Charakteristische am Median ist, daß genau 50% der Beobachtungen $X_1, \ldots, X_n$ $<$ (n gerade) oder $\leq$ (n ungerade) $\tilde{X}_{0,5}$ sind. Das heißt, daß der Median von Ausreißern kaum beeinflußt wird. Er ist ein erwartungstreuer Schätzer für den theoretischen Median $x_{0,5}$ mit $F(x_{0,5}) = 0,5$.

⊙ Das p-te empirische Quantil:
Ist $X_{(1)} \leq \cdots \leq X_{(n)}$ die geordnete Stichprobe zu $X_1, \ldots, X_n$, so bezeichnet man den folgenden Wert als p-tes empirisches Quantil $(0 < p < 1)$:

$$\tilde{X}_p = \begin{cases} \frac{1}{2}\left(X_{(np)} + X_{(np+1)}\right) & \text{falls} \quad np \quad \text{ganzzahlig} \\ X_{([np]+1)} & \text{sonst} \end{cases}$$

3.2 Streuungsmaße

⊙ Die empirische Standardabweichung: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})}$
ist die Wurzel aus der empirischen Varianz S^2. Diese ist ein erwartungstreuer und konsistenter Schätzer für die Varianz der Verteilung σ^2. Der Schätzer für die Varianz des Mittelwertes lautet S^2/n.

⊙ Die Spannweite (Range):
Sind $X_{(1)}, \ldots, X_{(n)}$ die der Größe nach geordneten Beobachtungswerte, so entspricht die Spannweite der Differenz des größten und kleinsten Wertes: $R = X_{(n)} - X_{(1)}$. Sie gibt also den Wertebereich an, in dem die Beobachtungen liegen.

⊙ Der Quartilsabstand (Interquartile Range): $IQR = \tilde{X}_{0,75} - \tilde{X}_{0,25}$
Mit dem IQR kann bei Vorliegen einer Normalverteilung $N(\mu, \sigma^2)$ die Standardabweichung σ geschätzt werden durch:

$$\tilde{\sigma} = \frac{IQR}{1,349}$$

Dieser Schätzer ist relativ robust gegenüber Verletzungen der Normalverteilungsannahme.

3.3 Diagnoseplots

3.3.1 Boxplot

Ein Boxplot ist ein Instrument der graphischen Datenanalyse. Die vertikale Achse steht für die beobachtete Responsevariable. Auf dieser Skala werden die Quantile $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ durch horizontale Striche aufgetragen. Durch Verbinden der Enden dieser Striche entsteht ein Rechteck, die sogenannte Box. Sie enthält laut Definition 50% aller Beobachtungen und ihre Länge entspricht dem IQR. An der Stelle des Medians wird die Box unterteilt. Schließlich werden die Werte $\min\left(\tilde{x}_{0,75} + 1,5 \times IQR, x_{(n)}\right)$ und $\max\left(\tilde{x}_{0,25} - 1,5 \times IQR, x_{(1)}\right)$ aufgetragen und durch eine Linie mit der Box verbunden. Ausreißer, die über $\tilde{x}_{0,75} + 1,5 \times IQR$ bzw. unter $\tilde{x}_{0,25} - 1,5 \times IQR$ liegen, sind durch Kreise gekennzeichnet. Extreme Ausreißer, die über $\tilde{x}_{0,75} + 3 \times IQR$ oder unter $\tilde{x}_{0,25} - 3 \times IQR$ liegen sind durch Sterne veranschaulicht.

Dieser Plot enthält also gleichzeitig ein Lokationsmaß , den Median, ein Streuungsmaß , den IQR, Hinweise auf Symmetrie oder Schiefe der Verteilung sowie mögliche Ausreißer. Abbildung 3.1 zeigt ein Beispiel für einen Plot, welcher sechs Boxplots nebeneinander enthält. Diese Art der Darstellung dient dazu die Verteilungen verschiedener Gruppen zu analysieren und miteinander zu vergleichen.

Für weitere Betrachtungen von Boxplots sei auf Falk et al. [3], Hartung [6] und Stadlober [22] verwiesen.

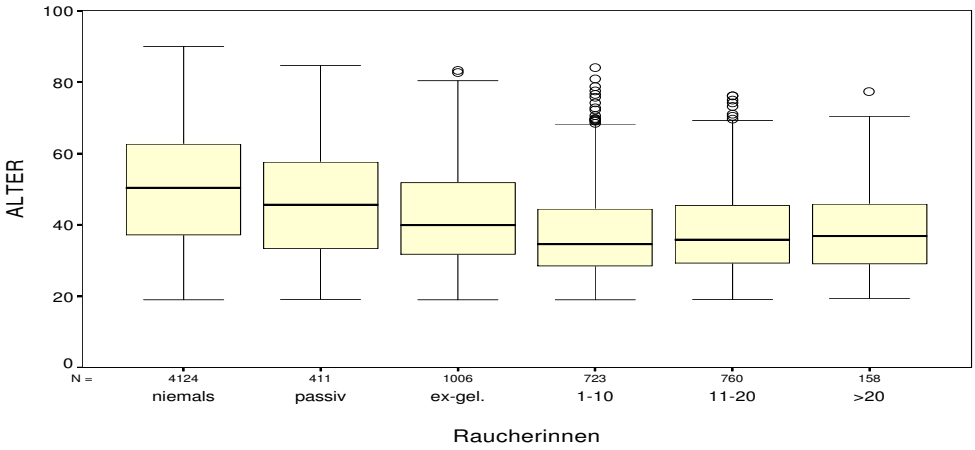


Abbildung 3.1: Boxplot

3.3.2 Histogramm

Um eine Vorstellung von der Verteilung der Beobachtungen innerhalb ihres Variationsbereiches zu erhalten, zerlegen wir das Intervall $[x_{(1)}, x_{(n)}]$ in d disjunkte Zellen

$$(a_0, a_1], (a_1, a_2], \ldots, (a_{d-1}, a_d],$$

die wir mit $I_1, \ldots, I_d$ bezeichnen; dabei ist $a_0 < a_1 < \cdots < a_d, a_0 < x_{(1)} \leq x_{(n)} \leq a_d$. Setzen wir nun n_s:= Anzahl der Daten unter $x_1, \ldots, x_n$, die in I_s liegen, d.h. $n_s = \#\{x_i, i = 1, \ldots, n : x_i \in I_s\}$ und tragen wir n_s über I_s ab, so erhalten wir ein Histogramm (siehe Abbildung 3.2). In dieser Arbeit werden die Klassenbreiten jeweils konstant gewählt.

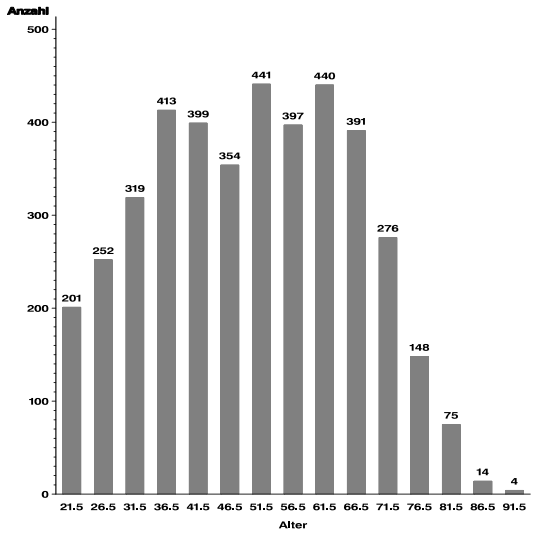


Abbildung 3.2: Histogramm

3.3.3 Blockhistogramme

Blockhistogramme sind 2-dimensionale Histogramme, die optisch 3-dimensional aufbereitet sind. Die Häufigkeiten sind nicht durch Balken, sondern durch Pfeiler dargestellt. Blockhistogramme bieten die Möglichkeit zusätzlich ein drittes Merkmal in ein Histogramm zu integrieren, wie Abbildung 3.3 zeigt. Die Häufigkeitsverteilungen bzgl. der beiden Referenzklassen können dank der 3-Dimensionalität in einer Graphik übersichtlich dargestellt werden.

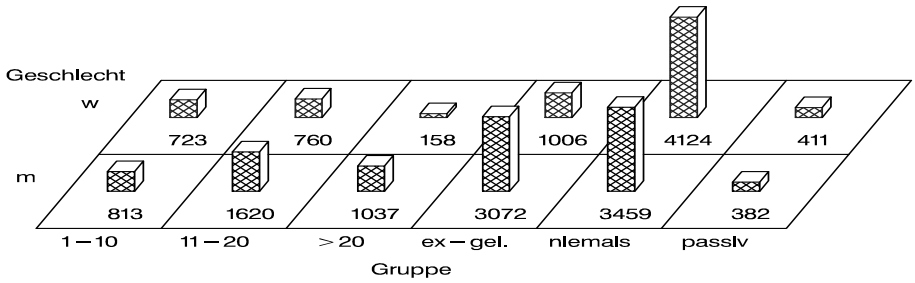


Abbildung 3.3: Blockhistogramm

3.4 Kovarianzanalyse

Wir verwenden die Kovarianzanalyse in Kapitel 4, um die adjustierten Mittelwerte der Lungenfunktionsparameter innerhalb der einzelnen Gruppen zu berechnen. Die Mittelwerte der Lungenfunktionsparameter werden nach den sogenannten Kovariaten Alter, Größe und Gewicht adjustiert, um mögliche Unterschiede in den Verteilungen dieser drei Variablen auszugleichen.

3.4.1 Voraussetzung

Voraussetzung, um eine Kovarianzanalyse durchzuführen, ist die Parallelität der Steigungen der Regressionsgeraden in den beiden miteinander zu vergleichenden Gruppen. Zu diesem Zweck wird folgendes Modell betrachtet:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{j=1}^{k-1} \beta_{p+j} z_j + \sum_{i=1}^{p} \sum_{j=1}^{k-1} \gamma_{ij} x_i z_j + \epsilon$$

In diesem Modell gibt es p Kovariate $x_1, \ldots, x_p$ und k Gruppen. Der Ausdruck $\sum_{i=1}^{p} \sum_{j=1}^{k-1} \gamma_{ij} x_i z_j$ beschreibt mögliche Wechselwirkungen zwischen den Gruppen. Um nun die Annahme der Parallelität der Regressionskurven zu überprüfen wird folgende Hypothese getestet:

$$H_0 : \gamma_{ij} = 0 \quad \text{für alle} \quad i = 1, \ldots, p;\ j = 1, \ldots, k-1$$

gegen

$$H_1 : \gamma_{ij} \neq 0 \quad \text{für mindestens ein} \quad i = 1, \ldots, p;\ j = 1, \ldots, k-1$$

Falls H_0 nicht verworfen werden kann, kann davon ausgegangen werden, daß die Regressionskurven parallel sind.

3.4.2 Theorie für p Kovariable und k Gruppen

Das Modell, um diese Situation zu beschreiben sieht folgendermaßen aus:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \beta_{p+1} z_1 + \cdots + \beta_{p+k-1} z_{k-1} + \epsilon \tag{3.1}$$

mit p Kovariaten $x_1, \ldots, x_p$ und k Gruppen, welche durch $k-1$ Dummy-Variablen $z_1, \ldots, z_{k-1}$ repräsentiert sind. Eine Möglichkeit, um diese Dummy-Variablen zu definieren ist folgende:

$$z_j = \begin{cases} 1 & \text{falls Gruppe} \quad j \\ 0 & \text{sonst} \end{cases} \qquad j = 1, \ldots, k-1$$

Damit können nun Regressionsgleichungen für die k Gruppen durch die entsprechenden Kombinationen der Werte der z's berechnet werden. Diese Gleichungen beschreiben parallele Hyperebenen im $\mathbb{R}^{p+1}$.

Die adjustierten Mittelwerte für die einzelnen Gruppen erhält man, indem der Mittelwert der aus der Regressionsgleichung vorhergesagten Y-Werte berechnet wird. Zur Berechnung der Y-Werte werden die auf den gesamten Daten der Gruppen basierenden Mittelwerte der Kovariaten $x_1, \ldots, x_p$ verwendet.

$$\begin{aligned} \overline{Y}_j(adj) &= (\tilde{\beta}_0 + \tilde{\beta}_{p+j}) + \tilde{\beta}_1 \overline{X}_1 + \cdots + \tilde{\beta}_p \overline{X}_p \quad j = 1, \ldots, k-1 \\ \overline{Y}_k(adj) &= \tilde{\beta}_0 + \tilde{\beta}_1 \overline{X}_1 + \cdots + \tilde{\beta}_p \overline{X}_p \end{aligned}$$

Um festzustellen, ob zumindest einer der k adjustierten Mittelwerte signifikant von Null verschieden ist, testen wir die Nullhypothese:

$$H_0 : \beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+k-1} = 0$$

gegen

$$H_1 : \beta_i \neq 0 \quad \text{für mindestens ein} \quad i = p+1, \ldots, p+k-1$$

unter Verwendung eines multiplen-partiellen F-Tests mit $k-1$ und $n-p-k$ Freiheitsgraden basierend auf dem Modell (3.1).

Das Statistikprogramm *SAS* mit welchem die Kovarianzanalyse durchgeführt wird berechnet die adjustierten Mittelwerte und führt zusätzlich noch paarweise Vergleiche dieser Mittelwerte auf signifikante Unterschiede durch.

Für weitere Erläuterungen zur Kovarianzanalyse siehe Kleinbaum et al. [10].

3.5 t-Test

Im folgenden betrachten wir den Fall, daß zwei unabhängige Stichprobenvektoren $X_1, \ldots, X_n$ mit $X_i \sim N(\mu_x, \sigma_x)$ und $Y_1, \ldots, Y_m$ mit $Y_i \sim N(\mu_y, \sigma_y)$ gegeben sind.

3.5.1 Test für $\mu_x - \mu_y$, falls $\sigma_x = \sigma_y$, t-Test

Hypothese

$$H_0 : \mu_x - \mu_y = 0 \quad \text{gegen} \quad H_1 : \mu_x - \mu_y \neq 0$$

Teststatistik

$$T = \frac{\overline{X} - \overline{Y}}{S_p} \sqrt{\frac{nm}{n+m}} \sim t_{n+m-2}, \quad \text{falls} \quad \mu_x = \mu_y,$$

mit

$$S_p^2 = \frac{1}{n+m-2} \left((n-1)S_x^2 + (m-1)S_y^2 \right) \quad \text{(gepoolte Varianz)},$$

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad S_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$$

Für $n = m$ gilt speziell

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{S_x^2 + S_y^2}}\sqrt{n} \sim t_{2n-2}, \quad \text{falls} \quad \mu_x = \mu_y.$$

Kritische Bereiche K

$$K = (-\infty, -t_{n+m-2;1-\alpha/2}) \cup (t_{n+m-2;1-\alpha/2,\infty})$$

Die Hypothese H_0 wird abgelehnt, falls $t \in K$. Die Quantile $t_{n+m-2;\alpha}$ sind den entsprechenden Tabellen zu entnehmen.

3.5.2 Test für $\mu_x - \mu_y$, falls $\sigma_x \neq \sigma_y$, Approximativer t-Test

Hypothese

$$H_0 : \mu_x - \mu_y = 0 \quad \text{gegen} \quad H_1 : \mu_x - \mu_y \neq 0$$

Teststatistik

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \overset{\text{appr}}{\sim} t_\nu, \quad \text{falls} \quad \mu_x = \mu_y,$$

mit einem angenäherten Freiheitsgrad von

$$\nu = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\left(\frac{1}{n-1}\left(\frac{S_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_y^2}{m}\right)^2\right)}$$

Kritische Bereiche K

$$K = (-\infty, -t_{\nu;1-\alpha/2}) \cup (t_{\nu;1-\alpha/2,\infty})$$

Die Hypothese H_0 wird wiederum abgelehnt, falls $t \in K$.

Kapitel 4

Voranalyse der Daten

4.1 Alters-,Gewichts-,Größenverteilung bei Frauen

Die folgenden Boxplots und Histogramme dienen zur optischen Kontrolle der jeweiligen Verteilungen. Besonders durch die Darstellung der Altersverteilung läßt sich feststellen, ab welcher Altersgruppe die Daten auszudünnen beginnen und ein Vergleich der Gruppen fragwürdig wird.

4.1.1 Altersverteilung

Die Einteilung der Klassen in den Histogrammen erfolgt analog zur Einteilung der Altersklassen bei der späteren Auswertung und graphischen Darstellung durch die Medianplots.

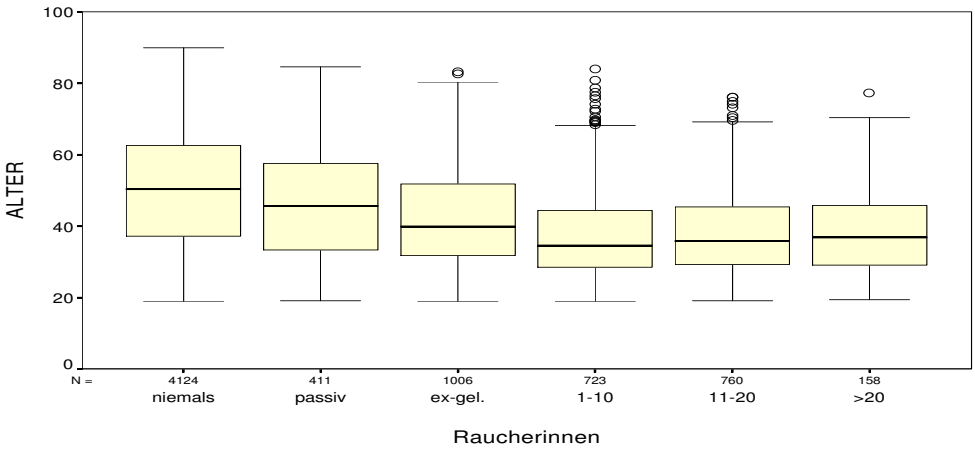


Abbildung 4.1: Boxplot Altersverteilung

Durch die Lage der Mediane und der Grenzen der Boxplots sieht man, daß die Altersverteilung in den Gruppen der Niemals- und Passivraucherinnen von jenen der

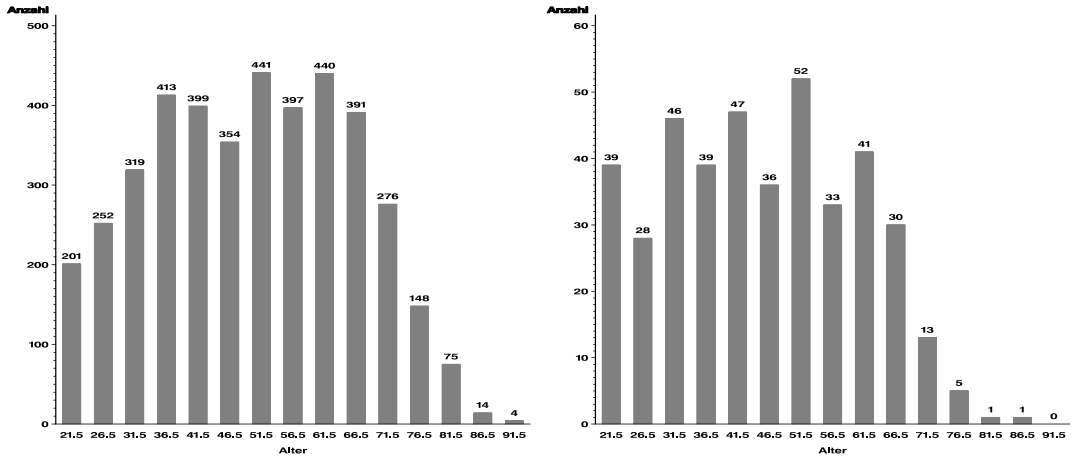


Abbildung 4.2: Niemals- und Passivraucherinnen

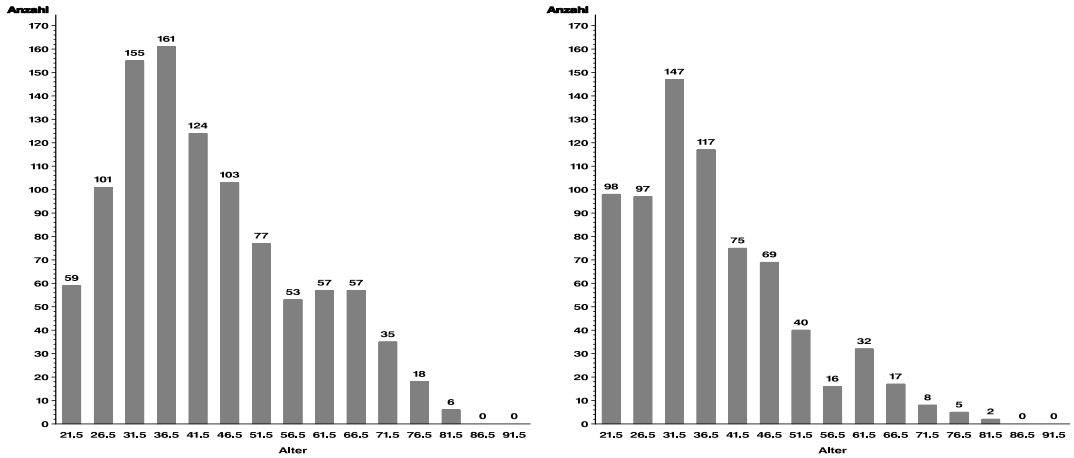


Abbildung 4.3: Ex_gelegentliche- und leichte Raucherinnen

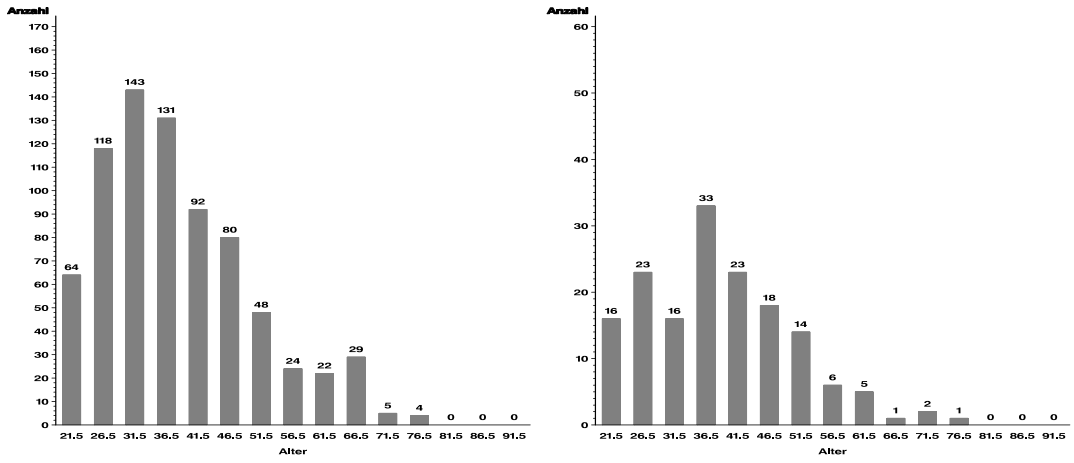


Abbildung 4.4: Mittlere- und starke Raucherinnen

Raucherninnengruppen abweicht. Die Niemals- und Passivraucherinnen haben im Vergleich zu den anderen Raucherinnengruppen ein höheres Durchschnittsalter und einen größeren Streubereich, ersichtlich an der Länge der Boxes (= IQR).

	Alter: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[m]$	49,91	45,29	42,97	37,38	38,27	38,70
$s[m]$	15,77	14,98	14,32	12,66	11,97	11,61

Die Mittelwerte und Standardabweichungen geben die unterschiedliche Altersstruktur der Gruppen wieder.

In den einzelnen Histogrammen sieht man sehr deutlich die unterschiedliche Altersverteilung der Probandinnen.Vor allem ist ersichtlich, daß bei leichten und mittleren Raucherinnen ab einem Alter von 70 Jahren bzw. bei starken Raucherinnen schon ab 55 Jahren nur mehr eine geringe Zahl von Beobachtungen vorhanden sind.

4.1.2 Größenverteilung

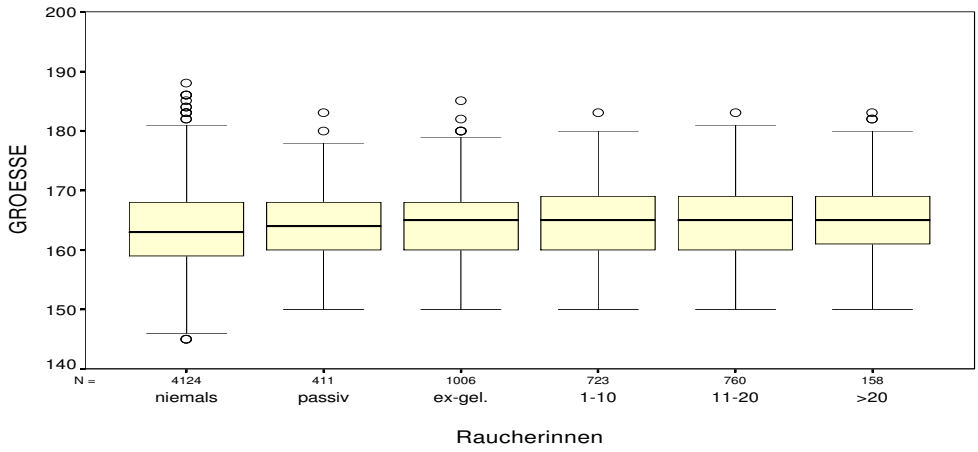


Abbildung 4.5: Boxplot Größenverteilung

Bei der Datenselektion wurden alle Frauen mit einer Körpergröße von unter 1,45 m ausgesondert gemäß dem Ziel, eine möglichst homogene Datenmenge mit wenigen extremen Werten zu erhalten. Aus den Boxplots ist ersichtlich, daß nur sehr wenige Ausreißer (nach oben) und keine extremen Ausreißer vorhanden sind.

In der folgenden Tabelle sind die Mittelwerte und Standardabweichungen für die einzelnen Gruppen angeführt.

	Größe: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[m]$	1,63	1,64	1,65	1,65	1,65	1,66
$s[m]$	0,06	0,07	0,06	0,06	0,06	0,06

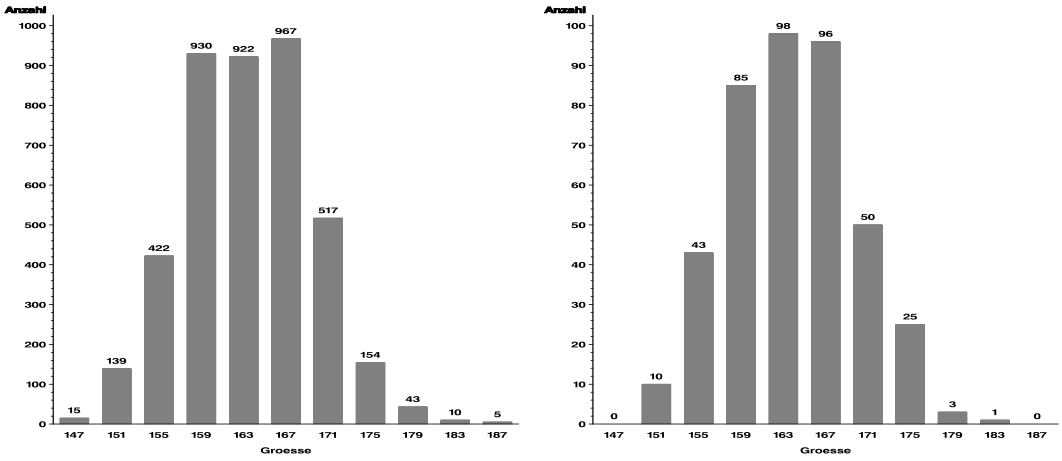


Abbildung 4.6: Niemals- und Passivraucherinnen

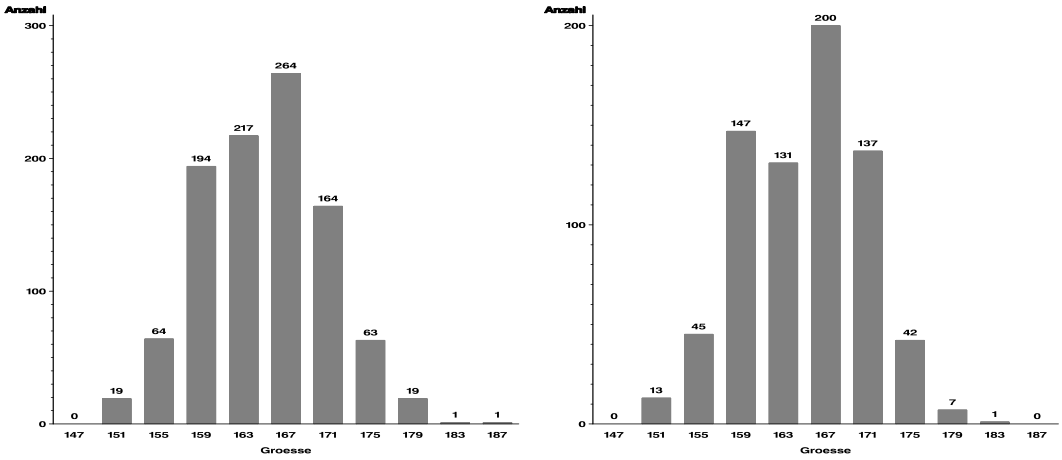


Abbildung 4.7: Ex_gelegentliche- und leichte Raucherinnen

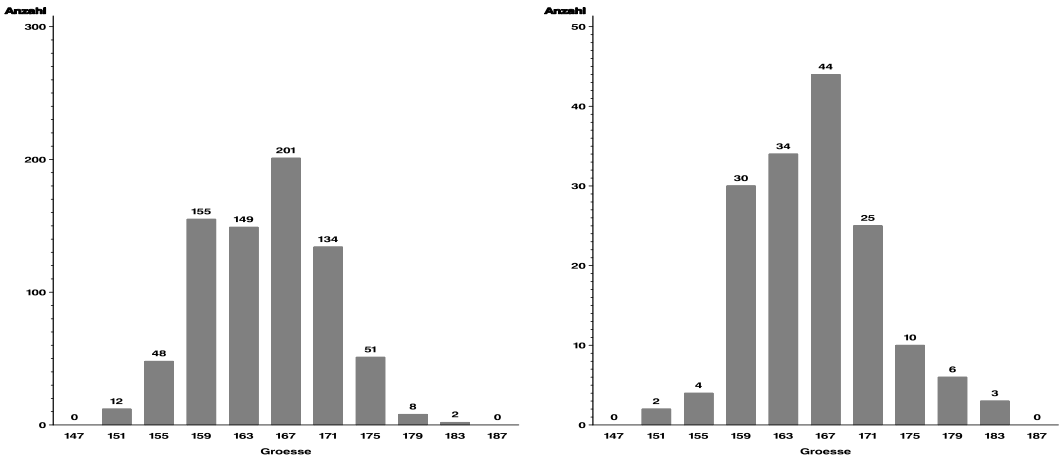


Abbildung 4.8: Mittlere- und starke Raucherinnen

Die Mittelwerte und die Standardabweichungen geben die Homogenität der Gruppen wieder.

An den Histogrammen ist die weitgehend symmetrische Verteilung der Größen in den Gruppen zu sehen.

4.1.3 Gewichtsverteilung

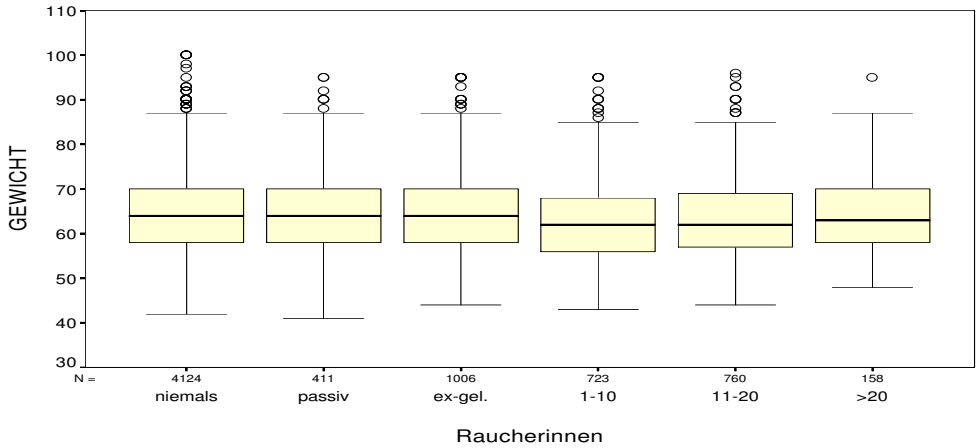


Abbildung 4.9: Boxplot Gewichtsverteilung

	Gewicht: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[kg]$	64,6	64,5	64,7	62,9	63,4	64,7
$s[kg]$	8,8	9,1	9,1	8,8	9,0	9,1

Bei der Datenselektion wurden alle Probanden mit einem Relativgewicht von unter 0,8 und über 1,3448 entfernt. In den Boxplots und in der Tabelle ist die daraus resultierende Homogenität der Gruppen bezüglich des Gewichtes ersichtlich.

In den Histogrammen ist vor allem die linkssteile Verteilung der Gewichtsklassen bei den Frauen im Gegensatz zu einer durchaus symmetrischen Verteilung der Gewichtsklassen bei den Männern (siehe Abschnitt 4.2.3) anzumerken. Dieser geschlechtsspezifischer Unterschied ist wahrscheinlich auf das Bestreben der Frauen in unserer Kultur zurückzuführen eine schlanke Figur zu haben.

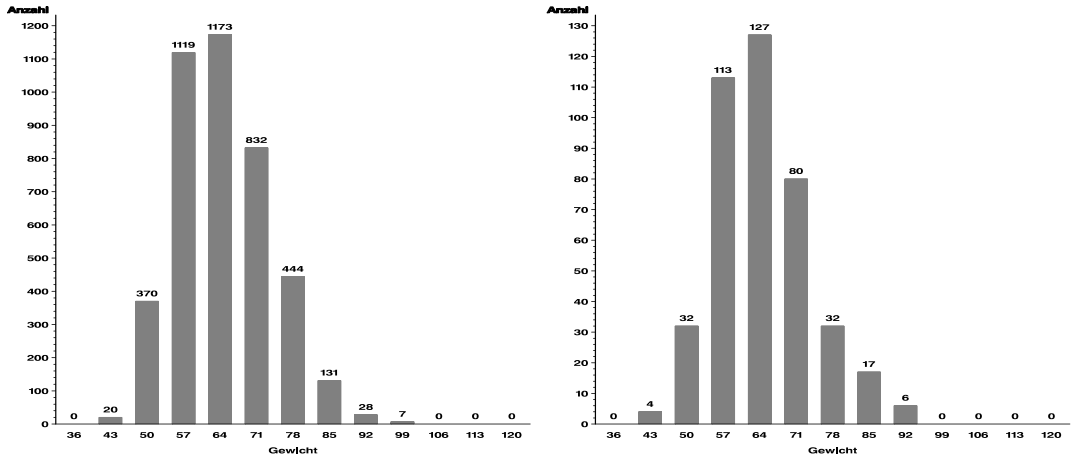


Abbildung 4.10: Niemals- und Passivraucherinnen

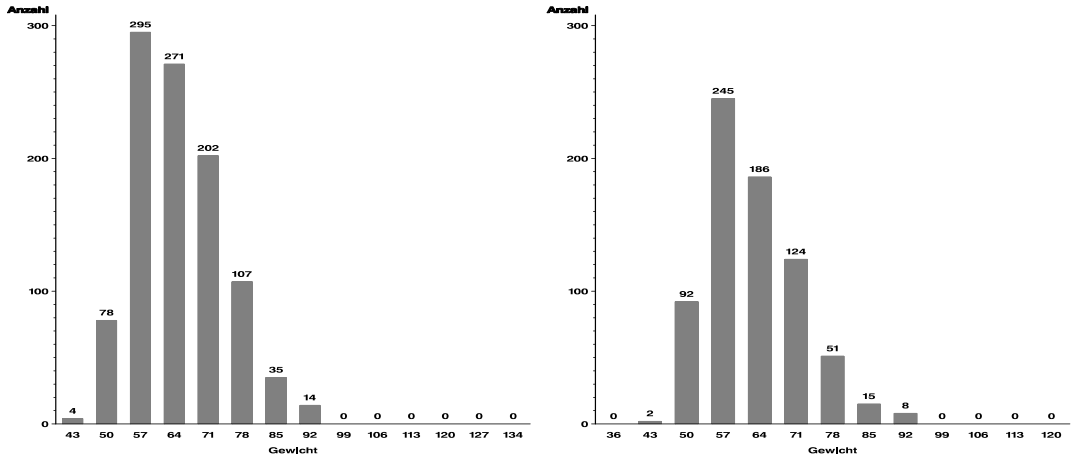


Abbildung 4.11: Ex_gelegentliche- und leichte Raucherinnen

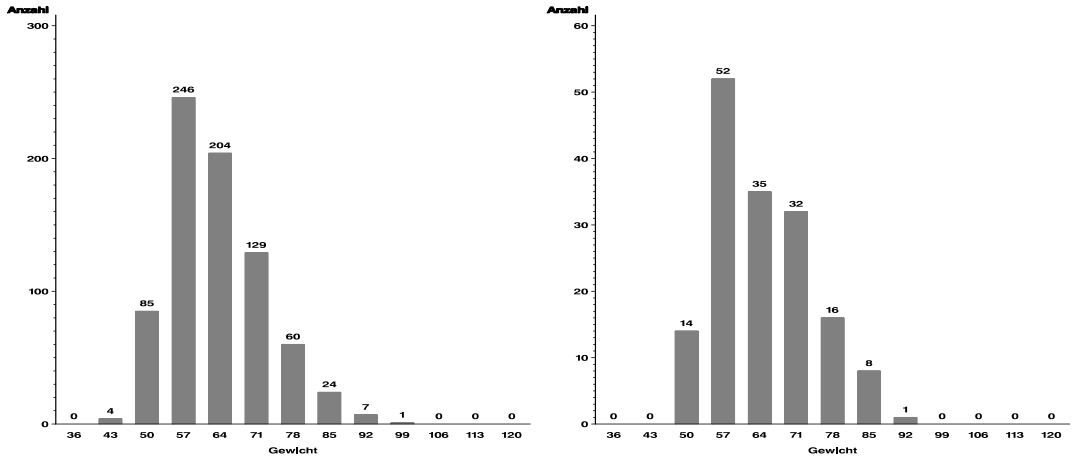


Abbildung 4.12: Mittlere- und starke Raucherinnen

4.2 Alters-,Gewichts-Größenverteilung bei Männern

Mit Boxplots und Histogrammen werden die Verteilungen in den Gruppen überprüft. Wichtig ist wiederum die Besetzung der höheren Altersklassen mit Probanden und die Homogenität der Gruppen bezüglich der Gewichts- und Größenverteilung.

4.2.1 Altersverteilung

Die Einteilung der Altersklassen in 5-Jahres Intervallen erfolgt analog zu späteren Auswertungen. Für die Interpretation der Auswertungen ist entscheidend wieviele Beobachtungen insbesondere in den höheren Altersklassen vorliegen.

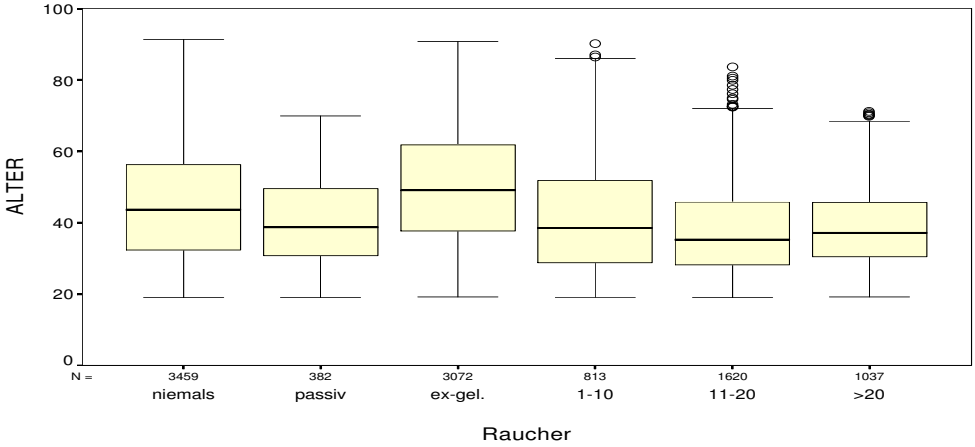


Abbildung 4.13: Boxplot Altersverteilung

Im Boxplot sieht man, daß die Altersverteilung der Ex- und gelegentlichen Raucher am stärksten (nach oben) von den anderen abweicht. Eine etwas geringere Abweichung ist

noch bei den Niemalsrauchern gegeben. Bei den Passivrauchern und den drei Raucher-gruppen ist eine geringere Anzahl von Probanden im höheren Altersbereich erkennbar.

	Alter: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[m]$	45,17	39,99	49,54	41,50	38,17	38,77
$s[m]$	15,53	11,71	14,86	15,70	12,87	10,90

Die Mittelwerte und Standardabweichungen bestätigen das Bild, das wir bereits in den Boxplots über die unterschiedlichen Altersverteilungen erhalten haben.

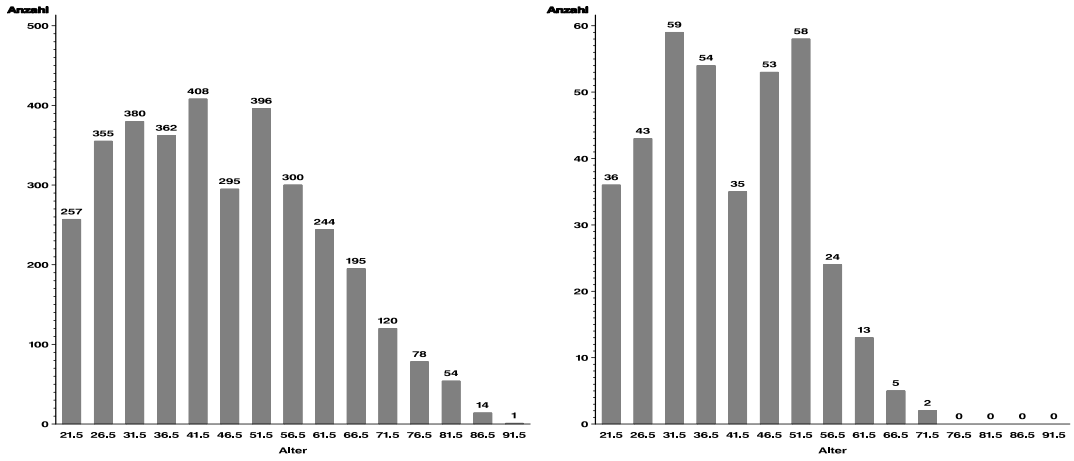


Abbildung 4.14: Niemals- und Passivraucher

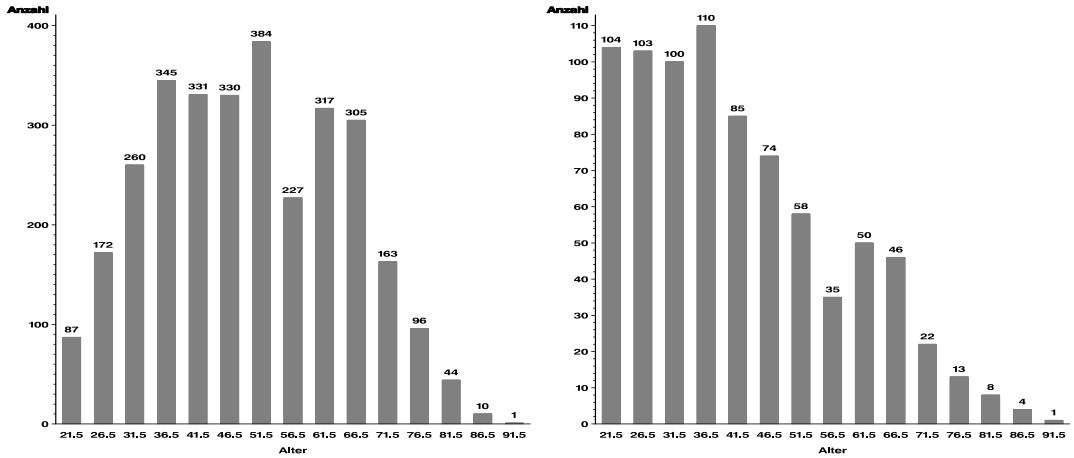


Abbildung 4.15: Ex_gelegentliche- und leichte Raucher

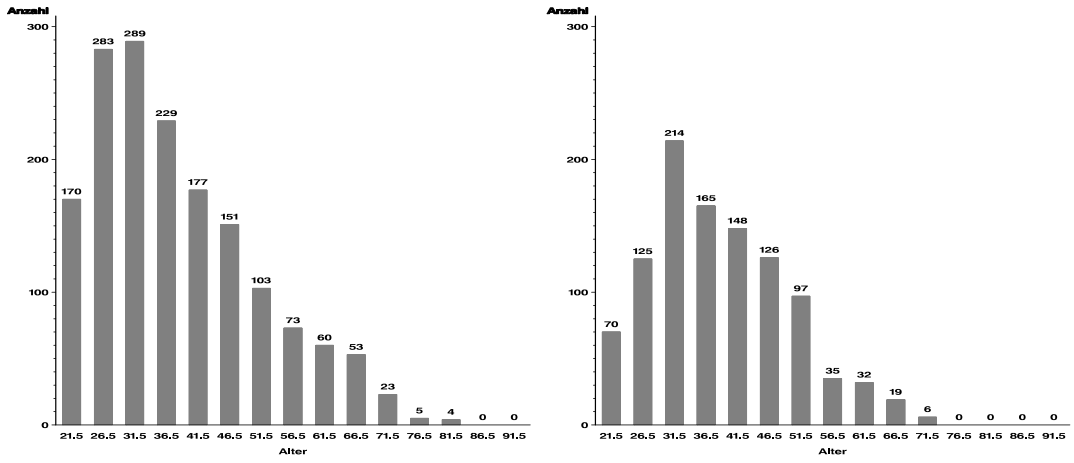


Abbildung 4.16: Mittlere- und starke Raucher

Aus den Histogrammen der Rauchergruppen ist ersichtlich, daß bei den leichten und mittleren Rauchern ab einem Alter von etwa 70 bzw. bei den starken Rauchern schon ab etwa 55 Jahren nur mehr eine geringe Zahl von Probanden zur Auswertung zur Verfügung stehen.

4.2.2 Größenverteilung

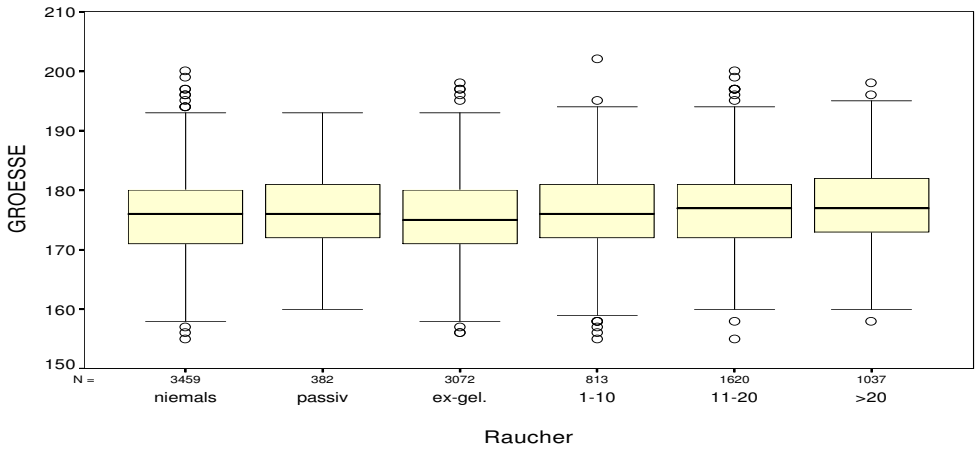


Abbildung 4.17: Boxplot Größenverteilung

Bei der Datenselektion wurden alle Probanden mit einer Größe von unter 1,55 m ausselektiert, um eine Verzerrung der Auswertungen durch extrem kleine Personen zu vermeiden. In den Boxplots ist auch schön die Übereinstimmung der Größenverteilungen zu sehen.

	Größe: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[m]$	1,76	1,76	1,75	1,77	1,77	1,77
$s[m]$	0,07	0,06	0,06	0,07	0,07	0,06

Die Mittelwerte und Standardabweichungen bestätigen die Homogenität der Gruppen untereinander.

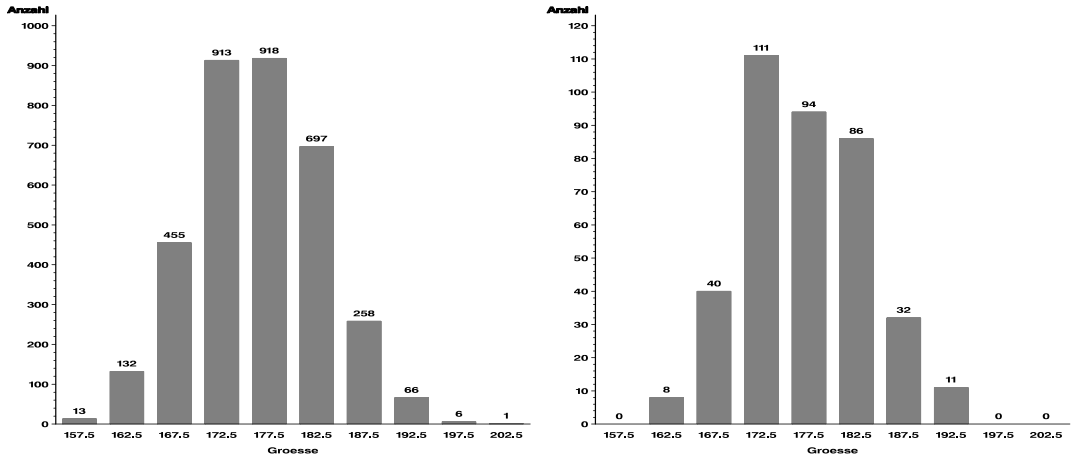


Abbildung 4.18: Niemals- und Passivraucher

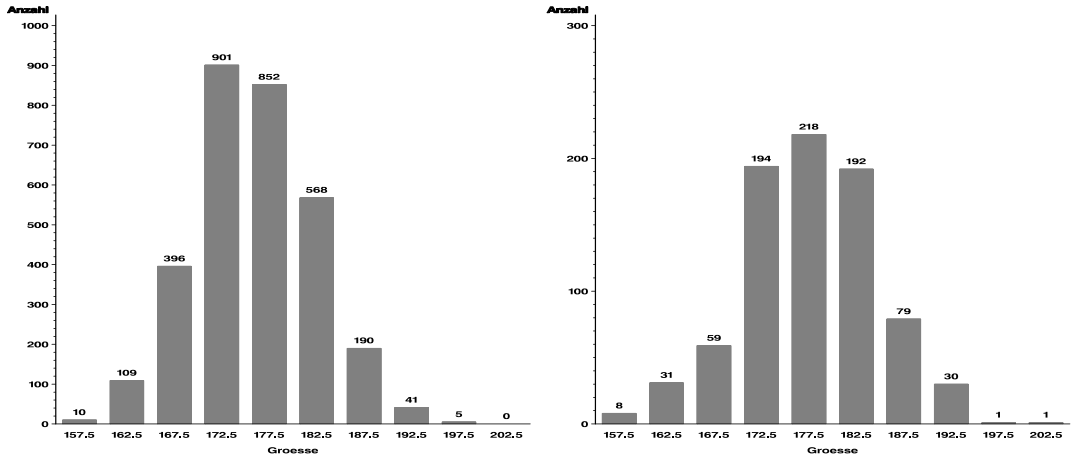


Abbildung 4.19: Ex_gelegentliche- und leichte Raucher

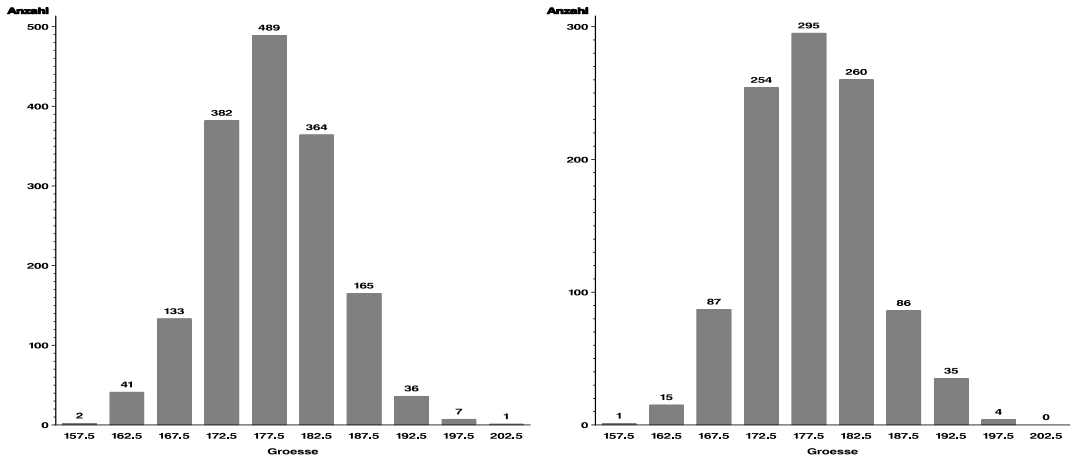


Abbildung 4.20: Mittlere- und starke Raucher

Die symmetrische Verteilung der Größen ist den Histogrammen zu entnehmen.

4.2.3 Gewichtsverteilung

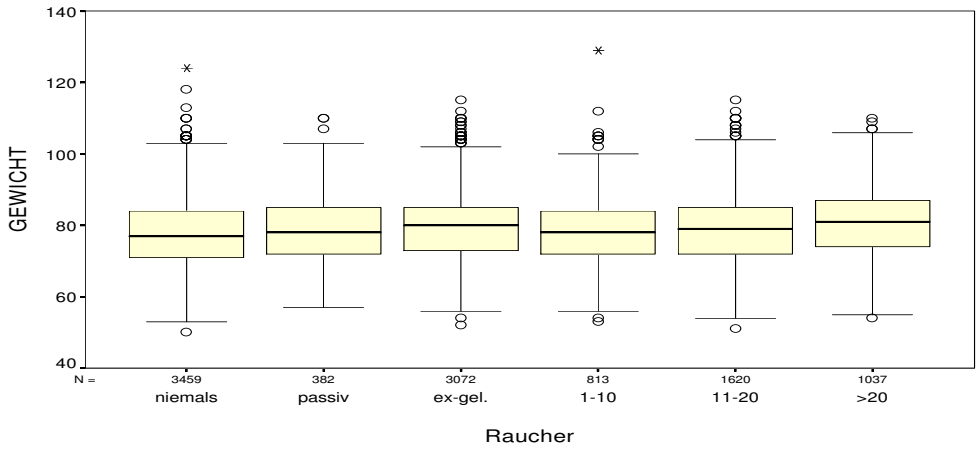


Abbildung 4.21: Boxplot Größenverteilung

	Gewicht: Mittelwert und Standardabweichung					
	nie	passiv	ex-gel	1-10	11-20	>20
$\overline{x}[m]$	77,8	78,7	79,4	78,4	79,2	81.0
$s[m]$	9,0	9,4	9,0	9,4	9,6	9,6

Bei der Datenselektion wurden alle Probanden mit einem Relativgewicht unter 0,8548 und über 1,282 von weiteren Untersuchungen ausgeschlossen. Das Ziel ist eine möglichst große Homogenität der Gruppen untereinander. Trotz dieser Selektion sind, wie im Boxplot ersichtlich, immer noch einige auch extreme Ausreißer vorhanden. Anhand des Boxplots und der Tabelle erkennt man eine gute Übereinstimmung der Gewichtsverteilungen. Zu bemerken ist aber doch, daß das mittlere Gewicht in der Gruppe der Niemalsraucher etwas geringer ist als in den restlichen Gruppen. Insbesondere ist das mittlere Gewicht in der Gruppe der schweren Raucher um mehr als 3 kg höher. Diese Unterschiede sind vielleicht auf die unterschiedlichen Altersverteilungen zurückzuführen. Man kann sagen, daß die Raucher eher jünger und schwerer sind als die Niemalsraucher. Zu erwähnen ist noch, daß solche Unterschiede bei den Frauen nicht erkennbar sind (siehe Abschnitt 4.1.3).

Die Histogramme zeigen im Gegensatz zu den Gewichtsverteilungen bei den Frauen keine Tendenz zu den leichteren Gewichtsklassen. Die Daten sind symmetrisch und als annähernd normalverteilt anzusehen.

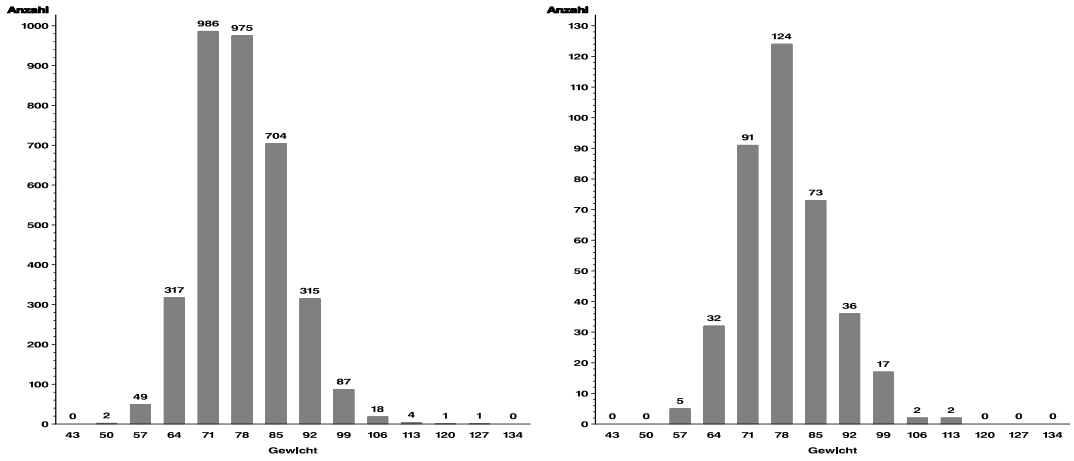


Abbildung 4.22: Niemals- und Passivraucher

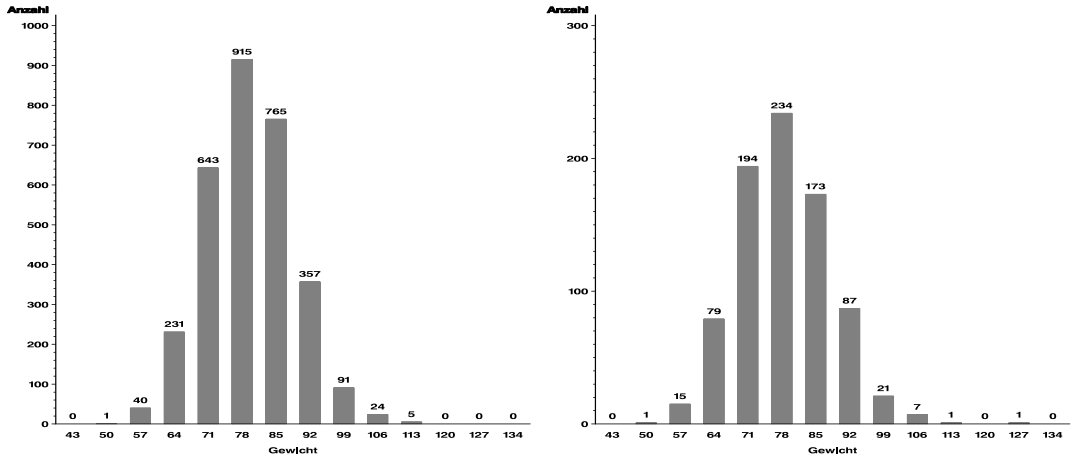


Abbildung 4.23: Ex- gelegentliche Raucher und leichte Raucher

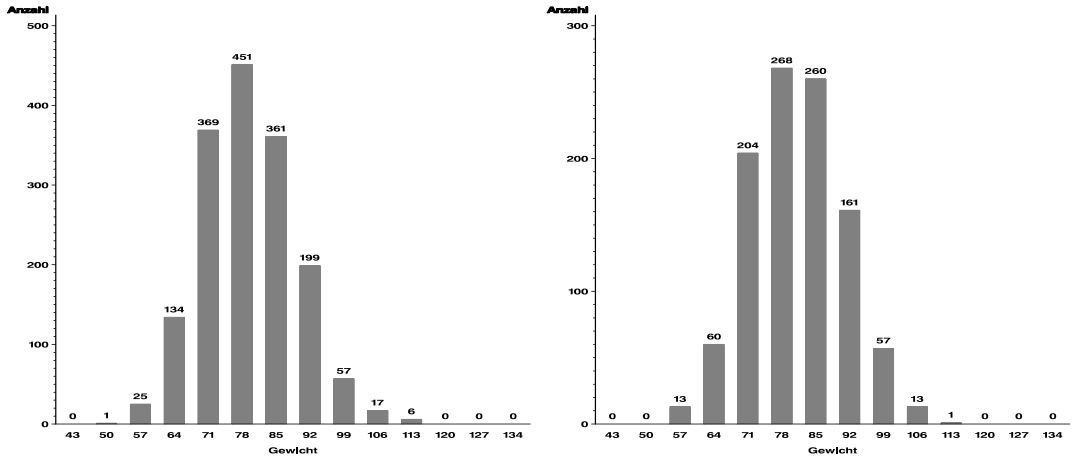


Abbildung 4.24: Mittlere- und starke Raucher

4.3 Kovarianzanalyse

Mit der Kovarianzanalyse wird nun untersucht, ob es zwischen den insgesamt zwölf Untergruppen signifikante Unterschiede bezüglich der verschiedenen Lungenfunktionsparameter gibt. Bekannt ist, daß aufgrund der biologischen Voraussetzungen Unterschiede zwischen den Geschlechtern bestehen. Interessanter ist, ob der Einfluß des Rauchverhaltens auf die Lungenfunktion nachweisbar ist. Insgesamt werden **17.565 Probanden** in die Kovarianzanalyse mit einbezogen.

Um die Mittelwerte der Lungenfunktionsparameter der Gruppen miteinander vergleichen und paarweise Tests auf signifikante Unterschiede durchführen zu können, müssen die bezüglich der möglichen Einflußvariablen Alter, Größe und Gewicht adjustierten Mittelwerte berechnet werden. Ergeben sich signifikante Unterschiede der adjustierten Mittelwerte, dann kann man von einem statistisch nachweisbaren Einfluß des Rauchverhaltens sprechen.

Eine Voraussetzung um die adjustierten Mittelwerte mit dem Verfahren der Kovarianzanalyse richtig interpretieren zu können, ist die Annahme der Parallelität der Regressionskurven. Da diese Voraussetzung zwischen Frauen und Männern nicht erfüllt ist, sind die entsprechenden paarweisen Vergleiche zwischen den adjustierten Mittelwerten kaum interpretierbar. Die Unterschiede zwischen den Geschlechtern werden in den späteren Kapiteln noch genauer spezifiziert.

Die Ergebnisse der Kovarianzanalyse sind in den Abbildungen 4.25 bis 4.36 dokumentiert. Die Abbildungen zeigen die Berechnung der adjustierten Mittelwerte und eine Auflistung der paarweisen Vergleiche auf signifikante Unterschiede, sowie unter ***CELL MEANS*** eine Auflistung der nicht adjustierten Mittelwerte und unter ***ANALYSIS OF VARIANCE*** den von den jeweiligen Variablen erklärten Varianzanteil und deren Signifikanz. Zuerst werden jeweils die Kovariablen in die Analyse aufgenommen, dann die beiden Faktoren und schließlich wird noch eine mögliche Wechselwirkung zwischen den beiden Faktoren untersucht.

Als Faktoren werden SEX und SMOKE sowie als Kovariablen Alter, Größe und Gewicht in die Kovarianzanalyse mit aufgenommen. Die beiden Faktoren SEX (0...Frauen; 1...Männer) und SMOKE (1...Niemalsraucher; 2...Passivraucher; 3...Ex-gel. Raucher; 4...leichte Raucher; 5...mittlere Raucher; 6...schwere Raucher) ergeben zusammen die zwölf Untergruppen. Um den Einfluß von unterschiedlichen Alters-, Größen- und Gewichtsverteilungen zwischen den Gruppen zu eliminieren, werden diese drei Parameter als Kovariablen mit in die Analyse aufgenommen. Aus den vorherigen Abschnitten dieses Kapitels wissen wir, daß zum Teil unterschiedliche Altersverteilungen vorliegen, während die Größen- und Gewichtsverteilungen innerhalb der Geschlechter weitgehend homogen sind. Bei den Parametern Alter, Größe und Gewicht ist vor allem interessant zu untersuchen, wie groß deren Anteil an der Gesamtvarianz ist.

Bei der Betrachtung der adjustierten Mittelwerte muß darauf geachtet werden, daß nicht die absoluten Unterschiede der Mittelwerte zueinander berurteilt werden können. Dies ist bedingt durch die gemeinsame Adjustierung von Frauen und Männern und andererseits kann ein Mittelwert bei mit dem Alter kleiner werdenden Werten, wie sie bei den Lungenfunktionsparametern auftreten, nicht durch einen Mittelwert, berechnet über den gesamten Altersbereich, charakterisiert werden. Wichtig sind die relativen Unterschiede der Mittelwerte innerhalb der Frauen- und Männergruppen untereinander und ob eine Signifikanz dieser Unterschiede vorliegt.

4.3.1 Ergebnisse: FVC und FEV_1

Die Auswertungen dieser beiden Parameter sind in den Abbildungen 4.25 bis 4.27 dokumentiert. Liegt bei den paarweisen Vergleichen der p-Wert unter 1% so sprechen wir von einem *hoch signifikanten* Unterschied, liegt der p-Wert zwischen 1% und 5% so sprechen wir von einem *signifikanten* Unterschied.

Frauen: Beim Vergleich der nicht adjustierten mit den adjustierten Mittelwerten, ist sofort zu erkennen, welchen Einfluß eine unterschiedliche Altersverteilung auf die Mittelwerte hat. So ist z.B. der nicht adjustierte Mittelwert der Niemalsraucher am kleinsten, während nach der Adjustierung der Mittelwert der schweren Raucher am kleinsten wird. In Folge werden nur mehr die adjustierten Mittelwerte analysiert.

Sowohl bei FVC als auch bei FEV_1 ist der Mittelwert bei den Ex- gel. Raucherinnen am höchsten und *signifikant höher* als jener der Niemalsraucherinnen. Die Mittelwerte der Niemals- und leichten Raucherinnen stimmen in etwa überein. Am niedrigsten ist er jeweils in der Gruppe der schweren Raucherinnen. Der adjustierte Mittelwert der schweren Raucherinnen ist *hoch signifikant kleiner* als jener der Ex-gel. Raucherinnen und *signifikant kleiner* als jene der Niemals- und leichten Raucherinnen.

Männer: Im Gegensatz zu den Frauen, befinden sich die höchsten Mittelwerte jetzt in der Gruppe der Niemalsraucher und bei FVC auch in der Gruppe der leichten Raucher. Bei beiden Parametern haben die schweren Raucher wiederum die *signifikant*

kleinsten Werte und die adjustierten Mittelwerte der Niemalsraucher sind jeweils *hoch signifikant höher* als jene der mittleren und schweren Raucher.

Frauen und Männer: Die Kovarianzanalyse zeigt, daß bei FVC die Größe wichtiger ist als das Alter und das Gewicht *keinen signifikanten* Beitrag zur Erklärung der Varianz liefert. Bei FEV_1 hingegen erklärt das Alter einen größeren Teil der Varianz und der Einfluß des Gewichtes ist *signifikant*. Das Gewicht erklärt aber immer noch einen relativ geringen Teil der Varianz. Die beiden Faktoren SEX und SMOKE sind *hoch signifikant* an der Erklärung der Varianz beteiligt, wobei allerdings die Unterscheidung in Frauen und Männern den größten Teil erklärt. Anhand der *signifikanten Wechselwirkung* zwischen SEX und SMOKE ist auf einen *nicht parallelen* Verlauf der Abnahme der beiden Parameter FVC und FEV_1 bzgl. Frauen und Männern zu schließen.

4.3.2 Ergebnisse: PEF, MEF_{75}, MEF_{50} und MEF_{25}

Ein Vergleich der nicht adjustierten mit den adjustierten Mittelwerten zeigt die Notwendigkeit der Adjustierung, um die Mittelwerte miteinander vergleichen zu können. Es werden wieder nur die adjustierten Mittelwerte analysiert.

Frauen: Analog zu FVC und FEV_1 ist auch bei den vier Flußparametern (außer MEF_{25}) der Mittelwert der Ex-gel. Raucherinnen am größten, aber *nicht signifikant* im Vergleich zu den Niemalsraucherinnen. Bei MEF_{25} ist der Mittelwert bei den Niemalsraucherinnen am größten. Der kleinste Mittelwert ist bei allen Parametern in der Gruppe der schweren Raucherinnen zu finden. Zwischen den Mittelwerten der Niemals- und schweren Raucherinnen bestehen durchwegs *hoch signifikante* Unterschiede.

Männer: Die höchsten Mittelwerte befinden sich hier in den Gruppen der Niemals- und Passivraucher. Bei PEF ist allerdings der Mittelwert der Niemalsraucher kleiner als jener der Passiv- und *hoch signifikant kleiner* als jener der Ex-gel. Raucher. Der Mittelwert der schweren Raucher ist bei allen vier Flußparametern am kleinsten und außer zu den mittleren Rauchern für PEF, MEF_{75} und MEF_{50} *höchst signifikant kleiner* als diejenigen der anderen Gruppen.

Frauen und Männer: Hier zeigt sich, daß neben dem Alter und der Größe auch das Gewicht einen *signifikanten* Beitrag zur Erklärung der Varianz liefert. Bei PEF ist die Größe die wichtigste Kovariable, bei MEF_{75} sind Alter und Größe ungefähr gleich wichtig und bei MEF_{50} und MEF_{25} ist das Alter die bedeutendste Kovariable. Eine ähnliche Charakeristik zeigen die Faktoren SEX und SMOKE. Von PEF bis MEF_{25} wird der größte Teil der Varianz durch die Unterscheidung in Frauen und Männern erklärt, wobei der Faktor SMOKE immer wichtiger wird. Bei allen vier Flußparametern bestehen zudem Wechselwirkung zwischen den Faktoren SEX und SMOKE. Was wiederum bedeutet, daß die Unterschiede zwischen Männern und Frauen nicht in allen Rauchergruppen gleich groß sind.

4.3.3 Zusammenfassende Betrachtungen

Frauen: Die höchsten Mittelwerte befinden sich in der Gruppe der Ex-gel. Raucher. Bei FVC und FEV_1 lassen sich die sechs Rauchergruppen in eine Hälfte mit höheren Mittelwerten, Niemals-, Ex_gel.- und leichte Raucher, sowie in eine Hälfte mit kleineren Mittelwerten, Passiv-, mittlere und schwere Raucher, einteilen. Bei den Flußparametern befinden sich die höheren Mittelwerte in den Gruppen der Niemals-, Passiv- und Ex-gel. Raucher.

Die Gruppe der schweren Raucher hat bei allen sechs Lungenfunktionsparametern die kleinsten Mittelwerte. Das ist ein Hinweis auf den negativen Einfluß des Rauchens auf die Lunge. Bemerkenswerterweise sind die Mittelwerte bei den Ex-gel. Raucher durchwegs höher als jene der Niemalsraucher; bei FVC und FEV_1 sogar *signifikant höher*.

Männer: Die kleinsten Mittelwerte befinden sich auch hier in der Gruppe der schweren Raucher, was wiederum auf den negativen Einfluß des Rauchens hinweist. Die höchsten Mittelwerte verteilen sich mit einzelnen Abweichungen auf die Gruppen der Niemals-, Passiv-, Ex_gel.- und leichten Raucher.

Frauen und Männer: Bei den Volumsparametern FVC und FEV_1 sind die Größe und das Alter die entscheidenden Kovariablen. Bei den Flußparametern ist neben Größe und Alter auch das Gewicht *signifikant* an der Erklärung der Varianz beteiligt. Besonders bei MEF_{25} ist auch der Faktor SMOKE wichtig. Sowohl bei Frauen und Männern schneiden die schweren Raucher am schlechtesten ab. Zwischen den Gruppen der Niemals-, Passiv-, Ex_gel.- und leichten Raucher ist aufgrund der Mittelwerte keine klare Unterscheidung zu treffen. Bei den Passiv- und leichten Rauchern ist bei einigen Parametern ein stärkerer Abfall im Vergleich zu den Niemals- und Ex-gel. Rauchern festzustellen.

4.3.4 Auswertungen: FVC

* * * C E L L M E A N S * * *

FVC
by SEX
SMOKE

SEX	SMOKE 1	2	3	4	5	6
0	3,40 (4124)	3,48 (411)	3,78 (1006)	3,90 (723)	3,82 (760)	3,79 (158)
1	4,92 (3459)	5,07 (382)	4,72 (3072)	5,11 (813)	5,12 (1620)	5,09 (1037)

* * * A N A L Y S I S O F V A R I A N C E * * *

FVC
by SEX
SMOKE
with ALTER
GROESSE
GEWICHT

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	15827,973	3	5275,991	15450,699	,000
ALTER	2847,019	1	2847,019	8337,473	,000
GROESSE	3572,024	1	3572,024	10460,644	,000
GEWICHT	1,079	1	1,079	3,159	,076
Main Effects	765,887	6	127,648	373,816	,000
SEX	737,036	1	737,036	2158,405	,000
SMOKE	22,866	5	4,573	13,392	,00
2-Way Interactions	7,559	5	1,512	4,427	,001
SEX SMOKE	7,559	5	1,512	4,427	,001
Explained	16601,419	14	1185,816	3472,652	,000
Residual	5992,845	17550	,341		
Total	22594,263	17564	1,286		

Abbildung 4.25: FVC: Kovarianzanalyse

General Linear Models Procedure
Least Squares Means

Z	FVC LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	LSMEAN Number
F1: nie	4.02033883	0.01071835	0.0001	1
F2: pas	3.92909634	0.02936009	0.0001	2
F3: ex-	4.09564700	0.01912179	0.0001	3
F4: 1-1	4.02509675	0.02259912	0.0001	4
F5: 11-	3.97391658	0.02202323	0.0001	5
F6: >20	3.90671663	0.04677870	0.0001	6
M1: nie	4.65371199	0.01052587	0.0001	7
M2: pas	4.58805780	0.03016203	0.0001	8
M3: ex-	4.61941373	0.01130111	0.0001	9
M4: 1-1	4.66156869	0.02090286	0.0001	10
M5: 11-	4.56216142	0.01517421	0.0001	11
M6: >20	4.53693029	0.01878220	0.0001	12

Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

i/j	2	3	4	5	6	7	8	9	10	11	12
1	0.0026	0.0003	0.8421	0.0467	0.0168	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	.	0.0001	0.0080	0.2112	0.6827	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3		.	0.0135	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
4			.	0.0919	0.0211	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
5				.	0.1885	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
6					.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
7						.	0.0375	0.0186	0.7305	0.0001	0.0001
8							.	0.3247	0.0426	0.4360	0.1440
9								.	0.0691	0.0017	0.0001
10									.	0.0001	0.0001
11										.	0.2780
12											.

Abbildung 4.26: FVC: Adjustierte Mittelwerte

4.3.5 Auswertungen: FEV_1

* * * C E L L M E A N S * * *

FEV1
by SEX
SMOKE

SEX	SMOKE 1	2	3	4	5	6
0	2,86 (4124)	2,95 (411)	3,19 (1006)	3,31 (723)	3,19 (760)	3,15 (158)
1	4,09 (3459)	4,24 (382)	3,87 (3072)	4,22 (813)	4,20 (1620)	4,13 (1037)

* * * A N A L Y S I S O F V A R I A N C E * * *

FEV1
by SEX
SMOKE
with ALTER
GROESSE
GEWICHT

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	11206,746	3	3735,582	14676,244	,000
ALTER	3002,486	1	3002,486	11796,079	,000
GROESSE	1986,452	1	1986,452	7804,315	,000
GEWICHT	1,608	1	1,608	6,317	,012
Main Effects	554,010	6	92,335	362,763	,000
SEX	512,678	1	512,678	2014,195	,000
SMOKE	64,867	5	12,973	50,970	,00
2-Way Interactions	7,438	5	1,488	5,844	,000
SEX SMOKE	7,438	5	1,488	5,844	,000
Explained	11768,194	14	840,585	3302,467	,000
Residual	4467,046	17550	,255		
Total	16235,240	17564	,924		

Abbildung 4.27: FEV_1: Kovarianzanalyse

General Linear Models Procedure
Least Squares Means

Z	FEV1 LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	LSMEAN Number
F1: nie	3.35772280	0.00925383	0.0001	1
F2: pas	3.28183785	0.02534843	0.0001	2
F3: ex-	3.40118704	0.01650906	0.0001	3
F4: 1-1	3.33587169	0.01951126	0.0001	4
F5: 11-	3.24720141	0.01901406	0.0001	5
F6: >20	3.18436840	0.04038703	0.0001	6
M1: nie	3.89473954	0.00908766	0.0001	7
M2: pas	3.84867864	0.02604080	0.0001	8
M3: ex-	3.82968177	0.00975697	0.0001	9
M4: 1-1	3.85772171	0.01804677	0.0001	10
M5: 11-	3.72759871	0.01310086	0.0001	11
M6: >20	3.67113192	0.01621587	0.0001	12

Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

i/j	2	3	4	5	6	7	8	9	10	11	12
1	0.0037	0.0148	0.2892	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	.	0.0001	0.0838	0.2631	0.0392	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3		.	0.0081	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
4			.	0.0007	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
5				.	0.1544	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
6					.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
7						.	0.0909	0.0001	0.0601	0.0001	0.0001
8							.	0.4895	0.7727	0.0001	0.0001
9								.	0.1614	0.0001	0.0001
10									.	0.0001	0.0001
11										.	0.0049
12											.

Abbildung 4.28: FEV_1: Adjustierte Mittelwerte

4.3.6 Auswertungen: PEF

```
                    * * *  C E L L   M E A N S  * * *
              PEF
           by SEX
              SMOKE

        SMOKE
                1          2          3          4          5          6
SEX
      0      7,15       7,29       7,66       7,76       7,60       7,58
          (  4124)   (   411)   (  1006)   (   723)   (   760)   (   158)

      1     10,71      11,15      10,59      10,85      10,69      10,57
          (  3459)   (   382)   (  3072)   (   813)   (  1620)   (  1037)


          * * *  A N A L Y S I S   O F   V A R I A N C E  * * *

               PEF
           by  SEX
               SMOKE
           with ALTER
               GROESSE
               GEWICHT

                                    Sum of                  Mean              Sig
Source of Variation                Squares      DF          Square         F   of F

Covariates                      53361,852       3       17787,284  6647,444  ,000
   ALTER                         5480,083       1        5480,083  2048,011  ,000
   GROESSE                      10644,187       1       10644,187  3977,933  ,000
   GEWICHT                        766,835       1         766,835   286,581  ,000

Main Effects                    11364,210       6        1894,035   707,837  ,000
   SEX                          10631,983       1       10631,983  3973,373  ,000
   SMOKE                          652,192       5         130,438    48,747   ,00

2-Way Interactions                51,443       5          10,289     3,845  ,002
   SEX      SMOKE                 51,443       5          10,289     3,845  ,002

Explained                       64777,505      14        4626,965  1729,184  ,000

Residual                        46960,434   17550           2,676

Total                          111737,939   17564           6,362
```

Abbildung 4.29: PEF: Kovarianzanalyse

```
                    General Linear Models Procedure
                             Least Squares Means

                Z                  PEF          Std Err      Pr > |T|    LSMEAN
                                 LSMEAN           LSMEAN    H0:LSMEAN=0    Number

                F1: nie       8.0297753       0.0300039         0.0001         1
                F2: pas       7.9055565       0.0821877         0.0001         2
                F3: ex-       8.0849189       0.0535276         0.0001         3
                F4: 1-1       7.9108069       0.0632617         0.0001         4
                F5: 11-       7.7888643       0.0616497         0.0001         5
                F6: >20       7.7146961       0.1309477         0.0001         6

                M1: nie      10.3546204       0.0294651         0.0001         7
                M2: pas      10.4612070       0.0844326         0.0001         8
                M3: ex-      10.4624909       0.0316352         0.0001         9
                M4: 1-1      10.2255086       0.0585134         0.0001        10
                M5: 11-       9.8760316       0.0424772         0.0001        11
                M6: >20       9.7529779       0.0525770         0.0001        12

                          Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

 i/j      2        3        4        5        6        7        8        9       10       11       12
  1   0.1425   0.3404   0.0751   0.0002   0.0178   0.0001   0.0001   0.0001   0.0001   0.0001   0.0001
  2        .   0.0612   0.9587   0.2449   0.2130   0.0001   0.0001   0.0001   0.0001   0.0001   0.0001
  3            .        0.0294   0.0002   0.0082   0.0001   0.0001   0.0001   0.0001   0.0001   0.0001
  4                     .        0.1514   0.1723   0.0001   0.0001   0.0001   0.0001   0.0001   0.0001
  5                              .        0.6041   0.0001   0.0001   0.0001   0.0001   0.0001   0.0001
  6                                       .        0.0001   0.0001   0.0001   0.0001   0.0001   0.0001

  7                                                .        0.2275   0.0082   0.0432   0.0001   0.0001
  8                                                         .        0.9885   0.0202   0.0001   0.0001
  9                                                                  .        0.0003   0.0001   0.0001
 10                                                                           .        0.0001   0.0001
 11                                                                                    .        0.0588
 12                                                                                             .
```

Abbildung 4.30: PEF: Adjustierte Mittelwerte

4.3.7 Auswertungen: MEF_{75}

```
                    * * *  C E L L   M E A N S  * * *
               MEF_75
            by SEX
               SMOKE

          SMOKE
                 1          2          3          4          5          6
SEX
        0      6,12       6,28       6,51       6,58       6,42       6,35
           (  4124)   (   411)   (  1006)   (   723)   (   760)   (   158)

        1      8,73       9,02       8,54       8,75       8,65       8,58
           (  3459)   (   382)   (  3072)   (   813)   (  1620)   (  1037)


          * * *  A N A L Y S I S   O F   V A R I A N C E  * * *

               MEF_75
            by SEX
               SMOKE
            with ALTER
               GROESSE
               GEWICHT
```

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	29645,682	3	9881,894	3527,943	,00
ALTER	4385,248	1	4385,248	1565,581	,000
GROESSE	4489,486	1	4489,486	1602,795	,000
GEWICHT	780,057	1	780,057	278,489	,000
Main Effects	5808,404	6	968,067	345,611	,000
SEX	5400,927	1	5400,927	1928,190	,000
SMOKE	483,866	5	96,773	34,549	,00
2-Way Interactions	23,589	5	4,718	1,684	,135
SEX SMOKE	23,589	5	4,718	1,684	,135
Explained	35477,675	14	2534,120	904,708	,000
Residual	49158,168	17550	2,801		
Total	84635,843	17564	4,819		

Abbildung 4.31: MEF_{75}: Kovarianzanalyse

General Linear Models Procedure
Least Squares Means

Z	MEF_75 LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	LSMEAN Number
F1: nie	6.79083185	0.03069793	0.0001	1
F2: pas	6.72991996	0.08408891	0.0001	2
F3: ex-	6.81469402	0.05476586	0.0001	3
F4: 1-1	6.66716303	0.06472513	0.0001	4
F5: 11-	6.53697297	0.06307575	0.0001	5
F6: >20	6.42098803	0.13397677	0.0001	6
M1: nie	8.47909547	0.03014668	0.0001	7
M2: pas	8.48801437	0.08638570	0.0001	8
M3: ex-	8.46407482	0.03236700	0.0001	9
M4: 1-1	8.27638802	0.05986695	0.0001	10
M5: 11-	8.01219301	0.04345977	0.0001	11
M6: >20	7.93654676	0.05379326	0.0001	12

Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

i/j	2	3	4	5	6	7	8	9	10	11	12
1	0.4822	0.6868	0.0706	0.0001	0.0066	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	.	0.3871	0.5449	0.0602	0.0488	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3		.	0.0712	0.0006	0.0060	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
4			.	0.1343	0.0940	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
5				.	0.4281	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
6					.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
7						.	0.9214	0.7189	0.0019	0.0001	0.0001
8							.	0.7929	0.0416	0.0001	0.0001
9								.	0.0047	0.0001	0.0001
10									.	0.0002	0.0001
11										.	0.2561
12											.

Abbildung 4.32: MEF_{75}: Adjustierte Mittelwerte

4.3.8 Auswertungen: MEF_{50}

```
                    * * *  C E L L   M E A N S  * * *
               MEF_50
            by SEX
               SMOKE

          SMOKE
                  1          2          3          4          5          6
SEX
        0      3,99       4,12       4,34       4,40       4,22       4,20
           (  4124)   (   411)   (  1006)   (   723)   (   760)   (   158)

        1      5,32       5,53       5,00       5,32       5,23       5,12
           (  3459)   (   382)   (  3072)   (   813)   (  1620)   (  1037)


          * * *  A N A L Y S I S   O F   V A R I A N C E  * * *

               MEF_50
            by SEX
               SMOKE
            with ALTER
               GROESSE
               GEWICHT
```

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	12564,661	3	4188,220	2210,598	,00
ALTER	5394,276	1	5394,276	2847,170	,000
GROESSE	796,641	1	796,641	420,478	,000
GEWICHT	233,852	1	233,852	123,430	,000
Main Effects	1251,290	6	208,548	110,075	,000
SEX	960,672	1	960,672	507,055	,000
SMOKE	396,339	5	79,268	41,839	,00
2-Way Interactions	22,175	5	4,435	2,341	,039
SEX SMOKE	22,175	5	4,435	2,341	,039
Explained	13838,126	14	988,438	521,710	,000
Residual	33250,398	17550	1,895		
Total	47088,524	17564	2,681		

Abbildung 4.33: MEF_{50}: Kovarianzanalyse

General Linear Models Procedure
Least Squares Means

Z	MEF_50 LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	LSMEAN Number
F1: nie	4.43795177	0.02524700	0.0001	1
F2: pas	4.35885087	0.06915750	0.0001	2
F3: ex-	4.45526123	0.04504126	0.0001	3
F4: 1-1	4.28886638	0.05323209	0.0001	4
F5: 11-	4.14599384	0.05187558	0.0001	5
F6: >20	4.11462535	0.11018693	0.0001	6
M1: nie	5.18999538	0.02479363	0.0001	7
M2: pas	5.14768428	0.07104646	0.0001	8
M3: ex-	5.04948741	0.02661969	0.0001	9
M4: 1-1	5.00054187	0.04923656	0.0001	10
M5: 11-	4.75841775	0.03574276	0.0001	11
M6: >20	4.65270765	0.04424136	0.0001	12

Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

i/j	2	3	4	5	6	7	8	9	10	11	12
1	0.2671	0.7221	0.0080	0.0001	0.0039	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	.	0.2317	0.4117	0.0117	0.0582	0.0001	0.0001	0.0001	0.0001	0.0001	0.0004
3		.	0.0134	0.0001	0.0038	0.0001	0.0001	0.0001	0.0001	0.0001	0.0024
4			.	0.0457	0.1496	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
5				.	0.7944	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
6					.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
7						.	0.5691	0.0001	0.0004	0.0001	0.0001
8							.	0.1904	0.0849	0.0001	0.0001
9								.	0.3702	0.0001	0.0001
10									.	0.0001	0.0001
11										.	0.0537
12											.

Abbildung 4.34: MEF_{50}: Adjustierte Mittelwerte

4.3.9 Auswertungen: MEF_{25}

```
                    * * *  C E L L   M E A N S  * * *
              MEF_25
           by SEX
              SMOKE

          SMOKE
                 1          2          3          4          5          6
SEX
        0      1,45       1,56       1,68       1,83       1,64       1,61
           (  4124)   (   411)   (  1006)   (   723)   (   760)   (   158)

        1      1,99       2,17       1,75       2,05       1,98       1,86
           (  3459)   (   382)   (  3072)   (   813)   (  1620)   (  1037)


          * * *  A N A L Y S I S   O F   V A R I A N C E  * * *

               MEF_25
           by  SEX
               SMOKE
           with ALTER
               GROESSE
               GEWICHT
```

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	4540,970	3	1513,657	3916,408	,00
ALTER	2773,320	1	2773,320	7175,639	,000
GROESSE	229,911	1	229,911	594,867	,000
GEWICHT	4,039	1	4,039	10,451	,001
Main Effects	210,267	6	35,044	90,673	,000
SEX	94,240	1	94,240	243,834	,000
SMOKE	140,659	5	28,132	72,788	,00
2-Way Interactions	6,293	5	1,259	3,257	,006
SEX SMOKE	6,293	5	1,259	3,257	,006
Explained	4757,530	14	339,824	879,253	,000
Residual	6782,918	17550	,386		
Total	11540,448	17564	,657		

Abbildung 4.35: MEF_{25}: Kovarianzanalyse

General Linear Models Procedure
Least Squares Means

Z	MEF_25 LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	LSMEAN Number
F1: nie	1.69287056	0.01140301	0.0001	1
F2: pas	1.66075604	0.03123555	0.0001	2
F3: ex-	1.68851313	0.02034325	0.0001	3
F4: 1-1	1.66738304	0.02404271	0.0001	4
F5: 11-	1.50531275	0.02343003	0.0001	5
F6: >20	1.47708142	0.04976683	0.0001	6
M1: nie	1.93751589	0.01119824	0.0001	7
M2: pas	1.95710355	0.03208871	0.0001	8
M3: ex-	1.84715391	0.01202300	0.0001	9
M4: 1-1	1.87555463	0.02223809	0.0001	10
M5: 11-	1.70676845	0.01614351	0.0001	11
M6: >20	1.60896217	0.01998197	0.0001	12

Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

i/j	2	3	4	5	6	7	8	9	10	11	12
1	0.3185	0.8429	0.3157	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.5104	0.0005
2	.	0.4458	0.8633	0.0001	0.0016	0.0001	0.0001	0.0001	0.0001	0.1995	0.1708
3		.	0.4867	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.4953	0.0067
4			.	0.0001	0.0005	0.0001	0.0001	0.0001	0.0001	0.1840	0.0681
5				.	0.6036	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010
6					.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0146
7						.	0.5595	0.0001	0.0107	0.0001	0.0001
8							.	0.0012	0.0345	0.0001	0.0001
9								.	0.2497	0.0001	0.0001
10									.	0.0001	0.0001
11										.	0.0001
12											.

Abbildung 4.36: MEF_{25}: Adjustierte Mittelwerte

Kapitel 5

Statistische Grundlagen II

Die statistischen Grundlagen zur Analyse der Lungenfunktionsparameter werden in diesem Kapitel zusammengefaßt.

5.1 Korrelationskoeffizient

Um festzustellen, inwieweit verschiedene Parameter voneinander abhängig sind, wurde jeweils der Korrelationskoeffzient berechnet. Die Korrelation gibt in erster Linie den Grad des linearen Zusammenhangs der Parameter wieder.

Um zur Korrelation zu kommen, muß zuerst der Begriff der *Kovarianz* definiert werden:

$$Cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Sie mißt die Stärke des Zusammenhangs zweier Zufallsvariabler X und Y.

Eine normierte Form der Kovarianz stellt die *Korrelation* dar:

$$Corr(X,Y) = \rho = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

Die Korrelation liegt immer zwischen -1 und $+1$. Ist $\rho = 0$, so sind X und Y unkorreliert. In Fall normalverteilter Zufallsvariablen ist dies mit der Unabhängigkeit gleichzusetzen. Ist $\rho = \pm 1$, so besteht ein exakter linearer Zusammenhang zwischen X und Y.

5.1.1 Der Pearson'sche Korrelationskoeffzient

Der Pearson'sche Korrelationskoeffzient eignet sich für metrisch verteilte Zufallsvariable X und Y, wenn man annimmt, daß sie normalverteilt sind, d.h. (X,Y) ist bivariat normalverteilt mit $E(X) = \mu_x$, $E(Y) = \mu_y$, $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$, $Corr(X,Y) = \rho$. Aus

einer Stichprobe von Umfang n: $(X_1, Y_1), \ldots, (X_n, Y_n)$ kann die Korrelation $Corr(X, Y)$ durch den Pearson'schen Korrelationskoeffzienten geschätzt werden:

$$\begin{aligned} R = \frac{S_{XY}^2}{\sqrt{S_X^2 \cdot S_Y^2}} &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2 \cdot \sum_{i=1}^n (Y_i - \overline{Y})^2}} \\ &= \frac{\sum_{i=1}^n X_i Y_i - n \cdot \overline{X} \cdot \overline{Y}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)\left(\sum_{i=1}^n Y_i^2 - n\overline{Y}^2\right)}} \end{aligned}$$

Man kann mit Hilfe des Pearson'schen Korrelationskoeffzienten einen Test der Hypothese

$$H_0 : \rho = 0, \quad \text{gegen} \quad H_1 : \rho \neq 0$$

durchführen, der einen Test auf Unabhängigkeit der Variablen darstellt. Unter der Nullhypothese gilt:

$$T = \frac{R \cdot \sqrt{n-2}}{\sqrt{1 - R^2}} \sim t_{n-2}$$

H_0 wird verworfen, falls $\mid t \mid > t_{n-2;1-\frac{\alpha}{2}}$ (siehe auch Friedl [5]).

5.2 Multiple lineare Regression

In komplexen Situationen, wie in unserem Fall, wo die abhängige Variable Y nicht nur von einer unabhängigen Variablen x, sondern von p unabhängigen Variablen $x_1, \ldots, x_p$ linear abhängt, verwendet man die multiple lineare Regression zur Erstellung eines Modells, welches folgendes Aussehen hat:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

In Matrixnotation: $y = X\beta + \epsilon$
Setzen wir für den $(n \times 1)$-Vektor der Regressanden:

$$y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

Für den $((p+1) \times 1)$-Vektor der unbekannten Parameter:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Für den $(n \times 1)$-Fehlervektor:

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Dabei bezeichnet die Zufallsvariable ϵ_i den nicht beobachtbaren statistischen Fehler, für den angenommen wird: $E(\epsilon_i) = 0, Var(\epsilon_i) = \sigma^2$ (unbekannt und konstant) und $Cov(\epsilon_i, \epsilon_j) = 0$ für $i \neq j$.
Unter der Annahme $Y_i \overset{\text{iid}}{\sim} N(\mu(x_i), \sigma^2)$, mit σ^2 unbekannt, folgt:

$$\epsilon_{\mathbf{i}} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Sowie für die $(n \times (p+1))$ Designmatrix X:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

5.2.1 Normalgleichungen

Die Methode der kleinsten Quadrate besteht darin, die Hyperebene bzw. den Vektor β so zu bestimmen, daß die Summe der Quadrate der Residuen minimiert wird:

$$R(\beta) := \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_p x_{pi})^2 = \min! \tag{5.1}$$

Um den Kleinste-Quadrate-Schätzer für β zu berechnen, ist es notwendig die sogenannten Normalgleichungen zu lösen. Diese ergeben sich, wenn man die partiellen Ableitungen von (5.1) nach β bildet, welche sich in Matrixnotation folgendermaßen darstellen lassen:

$$X^T X \beta = X^T y \tag{5.2}$$

Ist die $((p+1) \times (p+1))$-Matrix $X^T X$ invertierbar, so ist

$$\hat{\beta} := (X^T X)^{-1} X^T y$$

eine Lösung von (5.2). Der Vektor $\hat{\beta} \in \mathbb{R}^{p+1}$ heißt Kleinste-Quadrat-Schätzer.

Eigenschaften von β

- falls $E(y) = X\beta$ folgt: $E(\hat{\beta}) = \beta$
- falls $Var(y) = \sigma^2 I_n$ folgt: $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$
- falls $y \sim N(X\beta, \sigma^2 I_n)$ folgt: $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

Der Vorhersagevektor ist gegeben durch:

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy$$

mit $H = X(X^T X)^{-1} X^T$.

H heißt Hat-Matrix und ist eine symmetrische ($H = H^T$) und idempotente ($H = H^2$) $n \times n$-Matrix. Somit folgt für den Vektor der beobachtbaren Fehler (Residuenvektor) die Darstellung:

$$r = y - \hat{y} = (I_n - H)y$$

Eigenschaften von r

- falls $E(y) = X\beta$ folgt: $E(r) = 0$
- falls $Var(y) = \sigma^2 I_n$ folgt: $Var(r) = \sigma^2 (I_n - H)$
- falls $y \sim N(X\beta, \sigma^2 I_n)$ folgt: $r \sim N(0, \sigma^2 (I_n - H))$

5.2.2 Streuungszerlegung

Um ein Maß für die Güte unserer Vorhersage $\hat{y}$ von y herzuleiten, verwenden wir die folgende Streuungszerlegung der Regressionsanalyse:

$$\begin{aligned} SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_n\right)^2 &= \sum_{i=1}^{n} \left(\left(Y_i - \hat{Y}_i\right) + \left(\hat{Y}_i - \overline{Y_n}\right)\right)^2 \\ &= \sum_{i=1}^{n} \hat{\epsilon}_i^2 + \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y_n}\right)^2 = SSE + \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y_n}\right)^2 \end{aligned}$$

wobei $\overline{Y_n} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ das arithmethische Mittel der Beobachtungen $Y_1, \ldots, Y_n$ ist. Also gilt:

$$Gesamtstreuung\ (SST) = Reststreuung\ (SSE) + erklärte\ Streuung\ (SSR)$$

Die Gesamtstreuung (SST ...total sum of squares) $\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^2$ beschreibt die totale Variabilität in Y und wird in zwei Komponenten zerlegt. In die Regressions-Quadratsumme (SSR ...regression sum of squares) $\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}_n\right)^2$, welche die durch das Modell erklärte Variabilität beschreibt und in die Fehler-Quadratsumme (SSE ...error sum of squares) $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$, welche die durch das Modell nicht erklärte Variabilität beschreibt.

5.2.3 Bestimmtheitsmaá

R^2... **multipler Regressionskoeffizient**

Ausgehend von der oben vorgenommenen Streuungszerlegung läßt sich nun ein Maß für die Güte der Anpassung des Regressionsmodells an die Daten angeben. Das Bestimmtheitsmaß R^2 (coefficient of determination) ist der Quotient aus SSR und SST und läßt sich folgendermaßen motivieren. Der Beitrag aller Variablen im Modell wird durch SSR gemessen und hat als oberes Limit SST. Folglich quantifiziert das Bestimmtheitsmaß den Beitrag der Variablen.

$$R^2 := \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

R^2 liegt immer zwischen 0 und 1. Ein Wert nahe bei 1 spricht für eine gute Anpassung des Modells. R^2 gibt den Anteil der durch das Modell erklärten Varianz an.

Da die Residuenquadratsumme SSR i.a. allein aufgrund einer größeren Anzahl von erklärenden Variablen kleiner werden wird, wächst dementsprechend das Bestimmtheitsmaß . Diesen Umstand berücksichtigt das sogenannte adjustierte Bestimmtheitsmaß R^2_{adj}

$$R^2_{adj} := 1 - \frac{(n-1)}{(n-p)}(1 - R^2)$$

5.2.4 Hypothesentests

Um festzustellen, ob eine Variable x_j einen signifikanten Beitrag zur Erklärung der Variabilität der Daten im Modell leistet, wird die folgende Hypothese geprüft:

$$H_0 : \beta_j = 0, \quad j = 1, \ldots, p \quad gegen \quad H_1 : \beta_j \neq 0$$

Kann H_0 für ein spezielles $1 \leq j \leq p$ verworfen werden, so scheint der Einfluß der j-ten Variablen im Modell sehr gering zu sein. Deswegen sollten auch diese erklärenden Größen nicht in das Modell aufgenommen werden. Unter der Annahme der Nullhypothese gilt für die Verteilung von $\hat{\beta}_j$:

$$\hat{\beta} \sim N\left(0, \frac{\sigma^2}{x_j^T x_j}\right)$$

Weiters gibt es den Zusammenhang:

$$\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

mit $S^2 = \frac{1}{n-p-1}(y - X\hat{\beta})^T(y - X\hat{\beta}) = \frac{SSE}{n-p-1}$ als Schätzer von σ^2.

Aus diesen Voraussetzungen läßt sich nun die Teststatistik T definieren:

$$T = \frac{\hat{\beta}_j\sqrt{x_j^T x_j}/\sigma}{s/\sigma} = \frac{\hat{\beta}_j\sqrt{x_j^T x_j}}{s} = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim t_{n-p-1} \quad (\text{unter } H_0)$$

Mit dem t-Test (siehe 3.5) ist es nun möglich die Relevanz einzelner Komponenten von β zu testen.

Man kann jedoch auch alle Komponenten von β, die zu erklärenden Variablen gehören, auf einmal testen. Dies entspricht einem Test auf Modelladäquatheit. Die Nullhypothese lautet somit:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0 \quad \text{gegen} \quad H_1 : \exists\beta_j, \quad j \in (1, \ldots, p) : \beta_j \neq 0$$

Aus dem Satz von Cochran folgt die Unabhängigkeit der beiden χ^2-verteilten Fehlerquadratsummen SSR mit p Freiheitsgraden und SSE mit $n-p-1$ Freiheitsgraden. Unter der Nullhypothese gilt:

$$F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{MSR}{MSE} \sim F_{p,n-p-1}$$

Ist $F > F_{p,n-p-1;1-\alpha}$ ($1-\alpha$-Quantil der F-Verteilung), so wird H_0 verworfen.

5.2.5 Variablenselektion

Die Variablenselektion erfolgt immer nach der Rückwärts-(backward-) Methode. Dabei fängt man mit der Lösung an, die alle unabhängigen Variablen enthält und schließt dann

jeweils die unabhängige Variable mit dem kleinsten partiellen Korrelationskoeffizienten aus, soweit der zugehörige Regressionskoeffizient nicht signifikant ist (wobei hier ein Signifikanzniveau von $p = 0,01$ zugrunde gelegt wird). Treten gleichzeitig mehrere nicht signifikante $p-$Werte auf, so dürfen die dazugehörigen unabhängigen Variablen trotzdem nicht gleichzeitig aus dem Modell entfernt werden, sondern immer nur jene mit dem kleinsten partiellen Korrelationskoeffizienten. Die Reduktion des Regressionsmodells um eine einzige Einflußgröße kann nämlich die anderen Koeffizienten wesentlich beeinflussen.

Daraus, daß eine oder mehrere unabhängige Variablen aus dem Modell entfernt werden, ist nicht zu schließen, daß diese Variablen keinen Einfluß auf die abhängige Variable haben. Richtig ist, daß die vollständige Menge der unabhängigen Variablen den linearen Zusammenhang nicht wesentlich besser beschreibt als die reduzierte Menge der unabhängigen Variablen.

5.2.6 Analyse der Residuen

Zur Beurteilung der Voraussetzungen müssen wir stets prüfen, ob die Residuen annähernd als Stichprobe einer Normalverteilung gelten können. Die optische Kontrolle dieser Voraussetzungen erfolgt mittels standardisiertem Residuenplot, Histogramm der stand. Residuen und Normal Probability Plot. Darüberhinaus wird noch ein Test auf Normalverteilung der stand. Residuen durchgeführt. Für Stichprobenumfänge ≤ 2000 wird der Test nach $Shapiro-Wilk$ (siehe [18] und [19]) und für Stichprobenumfänge > 2000 wird die übliche $Kolmogorov-Smirnov-Statistik$ (siehe [5] und [6]) berechnet.

Test nach Shapiro-Wilk

$Z_{(.)} = (Z_{(1)}, \ldots, Z_{(n)})$ sei eine geordnete Stichprobe vom Umfang n aus einer Standard-Normalverteilung. $c^T = \left(E\left(Z_{(1)}\right), \ldots, E\left(Z_{(n)}\right)\right)$ sei der Vektor der Erwartungswerte der Ordnungsstatistiken $Z_{(i)}$. Angenommen $X^T = (X_1, \ldots, X_n)$ ist der Vektor der zu testenden zufälligen Stichprobe und $X_{(.)} = \left(X_{(1)}, \ldots, X_{(n)}\right)$ die zugehörige geordnete Stichprobe. V sei die Kovarianzmatrix mit $Cov(X_i, X_j) = v_{ij} \quad i, j = 1, \ldots, n$.

Das Ziel ist es nun einen Test herzuleiten, der die Hypothese, daß diese Stichprobe einer Normalverteilung mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2 entnommen wurde, überprüft.

Falls die X_i eine normalverteilten Grundgesamtheit entstammen, dann können die X_i folgendermaßen beschrieben werden:

$$X_i = \mu + \sigma Z_i \quad i = 1, \ldots, n.$$

Für symmetrische Verteilungen sind die besten linearen unverzerrten Schätzer für μ und σ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{und} \quad \hat{\sigma} = \frac{c^T V^{-1} x}{c^T V^{-1} c}$$

und $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$ sei der übliche symmetrische unverzerrte Schätzer von $(n-1)\sigma^2$. Die W-Teststatistik auf Normalverteilung ist nun definiert als:

$$W = \frac{(w^T x)^2}{S^2} = \frac{(\sum_{i=1}^{n} w_i X_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2},$$

mit

$$w^T = (w_1, \ldots, w_n) = \frac{c^T V^{-1}}{\sqrt{c^T V^{-1} V^{-1} c}}.$$

W kann als quadrierter Korrelationskoeffizient zwischen den Ordungsstatistiken $X_{(i)}$ und den Gewichten w_i gesehen werden. Die Gewichte w_i sind optimal für den gewichteten Kleinste-Quadrate-Schätzer von σ^2 unter der Hypothese, daß die stand. Residuen normalverteilt sind. Die Werte von W realisieren im Intervall $[0, 1]$. Kleine Werte von W führen zu einer Ablehnung der Annahme der Normalverteilung (siehe Shapiro et al. [19]).

Test nach Kolmogorov-Smirnov

Es seien $X_1, \ldots, X_n$ unabhängig ident verteilte Stichprobenvariable aus einer Grundgesamtheit mit unbekannter stetiger Verteilungsfunktion $F(x)$. Es soll nun überprüft werden, ob die Verteilungsfunktion $F(x)$ der Grundgesamtheit mit einer hypothetischen Verteilungsfunktion $F_0(x)$ übereinstimmt. In unserem Fall entspricht $F_0(x)$ der Normalverteilung $N(\mu_0, \sigma_0^2)$. Ist F_0 diskret, oder sind die Parameter nicht vollständig spezifiziert, so ist der KS-Test konservativ, d.h. das wahre Testniveau $P(H_0 \text{ ablehnen} \mid H_0 \text{ richtig}) < \alpha$. Der Test neigt also eher dazu, H_0 nicht zu verwerfen.

Es wird folgende Hypothese:

$$H_0 : F(x) = F_0(x) \quad \text{für alle} \quad x$$

gegen die Alternative

$$H_1 : F(x) \neq F_0(x) \quad \text{für wenigstens einen Wert von } x$$

mittels der Prüfgröße

$$\sqrt{n}D_n, \quad \text{mit} \quad D_n = \sup_x \mid F_0(x) - S_n(x) \mid,$$

getestet, wobei $S_n(x)$ die empirische Verteilungsfunktion der Beobachtungen $X_1, \ldots, X_n$ bezeichnet. $S_n(x)$ ist folgendermaßen definiert:

$$S_n(x) = \begin{cases} 0 & \text{falls } x < X_{(1)} \\ \frac{i}{n} & \text{falls } X_{(i)} \leq x < X_{(i+1)}, \quad i = 1, \ldots, n-1 \\ 1 & \text{falls } X_{(n)} \leq x \end{cases}$$

Die Größe D_n gibt den größten vertikalen Abstand zwischen hypothetischer und empirischer Verteilungsfunktion an.

Die Hypothese H_0 wird nun zum Niveau α verworfen, wenn gilt:

$$\sqrt{n}D_n \geq d_{n;1-\alpha}$$

wobei die Quantile $d_{n;1-\alpha}$ den entsprechenden Tabellen entnommen werden können.

D_n ist unter H_0 verteilungsunabhängig, falls alle Parameter spezifiziert sind. Bei Schätzung der Parameter aus der Stichprobe geht diese Eigenschaft verloren.

Das Statistikprogramm SAS verwendet für μ_0 und σ_0^2 die aus der Stichprobe geschätzten Werte $\overline{X}$ und S^2. Daher ist der KS-Test nicht mehr direkt anwendbar. Als Teststatistik wird nun

$$\sqrt{n}L_n^{norm}, \text{ mit } L_n^{norm} = \sup_x \mid S_n(x) - \Phi\left(\frac{x - \overline{X}}{S}\right) \mid$$

verwendet, welcher folgende Nullhypohthese zugrundeliegt:

H_0**:** die Verteilung der Grundgesamtheit ist eine Normalverteilung $N(\mu, \sigma^2)$ (mit nicht festgelegten μ und σ).

Die Hypothese H_0 wird verworfen, wenn gilt:

$$\sqrt{n}L_n^{norm} \leq l_{n;1-\alpha}^{norm}$$

wobei für $l_{n;1-\alpha}^{norm}$ eigene Tabellen zur Verfügung stehen (Lilliefors-Test).

5.3 Diagnoseplots

5.3.1 Scatterplot

Zur Untersuchung des Zusammenhangs zwischen zwei Variablen X und Y ist der Scatterplot ein wichtiges Werkzeug. Sein Hauptvorteil besteht darin, daß alle Daten vollständig gezeigt werden. Unter einem Scatterplot zu den Daten $(x_1, y_1), \ldots, (x_n, y_n)$ im $\mathbb{R}^2$ versteht man einen Plot dieser Punkte im (x, y)-Koordinatensystem.

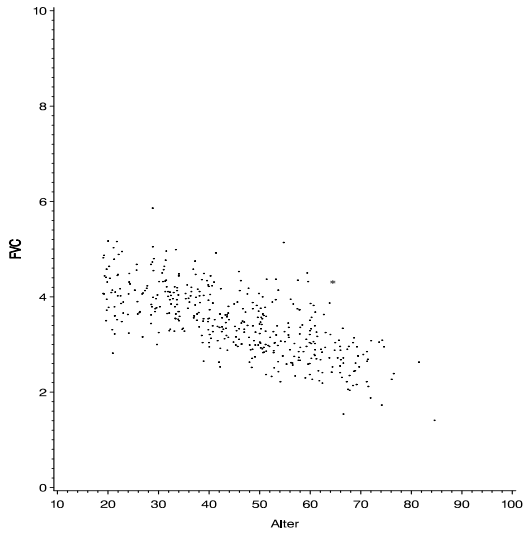


Abbildung 5.1: Scatterplot

Sind $(x_1, y_1), \ldots, (x_n, y_n)$ Realisationen von n unabhängigen Wiederholungen des Zufallsvektors (X, Y), so weist eine gewisse Linearität im Scatterplot auf einen betragsmäßig großen Korrelationskoeffizienten zwischen X und Y hin. Abbildung 5.1 zeigt ein Beispiel für einen Scatterplot.

5.3.2 Medianplot

Bei der Erstellung eines Medianplots wird folgendermaßen vorgegangen. Zuerst werden die Beobachtungen der x-Variablen in geeignete Klassen eingeteilt. Anschließend wird für jede der Klassen der Median $\hat{y}$ der Responsevariablen y berechnet. Beim Plot von y gegen x werden die Mediane als Datenpunkte eingezeichnet und schließlich durch Geraden verbunden. So ergibt sich durch den Plot ein ungefährer Eindruck des Kurvenverlaufs.

Zusätzlich zu den Medianen werden noch zur Wiedergabe der Variation der Beobachtungen die empirische Standardabweichung der Beobachtungen, welche für jede Klasse durch die Formel $\hat{\sigma} = IQR/1,349$ berechnet wird, eingezeichnet. Weiters ist die Modellvariabilität durch die Modellstandardabweichung repräsentiert. Wobei jeweils vom Median die zweifache Standardabweichung subtrahiert $(\hat{y} - 2\hat{\sigma})$ und als parallele Kurve zu den Medianen eingezeichnet wird.

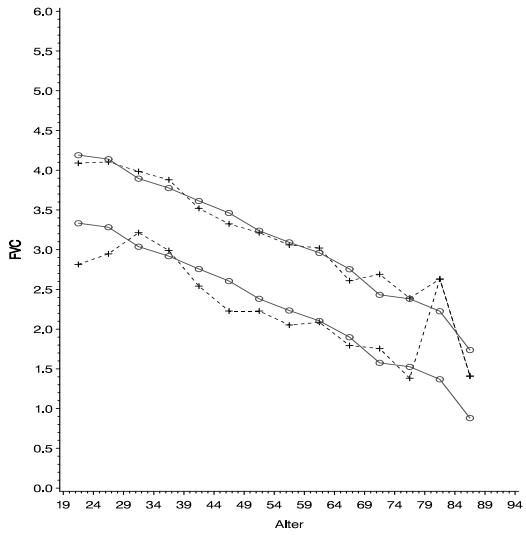


Abbildung 5.2: Medianplot

Abbildung 5.2 zeigt ein Beispiel für einen Medianplot des Parameters FVC gegen das Alter. Der Altersbereich von 19-94 Jahren wird in 15 Klassen mit je 5 Jahrgängen eingeteilt ($[19-24), [24-29), \ldots, [89-94)$).

Die beiden oberen Kurven geben die Mediane der Beobachtungen (+) und die Mediane der über das Modell geschätzten Werte $\hat{y}$ ($\circ$) wieder. Aus der Übereinstimmung dieser beiden Kurven läßt sich die Adäquatheit des verwendeten Modells erkennen. Je besser die Kurven übereinstimmen, desto genauer beschreibt das Modell die Beobachtungen. Die unteren beiden Kurven stehen für die zweifache untere Standardabweichung. Die Kurven der zweifachen unteren Standardabweichung der Beobachtungen (+) sollten wiederum in etwa mit jener der geschätzten Werte $\hat{y}$ ($\circ$) übereinstimmen. Wobei dazu anzumerken ist, daß die Variabilität und damit auch die Modellstreuung der geschätzten Werte aufgrund der Berücksichtigung der Einflußvariablen Größe im Modell im Schnitt etwas geringer sein sollte als jene der Beobachtungen.

5.3.3 Standardisierter Residuenplot

Bei diesem Plot werden jeweils die stand. Residuen r_i^* gegen das Alter geplottet. Gewöhnliche Residuen mit großen h_{ii} haben kleine Varianz und umgekehrt, wobei h_{ii} die Diagonalelemente der *Hat*-Matrix sind (siehe (5.2.1)). Um konstante Varianz der Residuen zu erhalten, werden die gewöhnlichen Residuen

$$r_i = y_i - \hat{y}_i \quad \text{mit} \quad Var(r_i) = \sigma^2(1 - h_{ii})$$

zu

$$r_i^* = \frac{y_i - \hat{y}_i}{s\sqrt{1 - h_{ii}}}$$

mit

$$s^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} r_i^2$$

transformiert. Nun gilt für alle $i = 1, \ldots, n$

$$E(r_i^*) = 0 \quad und \quad Var(r_i^*) = 1$$

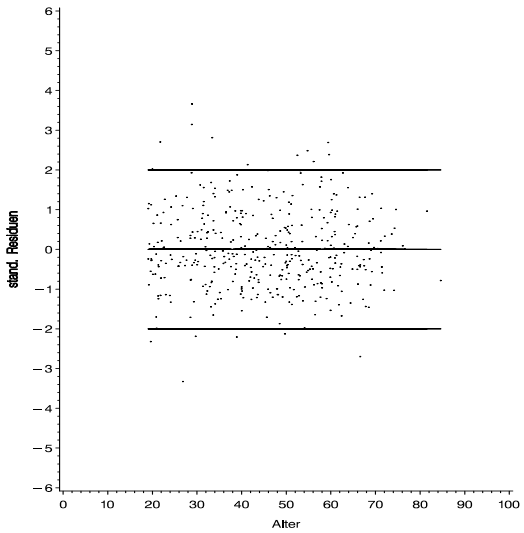


Abbildung 5.3: stand. Residuenplot

Abbildung 5.3 zeigt ein Beispiel für einen stand. Residuenplot. Die stand. Residuen sollten zufällig um die Nulllinie streuen und jeweils gleich viele Residuen (2,28%) sollten über bzw. unter der Linie für die zweifache Standardabweichung (=2) liegen.

5.3.4 Normal Probability Plot

Wir nehmen im folgenden an, daß die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig sind mit identischer Verteilungsfunktion F der Form

$$F(t) = G\left(\frac{t-\mu}{\sigma}\right), \qquad t \in \mathbb{R},$$

wobei G bekannt sei und in unserem Fall der Normalverteilung entspricht. Die Parameter μ und σ hingegen unbekannt. Die Zahl $\mu \in \mathbb{R}$ heißt *Lokationsparameter* und $\sigma > 0$ *Skalenparameter*. Aufgrund der Quantilstransformation (siehe 5.3.5) können wir annehmen, daß die Zufallsvariablen X_i die spezielle Stuktur $X_i = F^{-1}(U_i), \quad i = 1, \ldots, n$

besitzen, wobei $U_1, \ldots, U_n$ unabhängige und auf $(0,1)$ gleichverteilte Zufallsvariablen sind. Da $F^{-1} : (0,1) \to \mathbb{R}$ monoton steigend ist, bleibt die Ordnung der $U_{(i)}$ unter F^{-1} erhalten, d.h. wir erhalten aus dem Lemma 5.3.5 für die Ordnungsstatistiken $X_{(i)}$ die Darstellung:

$$X_{(i)} = F^{-1}(U_{(i)}) = \sigma G^{-1}(U_{(i)}) + \mu, \qquad i = 1, \ldots, n$$

Plotten wir nun $X_{(k)}$ gegen $G^{-1}(k/(n+1))$, d.h. tragen wir im (x,y)-Koordinatensystem die Punkte

$$\left(G^{-1}(k/(n+1)), X_{(k)}\right), \qquad k = 1, \ldots, n$$

ab, so erhalten wir einen *Probability Plot*. Für $G = \Phi$ spricht man von einem *Normal Probability Plot* (*NPP*). Abbildung 5.4 zeigt ein Beispiel für einen *NPP*.

Zusätzlich wird im NPP noch die Gerade $s = \sigma t + \mu \sim St + \overline{X},\ t \in \mathbb{R}$ eingezeichnet. S bezeichnet die Standardabweichung und $\overline{X}$ das arithmetische Mittel der $X_1, \ldots, X_n$. Die eingetragenen Punkte sollten in etwa auf dieser Geraden liegen. Abweichungen von dieser Geraden geben Information über die Schiefe und Exzeß der Verteilung der Daten.

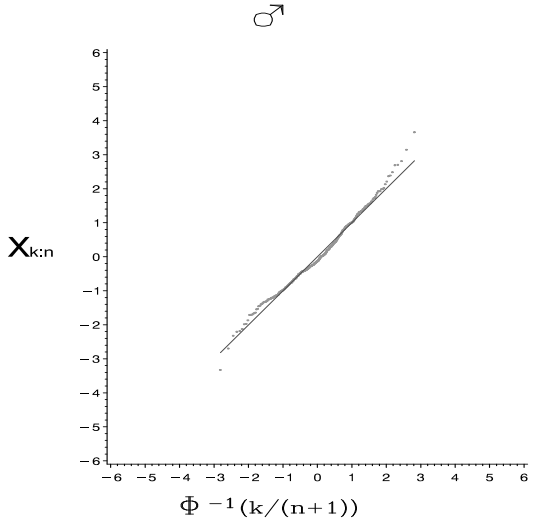


Abbildung 5.4: Normal Probability Plot

Abweichungen können folgendermaßen interpretiert werden:

1. Sind einige wenige Punkte an den Enden der Geraden weit entfernt, so kann dies auf Ausreißer bzgl. der Normalverteilung hinweisen.

2. Krümmungen an den Enden können folgendermaßen gedeutet werden: Zeigt am rechten (linken) Ende die Krümmung nach oben (unten), so hat die empirische Verteilung rechts (links) längere Schwänze (heavy tails) als die theoretische Verteilung. Ist rechts (links) eine Krümmung nach unten (oben), so weist dies auf kürzere Schwänze (light tails) rechts (links) hin.

3. Vergleicht man eine unsymmetrische empirische Verteilung mit einer symmetrischen theoretischen Verteilung, erhält man eine Kurve mit von links nach rechts steigender Krümmung (Daten sind rechtsschief) oder bei entsprechenden Daten umgekehrt.

4. Plateaus oder Sprünge in der Darstellung weisen auf hohe Datenkonzentrationen an einer Stelle oder fehlende Werte über einen größeren Bereich hin.

5.3.5 Quantilstransformation

Definition: Es sei F eine Verteilungsfunktion auf $\mathbb{R}$. Dann heißt

$$F^{-1}(U) := \inf\{t \in \mathbb{R} : F(t) \geq U\}, \quad U \in (0,1),$$

verallgemeinerte Inverse oder *Quantilfunktion* zu F.

Quantilstransformation

Es sei X eine Zufallsvariable mit Verteilungsfunktion F und U eine auf $(0,1)$ gleichverteilte Zufallsvariable. Dann gilt:

Für beliebiges F besitzt die Zufallsvariable $F^{-1}(U)$ die Verteilungsfunktion F:

$$P\left(F^{-1}(U) \leq t\right) = F(t), \qquad t \in \mathbb{R}$$

Dies ist die *Quantilstransformation*.

Lemma

Es sei Y eine Zufallsvariable mit Verteilungsfunktion G; $\sigma > 0$, $\mu \in \mathbb{R}$. Weiter sei F die Verteilungsfunktion zu $X := \sigma Y + \mu$, d.h. $F(t) = G((t-\mu)/\sigma)$, $t \in \mathbb{R}$. Dann gilt für $U \in (0,1)$:

$$F^{-1}(U) = \sigma G^{-1}(U) + \mu$$

5.4 Kovarianzanalyse

Bei der Analyse der Lungenfunktionsparameter wird versucht mit der Kovarianzanalyse (siehe Kapitel 3) Regressionsmodelle für Frauen und Männer gemeinsam zu erstellen. Das Regressionsmodell welches in diesem Zusammenhang verwendet wird hat folgendes Aussehen:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \beta_{p+2} SEX + \sum_{j=1}^{p} \gamma_j x_j SEX + \epsilon \quad i, j = 1, \ldots, p$$

Die Dummy-Variable SEX dient zur Unterscheidung der Frauen- und Männergruppe:

$$SEX = \begin{cases} 0 & \text{Frauengruppe} \\ 1 & \text{Männergruppe} \end{cases}$$

Der Ausdruck $\sum_{j=1}^{p} \gamma_j x_j SEX$ beschreibt Wechselwirkungen bezüglich der erklärenden Variablen zwischen den Gruppen.

Die Erstellung der Regressionsgleichung erfolgt nach der bereits beschriebenen Rückwärts-Methode. Bei der Interpretation der Gleichung ergeben sich drei Möglichkeiten:

1. Die Regressionskurven stimmen überein ($\beta_{p+2} = \gamma_j = 0 \quad j = 1, \ldots, p$
2. Die Regressionskurven sind parallel aber stimmen nicht überein ($\gamma_j = 0 \quad j = 1, \ldots, p$, aber $\beta_{p+2} \neq 0$)
3. Die Regressionkurven sind nicht parallel ($\gamma_j \neq 0, \quad j = 1, \ldots, p$)

Kapitel 6

FVC

In diesem Kapitel werden Modelle für den Parameter FVC bzgl. aller sechs Rauchergruppen erstellt. Frauen und Männer werden immer parallel untersucht. Bei zwei nebeneinander abgebildeten Graphiken, werden in der linken Graphik die Daten bzgl. der Frauen wiedergegeben und in der rechten die Daten bzgl. der Männer. Dementsprechend werden in den einzelnen Untergruppen vor allem Unterschiede zwischen den Geschlechtern verdeutlicht. Die Gruppen der Niemals- und schweren Raucher werden dabei detaillierter analysiert, um vor allem Unterschiede zwischen diesen beiden Gruppen herauszuarbeiten.

Folgende Tabelle gibt den Korrelationskoeffizienten (nach Pearson) zwischen FVC und FEV_1 innerhalb der Untergruppen an:

	FEV_1					
	nie	passiv	ex-gel	1-10	11-20	>20
FVC	0,949	0,957	0,945	0,919	0,931	0,928

Daraus ist ersichtlich, daß zwischen FVC und FEV_1 eine sehr hohe Korrelation besteht. Demzufolge werden ähnliche Modelltypen für FVC und für FEV_1 in Frage kommen.

Das Modell, welches hier verwendet wird, um die Lungenfunktionsparameter zu beschreiben, entspricht in der Wahl der erklärenden Variablen den Modellen der Arbeiten von Kummer [11] und Rapatz [15]. Für FVC sieht das Modell folgendermaßen aus:

$$FVC = \hat{\beta}_0 + \hat{\beta}_1 H + \hat{\beta}_2 AH + \hat{\beta}_3 \ln(A)$$

Die Kurzbezeichnungen für die erklärenden Variablen sind:

AH: Alter×Größe; **H**: Größe; **ln(A)**: ln(Alter)

Die Korrelationskoeffizienten in den folgenden Tabellen zeigen die Stärke des Zusammenhangs zwischen den erklärenden Variablen und FVC bei Frauen und Männern:

	FVC-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
H	0,564	0,613	0,613	0,489	0,555	**0,628**
AH	-0,667	-0,600	-0,600	**-0,508**	-0,555	-0,479
ln(A)	**-0,677**	**-0,622**	**-0,622**	-0,505	**-0,583**	-0,528

Hier zeigt sich, daß der stärkste Zusammenhang mit der Variablen ln(A) gegeben ist. Das ist deswegen auch interessant, weil dadurch die Nichtlinearität, die sich beschleunigende Abnahme der Lungenfunktionswerte mit zunehmendem Alter beschrieben wird. Auffallend sind auch die etwas geringeren Korrelationen in den drei Rauchergruppen, welche teilweise durch die unterschiedlichen Stichprobenumfänge zu erklären sind. Die hohe Korrelation der FVC-Werte der schweren Raucherinnen mit der Größe weist darauf hin, daß diese Werte stark von der Größe beeinflußt sind und beim Vergleich der schweren- mit den Niemalsraucherninnen, die Werte nach der Größe adjustiert werden müssen.

	FVC-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
H	**0,622**	**0,652**	0,605	**0,641**	0,597	0,543
AH	-0,582	-0,471	-0,624	-0,599	-0,588	-0,524
ln(A)	-0,603	-0,513	**-0,650**	-0,613	**-0,609**	**-0,566**

Bei den Männern hat keiner der Faktoren eine dominierende Bedeutung über alle Gruppen hinweg. Welche Auswirkung eine Adjustierung der FVC-Werte nach der Größe hat, ist wie bei den Frauen noch speziell zu untersuchen.

Im folgenden werden nun Regressionsmodelle für den Parameter FVC erstellt. Anhand der Scatterplots wird die Verteilung der Daten beurteilt, die Modellgleichungen für Frauen und Männer angegeben und die Güte der Modelle mit Medianplots, Residuenplots und Normal Probability Plots analysiert. Schließlich werden noch mit dem Verfahren der Kovarianzanalyse gemeinsame Modelle für Frauen und Männer erstellt. Dazu werden zusätzlich der Faktor SEX für Unterschiede zwischen Frauen und Männern und Faktoren, welche Wechselwirkungen mit den bereits im Regressionsmodell vorhandenen Parametern beschreiben eingeführt.

$$FVC = \hat{\beta}_0 + \hat{\beta}_1 H + \hat{\beta}_2 AH + \hat{\beta}_3 \ln(A) + \hat{\beta}_4 SEX + \hat{\beta}_5 HSEX + \hat{\beta}_6 AHSEX + \hat{\beta}_7 \ln(A) SEX$$

Wird nur der Faktor SEX in die neue Regressionsgleichung miteinbezogen, so verlaufen die Kurven der Mediane der geschätzten Werte für Frauen (SEX=0) und Männer (SEX=1) parallel. Werden jedoch auch Wechselwirkungsparameter in das Modell mitaufgenommen, so ist der Verlauf der beiden Modellkurven nicht mehr parallel.

6.1 Niemalsraucher

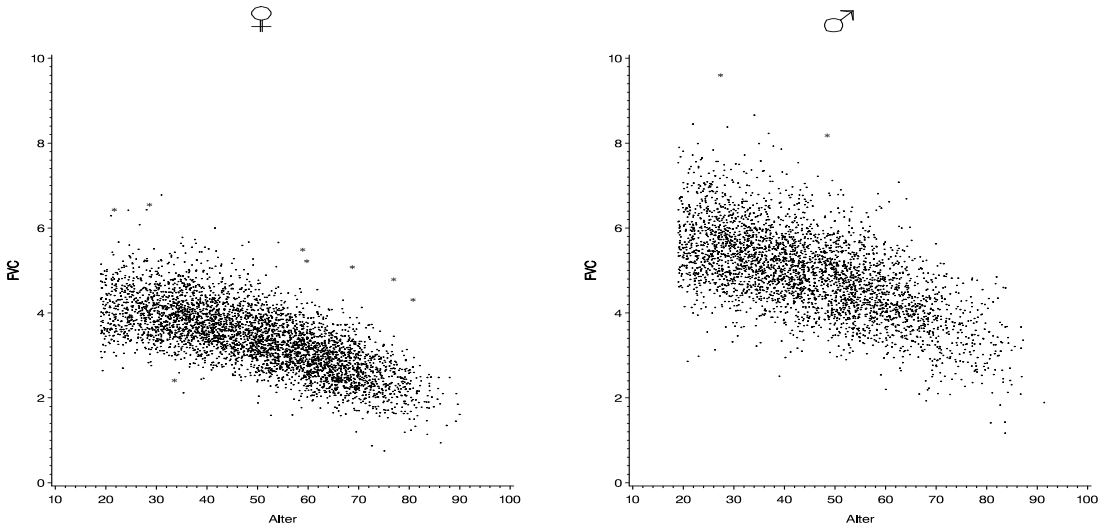


Abbildung 6.1: Scatterplot

Anhand der Scatterplots sieht man, daß die Werte der Frauen eine kompaktere Wolke bilden als jene der Männer. Das gilt, wie im folgenden gezeigt, nicht nur für die Gruppe der Niemalsraucher, sondern in gleicher Weise für die restlichen Untergruppen. Das drückt sich dadurch aus, daß der Schätzer s_e für die Standardabweichung der Residuen bei den Frauen im Schnitt um 25% kleiner ist als der vergleichbare Wert bei den Männern.

In Übereinstimmung mit den Arbeiten von Kummer [11] und Rapatz [15] wird die abhängige Variable vor Modellerstellung nicht transformiert, da die Voraussetzung der konstanten Varianz über den gesamten Altersbereich nicht verletzt ist. Bei den Frauen ist die Variabilität der Daten über den gesamten Altersbereich annähernd konstant, während bei den Männern eine leichte Abnahme derselben festzustellen ist. Wobei in oberen Altersbereichen eine Abnahme der Variabilität auf eine geringe Anzahl von Daten zurückzuführen ist.

Die Erstellung der Modellgleichungen erfolgt in zwei Schritten. Im ersten Schritt wird mit der multiplen linearen Regression ein Modell erstellt und die zugehörigen Residuen berechnet. Alle Werte mit einem stand. Residuum $\mid r_i^\star \mid \geq 4$ werden aus der Datenmenge entfernt. Im Scatterplot sind diese Werte durch Sterne gekennzeichnet. Damit sollen extreme Ausreißer, welche die Koeffizienten des Modells stark beeinflussen, erkannt und von weiteren Analysen ausgeschlossen werden. Im zweiten Schritt wird anhand der reduzierten Datenmenge wiederum mit der multiplen linearen Regression ein Modell erstellt. Diese Vorgangsweise wird im folgenden beibehalten. Die angegebenen Modellgleichnungen entsprechen jenen, welche im zweiten Schritt berechnet werden.

Modellgleichungen

Frauen: $n = 4124$

$$FVC = -10,799 + 6,728H - 0,041AH + 1,690\ln(A)$$
$$R^2 = 0,650 \quad s_e = 0,457$$

Männer: $n = 3459$

$$FVC = -11,273 + 8,075H - 0,034AH + 1,250\ln(A)$$
$$R^2 = 0,580 \quad s_e = 0,639$$

Die Modellgleichung der Frauen zeichnet sich durch einen höheren Bestimmtheitsgrad und eine geringere Modellstreuung aus. Auffällig an den Modellgleichungen ist, daß der Koeffizient von $\ln(A)$ bei den Frauen größer ist, als jener bei den Männern.

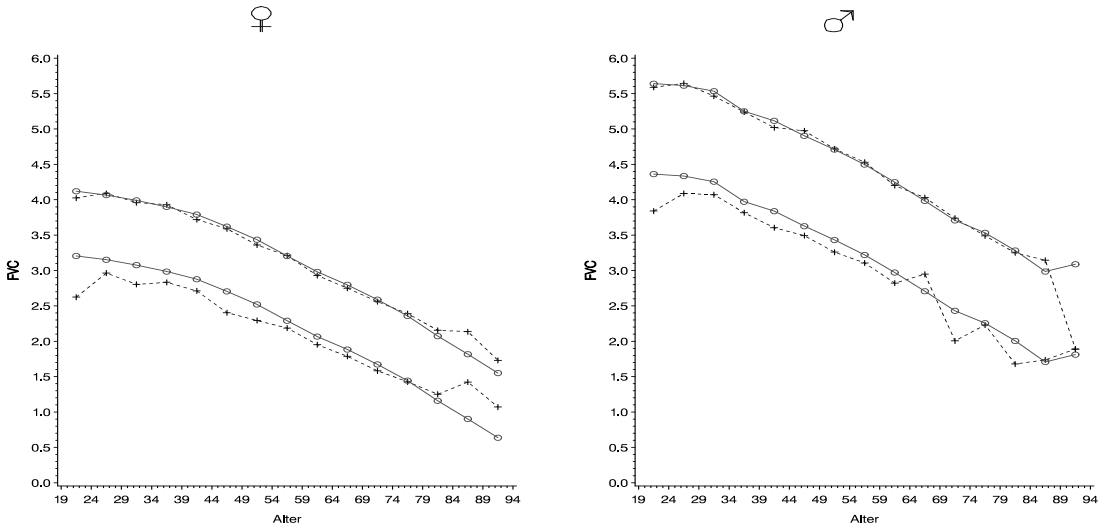


Abbildung 6.2: Medianplot

Die Medianplots der Gleichungen zeigen eine gute Anpassung an die erhobenen Daten. In der ersten Altersklasse [19-24) erfolgt eine leichte Überschätzung. Die Abweichungen in den letzten beiden Altersklassen sind nicht mehr sehr aussagekräftig, da hier nur mehr sehr wenige Daten in die Berechnung mit einbezogen werden. Weil die FVC-Werte nicht nach der Größe standardisiert sind, und auch keine spezielle Größenklasse ausgewählt wurde, ist die empirische Standardabweichung tendenziell bei allen Gruppen größer als die Modellstreuung. Während bei den Frauen die Mediane eine leichte Bogenform nachvollziehen, kann bei den Männern ab dreißig eine näherungsweise konstante Abnahme festgestellt werden. Bei den Frauen nimmt die FVC zwischen 30 und 55 um 2,8 ml/Jahr ab und zwischen 55 und 80 um 3,6 ml/Jahr. Das ergibt zwischen 30 und 80 eine Abnahme von 3,2 ml/Jahr. Bei den Männern erfolgt zwischen 30 und 80 ein Abnahme von 4,5 ml/Jahr.

Die stand. Residuenplots zeigen annähernd konstante Varianz über den gesamten Altersbereich.

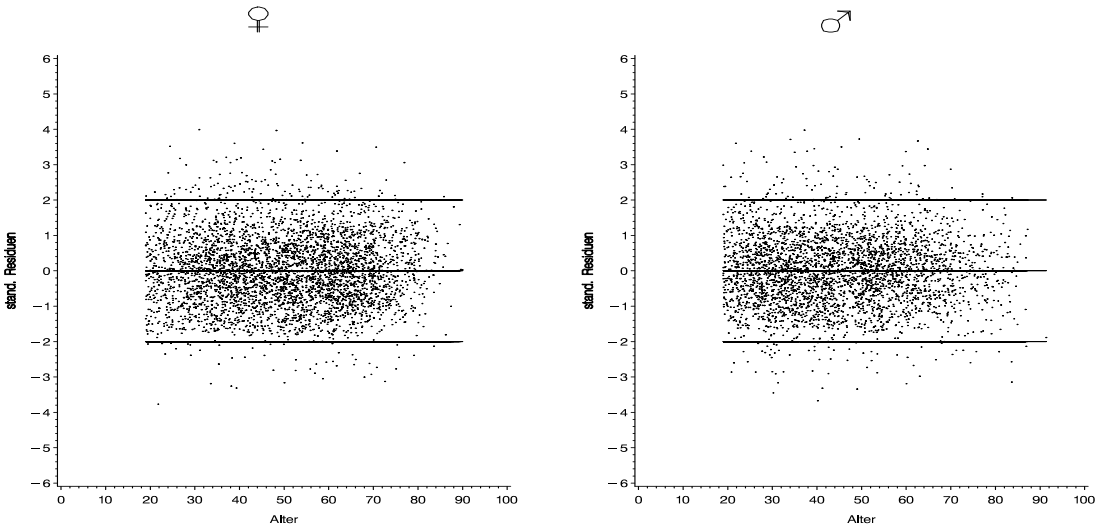


Abbildung 6.3: stand. Residuenplot

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2, 28\%)$	$< -2s_e(2, 28\%)$	K-S
Frauen	4116	51,65%	2,98%	1,17%	0,001
Männer	3457	50,30%	2,92%	1,82%	0,0175

Bei Erfüllung der Modellbedingungen sollten 50% der Residuen > 0 und jeweils 2,28% der Residuen $> 2s_e$ bzw. $< -2s_e$ sein. Schließlich ist noch das Ergebnis des Tests auf Normalverteilung angeführt. Bei mehr als 2000 Daten wird der Test nach Kolmogorov - Smirnov (was hier der Fall ist) und bei 2000 oder weniger Daten wird der Test nach Shapiro - Wilk ausgeführt, wobei der zugehörige p-Wert jeweils in der letzten Spalte angeführt ist.

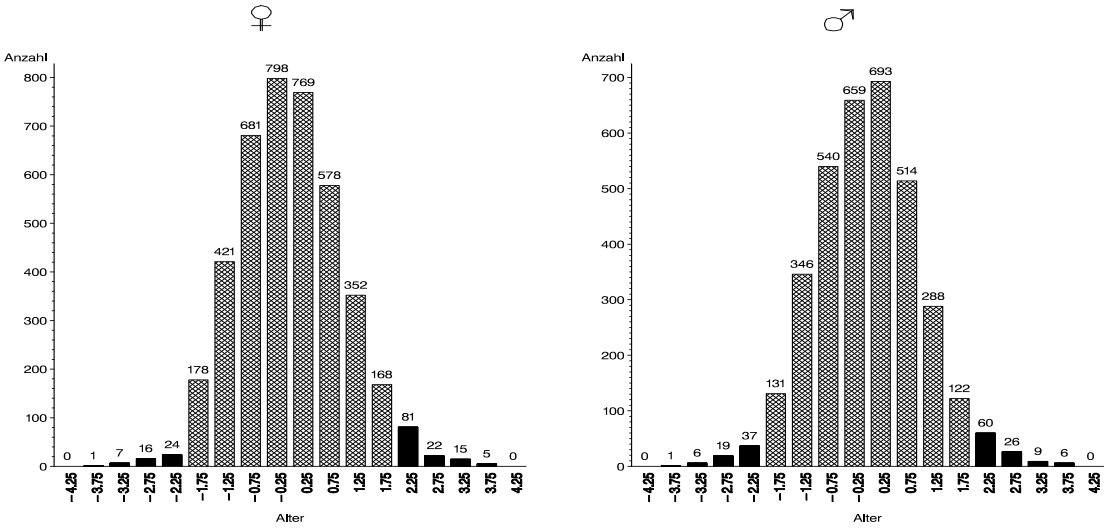


Abbildung 6.4: Histogramm der stand. Residuen

Man erkennt, daß die Residuen leicht rechtsschief verteilt sind. Insbesondere sind zuwenige Werte kleiner als $< -2s_e$. Die sich daraus ergebende Abweichung von der Normalverteilung ist im Normal Probability Plot klar ersichtlich. Bei einem Signifikanzniveau

von $\alpha \geq 2\%$ müßte die Annahme der Normalverteilung abgelehnt werden. Aufgrund der großen Anzahl von Probanden und der wenigen Werte, die für die extreme Abweichung verantwortlich sind, können die Abweichungen noch akzeptiert werden.

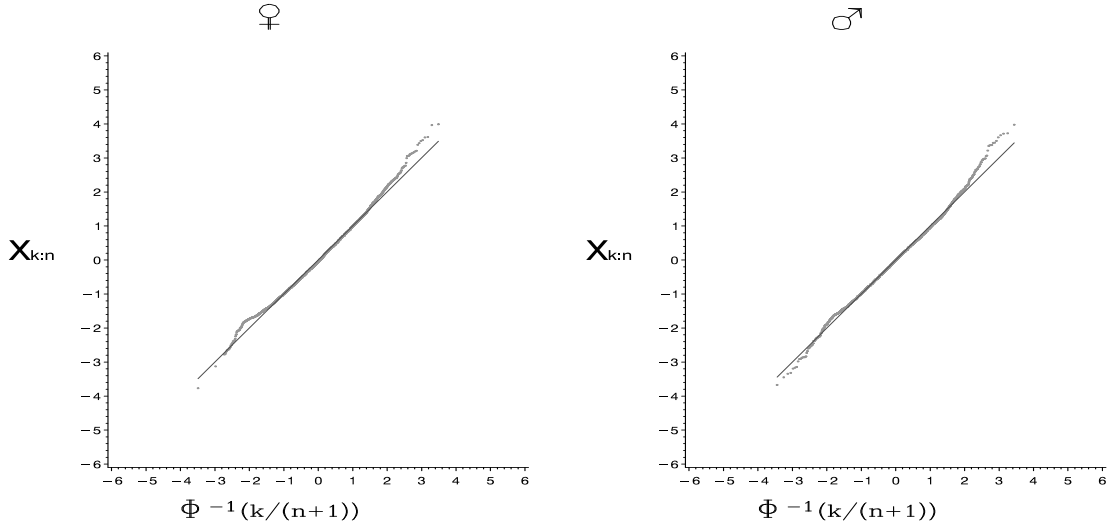


Abbildung 6.5: Normal Probability Plot der stand. Residuen

Modellgleichung Kovarianzanalyse: $n = 7583$

$$\begin{aligned} FVC &= -11,001 + 6,796H - 0,041AH + 1,721\ln(A) \\ &\quad +1,217HSEX + 0,008AHSEX - 0,532\ln(A)SEX \end{aligned}$$

$$R^2 = 0,777 \quad s_e = 0,545$$

Insgesamt werden 7583 Probanden in die Auswertung miteinbezogen. Interessanterweise werden die Unterschiede zwischen den Geschlechtern ausschließlich durch Wechselwirkungsterme beschrieben. Wobei die Männer vor allem aufgrund ihrer Körpergröße höhere FVC-Werte erreichen. Die erklärende Variable $\ln(A)$ mit Koeffizienten -0,532 beschreibt die eher lineare Abnahme der FVC-Werte bei den Männern.

6.2 Passivraucher

Die Werte bei den Männern streuen wiederum stärker als jene der Frauen. Der Altersbereich bei den Frauen erstreckt sich bis über 75 Jahre, während bei den Männern nur bis 65 Jahre genügend Werte vorliegen.

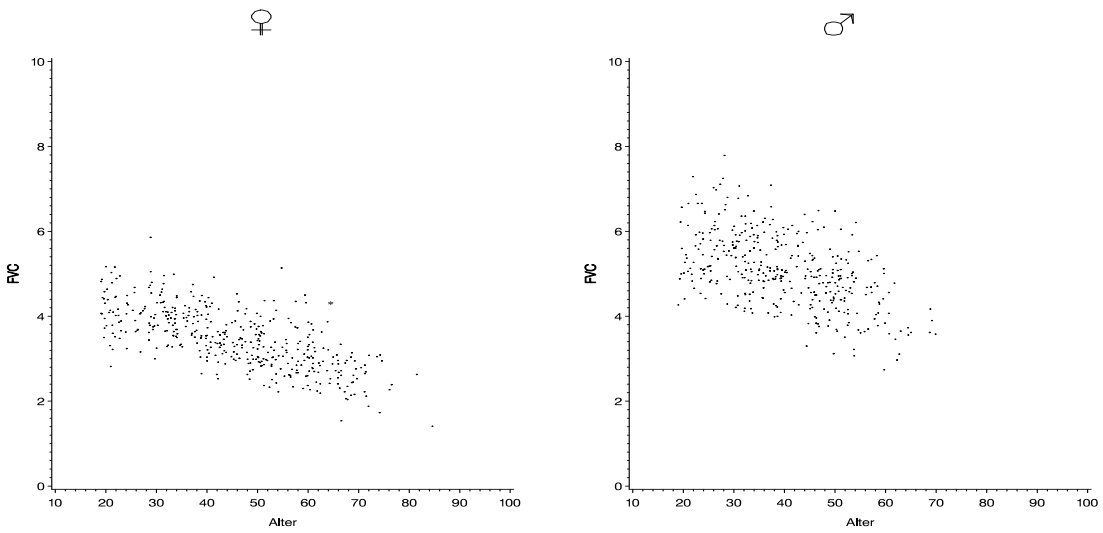


Abbildung 6.6: Scatterplot

Modellgleichungen

Frauen: $n = 411$

$$FVC = -4,383 + 5,584H - 0,017AH$$

$$R^2 = 0,634 \quad s_e = 0,428$$

Männer: $n = 382$

$$FVC = -13,144 + 8,405H - 0,042AH + 1,746\ln(A)$$

$$R^2 = 0,530 \quad s_e = 0,589$$

Die Modellgleichungen unterscheiden sich hier vor allem im Intercept und dadurch, daß bei den Frauen die erklärenden Variablen $\ln(A)$ nicht in das Modell mit aufgenommen wird. Es kann bei den Frauen gemäß dieses Modells von einer linearen Abnahme der Werte mit zunehmendem Alter ausgegangen werden. Der Bestimmtheitsgrad ist bei den Frauen größer und die Modellstreuung kleiner als bei den Männern.

Die Medianplots zeigen in beiden Fällen eine gute Anpassung an die Daten. Lediglich die Streuung der Werte bei den Männern wird nur ungenügend wiedergegeben. Was von den Modellgleichungen nur schlecht nachvollzogen wird ist, daß bei den Frauen und noch in viel stärkerem Maße bei den Männern bis dreißig ein Anstieg der körperlichen Leistungsfähigkeit und somit auch der Lungenkapazität erfolgt. In diesem Bereich erfolgt durch die Modelle eine Überschätzung der FVC-Werte.

Sowohl bei den Frauen, als auch bei den Männern ist die Abnahme der FVC über den Altersbereich annähernd konstant. Aufgrund des vergleichsweise großen Koeffizienten der Variablen $\ln(A)$ hätte man bei den Männern eigentlich einen anderen Verlauf des Medianplots erwartet. Bei den Frauen beträgt die Abnahme zwischen 30 und 75 Jahren im Schnitt 3,67 ml/Jahr, bei den Männern zwischen 30 und 65 Jahren im Schnitt 4,86 ml/Jahr.

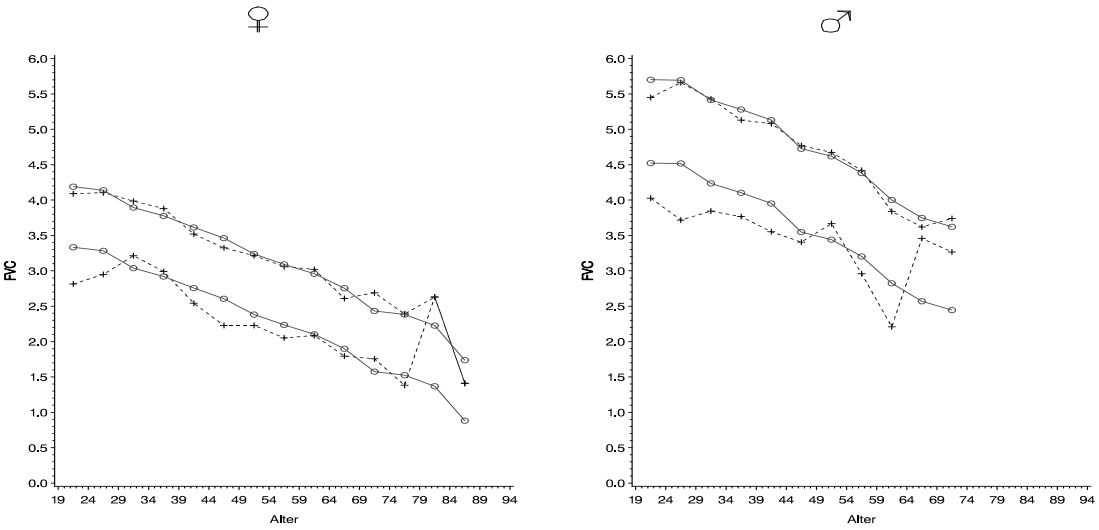


Abbildung 6.7: Medianplot

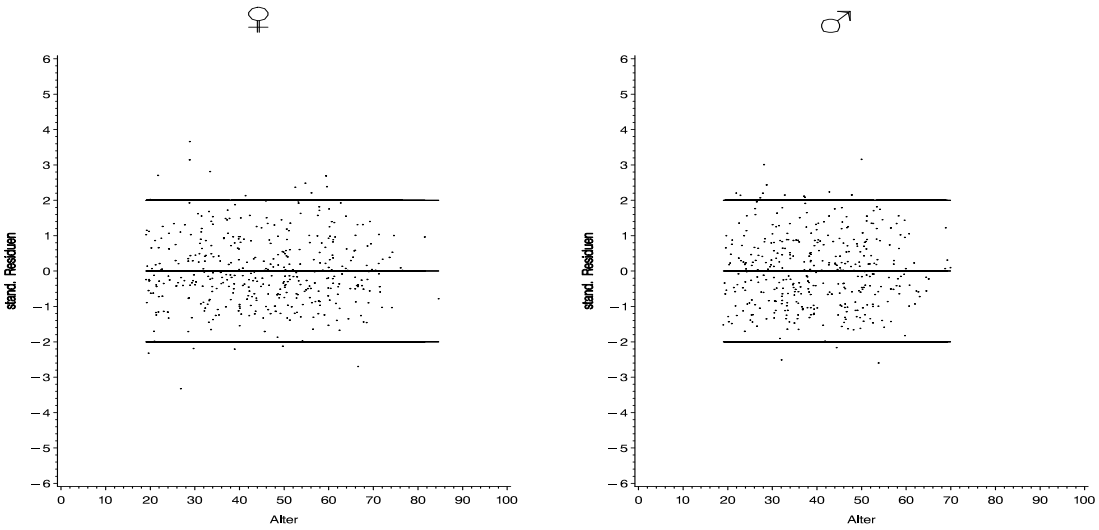


Abbildung 6.8: stand. Residuenplot

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	410	54,15%	2,68%	1,46%	0,4377
Männer	382	49,48%	3,14%	0,78%	0,5470

Hier wird in beiden Fällen der Test auf Normalverteilung nach Shapiro - Wilk durchgeführt. Sowohl bei den Frauen wie auch bei den Männern sind stärkere Abweichungen in den Schwänzen der Verteilung der Residuen vorhanden. Die Annahme der Normalverteilung wird aber in beiden Fällen nicht verworfen, was unter anderem auf die geringe Anzahl von Werten in den beiden Gruppen zurückzuführen ist.

Modellgleichung Kovarianzanalyse: $n = 793$

$$\begin{aligned} FVC &= -4,205 + 5,471H - 0,017AH - 8,939SEX \\ &\quad +2,934HSEX - 0,025AHSEX + 1,746\ln(A)SEX \end{aligned}$$

$$R^2 = 0,787 \quad s_e = 0,516$$

Das Ergebnis der Kovarianzanalyse bezieht hier einen konstanten Unterschied zwischen Frauen und Männern, sowie Wechselwirkungen in allen Variablen mit ein. Die Variable $\ln(A)$ wird in diesem Modell nicht mitaufgenommen. In Übereinstimmung mit den Modellgleichungen für Frauen und Männer getrennt, könnte also FVC bei den Frauen (SEX=0) nur mit linearen Termen erklärt werden. Hier bedeutet das positive Vorzeichen bezüglich der Wechselwirkung $\ln(A)$, daß die FVC-Werte bei den Männern nichtlinear vom Alter abhängig sind.

6.3 Ex-gelegentliche Raucher

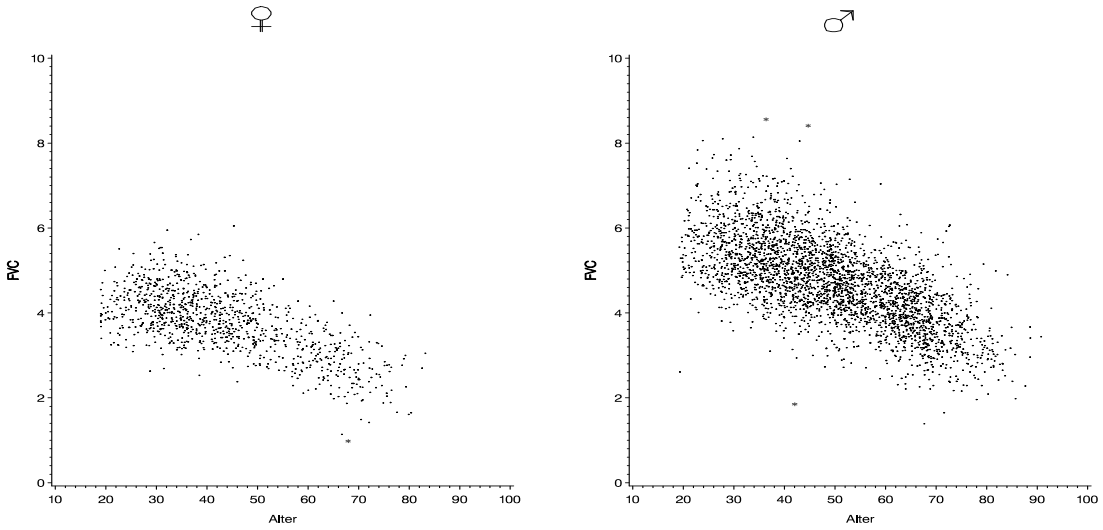


Abbildung 6.9: Scatterplot

Im Scatterplot ist diesmal bei den Frauen eine leichte Bogenform erkennbar. Die Werte der Männer streuen wiederum mehr, wobei allerdings darauf hingewiesen sei, daß in der Gruppe der Männer dreimal soviele Werte sind. In beiden Gruppen sind bis etwa 80 Jahren genügend Werte vorhanden.

Modellgleichungen

Frauen: $n = 1006$

$$FVC = -12,713 + 7,087H - 0,050AH + 2,255\ln(A)$$

$$R^2 = 0,602 \quad s_e = 0,473$$

Männer: $n = 3072$

$$FVC = -12,108 + 8,208H - 0,038AH + 1,491\ln(A)$$

$$R^2 = 0,603 \quad s_e = 0,624$$

Die Modellgleichungen sind einander hier sehr ähnlich. Der Intercept ist in beiden Fällen fast gleich groß . Deutlich wird der unterschiedliche Einfluß der Körpergröße durch die verschieden großen Koeffizienten der Variablen Größe wiedergegeben. Die offensichtliche Nichtlinearität im Scatterplot bei den Frauen ist auch im großen Koeffizienten bei der Variablen $\ln(A)$ erkennbar. Die Bestimmtheitsgrade sind gleich groß und das Modell der Männer hat eine größere Modellstreuung.

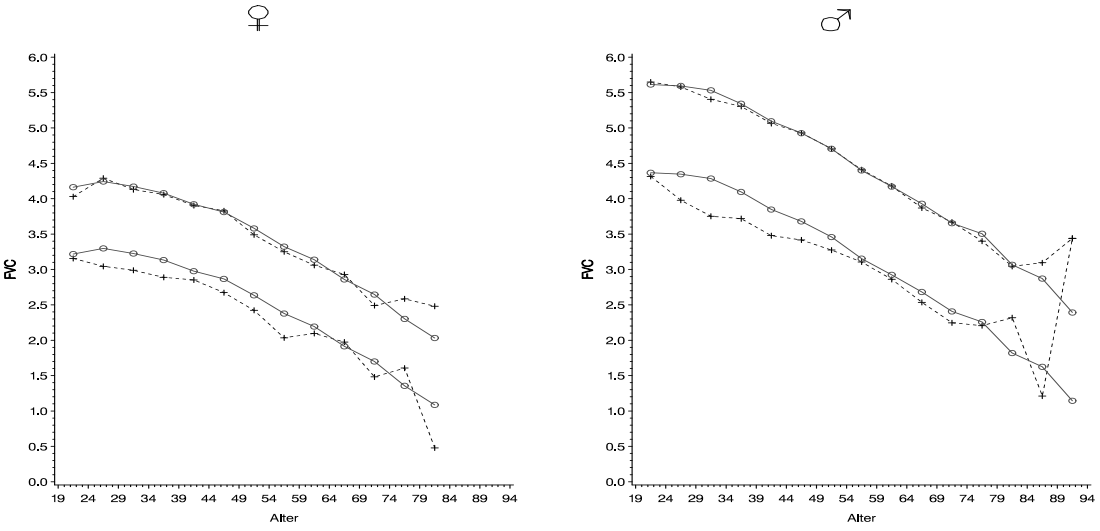


Abbildung 6.10: Medianplot

Die Medianplots zeigen bis 80 Jahren eine durchaus gute Anpassung. Bei den Männern ist auch im Altersbereich von 19 - 30 die Anpassung zufriedenstellend. Was auch darauf zurückzuführen ist, daß die Mediane ihr Maximum bereits in der Altersgruppe von [19-24) erreichen. Während bei den Frauen die bisher übliche Charakteristik beibehalten und das Maximum in der Altersgruppe der 24 - 29 jährigen zu finden ist.

Bei den Männern ist der durchschnittliche Abfall der FVC zwischen 30 und 80 Jahren 4,8 ml/Jahr. Bei den Frauen beträgt die Abnahme zwischen 30 und 50 Jahren 2,5 ml/Jahr, von 50 bis 75 Jahren aber 4,4 ml/Jahr. Im Schnitt sind das 3,56 ml/Jahr von 30 bis 75 Jahren.

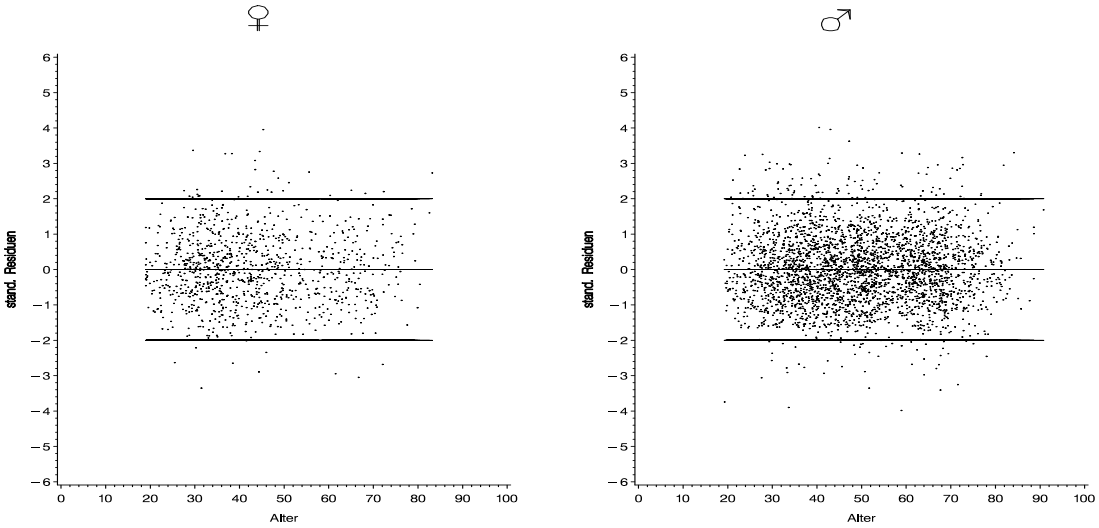


Abbildung 6.11: stand. Residuenplot

Die stand. Residuenplots zeigen keine groben Abweichungen von den Modellannahmen. Man erkennt aber, daß wieder zuviele Residuen größer als $2s_e$ sind.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	K-S/S-W
Frauen	1005	51,34%	3,08%	1,00%	0,3802
Männer	3069	51,68%	3,39%	1,53%	0,001

Durch die Abweichungen von der Normalverteilung in den Schwänzen der Verteilung der Residuen wird die Annahme der Normalverteilung bei den Männern abgelehnt. Es sind aber nur sehr wenige Werte für die Abweichungen verantwortlich.

Modellgleichung Kovarianzanalyse: $n = 4078$

$$\begin{aligned} FVC &= -12,315 + 6,963H - 0,049AH + 2,188\ln(A) \\ &\quad +1,311HSEX + 0,011AHSEX - 0,667\ln(A)SEX \end{aligned}$$

$$R^2 = 0,669 \quad s_e = 0,586$$

Die Unterschiede zwischen den Geschlechtern werden ausschließlich durch Wechselwirkungen in allen erklärenden Variablen beschrieben.

6.4 Raucher_leicht

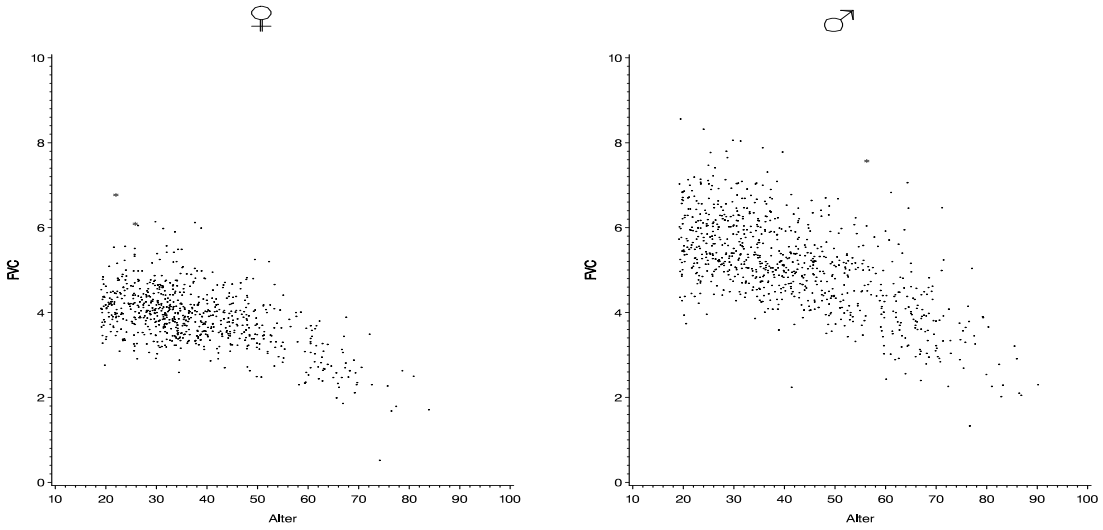


Abbildung 6.12: Scatterplot

In allen drei Rauchergruppen sind bei den Frauen die höheren Altersgruppen mit sehr wenigen Probanden oder gar nicht besetzt. Aufgrund der zugrunde liegenden Daten können nur Auswertungen bis ungefähr 65 Jahren vorgenommen werden. In der Gruppe der leichten Raucher sind bei den Männern bis 75 Jahren genügend Werte vorhanden. Die Scatterplots zeigen die bereits gewohnte Charakteristik mit einer ausgeprägteren Bogenform der Werte bei den Frauen als bei den Männern.

Modellgleichungen

Frauen: $n = 723$

$$FVC = -11,572 + 6,597H - 0,048AH + 2,122\ln(A)$$

$$R^2 = 0,495 \quad s_e = 0,484$$

Männer: $n = 813$

$$FVC = -12,815 + 8,672H - 0,039AH + 1,486\ln(A)$$

$$R^2 = 0,615 \quad s_e = 0,662$$

Bei den Modellgleichungen ist auffällig, daß neben der Modellstreuung bei den Frauen diesmal auch das Bestimmtheitsmaß geringer ist als jenes der Männer. Ansonsten geben die Gleichungen den Einfluß der geringen Körpergröße und der größeren Nichtlinearität bei den Frauen wieder.

Die Medianplots zeigen zu Beginn zum Teil eine leichte Überschätzung und in höheren Altersbereichen zum Teil Abweichungen nach oben und unten. Wie schon oben erwähnt

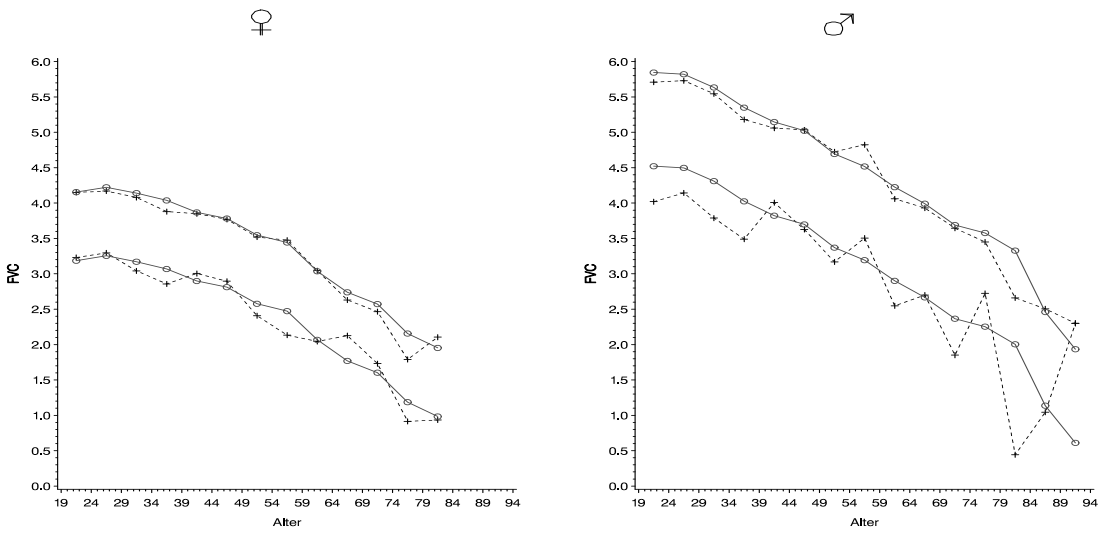


Abbildung 6.13: Medianplot

sind hier bei den Frauen nur bis 65 Jahren und bei den Männern bis 75 Jahren genügend Daten zur Auswertung vorhanden.

Bei den Frauen erfolgt von 30 bis 50 Jahren eine Abnahme der FVC von 2,25 ml/Jahr und zwischen 50 und 65 von 5,3 ml/Jahr, das sind von 30 bis 65 3,57 ml/Jahr. Bei den Männern erfolgt von 30 bis 75 eine durchschnittliche Abnahme von 3,67 ml/Jahr.

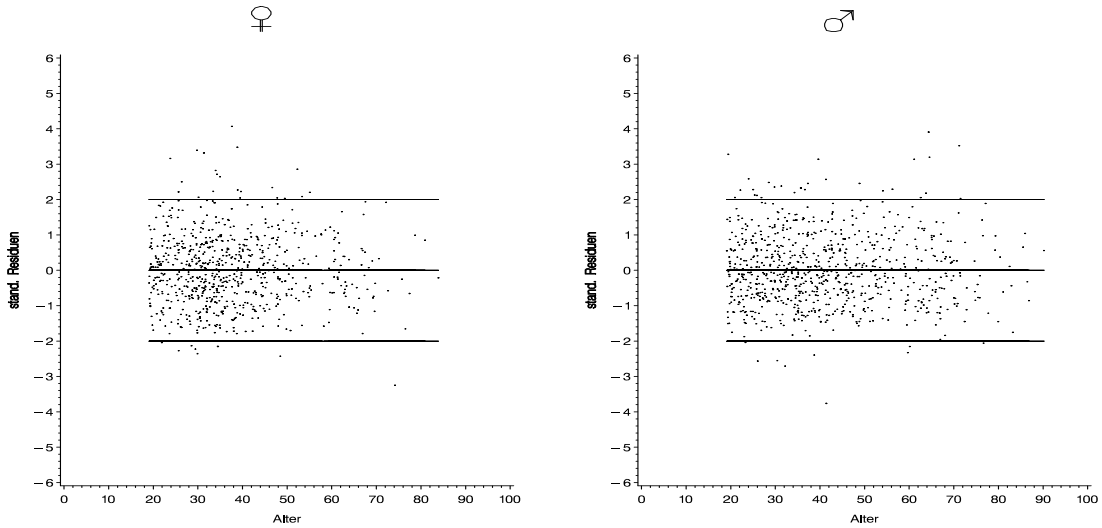


Abbildung 6.14: stand. Residuenplot

Anhand der stand. Residuenplots ist wieder eine zu große Anzahl von Residuen $> 2s_e$ feststellbar.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	721	51,18%	2,91%	1,24%	0,2462
Männer	812	53,70%	3,45%	1,23%	0,0985

Wie schon an den Residuenplots ersichtlich sind zu viele Residuen $> 2s_e$ und zuwenige $< -2s_e$. Trotz dieser Verzerrungen in den Schwänzen der Verteilung lehnt der Test nach Shapiro - Wilk die Annahme der Normalverteilung bei einem vorgebenen Niveau von $\alpha = 5\%$ nicht ab.

Modellgleichung Kovarianzanalyse: $n = 1536$

$$\begin{aligned} FVC &= -12,269 + 6,837H - 0,050AH + 2,244\ln(A) \\ &\quad +1,633HSEX + 0,012AHSEX - 0,819\ln(A)SEX \end{aligned}$$

$$R^2 = 0,715 \quad s_e = 0,583$$

Die Modellgleichung der Kovarianzanalyse betont noch einmal die Nichtlinearität bei den Frauen (SEX=0). Unterschiede zwischen den Geschlechtern werden diesmal durch Wechselwirkungen in allen drei Modellvariablen beschrieben.

6.5 Raucher_mittel

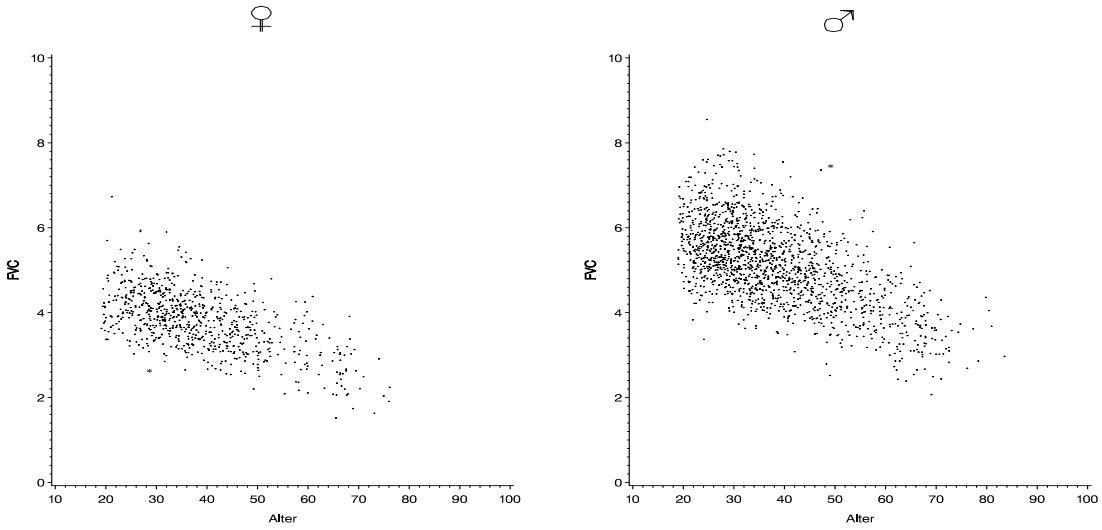


Abbildung 6.15: Scatterplot

Die Werte der Frauen streuen weniger als jene der Männer. Bei beiden sind keine klar ersichtlichen Nichtlinearitäten vorhanden. Bei den Frauen sind Auswertungen bis etwa 70 Jahren und bei Männern bis 75 Jahren möglich.

Modellgleichungen

Frauen: $n = 760$

$$FVC = -8,681 + 6,756H - 0,033AH + 0,962\ln(A)$$

$$R^2 = 0,564 \quad s_e = 0,466$$

Männer: $n = 1620$

$$FVC = -10,726 + 8,135H - 0,036AH + 1,087\ln(A)$$

$$R^2 = 0,586 \quad s_e = 0,620$$

Hier ist der Bestimmtheitsgrad bei den Frauen wiederum etwas kleiner als bei den Männern. Die Gleichungen stimmen bis auf Unterschiede bzgl. der Größe in etwa überein. Dies sollte dann später auch bei der Kovarianzanalyse zum Ausdruck kommen.

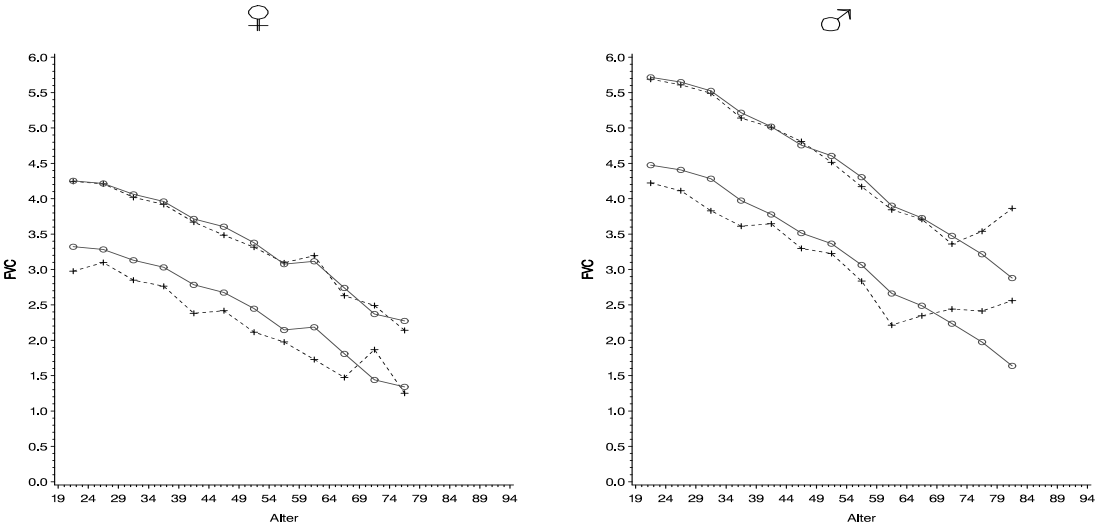


Abbildung 6.16: Medianplot

Die Medianplots zeigen eine gute Anpassung über den gesamten Altersbereich. Im Gegensatz zu allen bisher betrachteten Gruppen ist hier im Bereich von 19 bis 30 Jahren kein Anstieg mehr feststellbar. Interessanterweise ist im Medianplot der Frauen zwischen 59 und 64 Jahren ein Plateau vorhanden.

Die Abnahme der FVC Werte beträgt zwischen 30 und 70 Jahren bei den Frauen 3,87 ml/Jahr und bei den Männern 4,11 ml/Jahr.

Die stand. Residuenplots zeigen keine systematischen Abweichungen der stand. Residuen.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	759	52,44%	3,43%	1,19%	0,4334
Männer	1619	52,13%	3,46%	1,30%	<0,001

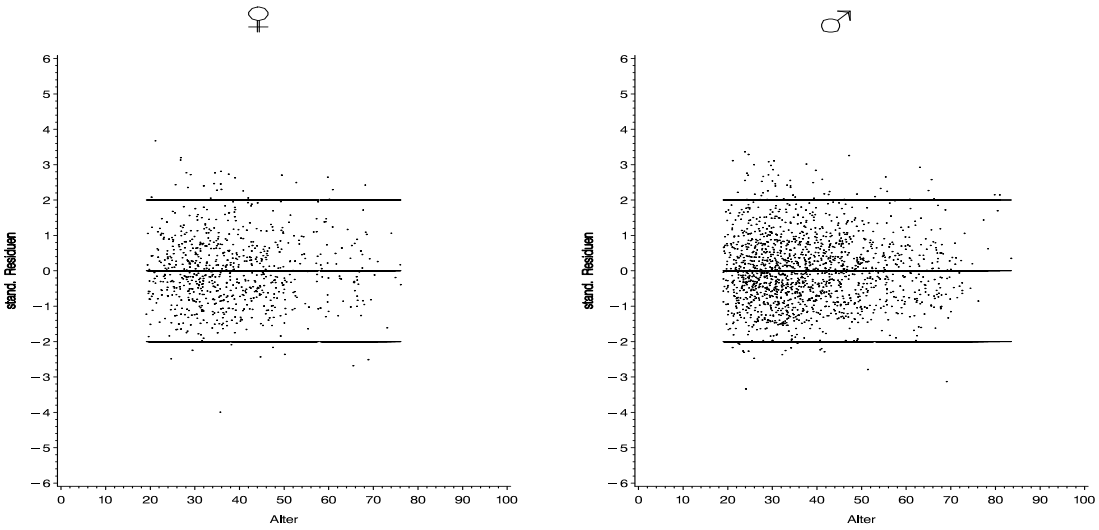


Abbildung 6.17: stand. Residuenplot

Zuviele Residuen sind $> 2s_e$ und zuwenige $< -2s_e$. Deshalb wird die Annahme der Normalverteilung bei den Männern nicht bestätigt. Im stand. Residuenplot ist hingegen keine Verletzung der Modellannahmen zu erkennen.

Modellgleichung Kovarianzanalyse: $n = 2380$

$$\begin{aligned} FVC &= -8,864 + 6,706H - 0,036AH + 1,082\ln(A) \\ &\quad -1,911SEX + 1,4541SEXH \end{aligned}$$

$$R^2 = 0,714 \quad s_e = 0,577$$

Der Unterschied zwischen Frauen und Männer läßt sich hier durch eine konstante Verschiebung und einer Wechselwirkung bezüglich der Größe beschreiben.

6.6 Raucher_schwer

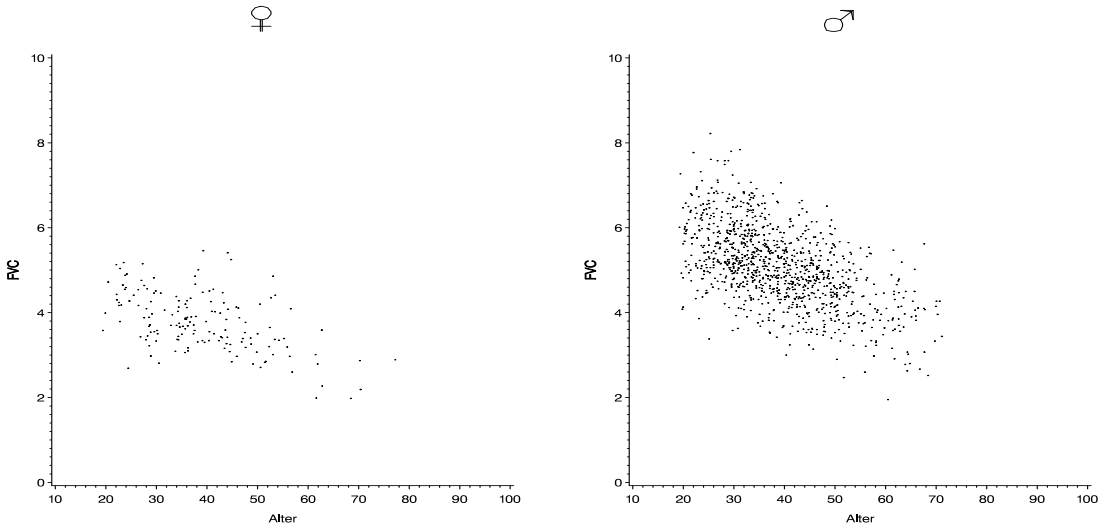


Abbildung 6.18: Scatterplot

Besonders bei den Frauen sind die Altersklassen ab 60 Jahre mit sehr wenig Werten besetzt. Bei den Männern hingegen kann das Modell bis etwa 70 Jahre mit ausreichend vielen Werten unterlegt werden.

Modellgleichungen

Frauen: $n = 158$

$$FVC = -6,063 + 6,500H - 0,014AH$$
$$R^2 = 0,522 \quad s_e = 0,478$$

Männer: $n = 1037$

$$FVC = -6,194 + 7,225H - 0,022AH$$
$$R^2 = 0,507 \quad s_e = 0,639$$

Sowohl bei den Frauen als auch bei den Männern liefert die Variable $\ln(A)$ keinen signifikanten Beitrag zur Erklärung der Gesamtstreuung. Darüberhinaus bestehen zwischen den beiden Modellgleichungen keine großen Unterschiede.

Im Medianplot bei den Frauen sind einige Abweichungen erkennbar, da nur bis 50 Jahre mehr als 10 Probanden pro Altersklasse vertreten sind. Das Modell der Männer stimmt in etwa mit den empirischen Medianen überein. Sehr deutlich ist bei den schweren Rauchern erkennbar, daß FVC bereits ab 20 Jahren beständig abnimmt. Die FVC nimmt bei den Frauen von 20 bis 60 Jahren mit 3,25 ml/Jahr ab, während bei den Männern die FVC zwischen 20 und 70 Jahren mit 4,2 ml/Jahr abnimmt.

Die stand. Residuenplots zeigen keine offensichtlichen Abweichungen von den Modellannahmen.

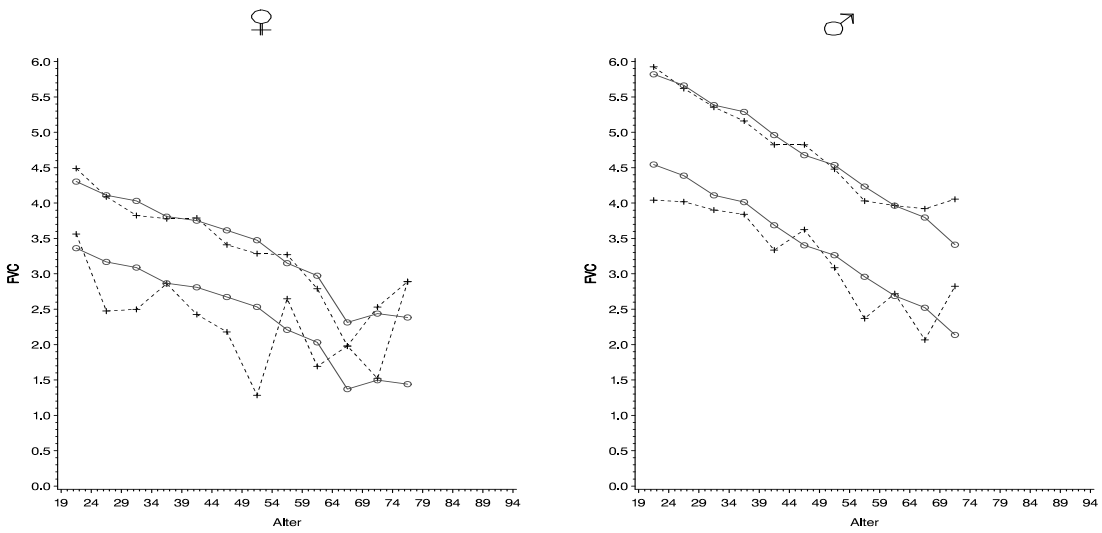


Abbildung 6.19: Medianplot

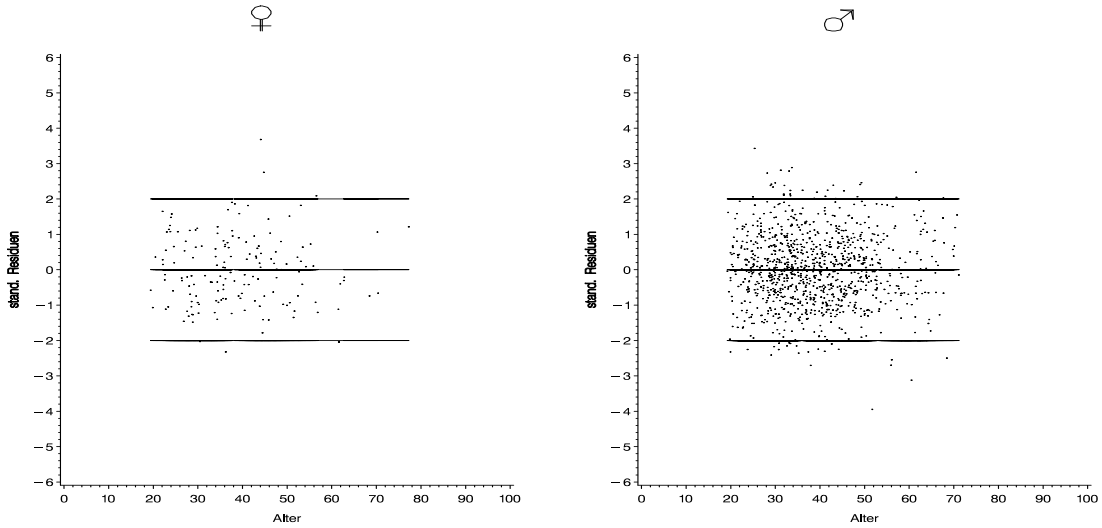


Abbildung 6.20: stand. Residuenplot

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	158	51,26%	1,90%	1,90%	0,3890
Männer	1037	50,05%	2,99%	2,22%	0,9682

Die Residuen sind hier annähernd symmetrisch und normalverteilt. Die Abweichungen sind bei den Frauen größer, aber in beiden Fällen wird die Annahme der Normalverteilung nicht verworfen.

Modellgleichung Kovarianzanalyse: $n = 1195$

$$\begin{aligned} FVC &= -12,158 + 7,812H - 0,036AH + 1,475\ln(A) \\ &\quad +2,288SEX - 0,477\ln(A)SEX \end{aligned}$$

$$R^2 = 0,613 \quad s_e = 0,614$$

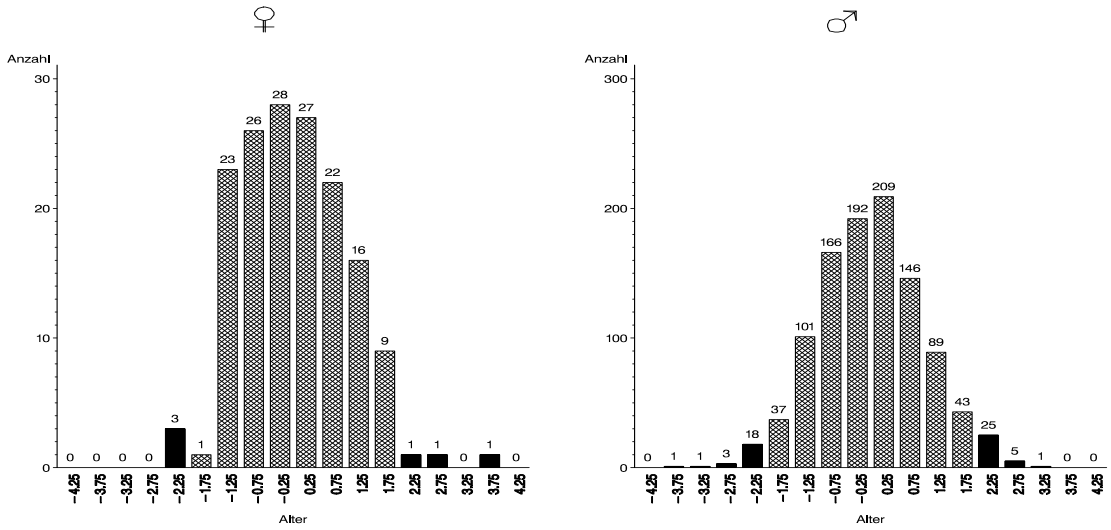


Abbildung 6.21: Histogramm der stand. Residuen

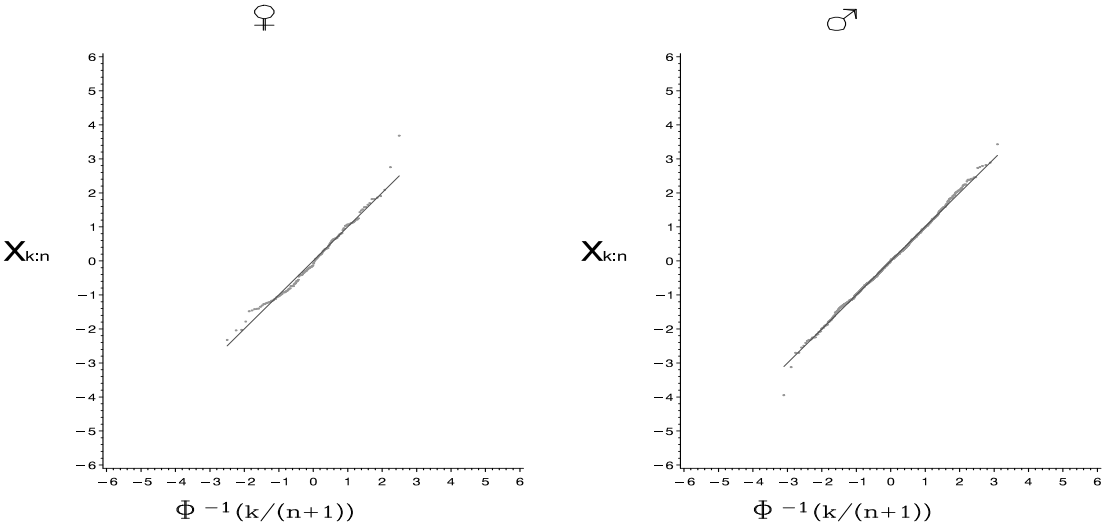


Abbildung 6.22: Normal Probability Plot der stand. Residuen

Das Ergebnis der Kovarianzanalyse zeigt einen konstanten Unterschied zwischen den Geschlechtern sowie eine Wechselwirkung in der Variablen $\ln(A)$. Im Gegensatz zu den Modellgleichungen getrennt für Frauen und Männer ist hier die Variable $\ln(A)$ wieder im Modell vertreten.

Kapitel 7

FEV_1

In diesem Kapitel werden Modelle für den Parameter FEV_1 erstellt. Frauen und Männer werden parallel untersucht. In den Untergruppen geht es wieder darum Unterschiede zwischen den Geschlechtern zu verdeutlichen. Wie schon zu Beginn des vorherigen Kapitels angeführt wird aufgrund der hohen Korrelation zwischen den Parametern FVC und FEV_1 in beiden Fällen dasselbe Modell verwendet.

Das Modell entspricht in der Wahl der erklärenden Variablen den Modellen der Arbeiten von Kummer [11] und Rapatz [15]. Für FEV_1 sieht das Modell folgendermaßen aus:

$$FEV_1 = \hat{\beta}_0 + \hat{\beta}_1 H + \hat{\beta}_2 AH + \hat{\beta}_3 \ln(A)$$

Die Kurzbezeichnungen für die erklärenden Variablen sind:

AH: Alter×Groesse; **H**: Größe; **ln(A)**: ln(Alter)

Die Korrelationskoeffzienten in den folgenden Tabellen zeigen die Stärke des Zusammenhangs zwischen den erklärenden Variablen und FEV_1 bei Frauen und Männern:

	FEV_1-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
H	0,518	0,510	0,433	0,429	0,483	0,568
AH	-0,724	-0,713	**-0,683**	**-0,590**	-0,634	-0,568
ln(A)	**-0,736**	**-0,741**	-0,679	**-0,590**	**-0,662**	**-0,615**

Bei FEV_1 tritt die Korrelation mit dem Alter noch stärker zutage, als bei FVC. Die Korrelation mit der Größe ist hier etwas geringer und der stärkste Zusammenhang ist wieder mit der Variablen $\ln(A)$ gegeben. Aufgrund der Abhängigkeit der FEV_1-Werte von der Größe, werden im Anschluß an die Analyse der einzelnen Untergruppen, die nach der Größe adjustierten Werte betrachtet.

	FEV_1-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
H	0,561	0,549	0,546	0,584	0,537	0,480
AH	-0,640	-0,537	-0,681	-0,650	-0,654	-0,577
ln(A)	**-0,653**	**-0,570**	**-0,695**	**-0,657**	**-0,672**	**-0,612**

Bei den Männern ergibt sich die gleiche Charakateristik der Korrelationen wie bei den Frauen.

Die Analyse des Parameters FEV_1 erfolgt in gleicher Weise wie jene des Parameters FVC.

7.1 Niemalsraucher

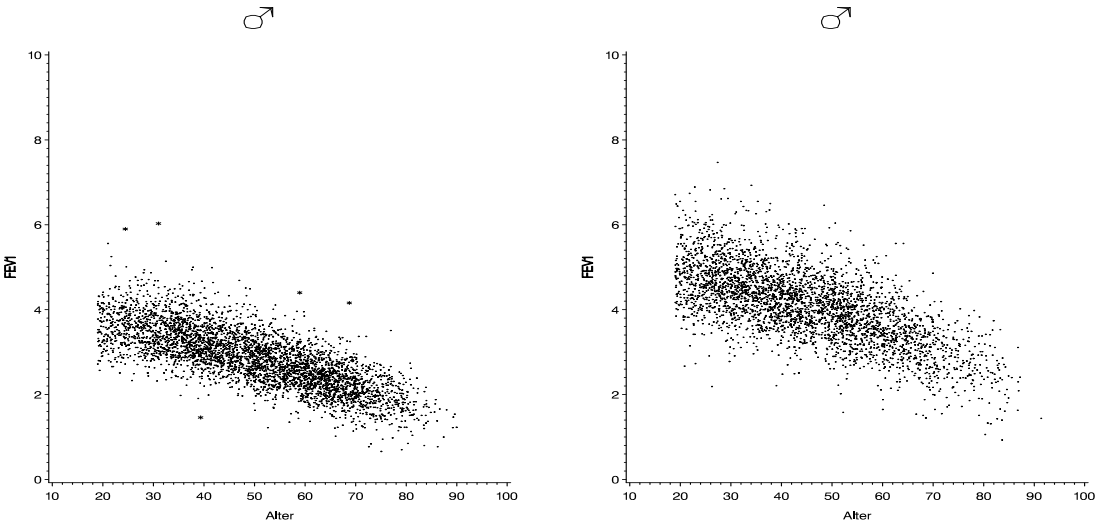


Abbildung 7.1: Scatterplot

Die Werte der Frauen sind im Schnitt wieder um 25% kleiner als jene der Männer. Die Daten der Frauen bilden eine kompaktere Wolke, was sich auch bei der Modellerstellung in einem höheren Bestimmtheitsgrad R^2 und geringerer Modellstreuung s_e niederschlägt.

Zur Modellerstellung werden wie im Kapitel zuvor die untransformierten FEV_1-Werte herangezogen. Die Regressionsmodelle werden analog zum Parameter FVC erstellt.

Modellgleichungen

Frauen: $n = 4124$

$$FEV_1 = -6,703 + 5,112H - 0,031AH + 0,968\ln(A)$$

$$R^2 = 0,682 \quad s_e = 0,390$$

Männer: $n = 3459$

$$FEV_1 = -8,694 + 6,252H - 0,033AH + 1,169\ln(A)$$

$$R^2 = 0,590 \quad s_e = 0,549$$

Wie oben anhand der Scatterplots vermutet, besitzen die Modellgleichungen der Frauen einen höheren Bestimmtheitsgrad und eine geringere Modellstreuung. Die Modelle für Frauen und Männer unterscheiden sich nur wenig.

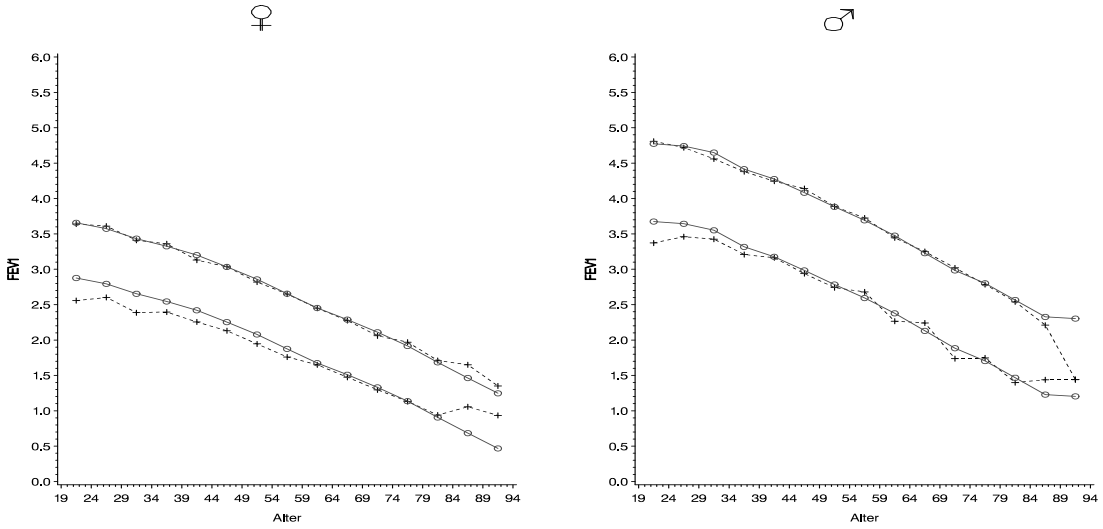


Abbildung 7.2: Medianplot

Beide Medianplots zeigen eine sehr gute Anpassung der Mediane der aus den Gleichungen prognostizierten Werte an die empirischen Datenmediane. Lediglich in den letzten Altersklassen sind, durch eine geringe Anzahl von Daten bedingt, leichte Abweichungen vorhanden. Obwohl die FEV_1-Werte nicht nach der Größe adjustiert sind, stimmen die Streuungen der Werte mit der Modellstreuung ab den mittleren Altersgruppen überein. Die Mediane beschreiben in beiden Fällen eine leichte Bogenform. Bei den Frauen nimmt FEV_1 zwischen 30 und 55 Jahren um 3 ml/Jahr und zwischen 55 und 80 Jahren um 3,8 ml/Jahr (3,4 ml/Jahr von 30 bis 80) ab. Bei den Männern zwischen 30 und 55 Jahren um 3,6 ml/Jahr und zwischen 55 und 80 Jahren um 4,4 ml/Jahr (4 ml/Jahr von 30 bis 80).

Die stand. Residuenplot bestätigen die Adäquatheit des Modells.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	K-S
Frauen	4119	50,87%	2,77%	1,65%	0,0341
Männer	3459	50,51%	2,60%	2,26%	0,0249

Die Residuen sind symmetrisch verteilt, wie auch aus dem Histogramm der stand. Residuen ersichtlich. Der Test auf Normalverteilung würde auf dem 5%-Niveau noch zur

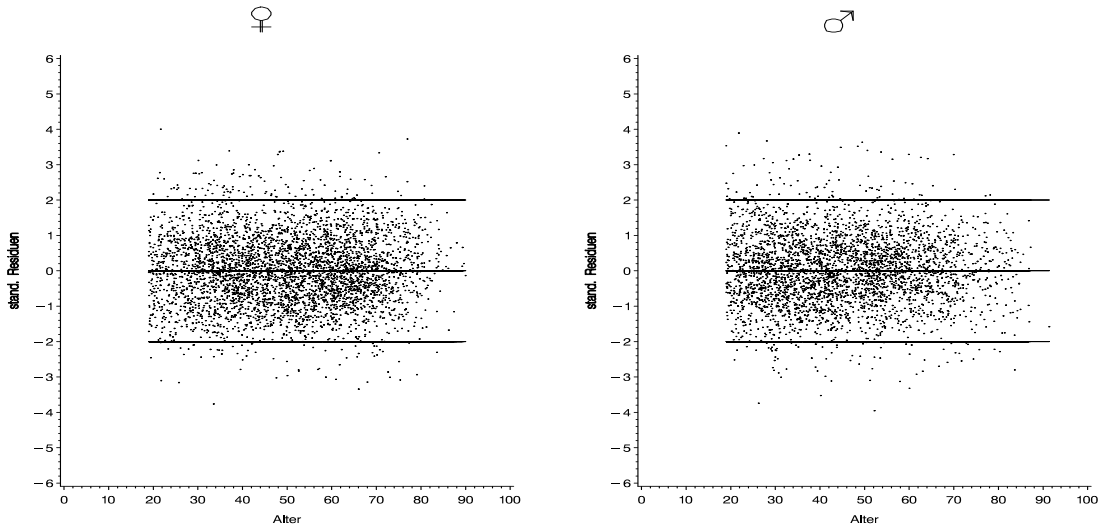


Abbildung 7.3: stand. Residuenplot

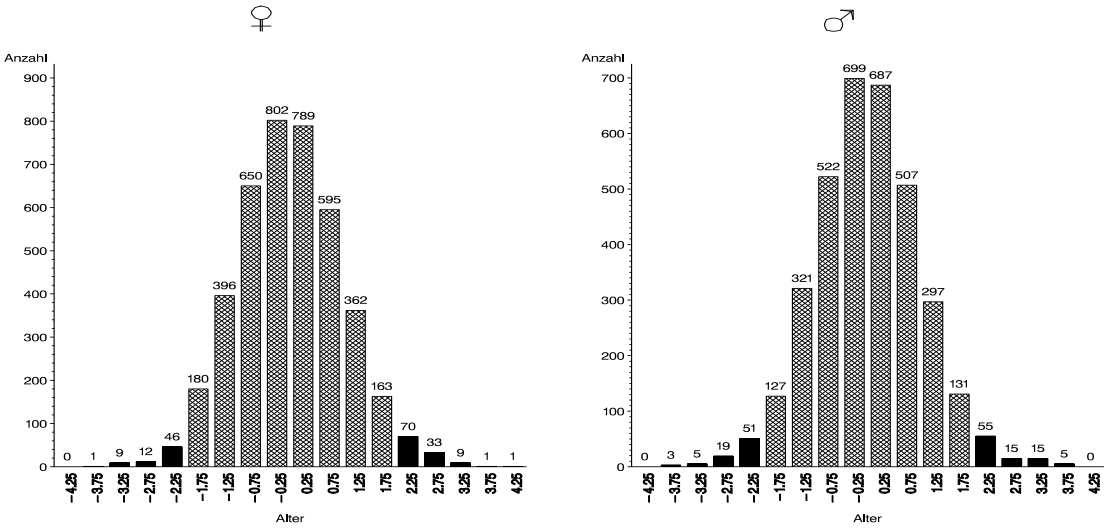


Abbildung 7.4: Histogramm der stand. Residuen

Ablehnung führen. Man sieht im Normal Probability Plot, daß vor allem Abweichungen in den Schwänzen für diese Ablehnung verantwortlich sind.

Modellgleichung Kovarianzanalyse: $n = 7583$

$$\begin{aligned} FEV_1 &= -7,259 + 5,265H - 0,032AH + 1,065\ln(A) \\ &\quad -0,897SEX + 0,861HSEX \end{aligned}$$

$$R^2 = 0,775 \quad s_e = 0,466$$

Die Unterschiede zwischen Frauen und Männern beruhen hier auf einer konstanten Verschiebung und einer Wechselwirkung bzgl. der Größe.

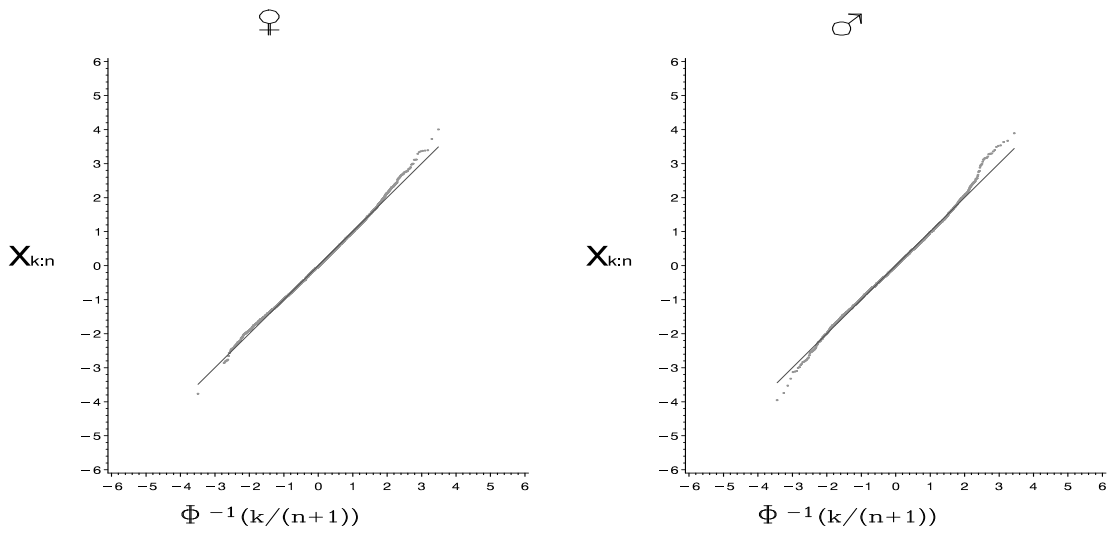


Abbildung 7.5: Normal Probability Plot der stand. Residuen

7.2 Passivraucher

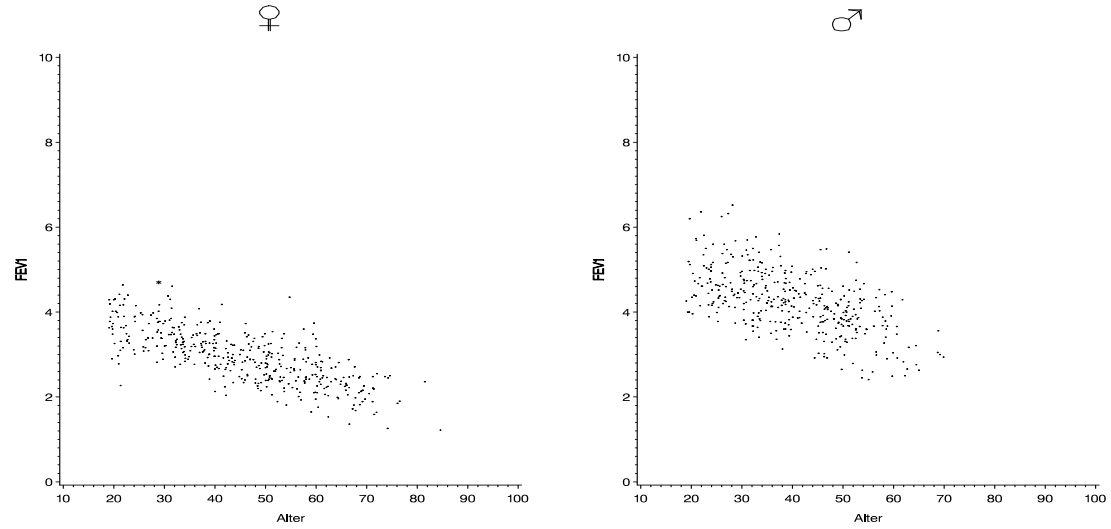


Abbildung 7.6: Scatterplot

Die Werte der Männer erstrecken sich von 19 bis etwa 65 Jahren, während bei den Frauen bis 75 Jahren Werte vorhanden sind. Anhand der Scatterplots sind keine nichtlinearen Verläufe, wie z.B. ein stärkerer Abfall in höheren Altersklassen feststellbar.

Modellgleichungen

Frauen: $n = 382$

$$FEV_1 = -3,223 + 4,553H - 0,017AH$$

$$R^2 = 0,681 \quad s_e = 0,358$$

Männer: $n = 411$

$$FEV_1 = -4,254 + 5,474H - 0,016AH$$

$$R^2 = 0,502 \quad s_e = 0,514$$

Die Modellgleichungen werden, wie schon bei FVC, ohne der Variablen $\ln(A)$ erstellt. Man kann hier durchaus einen konstanten Abfall über den gesamten Altersbereich annehmen.

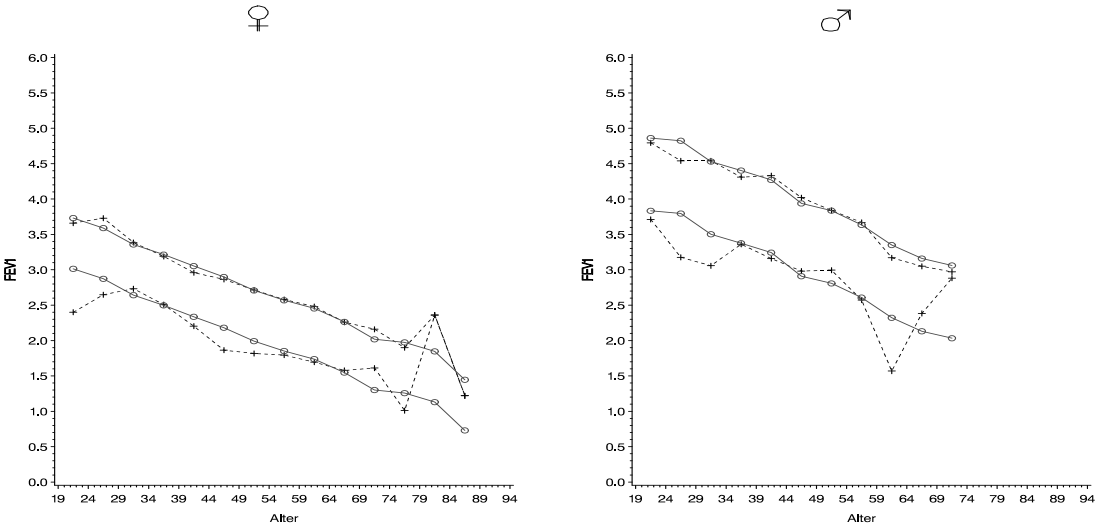


Abbildung 7.7: Medianplot

Die Medianplots zeigen bei den Frauen eine bessere Anpassung. Die Streuung wird vor allem bei den Männern zum Teil nur schlecht nachvollzogen. Bei Frauen und Männern kann die Abnahme der FEV_1-Werte als konstant angesehen werden. Bei den Frauen beträgt der durchschnittliche Abfall der FEV_1-Werte zwischen 20 und 75 Jahren 3,1 ml/Jahr, während die Männer zwischen 20 und 65 Jahren im Schnitt 3,56 ml/Jahr ihrer FEV_1 verlieren.

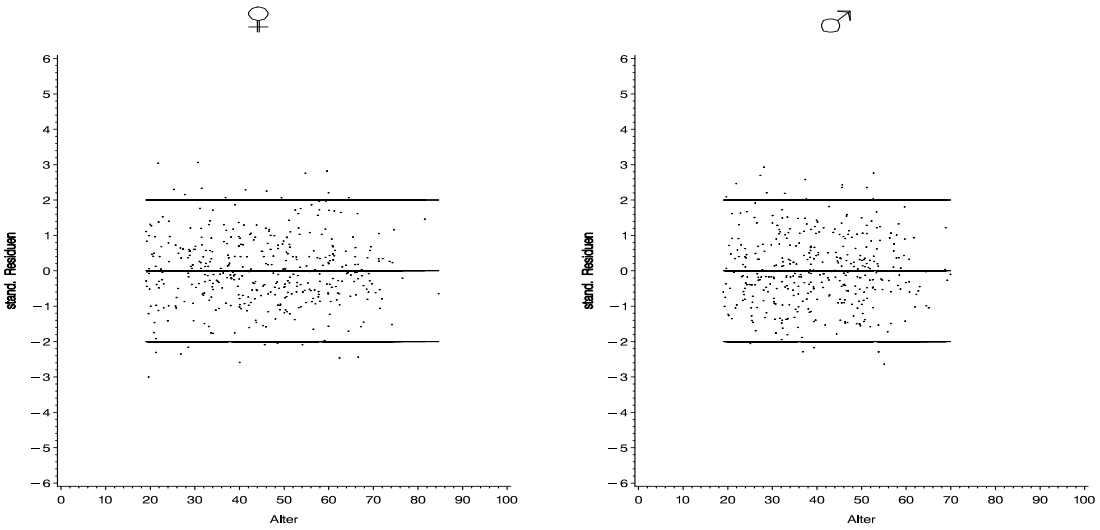


Abbildung 7.8: stand. Residuenplot

Der Residuenplot zeigt keine Abweichungen von den Modellannahmen.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	410	52,20%	3,41	2,93%	0,4460
Männer	382	52,88%	3,40%	1,31%	0,0764

Die Residuen der Frauen sind normalverteilt, bei den Männern sind leichte Abweichungen im oberen Bereich vorhanden.

Modellgleichung Kovarianzanalyse: $n = 793$

$$FEV_1 = -3,595 + 4,770H - 0,017AH + 0,353HSEX$$

$$R^2 = 0,778 \quad s_e = 0,444$$

In Übereinstimmung mit den bisherigen Ergebnissen hat hier die Variable $\ln(A)$ keinen signifikanten Anteil am Modell. Sowohl für Frauen als auch für Männer würde ein Modell mit linearen Variablen ausreichen, um die FEV_1-Werte zu beschreiben. Der Unterschied zwischen Frauen und Männern wird einzig durch eine Wechselwirkung bzgl. der Größe beschrieben.

7.3 Ex-gelegentliche Raucher

Beim Vergleich der beiden Datenwolken muß darauf geachtet werden, daß bei den Männern dreimal soviel Werte vorhanden sind, als bei den Frauen. Die Werte mit großen Residuen sind durch Sterne hervorgehoben. Bei den Frauen liegen bis 80 Jahren und bei den Männern bis 85 Jahren genügend Werte zur Analyse vor.

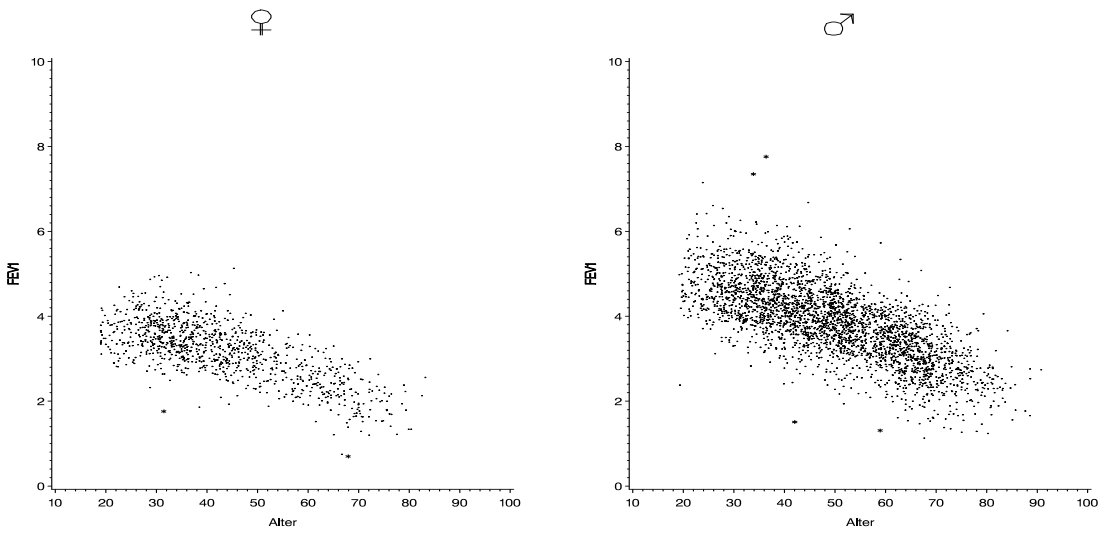


Abbildung 7.9: Scatterplot

Modellgleichungen

Frauen: $n = 1006$

$$FEV_1 = -8,989 + 5,445H - 0,043AH + 1,684\ln(A)$$

$$R^2 = 0,634 \quad s_e = 0,414$$

Männer: $n = 3072$

$$FEV_1 = -9,022 + 6,257H - 0,036AH + 1,312\ln(A)$$

$$R^2 = 0,623 \quad s_e = 0,535$$

Die Modellgleichungen stimmen in etwa überein. In beiden Fällen sind wieder größere Nichtlinearitäten, ersichtlich an den Koeffizienten der Variablen $\ln(A)$, vorhanden. Die Bestimmtheitsgrade sind fast gleich groß , lediglich die Modellstreuung ist bei den Männern größer.

Die Medianplots zeigen eine gute Übereinstimmung der Modellmediane mit den empirischen Medianen. Bei den Frauen ist der nichtlineare Verlauf der Mediane deutlich erkennbar. So nehmen die FEV_1-Werte bei den Frauen zwischen 30 und 55 Jahren um 3 ml/Jahr und zwischen 55 und 80 Jahren um 3,8 ml/Jahr ab (3,4 ml/Jahr von 30 bis 80). Die Werte der Männer nehmen zwischen 30 und 55 Jahren um 3,4 ml/Jahr und zwischen 55 und 80 Jahren um 3,8 ml/Jahr ab (3,6 ml/Jahr von 30 bis 80).

An den stand. Residuenplots sind wiederum keine systematischen Abweichungen feststellbar.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	K-S/S-W
Frauen	1004	51,10%	3,19%	1,59%	0,3392
Männer	3068	50,95%	2,71%	2,15%	0,001

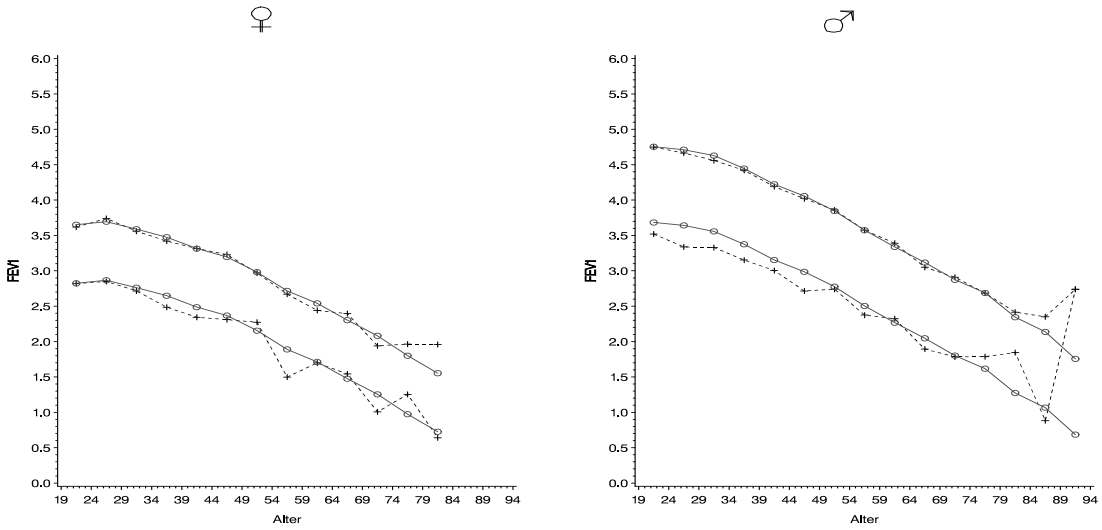


Abbildung 7.10: Medianplot

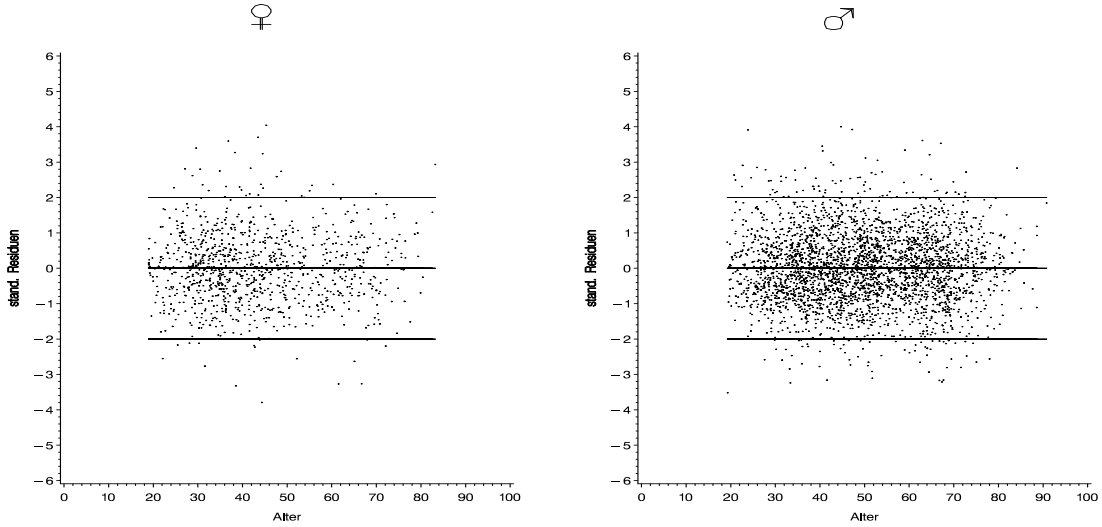


Abbildung 7.11: stand. Residuenplot

Obwohl aufgrund der Prozentwerte in der Tabelle angenommen werden könnte, daß die Residuen der Männer besser normalverteilt sind, ergibt der Test auf Normalverteilung eine Ablehnung.

Modellgleichung Kovarianzanalyse: $n = 4078$

$$\begin{aligned} FEV_1 &= -9,025 + 5,420H - 0,043AH + 1,713\ln(A) \\ &\quad +0,824HSEX + 0,007AHSEX - 0,394\ln(A)SEX \end{aligned}$$

$$R^2 = 0,666 \quad s_e = 0,506$$

Die Unterschiede zwischen den Geschlechtern beruhen auf Wechselwirkungen in allen drei erklärenden Variablen.

7.4 Raucher_leicht

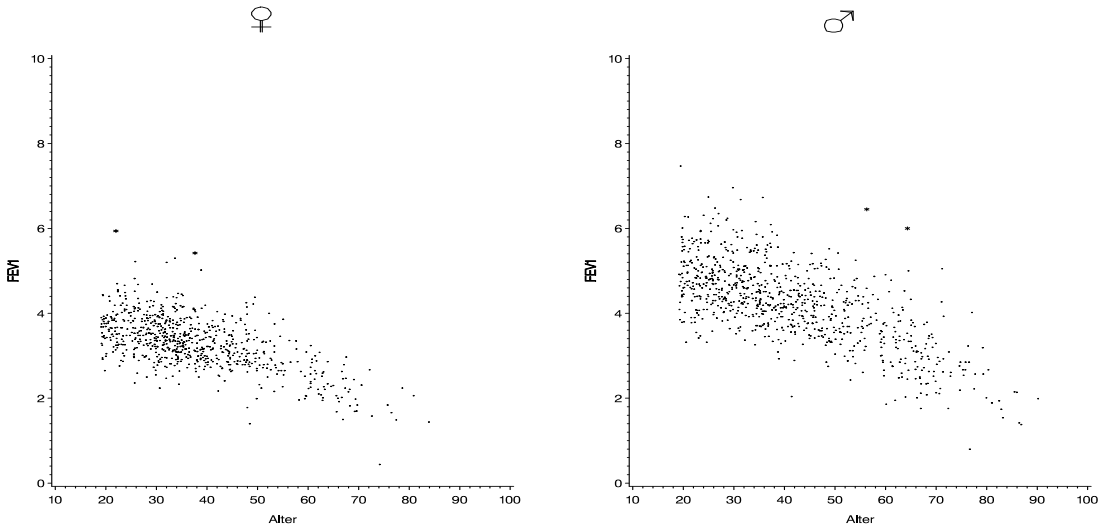


Abbildung 7.12: Scatterplot

In beiden Scatterplots sieht man die geringe Anzahl von Werten in den höheren Altersgruppen. Besonders bei den Frauen sind die höheren Altersgruppen schlecht besetzt. Dies könnte darauf zurückzuführen sein, daß die Männer eine längere Tradition beim Rauchen haben und Frauen erst in den letzten Jahrzehnten begonnen haben vermehrt zu rauchen.

Modellgleichungen

Frauen: $n = 723$

$$FEV_1 = -6,676 + 4,759H - 0,036AH + 1,212\ln(A)$$

$$R^2 = 0,513 \quad s_e = 0,426$$

Männer: $n = 813$

$$FEV_1 = -9,950 + 6,676H - 0,038AH + 1,409\ln(A)$$

$$R^2 = 0,629 \quad s_e = 0,568$$

Bei ungefähr gleich vielen Daten ist diesmal das Bestimmtheitsmaß bei den Männern höher. Die Nichtlinearität, ausgedrückt in der Variablen $\ln(A)$, ist bei beiden Gruppen ungefähr gleich groß .

Die Medianplots zeigen eine gute Übereinstimmung der Mediane. Bei den Männern ist wie schon bei FVC eine leichte Überschätzung in jüngeren Altersklassen festzustellen. In beiden Fällen stimmen die empirischen mit den Modellmedianen bis etwa 70 Jahren gut überein und zeigen einen leicht bogenförmigen Verlauf. Die FEV_1-Werte bei den Frauen

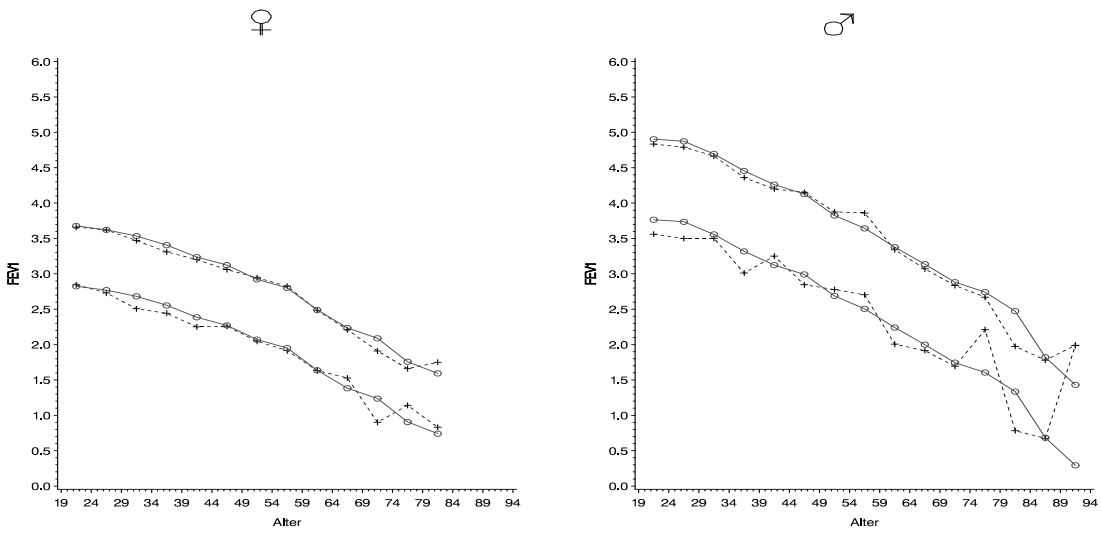


Abbildung 7.13: Medianplot

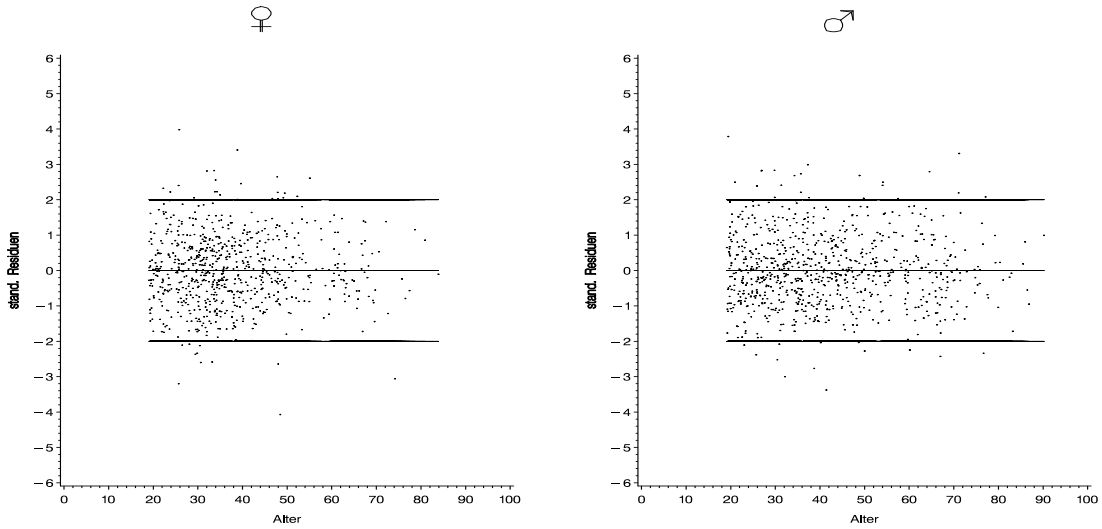


Abbildung 7.14: stand. Residuenplot

nehmen zwischen 30 und 50 Jahren um 2,75 ml/Jahr und zwischen 50 und 70 Jahren um 4,75 ml/Jahr ab (3,75 ml/Jahr von 30 bis 70). Die Werte der Männer nehmen zwischen 30 und 50 Jahren um 4 ml/Jahr und zwischen 50 und 70 Jahren um 5 ml/Jahr ab (4,5 ml/Jahr von 30 bis 70).

Im stand. Residuenplot ist die Ausdünnung der Daten im oberen Altersbereich ersichtlich. Eine graphische Beurteilung ergibt keine Abweichungen von den Modellannahmen. Wie aus der unten angeführten Tabelle ersichtlich, sind in beiden Fällen zuviele Residuen $> 2s_e$ und zuwenige Residuen $< -2s_e$.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	721	51,46%	3,05%	1,53%	0,9974
Männer	811	53,14%	2,71%	1,72%	0,4476

Trotz einiger Verzerrungen wird der Test auf Normalverteilung nicht verworfen.

Modellgleichung Kovarianzanalyse: $n = 1536$

$$\begin{aligned} FEV_1 &= -7,320 + 4,940H - 0,037AH + 1,342\ln(A) \\ &\quad -2,278SEX + 1,639HSEX \end{aligned}$$

$$R^2 = 0,692 \quad s_e = 0,508$$

Die Kovarianzanalyse erklärt die Unterschiede zwischen Frauen und Männern diesmal durch eine Parallelverschiebung und eine Wechselwirkung bzgl. der Größe.

7.5 Raucher_mittel

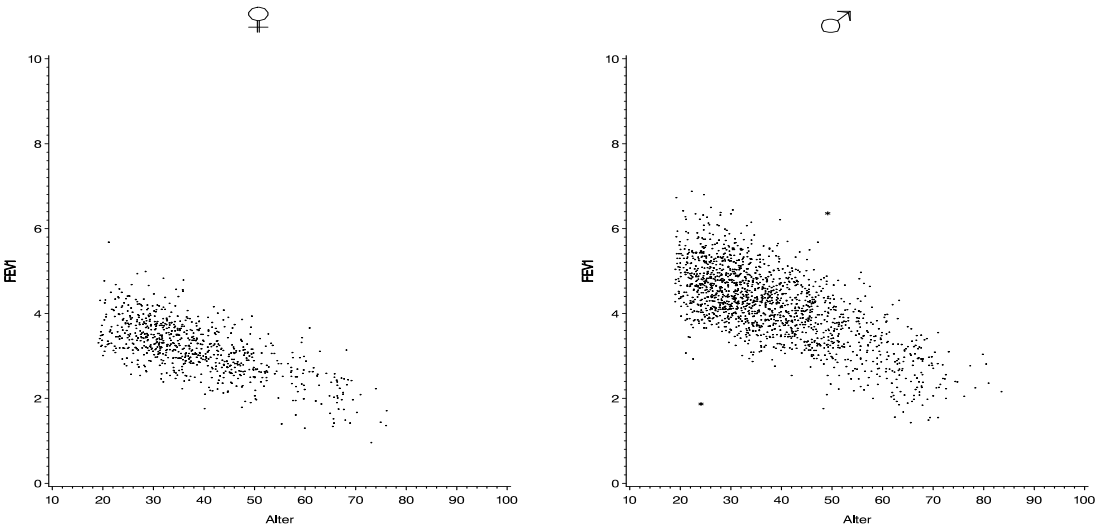


Abbildung 7.15: Scatterplot

Die Werte der Frauen streuen weniger als jene der Männer. In beiden Fällen scheinen die Werte mit den Alter konstant abzunehmen. Auswertungen sind bei den Frauen bis etwa 70 Jahren und bei den Männern bis etwa 75 Jahren möglich.

Modellgleichungen

Frauen: $n = 760$

$$FEV_1 = -3,416 + 4,761H - 0,020AH$$

$$R^2 = 0,587 \quad s_e = 0,409$$

Männer: $n = 1620$

$$FEV_1 = -7,507 + 6,135H - 0,036AH + 0,907\ln(A)$$

$$R^2 = 0,610 \quad s_e = 0,543$$

Die Modellgleichung der Frauen wird wie schon bei den Passivraucherinnen ohne der Variablen $\ln(A)$ erstellt. Es kann also bei den Frauen von einer konstanten Abnahme der Werte ausgegangen werden.

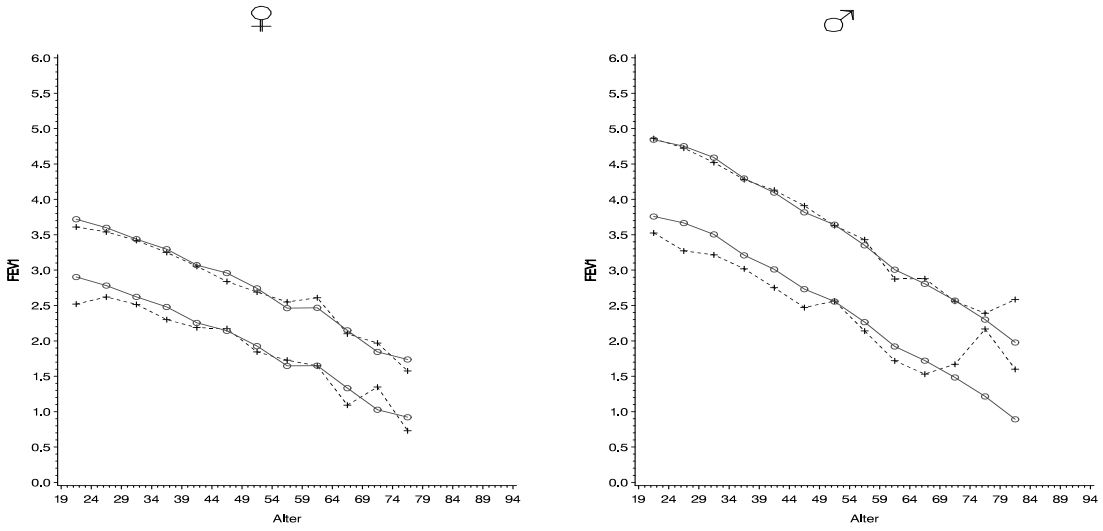


Abbildung 7.16: Medianplot

Die Medianplots zeigen eine gute Anpassung mit der üblichen Unterschätzung der empirischen Streuung durch die Modellstreuung. Bei den Frauen ist, wie bei den FVC-Daten, im Bereich um sechzig ein Plateau erkennbar. Bei beiden Geschlechtern ist bereits ab dem 20. Lebensjahr eine nahezu lineare Abnahme der FEV_1-Werte festzustellen. Im Gegensatz z.B. zu der Gruppe der Niemalsraucher, bei welcher die Werte erst ab 30 Jahren stärker zu fallen beginnen.

Die FEV_1-Werte bei den Frauen nehmen zwischen 30 und 70 Jahren um 3,2 ml/Jahr ab. Die Werte der Männer nehmen zwischen 30 und 75 Jahren um 4,36 ml/Jahr ab.

Der stand. Residuenplot bestätigt wiederum die Adäquatheit des Modells.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	760	49,47%	2,63%	2,76%	0,6153
Männer	1618	50,00%	2,97%	1,79%	0,4879

In Übereinstimmung mit dem Residuenplot wird trotz der Abweichungen in den Schwänzen der Verteilung der stand. Residuen die Annahme der Normalverteilung nicht verworfen.

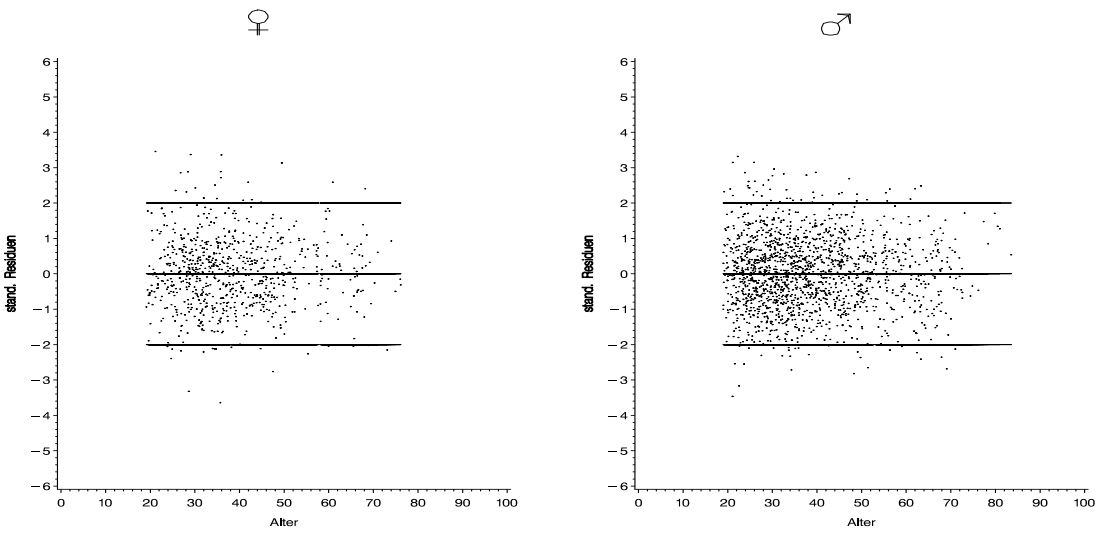


Abbildung 7.17: stand. Residuenplot

Modellgleichung Kovarianzanalyse: $n = 2378$

$$\begin{aligned} FEV_1 &= -7,157 + 5,699H - 0,035AH + 0,871\ln(A) \\ &\quad +0,269HSEX \end{aligned}$$

$$R^2 = 0,699 \quad s_e = 0,513$$

Wie schon aus den getrennten Modellgleichungen ersichtlich, ergibt auch die Kovarianzanalyse geringe Unterschiede zwischen den Geschlechtern. Die Unterschiede beruhen einzig auf einer Wechselwirkung bzgl. der Größe.

7.6 Raucher_schwer

Bei den Frauen sind insgesamt nur sehr wenige Werte vorhanden, die sich ab 60 Jahren ausdünnen. Bei den Männern sind bis etwa 70 Jahren genügend Werte vorhanden.

Modellgleichungen

Frauen: $n = 158$

$$FEV_1 = -4,417 + 5,206H - 0,017AH$$

$$R^2 = 0,607 \quad s_e = 0,378$$

Männer: $n = 1037$

$$FEV_1 = -4,196 + 5,605H - 0,023AH$$

$$R^2 = 0,515 \quad s_e = 0,575$$

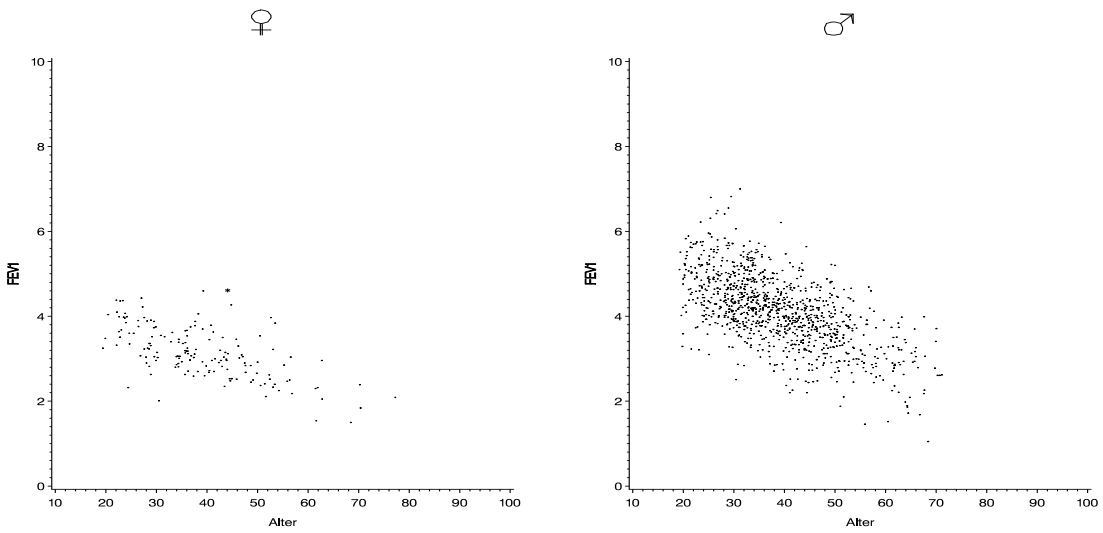


Abbildung 7.18: Scatterplot

Analog zu den Niemalsrauchern(innen) kann in beiden Fällen von einer konstanten Abnahme der Werte mit zunehmendem Alter ausgegangen werden, ersichtlich am Ausschluß der Variablen $\ln(A)$.

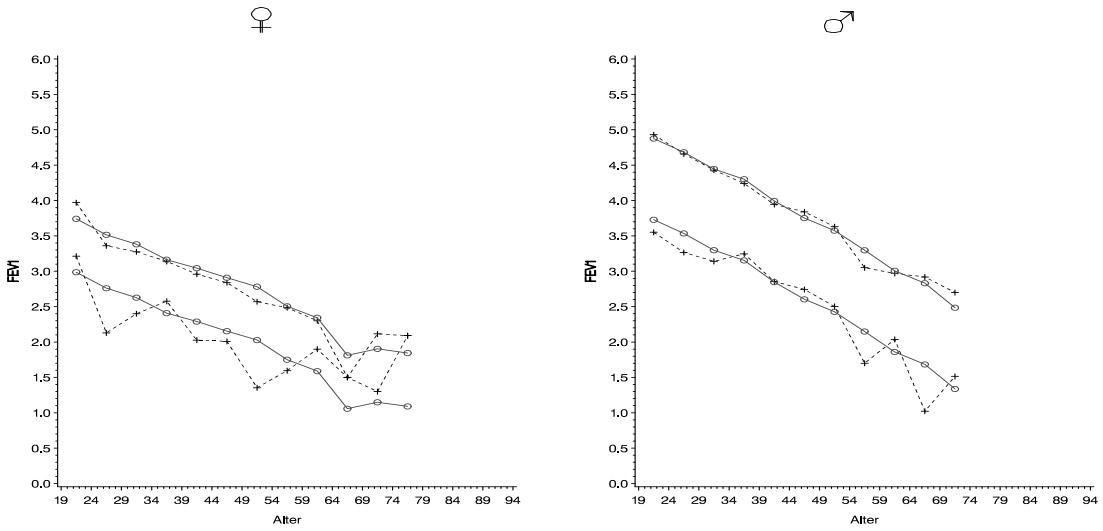


Abbildung 7.19: Medianplot

Aufgrund der geringen Anzahl von Daten weichen die Modellmediane bei den Frauen etwas stärker von den empirischen Medianen ab. Insbesondere ist die empirische Streuung starken Schwankungen unterworfen. Bei den Männern stimmen die Mediane gut überein und die Streuung wird, bis auf stärkere Schwankungen der empirischen Streuung im höheren Altersbereich, gut durch die Modellstreuung nachvollzogen.

In beiden Fällen nehmen die FEV_1-Werte bereits ab dem 20. Lebensjahr ab. So beträgt die Abnahme bei den Frauen zwischen 20 und 60 Jahren 3,5 ml/Jahr und bei den Männern zwischen 20 und 70 Jahren 4,4 ml/Jahr.

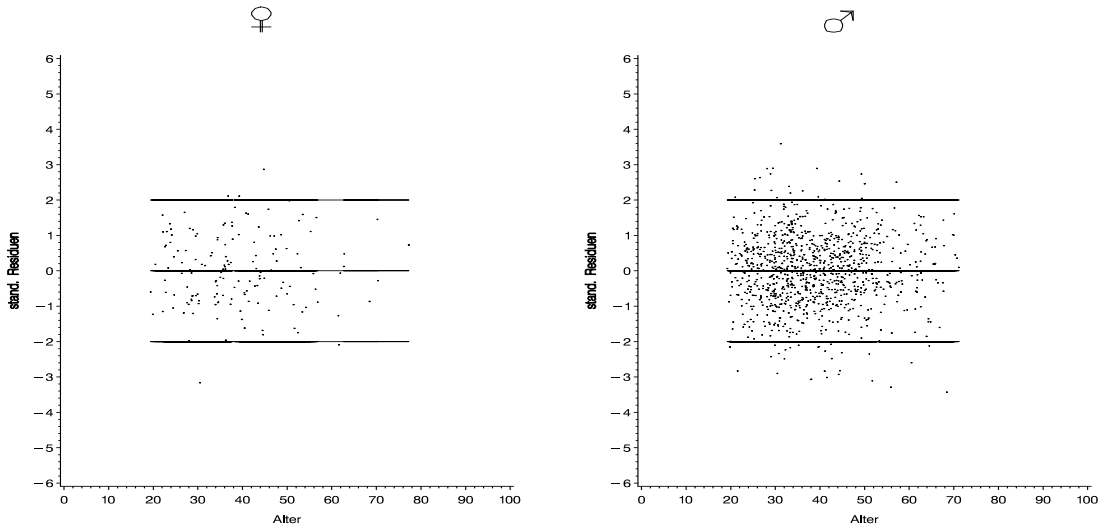


Abbildung 7.20: stand. Residuenplot

Der stand. Residuenplot spricht wiederum für die Güte des Modells.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	157	50,96%	1,91%	1,27%	0,8795
Männer	1037	48,22%	2,12%	2,79%	0,4923

In beiden Fällen wird die Annahme, daß die Residuen normalverteilt sind, nicht verworfen.

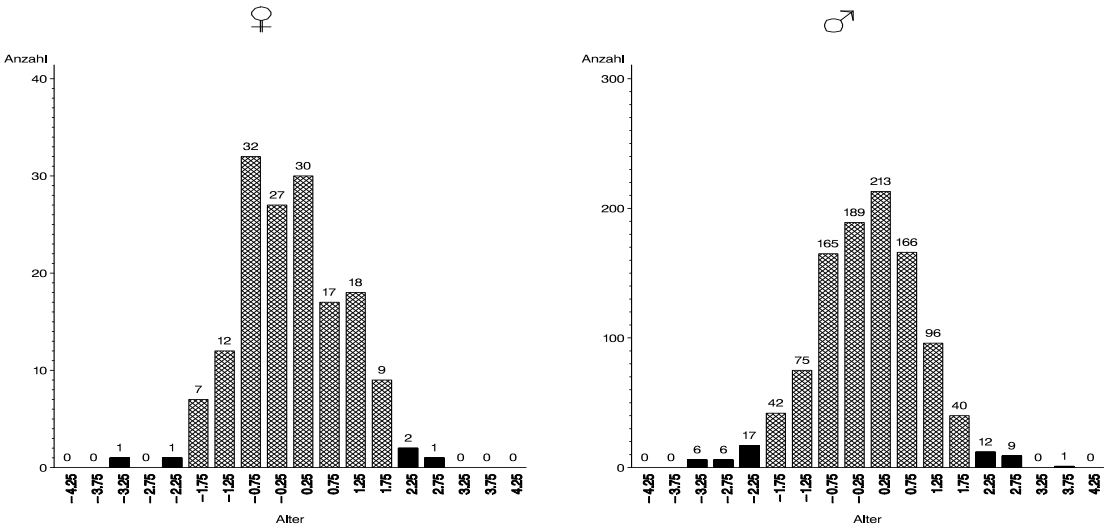


Abbildung 7.21: Histogramm der stand. Residuen

Im Histogramm der stand. Residuen und im Normal Probability Plot sind leichte Abweichungen feststellbar. So sieht man z.B. sehr schön aus dem Histogramm bei den Männern, wie eine zu große Anzahl von negativen Residuen im Bereich von -3,5 bis -3 eine entsprechenden Abweichung im Normal Probability Plot erkennen läßt.

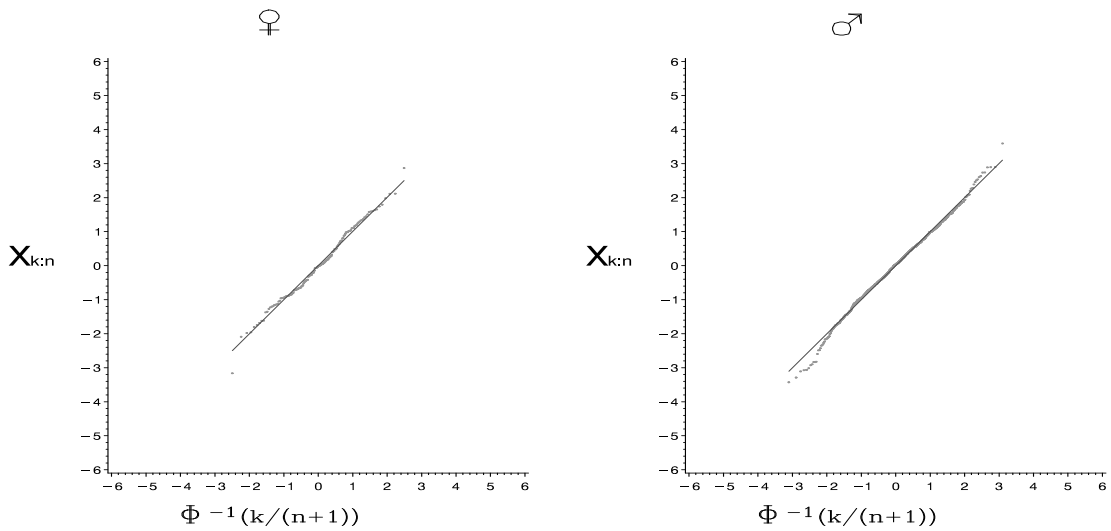


Abbildung 7.22: Normal Probability Plot der stand. Residuen

Modellgleichung Kovarianzanalyse: $n = 1194$

$$\begin{aligned} FEV_1 &= -7,620 + 5,868H - 0,036AH + 0,927\ln(A) \\ &\quad +0,260HSEX \end{aligned}$$

$$R^2 = 0,590 \quad s_e = 0,555$$

Die Kovarianzanalyse bestätigt wieder die geringen Unterschiede zwischen Frauen und Männern. Die Unterschiede sind nur in einer Wechselwirkung bzgl. der Größe begründet. Im Gegensatz zu den getrennten Modellgleichungen wird hier die Variable $\ln(A)$ nicht aus dem Modell entfernt.

Bei beiden Parametern FVC und FEV_1 bestehen in der Charakteristik der Modellgleichungen für Niemals- und starke Raucher(innen) deutliche Unterschiede. Bei den schweren Rauchern(innen) werden die Modellgleichungen ohne die Variable $\ln(A)$ erstellt und die Koeffizienten der verbleibenden erklärenden Variablen stimmen weitgehend überein. Aus der Kovarianzanalyse der schweren Raucher(innen) läßt sich ablesen, daß Unterschiede nur bzgl. der Größe vorhanden sind.

In der Gruppe der Niemalsraucher(innen) haben alle vorgeschlagenen erklärenden Variablen einen signifikanten Anteil an der Erklärung der Gesamtstreuung. Darüberhinaus unterscheiden sich die Modellgleichungen von Frauen und Männern stärker in den Koeffizienten. Ebenso ergibt die Kovarianzanalyse, daß die Unterschiede der Geschlechter nicht nur durch unterschiedliche Körpergrößen zu erklären sind.

Kapitel 8

Quotient FEV_1/FVC

Dieses Kapitel analysiert den Quotienten aus den beiden Parametern FEV_1 und FVC. Der Quotient FEV_1/FVC dient zusammen mit FEV_1 und FVC zur Diagnose von Lungenerkrankungen

Das Modell um FEV_1/FVC zu schätzen ist (siehe Kummer [11] und Rapatz [15]):

$$FEV_1/FVC = \hat{\beta}_0 + \hat{\beta}_1 H^2 + \hat{\beta}_2 \ln(A)$$

Die Kurzbezeichnungen für die erklärenden Variablen sind:
$\boldsymbol{H^2}$: Größe×Größe; **ln(A):** ln(*Alter*)

Die Korrelationskoeffzienten in den folgenden Tabellen zeigen die Stärke des Zusammenhangs zwischen den erklärenden Variablen und FEV_1/FVC bei Frauen und Männern:

	FEV_1/FVC-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
H^2	0,011	0,009	-0,066	-0,079	-0,018	-0,017
ln(A)	**-0,359**	**-0,334**	**-0,403**	**-0,312**	**-0,371**	**-0,365**

	FEV_1/FVC-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
H^2	-0,053	-0,120	0,001	-0,009	0,002	-0,019
ln(A)	**-0,225**	**-0,147**	**-0,312**	**-0,281**	**-0,378**	**-0,310**

Bei den Frauen wie auch bei den Männern ist die Korrelation des Parameters H^2 mit FEV_1/FVC sehr gering und nur bei den Niemalsrauchern (große Anzahl von Werten) *hoch signifikant* von Null verschieden. Es zeigt sich aber, daß eine Abhängigkeit zwischen dem Logarithmus des Alters und dem Parameter FEV_1/FVC besteht.

Die FEV_1/FVC-Werte werden jeweils in Prozent angegeben.

8.1 Niemalsraucher

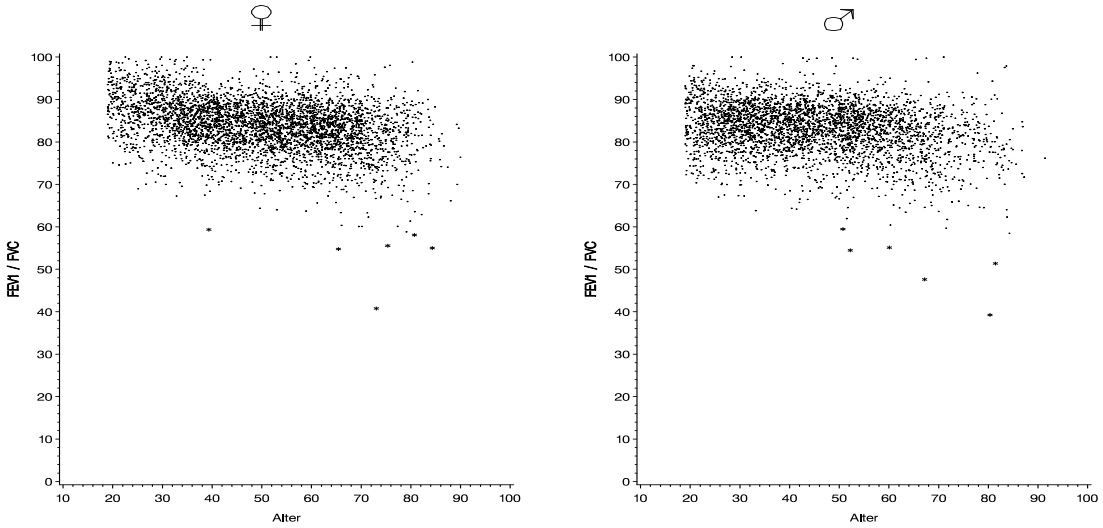


Abbildung 8.1: Scatterplot

Frauen wie Männer zeigen in etwa die gleiche Charakteristik, was die Verteilung der Werte betrifft. Mit zunehmendem Alter ist ein leichter Abfall zu erkennen.

Zur Modellerstellung werden die untransformierten FEV_1/FVC-Werte herangezogen. Die Regressionsmodelle werden analog zu den beiden vorherigen Kapiteln erstellt.

Modellgleichungen

Frauen: $n = 4118$

$$FEV_1/FVC = 117,57 - 3,18H^2 - 6,51\ln(A)$$

$$R^2 = 0,138 \quad s_e = 5,408$$

Männer: $n = 3453$

$$FEV_1/FVC = 111,70 - 3,92H^2 - 4,39\ln(A)$$

$$R^2 = 0,070 \quad s_e = 5,563$$

Besonders auffallend ist bei den Modellen für FEV_1/FVC der sehr kleine Bestimmtheitsgrad. Unterschiede in den Modellen sind im Intercept und in den Koeffizienten der Variablen $\ln(A)$ zu finden.

Sowohl bei den Frauen als auch bei den Männern zeigen die Medianplots eine gute Übereinstimmung der empirischen mit den Modellmedianen. Durch den geringen Einfluß der Größe stimmt auch die Modellstreuung mit der empirischen Streuung gut überein. Mit dem Alter ist ein leichter, nahezu linearer Abfall zu erkennen.

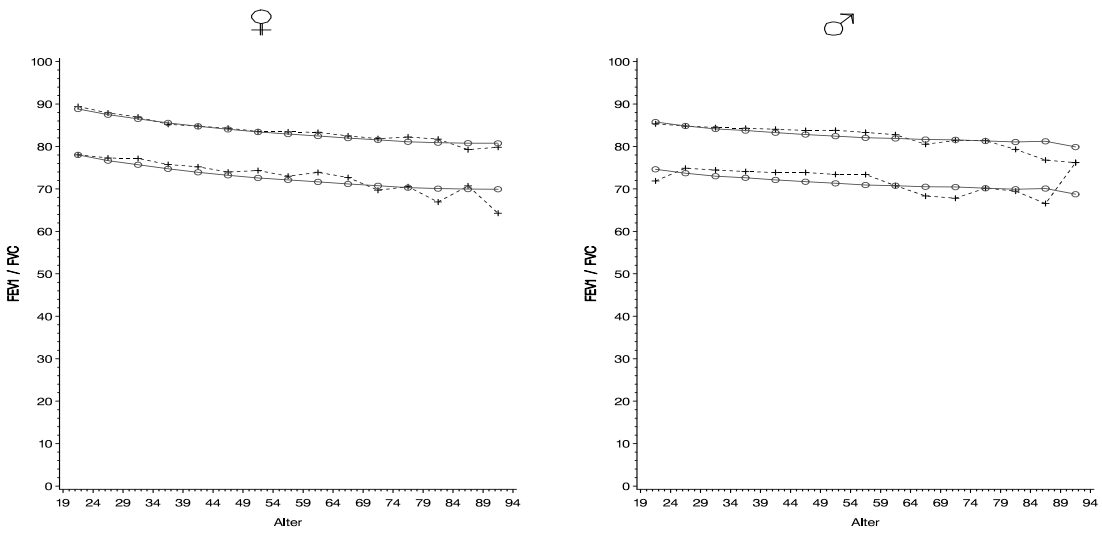


Abbildung 8.2: Medianplot

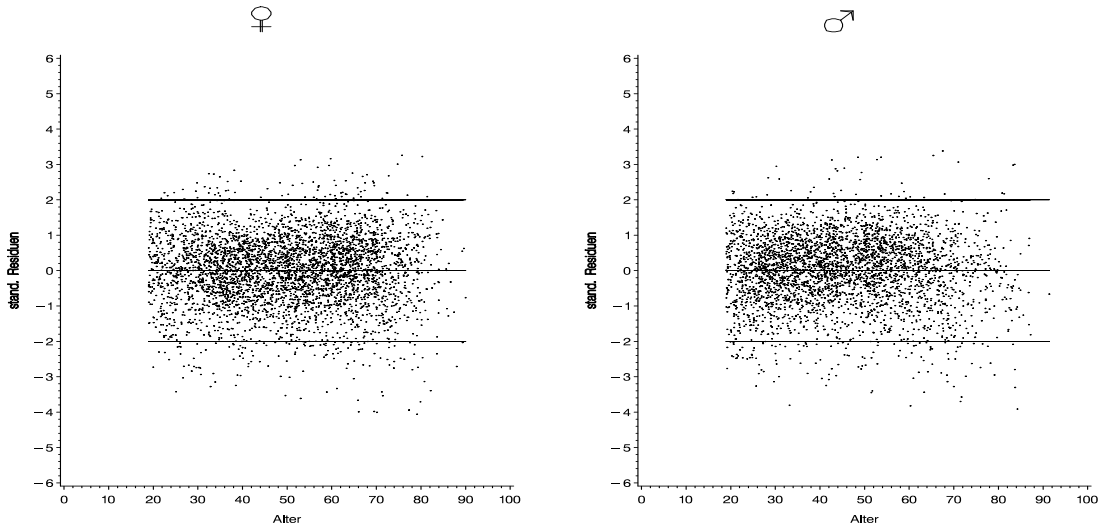


Abbildung 8.3: stand. Residuenplot

Bei Frauen und Männern nehmen die FEV_1/FVC-Werte ungefähr um 0,14% pro Jahr ab.

In den stand. Residuenplots sind keine auffälligen Abweichungen zu erkennen.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	K-S
Frauen	4118	48,01%	1,94%	3,40%	0,001
Männer	3453	45,99%	1,36%	3,76%	0,001

Bei Frauen und Männern sind zuviele Residuen kleiner als $-2s_e$ und die Residuen der Männer sind zusätzlich noch linksschief verteilt. Wie im Histogramm der stand. Residuen und im Normal Probability Plot schön zu sehen, sind vor allem Abweichungen im Bereich der negativen Residuen für die Ablehnung der Annahme der Normalverteilung verantwortlich.

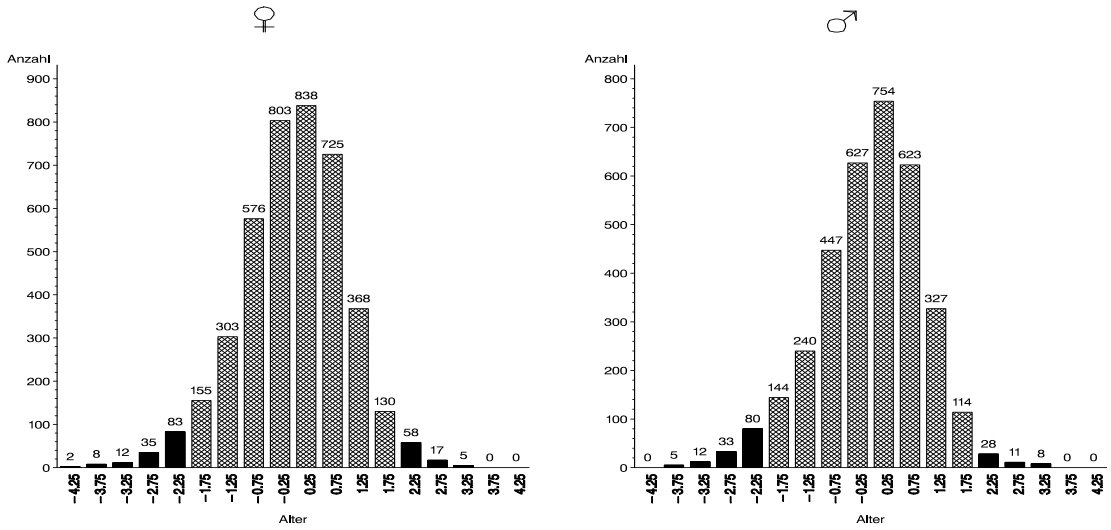


Abbildung 8.4: Histogramm der stand. Residuen

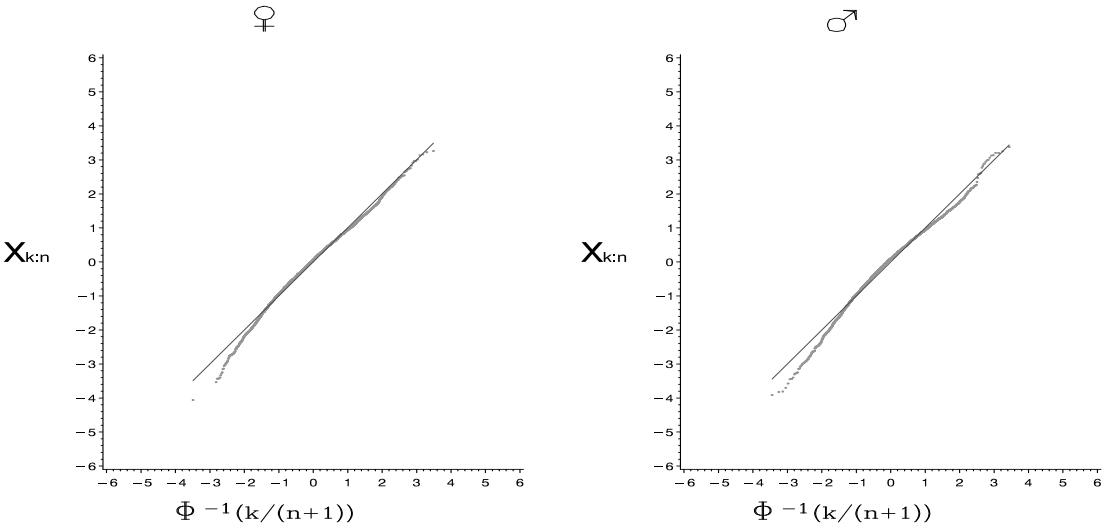


Abbildung 8.5: Normal Probability Plot der stand. Residuen

Modellgleichung Kovarianzanalyse: $n = 7571$

$$\begin{aligned} FEV_1 &= 118,86 - 3,57H^2 - 6,57\ln(A) \\ &\quad -8,51SEX + 2,26H^2SEX \end{aligned}$$

$$R^2 = 0,112 \quad s_e = 5,479$$

Die Unterschiede zwischen Frauen und Männern beruhen hier, in Übereinstimmung mit den Modellgleichungen für Frauen und Männer getrennt, auf einer konstanten Verschiebung und einer Wechselwirkung bzgl. H^2.

8.2 Passivraucher

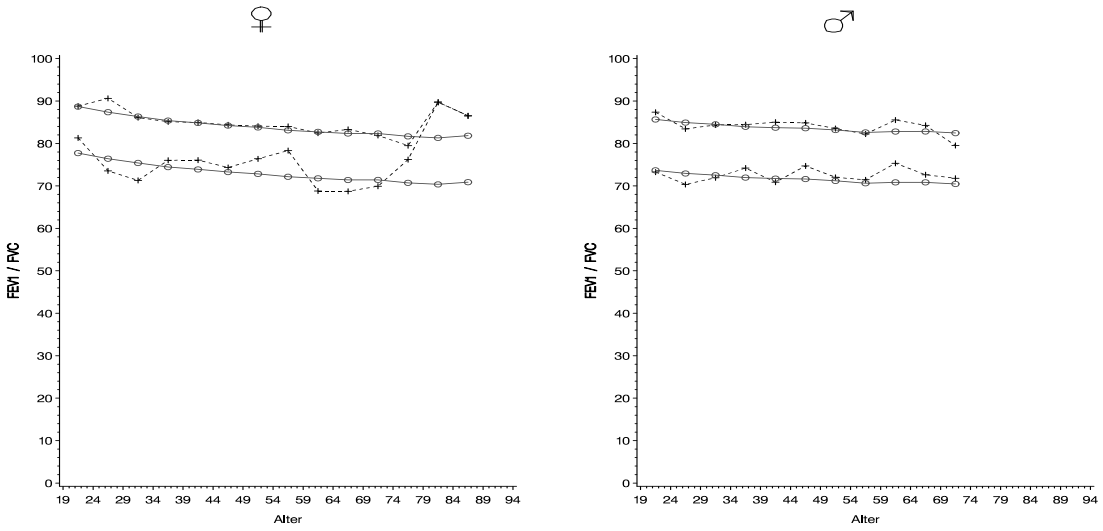


Abbildung 8.6: Medianplot

Die Mediane stimmen bei Frauen und Männern bis etwa 75 Jahren ganz gut überein. Die empirischen Streuungen variieren relativ stark. Die Quotienten der Frauen sinken im Schnitt um 0,15% pro Jahr, während die Quotienten der Männer, ausgehend von einem niedrigeren Anfangswert, um etwa 0,09% pro Jahr fallen.

Modellgleichungen

Frauen: $n = 411$

$$FEV_1/FVC = 105,01 - 5,40\ln(A)$$
$$R^2 = 0,117 \quad s_e = 5,474$$

Männer: $n = 382$

$$FEV_1/FVC = 113,27 - 4,84H^2 - 3,91\ln(A)$$
$$R^2 = 0,050 \quad s_e = 5,991$$

Bei den Frauen findet nur die Variable $\ln(A)$ Eingang in die Regressionsgleichung im Gegensatz zur Gleichung der Männer. Bei den Männern ist der Bestimmtheitsgrad wiederum sehr gering.

8.3 Ex-gelegentliche Raucher

Die Mediane stimmen hier in beiden Fällen bis in hohe Altersbereiche sehr gut und die Streuung gut überein. Zwischen den Geschlechtern sind keine wesentlichen Unterschiede

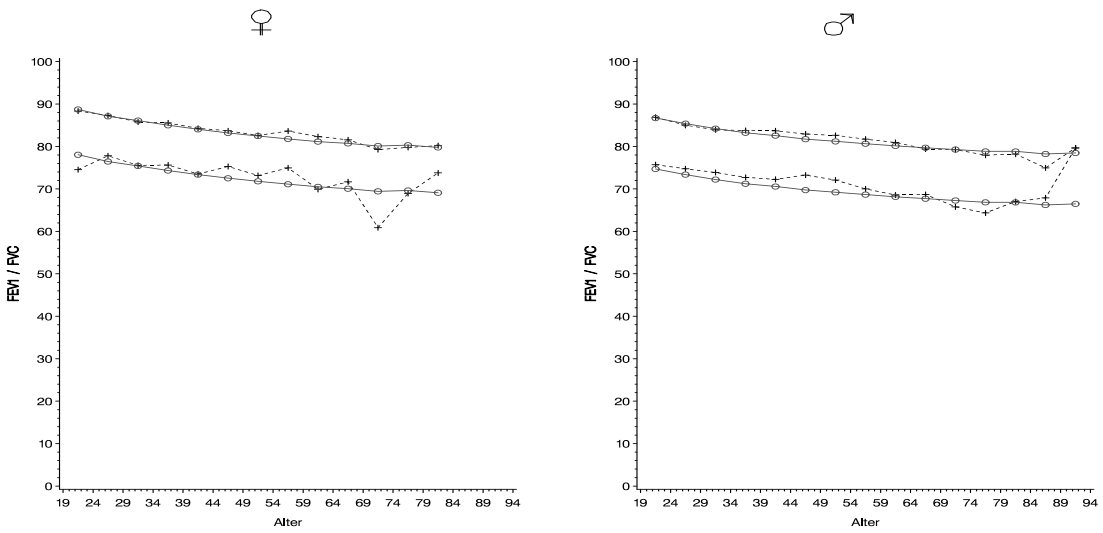


Abbildung 8.7: Medianplot

zu erkennen. Die FEV_1/FVC-Werte nehmen bei Frauen und Männern um etwa 0,13 - 0,15% pro Jahr ab.

Modellgleichungen

Frauen: $n = 1005$

$$FEV_1/FVC = 123,12 - 4,09H^2 - 7,51\ln(A)$$

$$R^2 = 0,177 \quad s_e = 5,336$$

Männer: $n = 3067$

$$FEV_1/FVC = 120,24 - 3,62H^2 - 7,10\ln(A)$$

$$R^2 = 0,112 \quad s_e = 6,000$$

In Übereinstimmung mit den Medianplots unterscheiden sich auch die Modellgleichungen kaum voneinander. Der Bestimmtheitgrad ist diesmal bei den Männern etwas höher im Vergleich zu den Niemals- und Passivrauchern.

8.4 Raucher_leicht

Die empirische Streuung variiert wieder etwas stärker zwischen den einzelnen Altersgruppen. Die Mediane stimmen bis etwa 75 Jahre ganz gut überein. Die FEV_1/FVC-Werte sinken im Schnitt um etwa 11% pro Jahr. Die Männer haben über die Jahre konstant etwas niedrigere Werte als die Frauen.

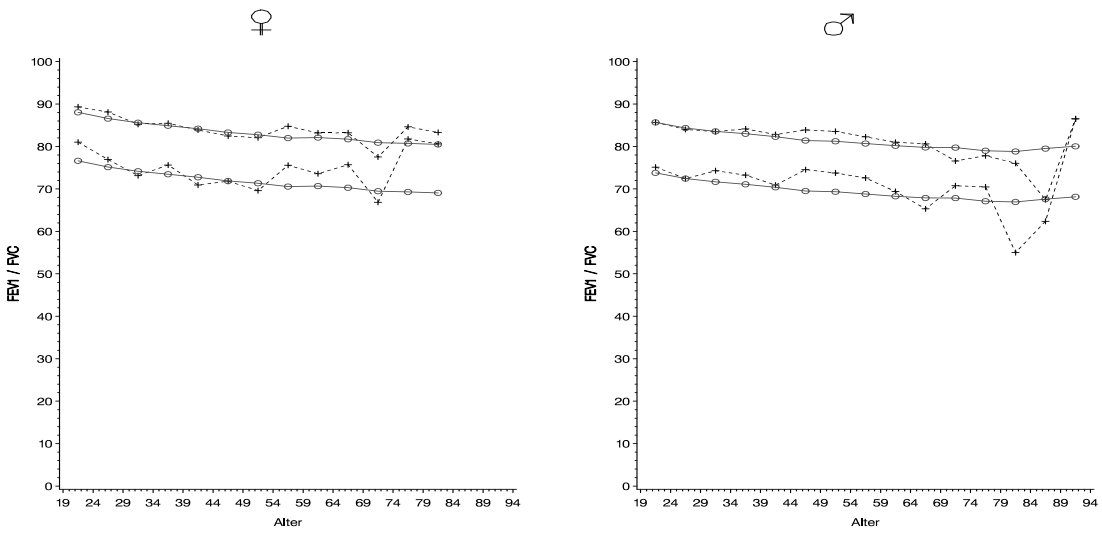


Abbildung 8.8: Medianplot

Modellgleichungen

Frauen: $n = 722$

$$FEV_1/FVC = 119,34 - 4,42H^2 - 6,25\ln(A)$$

$$R^2 = 0,115 \quad s_e = 5,714$$

Männer: $n = 810$

$$FEV_1/FVC = 114,36 - 3,43H^2 - 5,77\ln(A)$$

$$R^2 = 0,105 \quad s_e = 5,947$$

Im Intercept ist ein Unterschied feststellbar. Im Schnitt sind die FEV_1/FVC-Werte bei den Männern etwas niedriger. Mit ein Grund dafü könnte die durchschnittlich größere Körpergröße der Männer sein. Größere Menschen sind aufgrund anatomischer Voraussetzungen nicht in der Lage einen gleich großen Teil ihrer FVC innerhalb einer Sekunde auszuatmen, wie kleinere Menschen.

8.5 Raucher_mittel

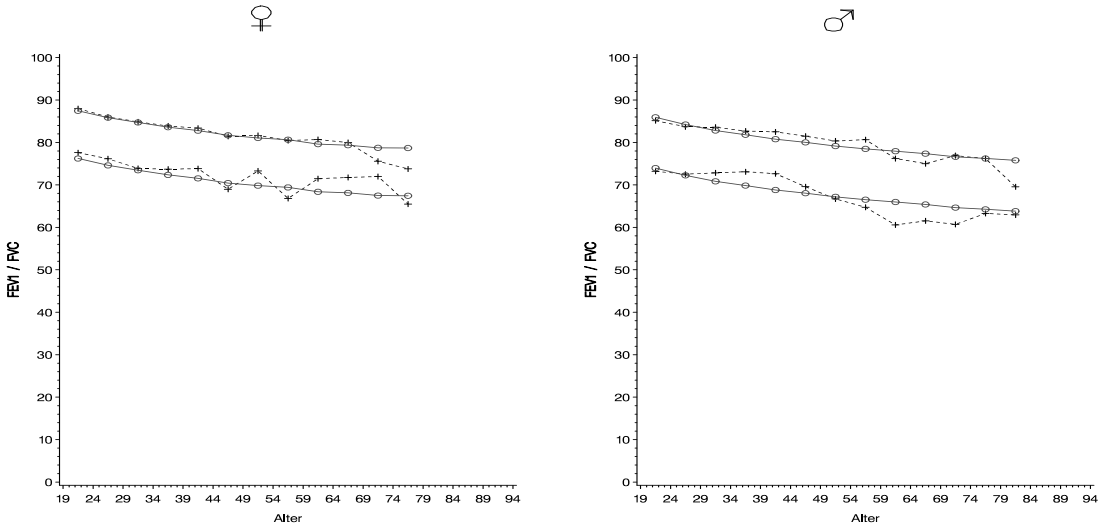


Abbildung 8.9: Medianplot

Die Mediane stimmen besonders bei den Frauen gut überein. Die empirische Streuung variiert wieder etwas stärker. Die Werte der Männer sind tendenziell etwas niedriger. Die Werte der Frauen sinken im Schnitt um 0,15% pro Jahr und die Werte der Männer um etwa 0,18% pro Jahr.

Modellgleichungen

Frauen: $n = 759$

$$FEV_1/FVC = 119,77 - 3,12H^2 - 7,70\ln(A)$$

$$R^2 = 0,142 \quad s_e = 5,622$$

Männer: $n = 1618$

$$FEV_1/FVC = 122,40 - 3,39H^2 - 8,34\ln(A)$$

$$R^2 = 0,158 \quad s_e = 5,987$$

Diesmal zeigt das Modell der Männer eine bessere Anpassung. Sonst unterscheiden sich die Modelle recht wenig voneinander.

8.6 Raucher_schwer

Trotz der wenigen Werte bei den Frauen ist, wie bei den Männern, mit zunehmendem Alter ein Abfall zu erkennen.

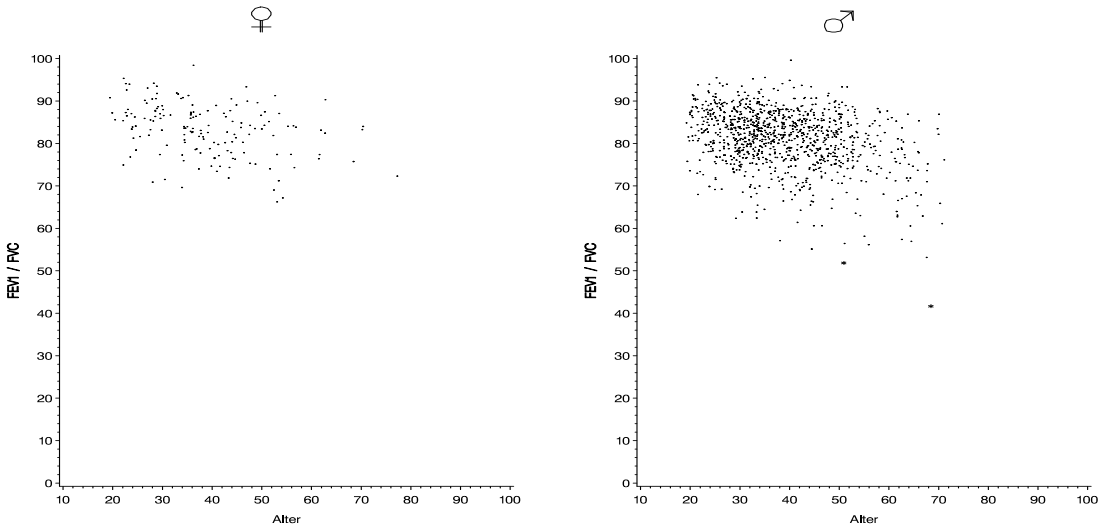


Abbildung 8.10: Scatterplot

Modellgleichungen

Frauen: $n = 158$

$$FEV_1/FVC = 123,15 - 3,74H^2 - 8,19\ln(A)$$

$$R^2 = 0,146 \quad s_e = 5,722$$

Männer: $n = 1035$

$$FEV_1/FVC = 116,70 - 2,54H^2 - 7,62\ln(A)$$

$$R^2 = 0,098 \quad s_e = 6,321$$

Der Unterschied ist wieder vor allem im Intercept zu finden. Bei allen sechs Untergruppen war zu sehen, daß dieses Modell einen sehr geringen Bestimmtheitsgrad aufweist. Der Grund dafür ist, daß der Quotient FEV_1/FVC nur gering mit den Variablen Alter, Größe und Gewicht, sowie Kombinationen und Transformationen dieser drei Variablen korreliert. Wie am Anfang des Kapitels dargelegt, ist der stärkste Zusammenhang noch mit dem Alter gegeben.

Die Anpassung ist diesmal bei den Männern besser als bei den Frauen. Bei den Frauen ist allerdings zu berücksichtigen, daß nur 158 Werte in die Auswertung eingehen. Frauen und Männer haben im Schnitt einen Abfall ihrer FEV_1/FVC-Werte von 0,15% pro Jahr zu erwarten. Aufgrund der FEV_1/FVC-Werte ist kein Unterschied zwischen Nichtrauchern und schweren Rauchern zu erkennen.

In den stand. Residuenplots sind bei den Männern zuwenig große positive und zuviele große negative Residuen zu erkennen.

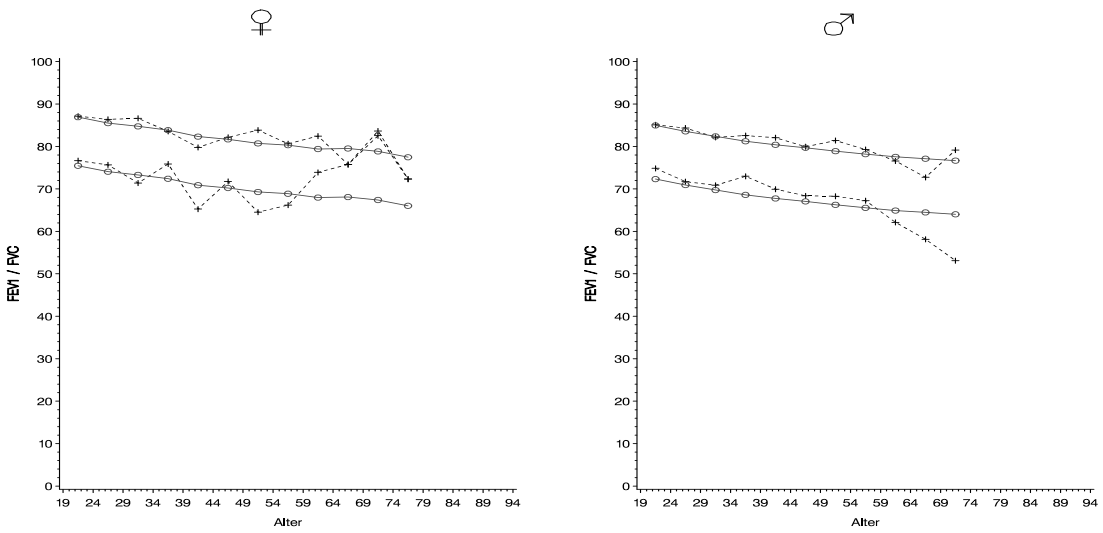


Abbildung 8.11: Medianplot

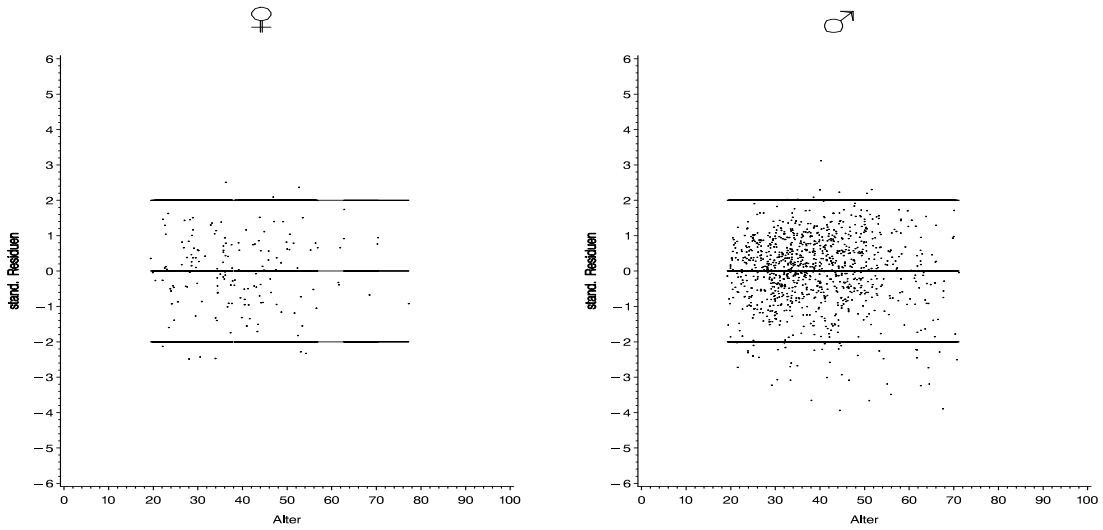


Abbildung 8.12: stand. Residuenplot

	Residuenanalyse				
	n	$> 0 (50\%)$	$> 2s_e (2, 28\%)$	$< -2s_e (2, 28\%)$	S-W
Frauen	158	46,84%	1,90%	3,80%	0,2269
Männer	1035	45,31%	0,87%	4,15%	0,001

Trotz einiger Verzerrungen wird wohl auch aufgrund der geringen Anzahl von Werten bei den Frauen die Annahme der Normalverteilung nicht verworfen. Bei den Männern hingegen sind im Histogramm der stand. Residuen und im Normal Probability Plot klare Abweichungen erkennbar und werden durch den Test auf Normalverteilung auch bestätigt.

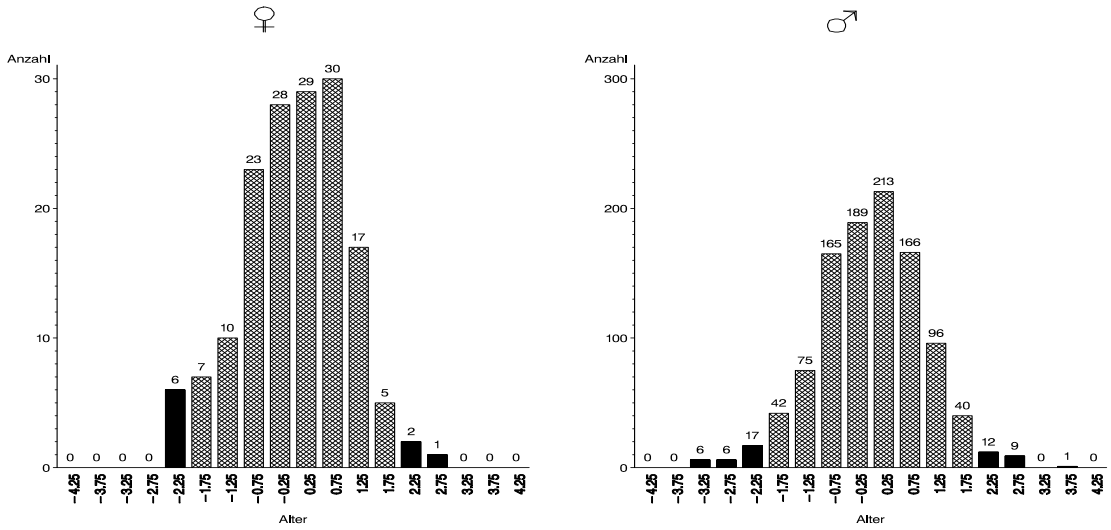


Abbildung 8.13: Histogramm der stand. Residuen

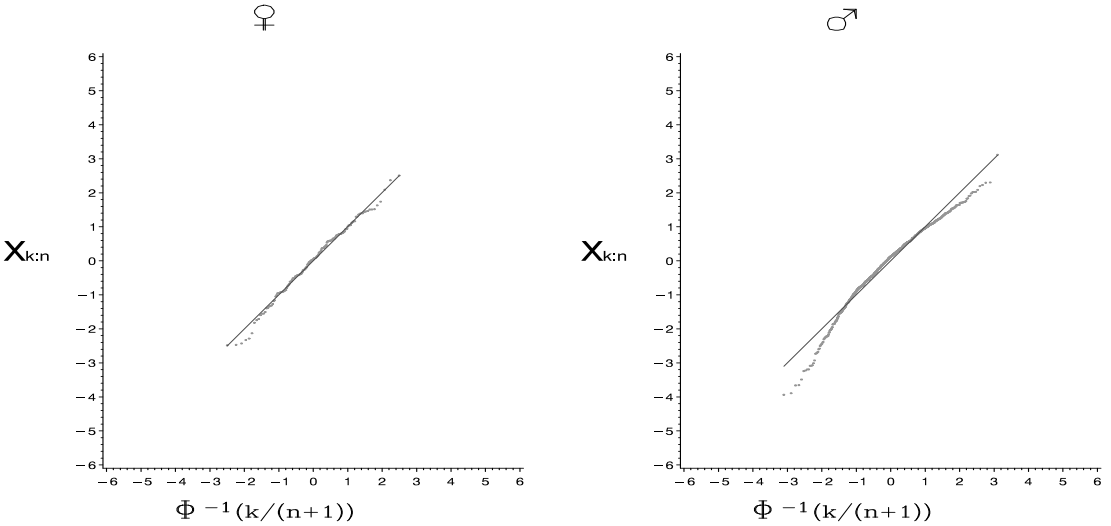


Abbildung 8.14: Normal Probability Plot der stand. Residuen

Modellgleichung Kovarianzanalyse: $n = 1193$

$$FEV_1 = 120,28 - 3,40H^2 - 7,83\ln(A)$$

$$R^2 = 0,112 \quad s_e = 6,246$$

Die Kovarianzanalyse ergibt in diesem Fall keine Unterscheidung zwischen Frauen und Männern.

Kapitel 9

Spitzenfluß PEF

In diesem Kapitel werden Modelle für den Parameter PEF erstellt. Frauen und Männer werden parallel untersucht. In den Untergruppen geht es wieder darum Unterschiede zwischen den Geschlechtern zu verdeutlichen.

In Übereinstimmung mit den Arbeiten von Kummer [11] und Rapatz [15] werden die Flußvolumsparameter vor der Modellerstellung einer Wurzeltransformation unterzogen. In den Medianplots, zur Analyse der Güte der Anpassung der vorhergesagten an die empirischen Werte, werden die rücktransformierten Werte eingetragen.

Um den Parameter PEF vorherzusagen, wird folgendes Modell verwendet:

$$\sqrt{PEF} = \hat{\beta}_0 + \hat{\beta}_1 \ln(H) + \hat{\beta}_2 A + \hat{\beta}_3 A^2$$

Die Kurzbezeichnungen für die erklärenden Variablen sind:
ln(H): ln(Größe) **A**: Alter $\boldsymbol{A^2}$: Alter×Alter

Die Korrelationskoeffzienten in den folgenden Tabellen zeigen die Stärke des Zusammenhangs zwischen den erklärenden Variablen und $\sqrt{PEF}$ bei Frauen und Männern:

	$\sqrt{PEF}$-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,367	0,372	0,306	0,273	0,285	**0,338**
A	-0,536	-0,457	-0,442	-0,380	-0,408	-0,300
A^2	**-0,558**	**-0,474**	**-0,475**	**-0,412**	**-0,423**	-0,322

	$\sqrt{PEF}$-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,393	**0,319**	0,412	0,423	0,376	0,339
A	-0,465	-0,201	-0,510	-0,451	-0,433	-0,398
A^2	**-0,496**	-0,225	**-0,533**	**-0,488**	**-0,459**	**-0,413**

Anhand der Korrelationstabellen ist zu erkennen, daß die $\sqrt{PEF}$-Werte quadratisch vom Alter abhängig sind. Einzig bei den starken Rauchern ist die Korrelation mit der Größe

etwa gleich stark. Insgesamt sind die Korrelationen der abhängigen Variablen mit den erklärenden Variablen geringer, im Vergleich zu den Modellen für FVC und FEV_1. Diese Tatsache spiegelt sich auch in den geringeren Bestimmtheitsgraden bei den Modellen der Flußvolumsparameter wieder.

9.1 Niemalsraucher

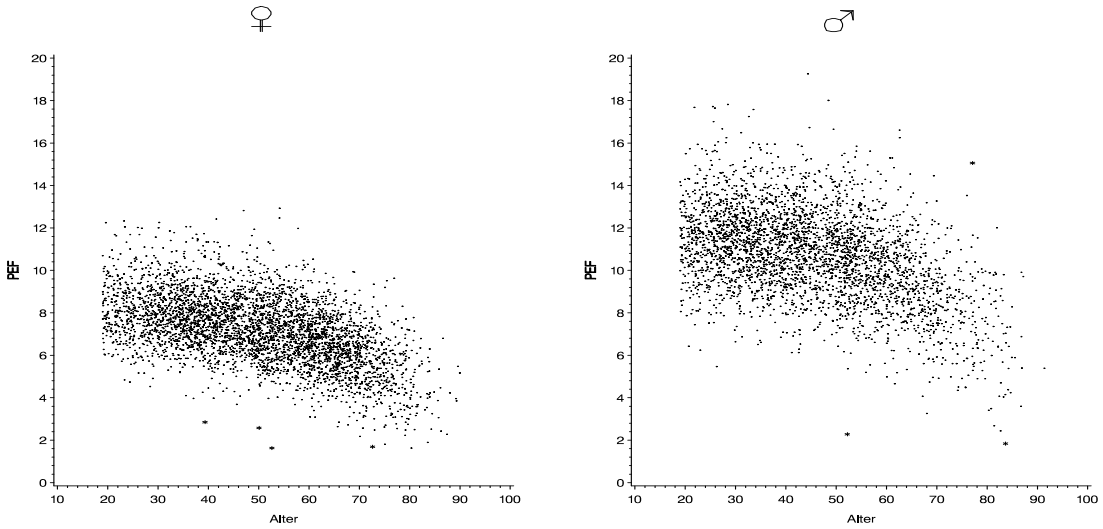


Abbildung 9.1: Scatterplot

Die Werte der Männer sind natürlich wieder etwas höher als jene der Frauen. Bis etwa 50 Jahre sinken die Werte nur leicht. Ab 50 Jahren ist in den Scatterplots ein sich beschleunigender Rückgang der PEF-Werte festzustellen.

Die Modellerstellung erfolgt wiederum in zwei Schritten. Jene Werte, die im ersten Schritt ($\mid r_i^\star \mid \geq 4$) ausselektioniert werden, sind im Scatterplot durch Sterne gekennzeichnet. Die angegebenen Modellgleichungen wurden aus der reduzierten Datenmenge berechnet.

Modellgleichungen

Frauen: $n = 4120$

$$\sqrt{PEF} = 1,77 + 1,79\ln(H) + 0,012A - 0,000207A^2$$

$$R^2 = 0,367 \quad s_e = 0,240$$

Männer: $n = 3456$

$$\sqrt{PEF} = 1,79 + 2,29\ln(H) + 0,017A - 0,000265A^2$$

$$R^2 = 0,325 \quad s_e = 0,269$$

Die Modellgleichungen für Frauen und Männer unterscheiden sich nur wenig voneinander. Die Koeffizienten der Variablen $\ln(H)$ zeigen den Einfluß der Körpergröße und die Koeffizienten der Altersvariablen geben den stärkeren Abfall der PEF-Werte bei den Männern mit zunehmendem Alter wieder. Der Bestimmtheitsgrad und die Modellstreuung stimmen in etwa überein.

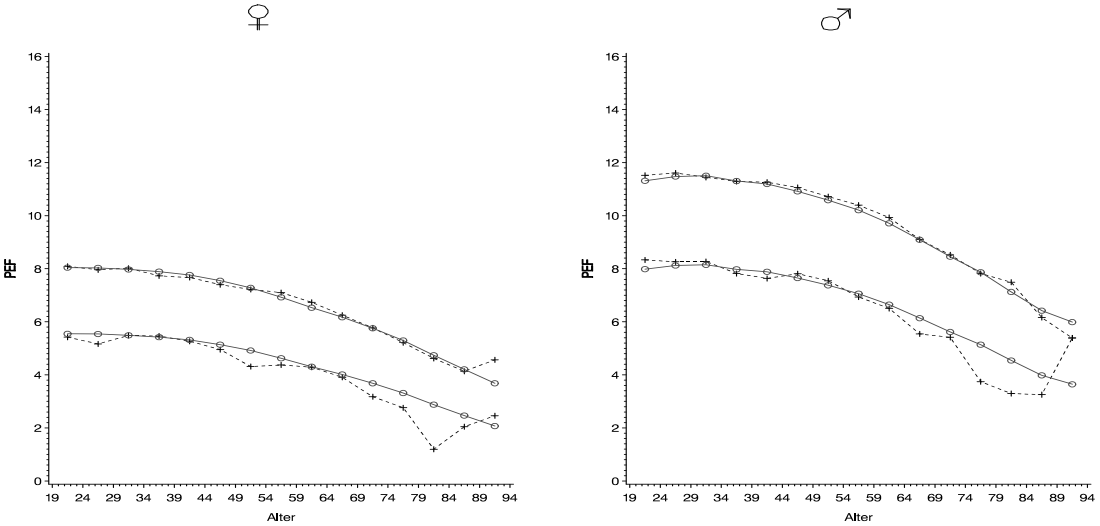


Abbildung 9.2: Medianplot

Die Medianplots zeigen durchwegs gute Übereinstimmung. In höheren Altersbereichen nimmt die empirische Streuung zu. Anhand der Mediane ist die sich beschleunigende Abnahme der PEF-Werte mit zunehmendem Alter klar ersichtlich. Bei den Frauen nehmen die PEF-Werte zwischen 20 und 50 Jahren um durchschnittlich 2 (ml/s)/Jahr und zwischen 50 und 80 Jahren um 8 (ml/s)/Jahr ab (5 (ml/s)/Jahr von 20 bis 80). Die PEF-Werte der Männer verringern sich um 2,67 (ml/s)/Jahr zwischen 20 und 50 Jahren und um 9,67 (ml/s)/Jahr zwischen 50 und 80 Jahren (6,17 (ml/s)/Jahr von 20 bis 80).

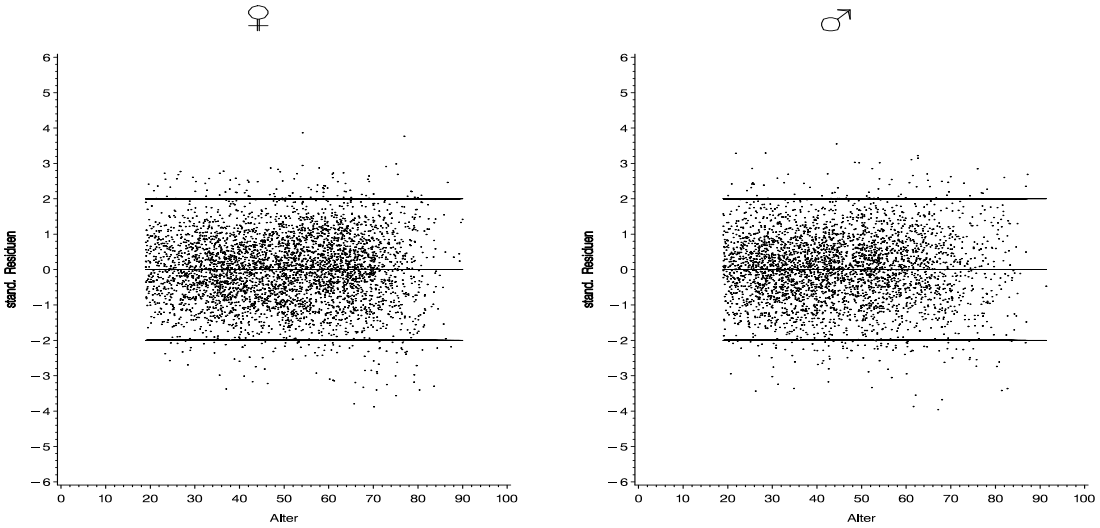


Abbildung 9.3: stand. Residuenplot

In den stand. Residuenplots ist bei den Frauen eine leichte Streuungszunahme der neg. Residuen mit zunehmendem Alter zu erkennen. Insgesamt ist die Modellvoraussetzung, daß die Residuen konstante Streuung besitzen sollen aber durchaus erfüllt.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	K-S
Frauen	4120	49,32%	2,42%	2,33%	0,2000
Männer	3456	48,44%	2,08%	2,92%	0,001

Die Residuen sind symmetrisch verteilt und bei den Frauen wird aufgrund des Tests die Annahme der Normalverteilung der stand. Residuen nicht verworfen. Bei den Männern sind im Bereich der negativen Residuen Abweichungen vorhanden, wodurch die Annahme der Normalverteilung der stand. Residuen abgelehnt wird. Die symmetrische Verteilung der stand. Residuen ist im Histogramm ersichtlich. Im Normal Probability Plot sind besonders bei den neg. Residuen Abweichungen zu erkennen.

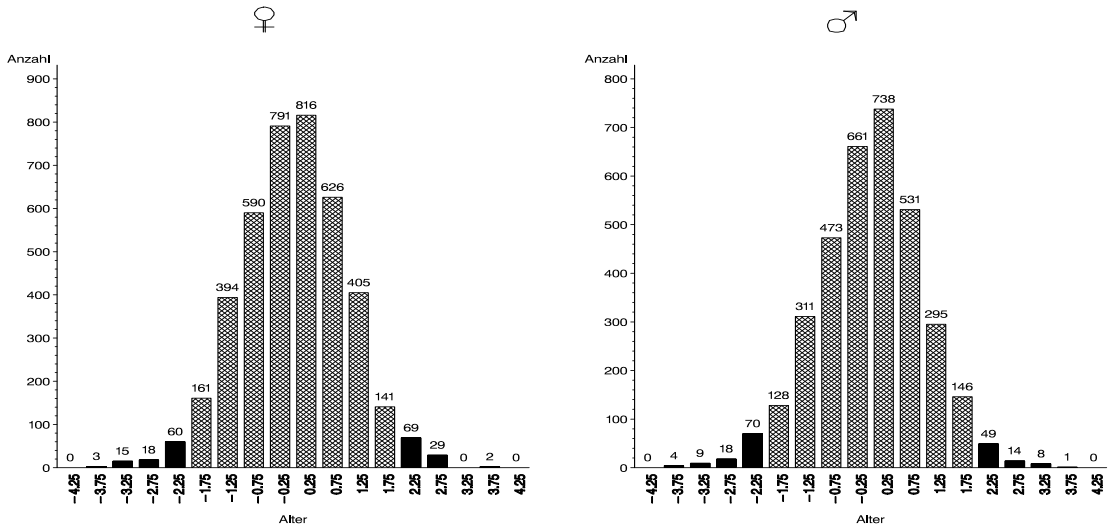


Abbildung 9.4: Histogramm der stand. Residuen

Modellgleichung Kovarianzanalyse: $n = 7580$

$$\begin{aligned}\sqrt{PEF} &= 1,76 + 3,57\ln(H) + 0,011A - 0,000205A^2 \\ &\quad +0,49\ln(H)SEX + 0,007ASEX - 0,000066A^2SEX\end{aligned}$$

$$R^2 = 0,655 \quad s_e = 0,255$$

Die Modellgleichung für Frauen und Männer gemeinsam zeichnet sich durch einen hohen Bestimmtheitsgrad aus. Unterschiede zwischen den Geschlechtern werden durch Wechselwirkungen in allen drei erklärenden Variablen beschrieben.

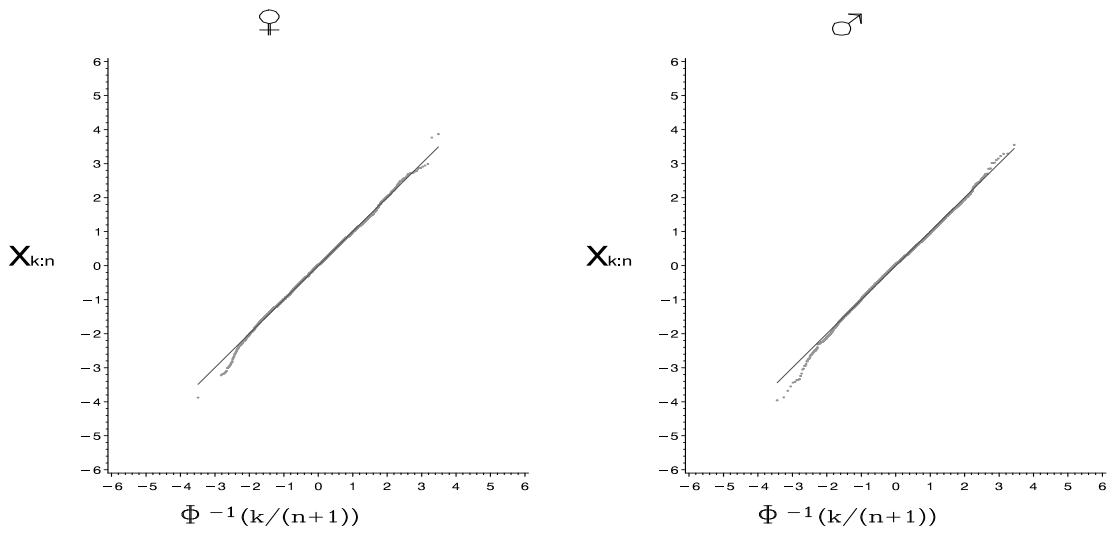


Abbildung 9.5: Normal Probability Plot der stand. Residuen

9.2 Passivraucher

Bei den Frauen sind bis etwa 70 Jahren und bei den Männern bis 60 Jahren hinreichend Werte zur Modellerstellung vorhanden. Die Modellmediane weichen bei den Frauen etwas von den emprischen Medianen ab. Bei den Männern variiert die empirische Streuung hingegen etwas stärker. Bei den Frauen ist eine sich leicht beschleunigende Abnahme der PEF-Werte festzustellen, während die PEF-Werte der Männer bis 40 nahezu konstant bleiben, um ab 40 Jahren aber umso stärker zu fallen.

Die Werte der Frauen sinken zwischen 20 und 50 um 2,67 (ml/s)/Jahr und zwischen 50 und 70 Jahren um 7 (ml/s)/Jahr (4,4 (ml/s)/Jahr von 20 bis 70). Die Werte der Männer beginnen erst ab 40 zu sinken. Von 40 bis 60 Jahren nehmen die PEF-Werte der Männer um 7,5 (ml/s)/Jahr ab.

Modellgleichungen

Frauen: $n = 411$

$$\sqrt{PEF} = 1,86 + 2,06 \ln(H) - 0,000082A^2$$

$$R^2 = 0,289 \quad s_e = 0,238$$

Männer: $n = 382$

$$\sqrt{PEF} = 1,75 + 2,15 \ln(H) + 0,022A - 0,000303A^2$$

$$R^2 = 0,142 \quad s_e = 0,255$$

Bei den Frauen hat die Variable Alter diesmal keinen signifikanten Einfluá. Der Bestimmtheitsgrad ist bei den Männern sehr klein.

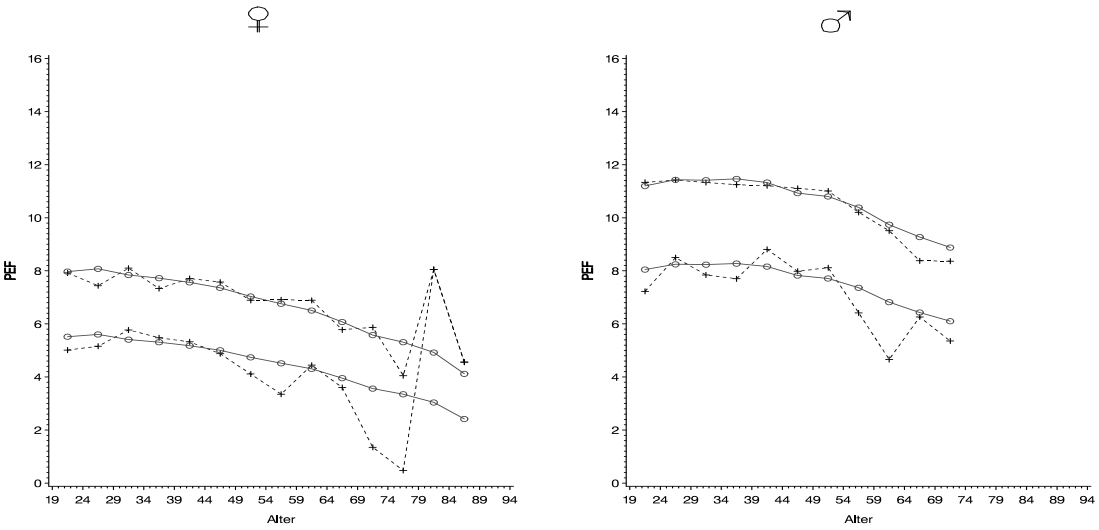


Abbildung 9.6: Medianplot

9.3 Ex-gelegentliche Raucher

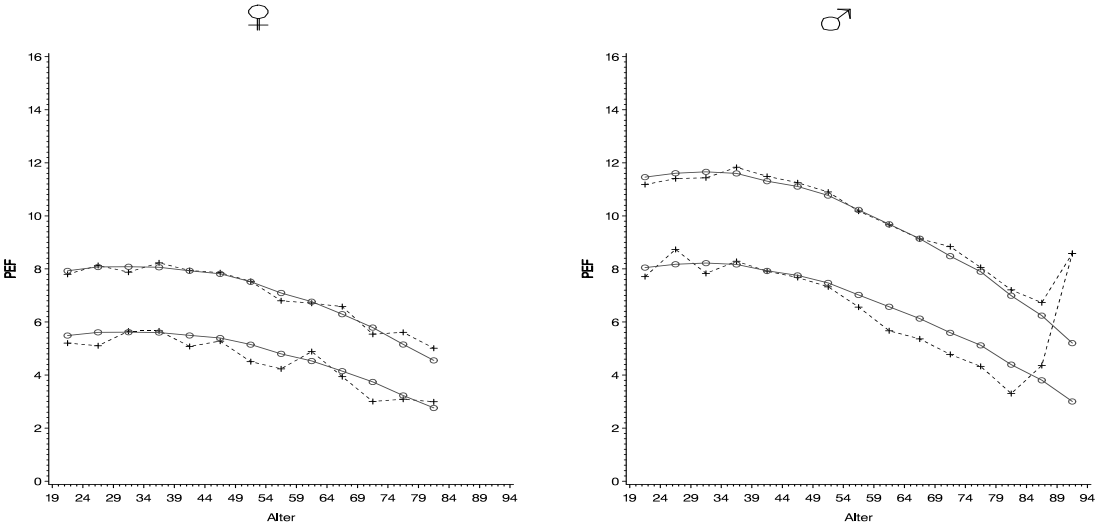


Abbildung 9.7: Medianplot

In beiden Fällen ist die Anpassung der Mediane sehr gut. Die empirische Streuung nimmt mit dem Alter zu. Deutlich ist auch zu erkennen, daß die PEF-Werte bis etwa 40 Jahren konstant bleiben, wenn nicht sogar bis 30 noch etwas steigen, um hernach, besonders bei den Männern, deutlich abzunehmen.

Die Werte der Frauen nehmen zwischen 20 und 50 Jahren um 1 (ml/s)/Jahr und zwischen 50 und 80 um 8,67 (ml/s)/Jahr ab (4,83 (ml/s)Jahr von 20 bis 80). Bei den Männern nehmen die Werte zwischen 20 und 50 Jahren um 1,67 (ml/s)/Jahr und zwischen 50 und 80 um 11.3 (ml/s)/Jahr ab (6,5 (ml/s)/Jahr von 20 bis 80).

Modellgleichungen

Frauen: $n = 1006$

$$\sqrt{PEF} = 1,68 + 1,70\ln(H) + 0,019A - 0,000284A^2$$

$$R^2 = 0,295 \quad s_e = 0,236$$

Männer: $n = 3066$

$$\sqrt{PEF} = 1,72 + 2,42\ln(H) + 0,018A - 0,000279A^2$$

$$R^2 = 0,363 \quad s_e = 0,273$$

Die Modellgleichungen zeigen den Einfluß der Größe auf die PEF-Werte. Ansonsten unterscheiden sich Frauen und Männer nicht sehr voneinander.

9.4 Raucher_leicht

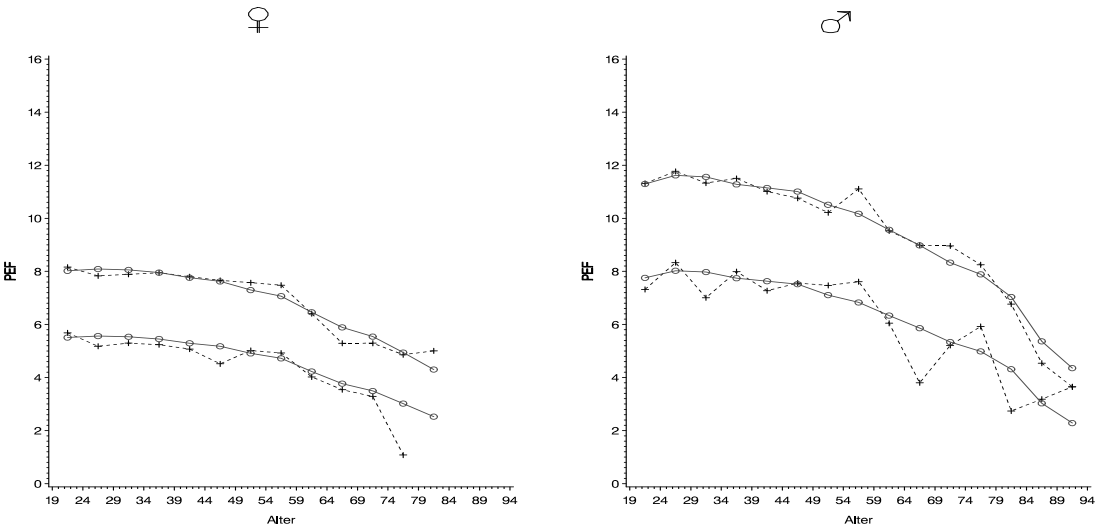


Abbildung 9.8: Medianplot

Bei Frauen und Männern sind bis etwa 70 Jahren genügend Werte zur Modellerstellung vorhanden. Die Modellmediane weichen in den höheren Altesbereichen etwas von den empirischen Medianen ab. Ebenso unterliegt die empirische Streuung größeren Schwankungen. Die PEF-Werte der Frauen bleiben bis 50 Jahren beinahe konstant. Jene der Männer erreichen um 30 Jahren ein Maximum.

Die Werte der Frauen nehmen zwischen 20 und 50 Jahren um 2,3 (ml/s)/Jahr und zwischen 50 und 70 um 8,5 (ml/s)/Jahr ab (4,8 (ml/s)/Jahr von 20 bis 70). Bei den Männern nehmen die Werte zwischen 20 und 50 Jahren um 3,3 (ml/s)/Jahr und zwischen 50 und 70 um 8,8 (ml/s)/Jahr ab (5,6 (ml/s)/Jahr von 20 bis 70).

Modellgleichungen

Frauen: $n = 722$

$$\sqrt{PEF} = 1,87 + 1,53\ln(H) + 0,014A - 0,000243A^2$$

$$R^2 = 0,206 \quad s_e = 0,243$$

Männer: $n = 813$

$$\sqrt{PEF} = 1,56 + 2,58\ln(H) + 0,020A - 0,000299A^2$$

$$R^2 = 0,334 \quad s_e = 0,288$$

Der Unterschied zwischen den Geschlechtern liegt vor allem in der Größe.

9.5 Raucher_mittel

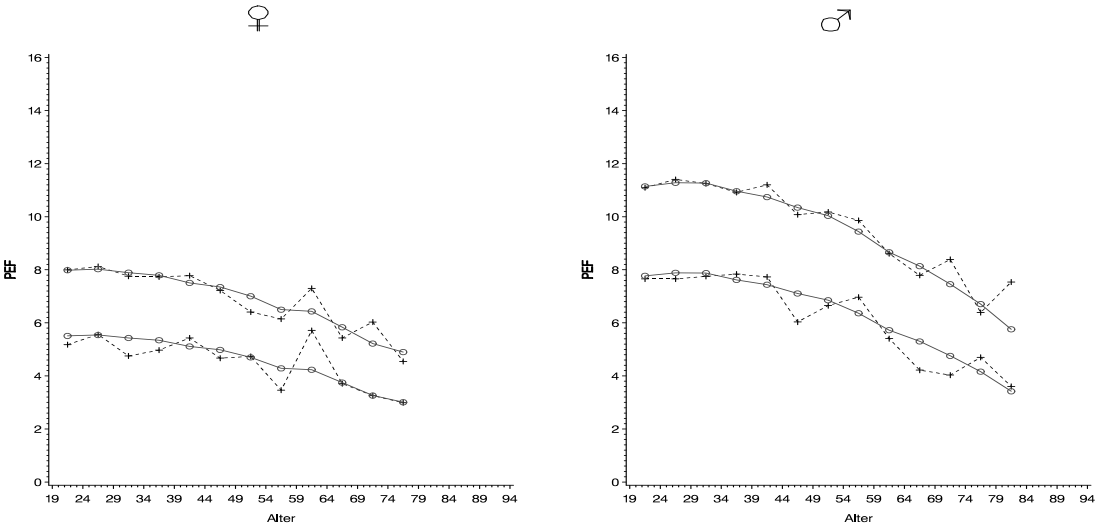


Abbildung 9.9: Medianplot

In den Medianplots sind wieder die schon bekannten Abweichungen in den höheren Altersgruppen zu sehen. Bei den Männern ist eine stärkere Krümmung zu erkennen.

Die Werte der Frauen verringern sich zwischen 20 und 50 Jahren um 3 (ml/s)/Jahr und zwischen 50 und 70 um 7,6 (ml/s)/Jahr (4,8 (ml/s)/Jahr von 20 bis 70). Bei den Männern nehmen die Werte zwischen 20 und 50 Jahren ebenfalls um 3 (ml/s)/Jahr und zwischen 50 und 70 um 8 (ml/s)/Jahr ab (6 (ml/s)/Jahr von 20 bis 70).

Modellgleichungen

Frauen: $n = 760$

$$\sqrt{PEF} = 2,08 + 1,64 \ln(H) - 0,000099A^2$$

$$R^2 = 0,222 \quad s_e = 0,240$$

Männer: $n = 1036$

$$\sqrt{PEF} = 1,74 + 2,51 \ln(H) + 0,013A - 0,000268A^2$$

$$R^2 = 0,240 \quad s_e = 0,279$$

Bei den Frauen fällt die Variable Alter, wie bei den Passivraucherinnen, aus dem Modell heraus.

9.6 Raucher_schwer

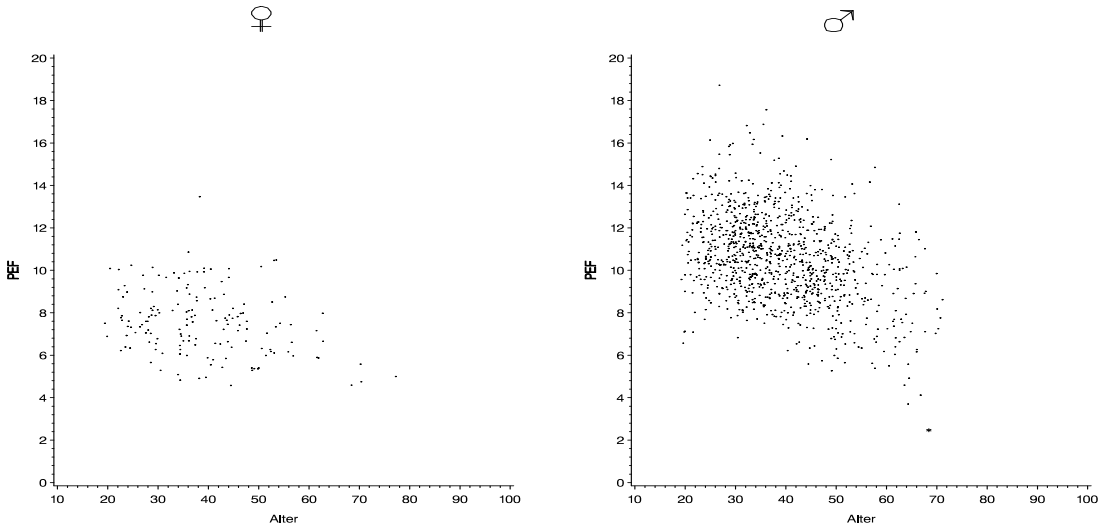


Abbildung 9.10: Scatterplot

Bei den Frauen sind nur bis knapp 60 Jahren und bei den bis 65 Jahren Werte zur Modellerstellung vorhanden.

Modellgleichungen

Frauen: $n = 158$

$$\sqrt{PEF} = 1,74 + 2,21 \ln(H) - 0,000072A^2$$

$$R^2 = 0,178 \quad s_e = 0,256$$

Männer: $n = 1036$

$$\sqrt{PEF} = 2,02 + 2,46 \ln(H) - 0,000120A^2$$

$$R^2 = 0,235 \quad s_e = 0,280$$

Sowohl bei den Frauen als auch bei den Männern hat die Variable Alter keinen signifikanten Einfluß auf die Modellerstellung.

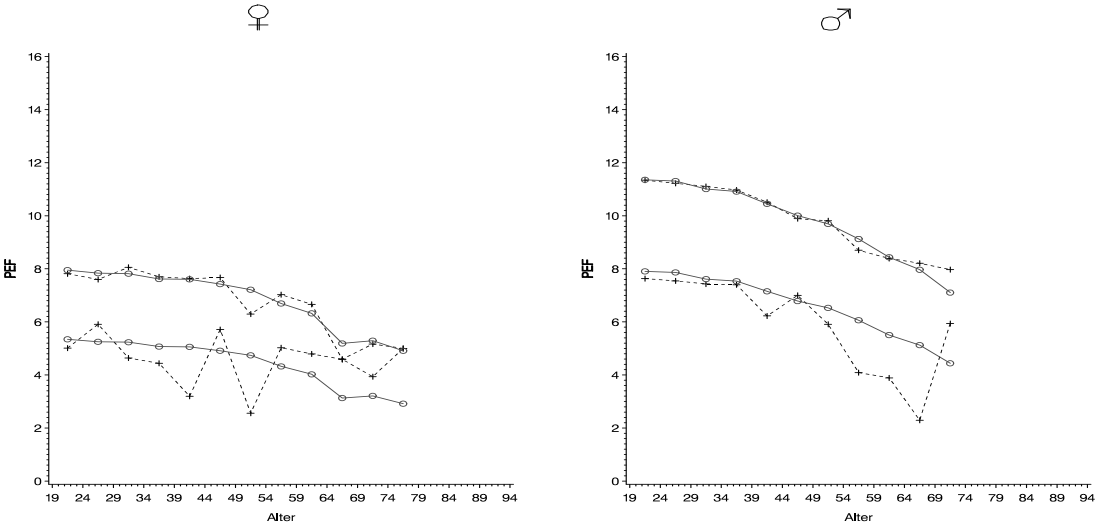


Abbildung 9.11: Medianplot

Die Mediane zeigen trotz der wenigen Werte bei den Frauen durchaus noch eine aktzeptable Anpassung. Die empirische Streuung unterscheidet sich doch schon wesentlich von der Modellstreuung.

Die PEF-Werte der Frauen sinken zwischen 20 und 60 Jahren um 4 (ml/s)/Jahr. Jene der Männer zwischen 20 und 65 um 8,9 (ml/s)/Jahr.

Die stand. Residuen genügen den Modellvoraussetzungen.

	Residuenanalyse				
	n	$> 0(50\%)$	$> 2s_e(2,28\%)$	$< -2s_e(2,28\%)$	S-W
Frauen	159	52,53%	1,27%	1,90%	0,9138
Männer	1037	48,46%	2,41%	2,61%	0,5685

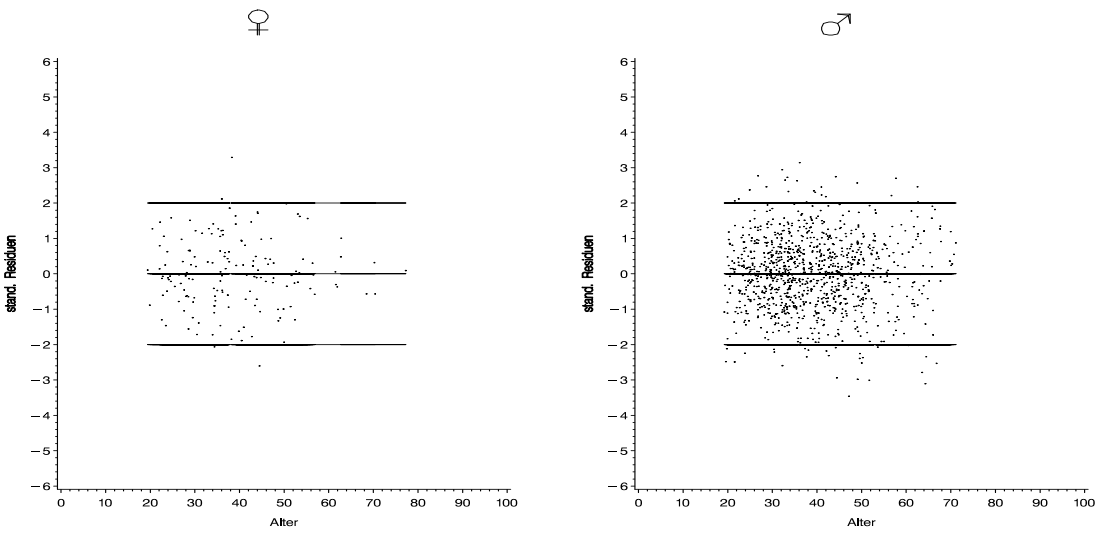


Abbildung 9.12: stand. Residuenplot

Die Annahme der Normalverteilung der Residuen wird in beiden Fällen nicht verworfen. Aufgrund der wenigen Werte in der Gruppe der Frauen fallen die vorhandenen Abweichungen nicht so ins Gewicht.

Modellgleichung Kovarianzanalyse: $n = 1194$

$$\begin{aligned}\sqrt{PEF} &= 1,41 + 2,49\ln(H) + 0,013A - 0,000264A^2 \\ &\quad +0,327SEX\end{aligned}$$

$$R^2 = 0,404 \quad s_e = 0,276$$

Hier tritt der Fall ein, daß sich die Modelle von Frauen und Männern nur um eine Konstante unterscheiden. d.h., daß die Regressionskurven einen parallelen Verlauf haben. Die PEF-Werte der Männer sind im Schnitt um 32,7 ml/s höher.

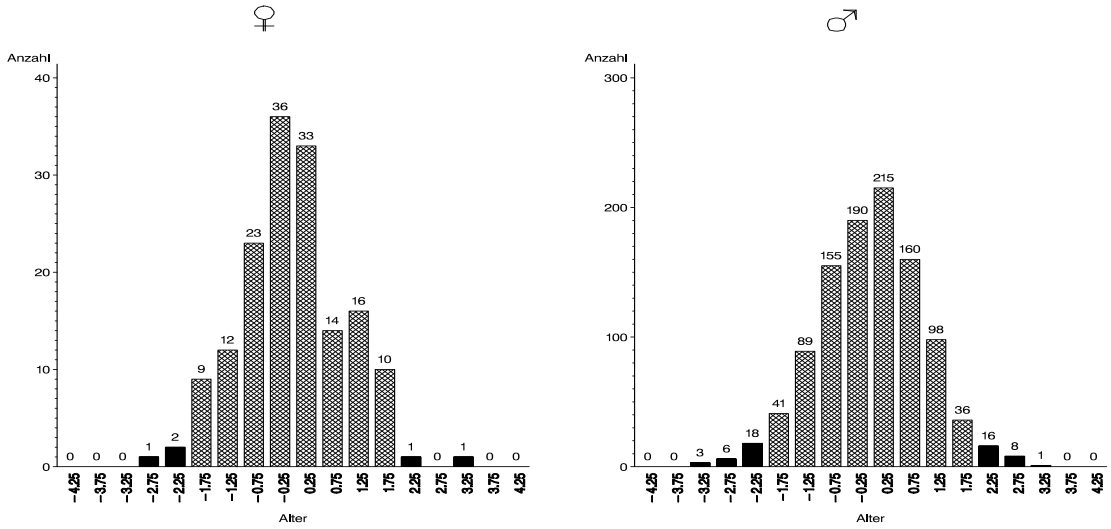


Abbildung 9.13: Histogramm der stand. Residuen

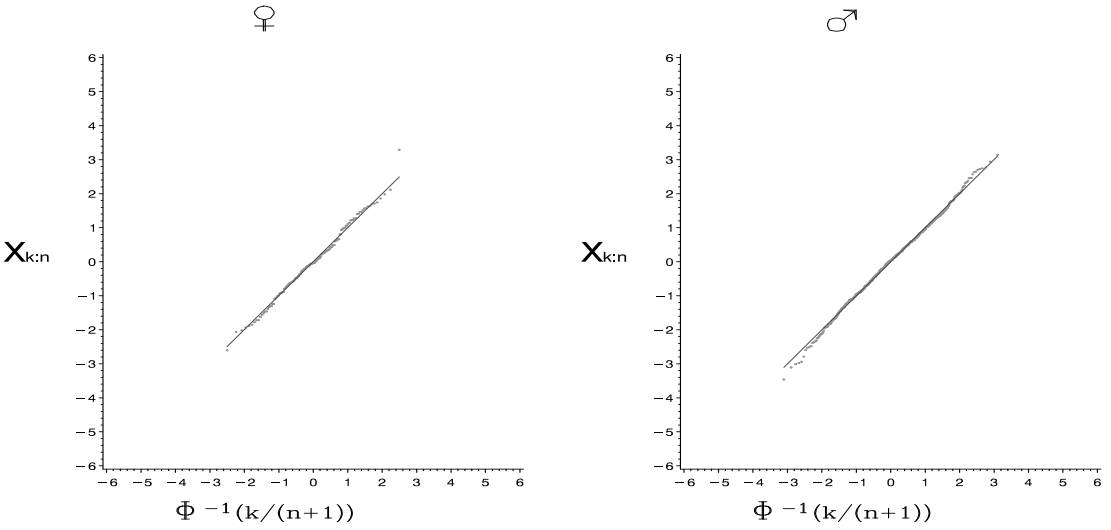


Abbildung 9.14: Normal Probability Plot der stand. Residuen

Kapitel 10

MEF_{75}, MEF_{50} und MEF_{25}

Die Analyse der drei Flußvolumsparameter MEF_{75}, MEF_{50}und MEF_{25} erfolgt in gekürzter Form. Die Charakteristiken entsprechen denen des Parameters PEF.

Die Flußvolumsparameter werden vor der Modellerstellung einer Wurzeltransformation unterzogen. In den Medianplots, zur Analyse der Güte der Anpassung der vorhergesagten an die empirischen Werte, werden die rücktransformierten Werte eingetragen.

Um die Flußvolumsparameter zu schätzen, wird der gleiche Modelltyp wie für die Variable PEF verwendet:

$$\sqrt{MEF} = \hat{\beta}_0 + \hat{\beta}_1 \ln(H) + \hat{\beta}_2 A + \hat{\beta}_3 A^2$$

Die Kurzbezeichnungen für die erklärenden Variablen sind:

H: Größe; **A**: Alter; $\boldsymbol{A^2}$: Alter×Alter

Die Korrelationskoeffzienten in den folgenden Tabellen zeigen die Stärke des Zusammenhangs zwischen den erklärenden Variablen und $\sqrt{MEF_{75}}$, $\sqrt{MEF_{50}}$ und $\sqrt{MEF_{25}}$ bei Frauen und Männern:

	$\sqrt{MEF_{75}}$-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,308	0,317	0,217	0,199	0,235	0,299
A	-0,472	-0,351	-0,399	-0,328	-0,406	-0,283
A^2	**-0,497**	**-0,366**	**-0,430**	**-0,353**	**-0,421**	**-0,300**

	$\sqrt{MEF_{75}}$-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,303	**0,204**	0,298	0,329	0,314	0,230
A	-0,356	-0,052	-0,438	-0,402	-0,430	-0,376
A^2	**-0,393**	-0,065	**-0,462**	**-0,440**	**-0,455**	**-0,394**

Sowohl bei den Frauen als auch bei den Männern sind die stärksten Korrelationen mit der Variablen A^2 gegeben. Einzig bei den Passivrauchern ist praktisch kein Zusammenhang mit dem Alter feststellbar.

	$\sqrt{MEF_{50}}$-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,271	0,268	0,188	0,147	0,219	0,191
A	-0,503	-0,449	-0,486	-0,376	-0,468	**-0,419**
A^2	**-0,517**	**-0,452**	**-0,502**	**-0,386**	**-0,480**	-0,418

	$\sqrt{MEF_{50}}$-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,251	0,161	0,244	0,276	0,266	0,195
A	-0,410	**-0,194**	-0,489	-0,473	-0,510	-0,430
A^2	**-0,436**	-0,188	**-0,503**	**-0,501**	**-0,525**	**-0,440**

Im Vergleich zu MEF_{75} ist hier die Abhängigkeit des Flußvolumsparameters MEF_{50} vom Alter noch stärker ausgeprägt.

	$\sqrt{MEF_{25}}$-Frauen					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,304	0,272	0,209	0,174	0,257	0,329
A	**-0,671**	**-0,640**	**-0,686**	**-0,575**	**-0,657**	**-0,621**
A^2	-0,658	-0,613	-0,683	-0,564	-0,645	-0,597

	$\sqrt{MEF_{25}}$-Männer					
	nie	passiv	ex-gel	1-10	11-20	>20
ln(H)	0,297	0,222	0,305	0,330	0,280	0,246
A	**-0,590**	**-0,493**	**-0,640**	**-0,645**	**-0,645**	**-0,578**
A^2	-0,589	-0,480	-0,638	**-0,645**	-0,641	-0,572

MEF_{25} ist stark vom Alter aber auch nicht unwesentlich von der Größe abhängig.

10.1 Niemalsraucher: MEF_{75}

Die Verteilung der MEF_{75}-Werte zeigt die gleiche Charakteristik wie jene der PEF-Werte. Bis etwa 50 Jahre sinken die Werte nur leicht, während sich ab diesem Zeitpunkt die Abnahme zunehmend beschleunigt.

Die Erstellung der Modellgleichungen erfolgt wiederum im schon bekannten Zweischritt-verfahren. Jene Werte, die an der Modellerstellung nicht teilnehmen, sind im Scatterplot durch Sterne hervorgehoben.

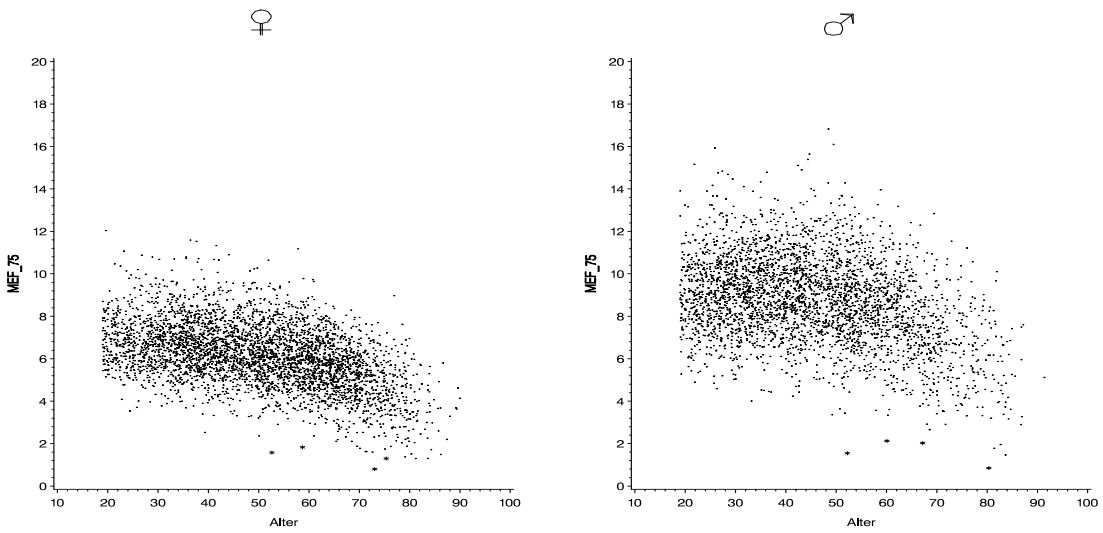


Abbildung 10.1: Scatterplot

Modellgleichungen

Frauen: $n = 4120$

$$\sqrt{MEF_{75}} = 1,66 + 1,44\ln(H) + 0,014A - 0,000218A^2$$

$$R^2 = 0,289 \quad s_e = 0,253$$

Männer: $n = 3455$

$$\sqrt{MEF_{75}} = 1,55 + 1,83\ln(H) + 0,024A - 0,000312A^2$$

$$R^2 = 0,220 \quad s_e = 0,303$$

Die Modellgleichungen für Frauen und Männer unterscheiden sich nur wenig voneinander. Bei den Männern bewirkt die Größe die durchschnittlich höheren Werte.

Die Medianplots zeigen die bei den Niemalsrauchern(innen) übliche gute Anpassung. Mit zunehmendem Alter nimmt die empirische Streuung im Vergleich zur Modellstreuung zu. Bei den Männern ist bis etwa 30 Jahre noch ein Anstieg der MEF_{75}-Werte zu erkennen. Mit 40 beginnen sie dann leicht und ab 50 Jahre stärker zu fallen.

Die MEF_{75}-Werte der Frauen nehmen zwischen 20 und 50 um 2 (ml/s)/Jahr und zwischen 50 und 80 um 7,3 (ml/s)/Jahr ab (4,67 (ml/s)/Jahr von 20 bis 80). Bei den Männern ist zwischen 50 und 80 eine durchschnittliche Abnahme von 8,67 (ml/s)/Jahr festzustellen.

Modellgleichung Kovarianzanalyse: $n = 7578$

$$\begin{aligned}\sqrt{MEF_{75}} &= 1,60 + 1,54\ln(H) + 0,014A - 0,000223A^2 \\ &\quad +0,23\ln(H)SEX + 0,009ASEX - 0,000087A^2SEX\end{aligned}$$

$$R^2 = 0,518 \quad s_e = 0,278$$

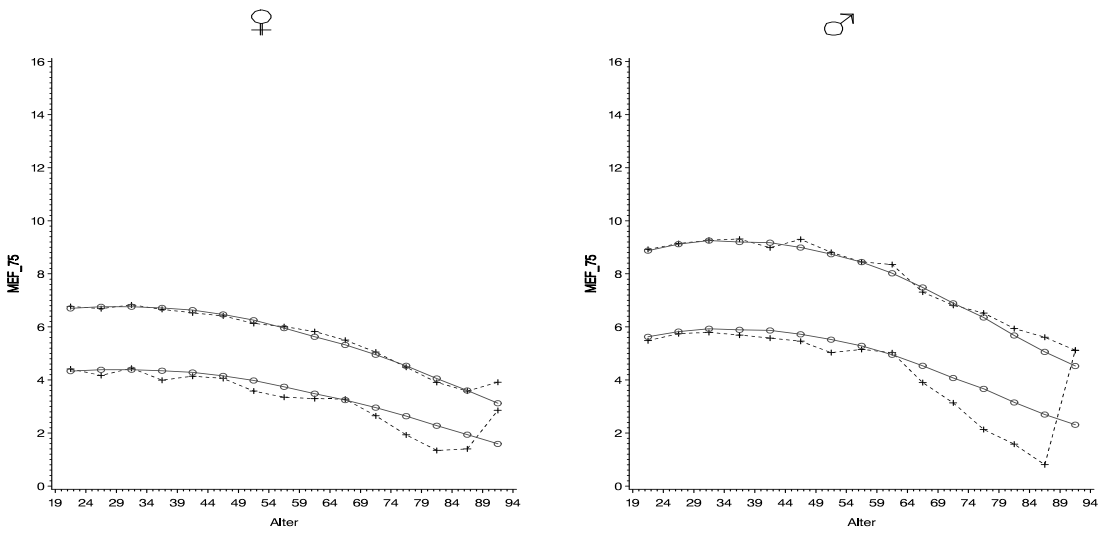


Abbildung 10.2: Medianplot

Wie schon bei den Modellgleichungen zeigt auch das Ergebnis der Kovarianzanalyse, daß die stärksten Unterschiede zwischen den Geschlechtern in der Größe zu suchen sind. Daneben sind noch Wechselwirkungen in den beiden Alterstermen vorhanden.

10.2 Niemalsraucher: MEF_{50}

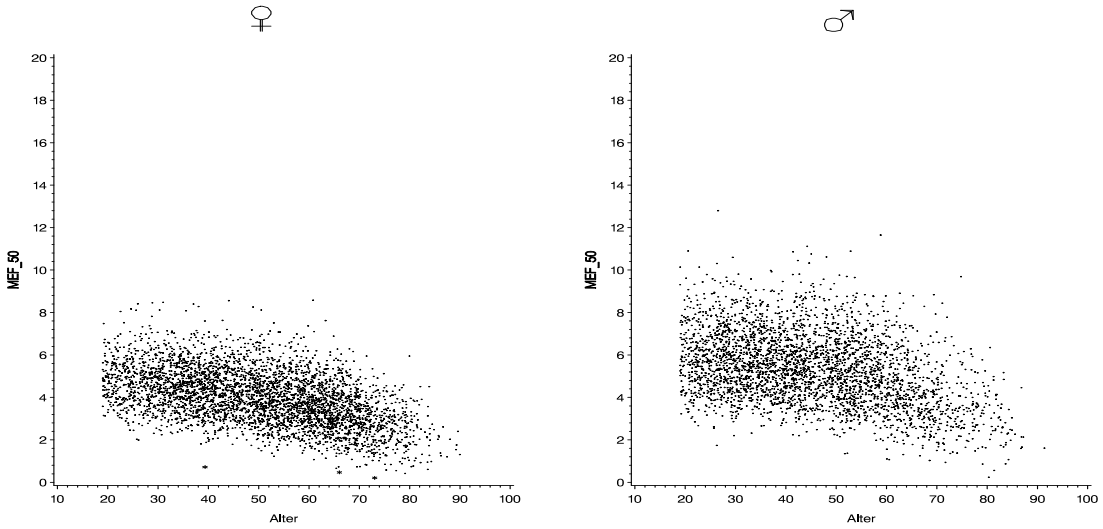


Abbildung 10.3: Scatterplot

Die MEF_{50}-Werte der Frauen und Männer unterscheiden sich nicht mehr allzusehr voneinander, da die Werte zum Teil schon in der Nähe ihrer unteren Schranke von $MEF_{50} = 0$ liegen.

Modellgleichungen

Frauen: $n = 4121$

$$\sqrt{MEF_{50}} = 1,54 + 1,13\ln(H) + 0,006A - 0,000160A^2$$

$$R^2 = 0,286 \quad s_e = 0,275$$

Männer: $n = 3459$

$$\sqrt{MEF_{50}} = 1,49 + 1,17\ln(H) + 0,016A - 0,000254A^2$$

$$R^2 = 0,215 \quad s_e = 0,325$$

Die Unterschiede zwischen Frauen und Männern sind hier vor allem in den Alterstermen zu finden.

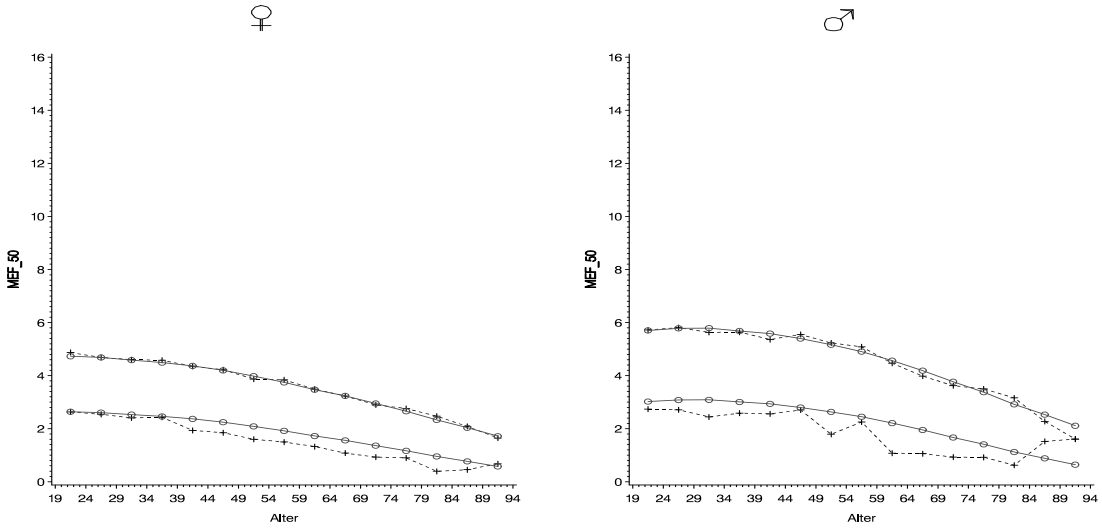


Abbildung 10.4: Medianplot

In den Medianplots sieht man die Zunahme der empirischen Streuung mit zunehmendem Alter. Die Werte der Frauen sinken ab 20 Jahre, während die Werte der Männer ihr Maximum, wie schon beim Parameter MEF_{75} etwa bei 30 Jahren haben.

Die MEF_{50}-Werte der Frauen sinken um 2,67 (ml/s)/Jahr zwischen 20 und 50 Jahren und um 5,3 (ml/s)/Jahr zwischen 50 und 80 Jahren (4 (ml/s)/Jahr von 20 bis 80). Jene der Männer sinken zwischen 20 und 50 um 2 (ml/s)/Jahr und zwischen 50 und 80 Jahren um 7 (ml/s)/Jahr (4,5 (ml/s)/Jahr von 20 bis 80).

Modellgleichung Kovarianzanalyse: $n = 7581$

$$\begin{aligned}\sqrt{MEF_{50}} &= 1,52 + 1,14\ln(H) + 0,007A - 0,000165A^2 \\ &\quad +0,008ASEX - 0,000084A^2SEX\end{aligned}$$

$$R^2 = 0,372 \quad s_e = 0,299$$

Die Kovarianalyse bestätigt, daß die Unterschiede zwischen den Geschlechtern in den Alterstermen zu suchen sind, während die Größe keinen signifikanten Beitrag zur Unterscheidung der Geschlechter liefert.

10.3 Niemalsraucher: MEF_{25}

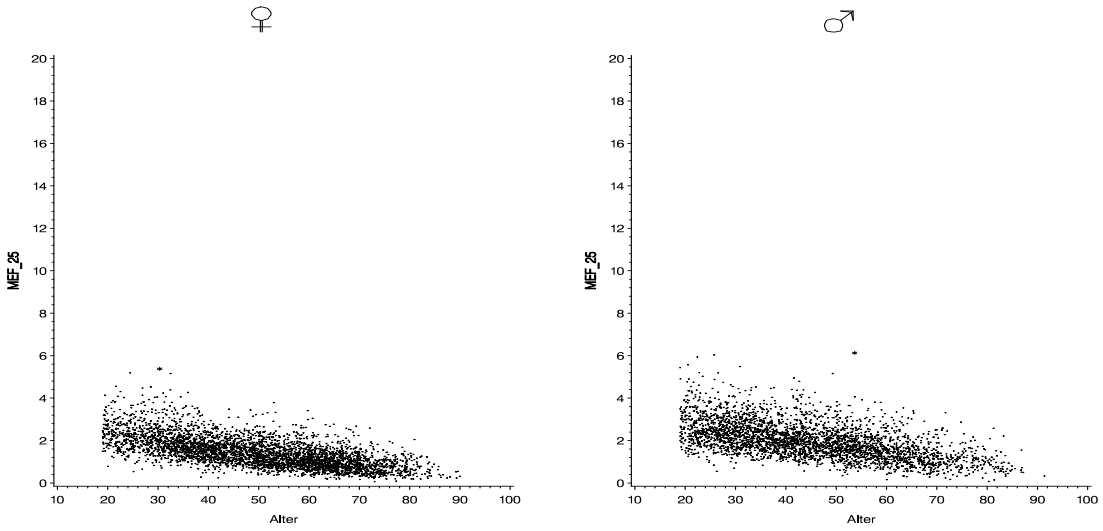


Abbildung 10.5: Scatterplot

Hier ist zwischen Frauen und Männern anhand des Scatterplots beinahe kein Unterschied zu erkennen. Die MEF_{25}-Werte liegen schon knapp am unteren Limit von $MEF_{25} = 0$.

Modellgleichungen

Frauen: $n = 4123$

$$\sqrt{MEF_{25}} = 1,41 + 0,86\ln(H) + 0,015A - 0,000035A^2$$

$$R^2 = 0,463 \quad s_e = 0,212$$

Männer: $n = 3456$

$$\sqrt{MEF_{25}} = 1,30 + 0,79\ln(H) - 0,005A - 0,000054A^2$$

$$R^2 = 0,361 \quad s_e = 0,233$$

Diesmal haben sogar die Frauen den größeren Koeffizienten bei der Variablen $\ln(H)$.

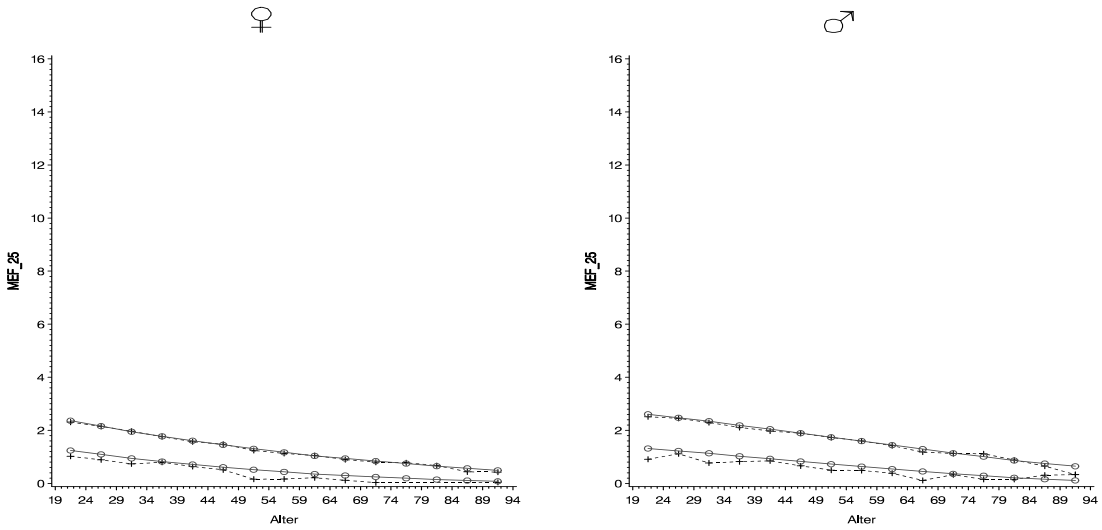


Abbildung 10.6: Medianplot

Die Medianplots zeigen die bei den Niemalsrauchern(innen) übliche gute Anpassung. Bei den Frauen und vor allem bei den Männern nehmen die MEF_{25}-Werte konstant ab dem 20 Lebensjahr ab.

Bei den Frauen beträgt die durchschnittliche Abnahme 2,83 (ml/s)/Jahr und bei den Männern 3,17 (ml/s)/Jahr.

Modellgleichung Kovarianzanalyse: $n = 7581$

$$\begin{aligned}\sqrt{MEF_{25}} &= 1,43 + 0,83\ln(H) - 0,015A - 0,000036A^2 \\ &\quad -0,16SEX + 0,010ASEX - 0,000089A^2SEX\end{aligned}$$

$$R^2 = 0,483 \quad s_e = 0,222$$

Unterschiede zwischen den Geschlechtern lassen sich durch eine konstante Parallelverschiebung der Regressionsgeraden und durch Wechselwirkungen in den Alterstermen erklären.

10.4 Passiv- Ex_gel- leichte- mittlere Raucher: MEF_{75}

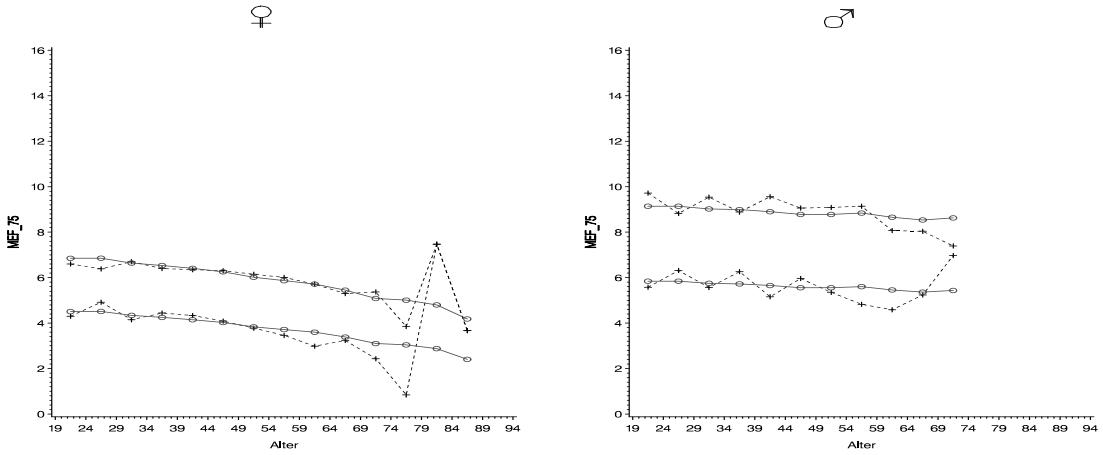


Abbildung 10.7: Passivraucher(innen): Medianplot

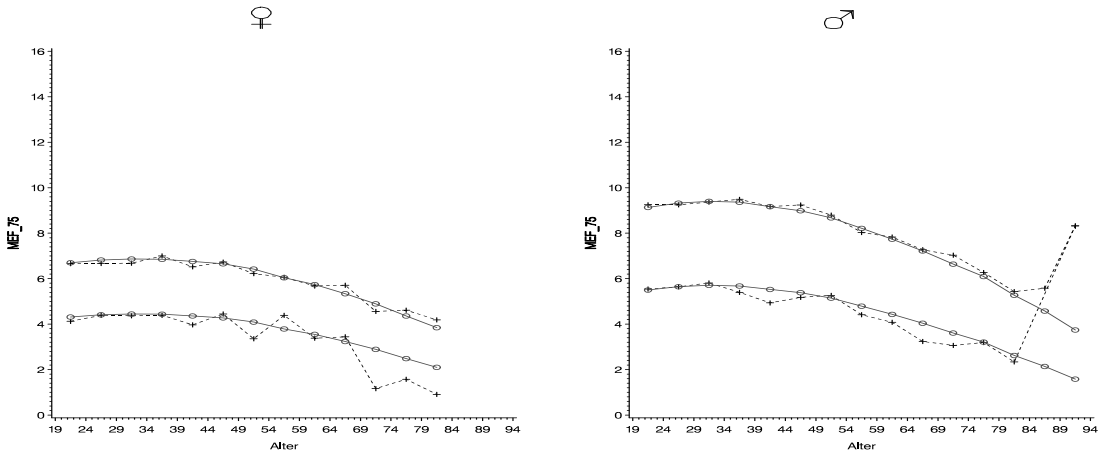


Abbildung 10.8: Ex-gel. Raucher(innen): Medianplot

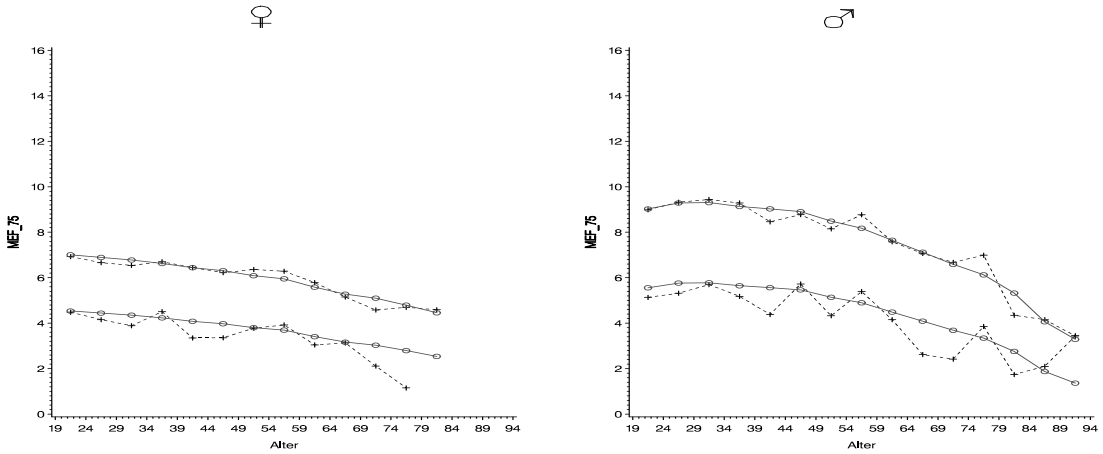


Abbildung 10.9: leichte Raucher(innen): Medianplot

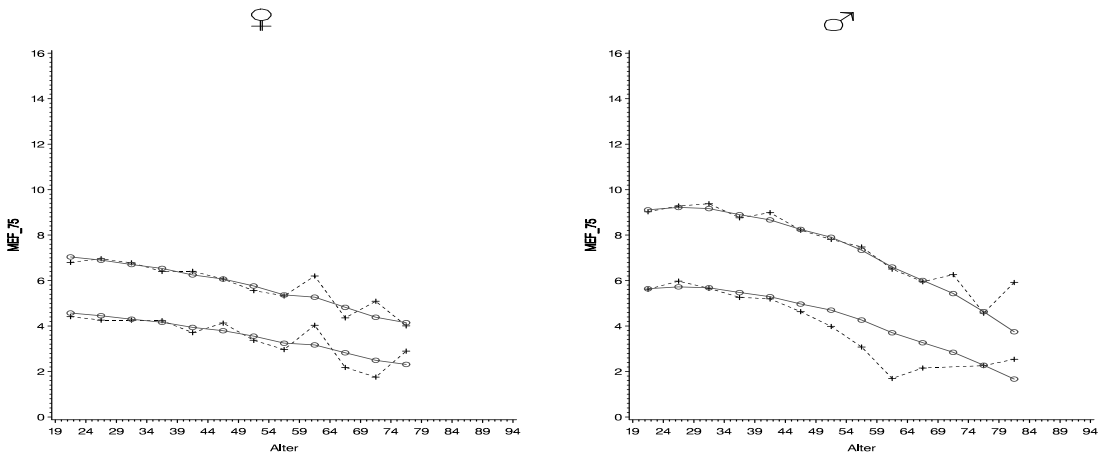


Abbildung 10.10: mittlere Raucher(innen): Medianplot

Bei den Frauen ist durchwegs eine Abnahme der MEF_{75}-Werte ab dem 20. Lebensjahr festzustellen. Einzig bei den Ex-gel. Raucherinnen ist ein Anstieg bis etwa 30 zu sehen. Der Medianplot der Passivraucher zeigt im Vergleich zu den anderen Gruppen einen deutlich unterschiedlichen Verlauf. Bei den Männern steigen die Werte bis etwa 30 Jahre an, um danach immer stärker abzunehmen.

Modellgleichungen: Frauen

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{75}} = 1{,}75 + 1{,}78\ln(H)$ - $0{,}000060A^2$	0,185	0,247
Ex-gel.	$\sqrt{MEF_{75}} = 1{,}76 + 1{,}09\ln(H) + 0{,}019$A - $0{,}000279A^2$	0,220	0,256
1-10	$\sqrt{MEF_{75}} = 2{,}14 + 1{,}07\ln(H)$ - $0{,}000078A^2$	0,130	0,259
11-20	$\sqrt{MEF_{75}} = 2{,}09 + 1{,}22\ln(H)$ - $0{,}000109A^2$	0,202	0,258

Bei den Frauen hat die Variable Alter, außer bei den Ex-gel. Raucherinnen keinen signifikanten Einfluß .

Modellgleichungen: Männer

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{75}} = 1{,}97 + 1{,}78\ln(H)$	0,042	0,303
Ex-gel.	$\sqrt{MEF_{75}} = 1{,}73 + 1{,}70\ln(H) + 0{,}021$A - $0{,}000307A^2$	0,252	0,339
1-10	$\sqrt{MEF_{75}} = 1{,}62 + 1{,}84\ln(H) + 0{,}022$A - $0{,}000315A^2$	0,245	0,323
11-20	$\sqrt{MEF_{75}} = 1{,}63 + 1{,}99\ln(H) + 0{,}019$A - $0{,}000339A^2$	0,256	0,322

In den Modellgleichungen der Männer zeigt sich auch die Abweichung der Passivraucher von den anderen Gruppen.

10.5 Passiv- Ex_gel- leichte- mittlere Raucher: MEF_{50}

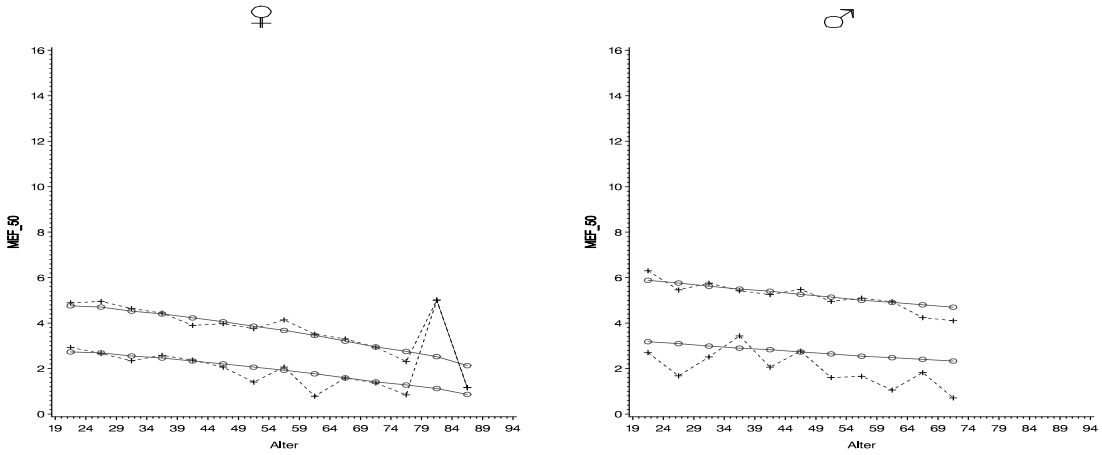


Abbildung 10.11: Passivraucher(innen): Medianplot

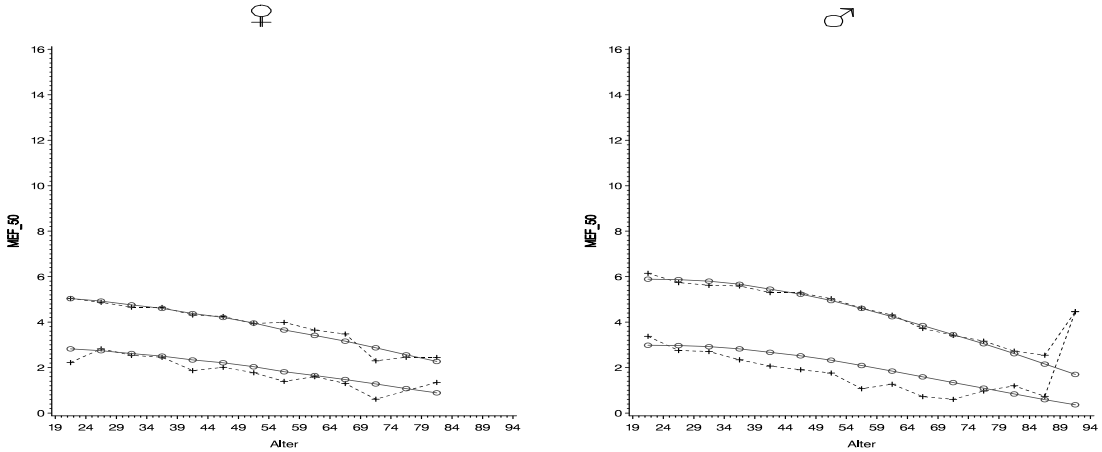


Abbildung 10.12: Ex-gel. Raucher(innen): Medianplot

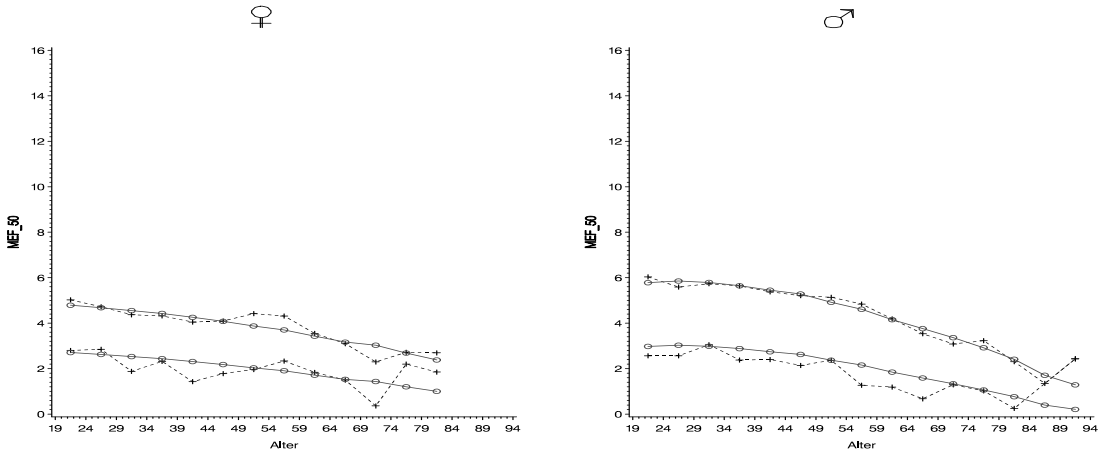


Abbildung 10.13: leichte Raucher(innen): Medianplot

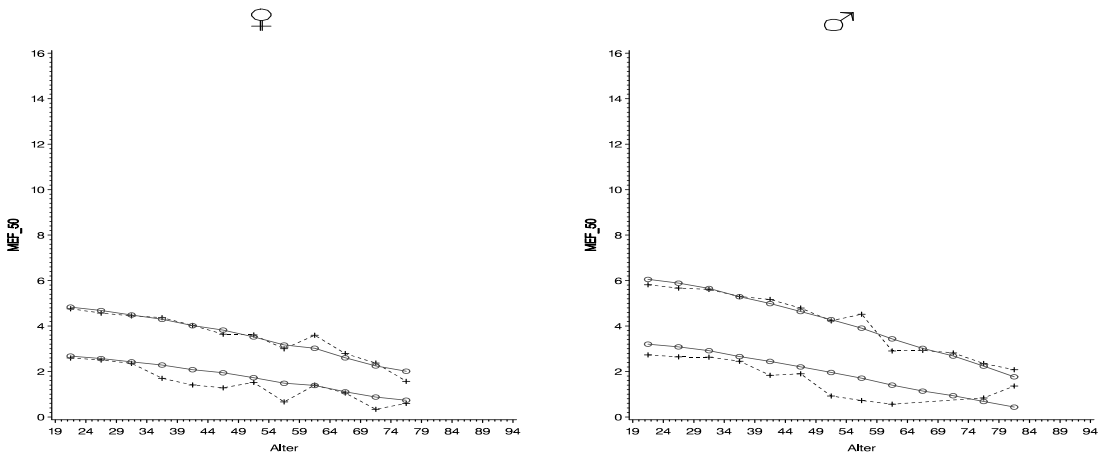


Abbildung 10.14: mittlere Raucher(innen): Medianplot

Die Frauen zeigen einen eher konstanten Kurvenverlauf. Bei den Männern haben die Passivraucher wieder einen abweichenden Verlauf. Die Ex-gel.- sowie die leichten Raucher beschreiben eine Bogenform mit einem Plateau bis etwa 30. Bei den mittleren Raucher sinken die Werte schon deutlich ab dem 20. Lebensjahr.

Modellgleichungen: Frauen

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{50}} = 1{,}56 + 1{,}31\ln(H)$ - $0{,}000089A^2$	0,227	0,264
Ex-gel.	$\sqrt{MEF_{50}} = 1{,}88 + 0{,}82\ln(H)$ - $0{,}000116A^2$	0,260	0,282
1-10	$\sqrt{MEF_{50}} = 2{,}24$ - $0{,}000102A^2$	0,150	0,271
11-20	$\sqrt{MEF_{50}} = 1{,}65 + 1{,}21\ln(H)$ - $0{,}000140A^2$	0,247	0,281

Die Variable Alter hat hier keinen signifikanten Einfluß auf die Modellerstellung. Bei den leichten Raucherinnen fällt auch der Größenterm aus dem Modell heraus.

Modellgleichungen: Männer

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{50}} = 2{,}54$ - $0{,}005$A	0,038	0,321
Ex-gel.	$\sqrt{MEF_{50}} = 1{,}84 + 0{,}81\ln(H) + 0{,}010$A - $0{,}000226A^2$	0,262	0,350
1-10	$\sqrt{MEF_{50}} = 1{,}54 + 1{,}13\ln(H) + 0{,}015$A - $0{,}000284A^2$	0,274	0,339
11-20	$\sqrt{MEF_{50}} = 1{,}82 + 1{,}25\ln(H)$ - $0{,}000173A^2$	0,291	0,335

Bei den Männern variieren die Modelle von Gruppe zu Gruppe. Bei den Passivrauchern reicht allein das Alter aus, um die Abnahme der MEF_{50}-Werte zu erklären.

10.6 Passiv- Ex_gel- leichte- mittlere Raucher: MEF_{25}

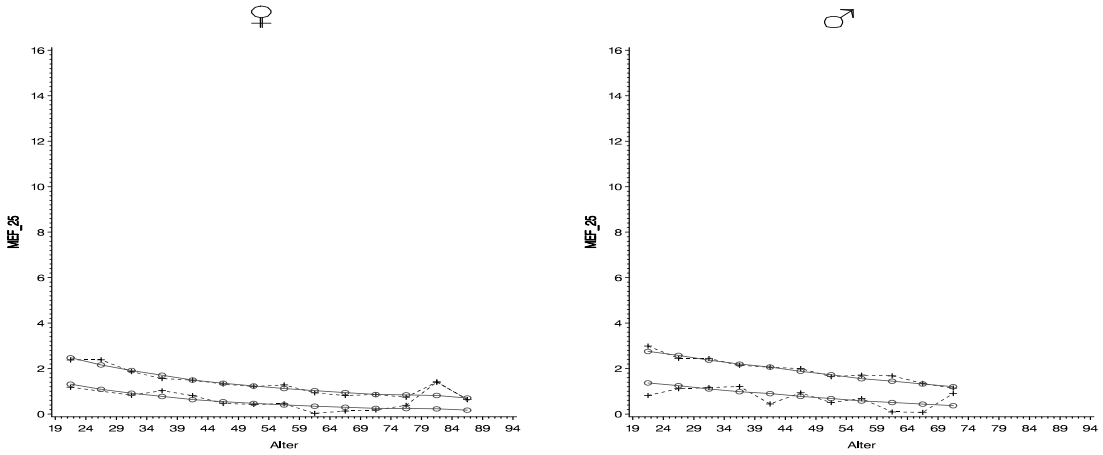


Abbildung 10.15: Passivraucher(innen): Medianplot

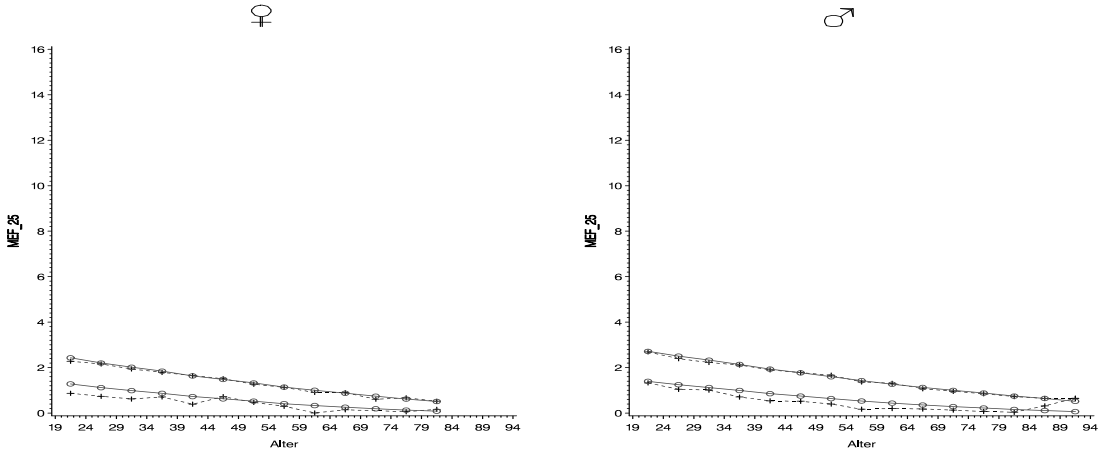


Abbildung 10.16: Ex-gel. Raucher(innen): Medianplot

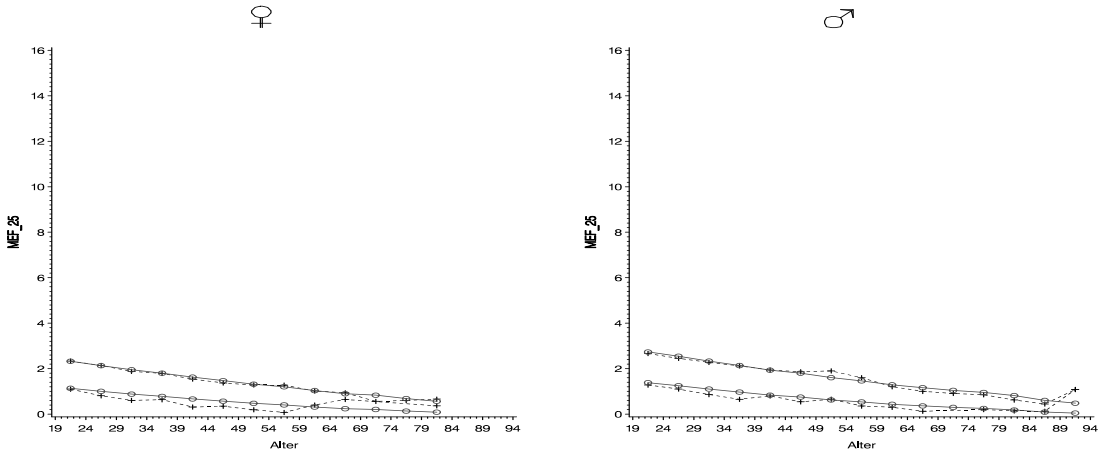


Abbildung 10.17: leichte Raucher(innen): Medianplot

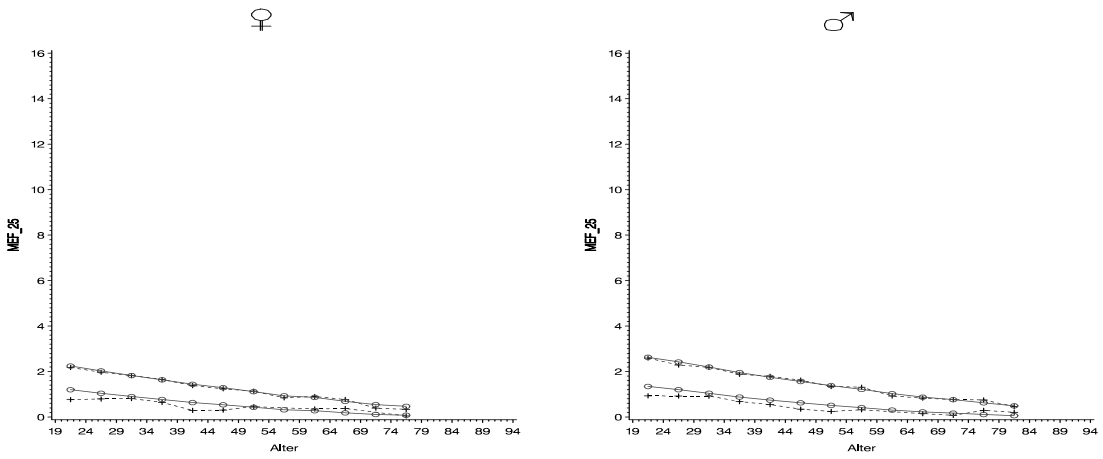


Abbildung 10.18: mittlere Raucher(innen): Medianplot

Für den Parameter MEF_{25} sehen die Medianplots für Frauen und Männer über alle Gruppen hinweg relativ ähnlich aus.

Modellgleichungen: Frauen

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{25}} = 1{,}57 + 0{,}92\ln(H)$ - 0,025A + $0{,}000139A^2$	0,451	0,214
Ex-gel.	$\sqrt{MEF_{25}} = 1{,}54 + 0{,}64\ln(H)$ - 0,014A	0,476	0,213
1-10	$\sqrt{MEF_{25}} = 1{,}80$ - 0,013A	0,330	0,230
11-20	$\sqrt{MEF_{25}} = 1{,}29 + 1{,}02\ln(H)$ - 0,014A	0,452	0,201

Wichtig zur Erklärung der Abnahme der MEF_{25}-Werte ist das Alter. Der quadratische Altersterm ist zur Erklärung dieser Abnahme nicht unbedingt notwendig. Bei den leichten Rauchern reicht sogar nur das Alter alleine zur Modellerstellung.

Modellgleichungen: Männer

Gruppe	Modellgleichung	R^2	s_e
passiv	$\sqrt{MEF_{25}} = 1{,}92$ - 0,012A	0,244	0,245
Ex-gel.	$\sqrt{MEF_{25}} = 1{,}51 + 0{,}71\ln(H)$ - 0,013A	0,421	0,233
1-10	$\sqrt{MEF_{25}} = 1{,}38 + 0{,}91\ln(H)$ - 0,012A	0,427	0,239
11-20	$\sqrt{MEF_{25}} = 1{,}50 + 0{,}76\ln(H)$ - 0,015A	0,432	0,230

Auch bei den Männern ist der quadratische Altersterm zur Modellerstellung nicht notwendig. Die Passivraucher weichen wiederum von den anderen Gruppen ab.

10.7 Raucher_schwer: MEF_{75}

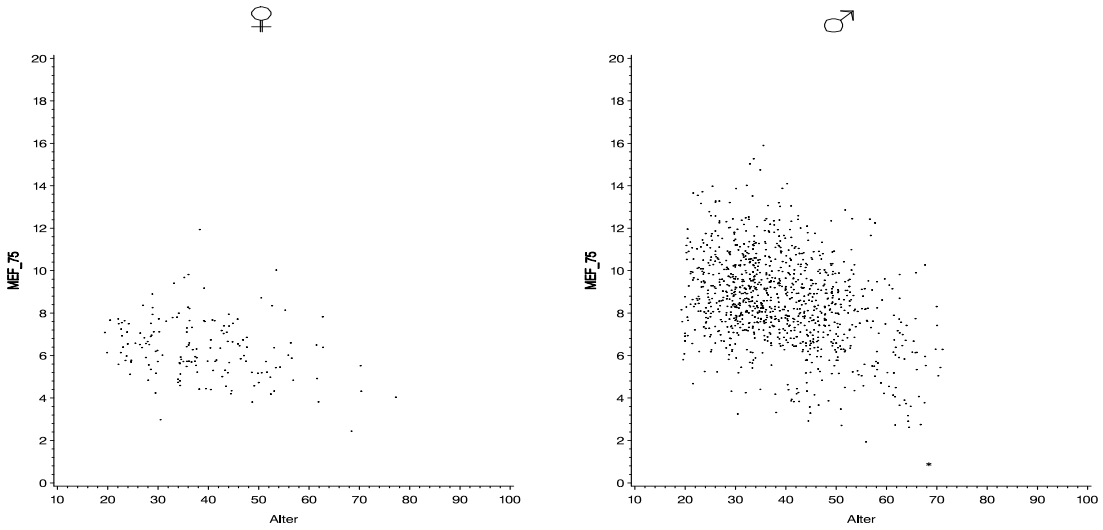


Abbildung 10.19: Scatterplot

Bei den Frauen sind nur bis etwa 60 Jahre Werte vorhanden. Die MEF_{75}-Werte der Männer streuen sehr stark. Die Werte der Frauen überdecken sich mit der unteren Hälfte der Werte bei den Männern.

Modellgleichungen

Frauen: $n = 158$

$$\sqrt{MEF_{75}} = 1,65 + 1,92\ln(H) - 0,000068A^2$$

$$R^2 = 0,146 \quad s_e = 0,261$$

Männer: $n = 1036$

$$\sqrt{MEF_{75}} = 1,78 + 1,75\ln(H) + 0,018A - 0,000347A^2$$

$$R^2 = 0,180 \quad s_e = 0,335$$

Das Modell der Frauen besteht nur aus jeweils einem Größen- und Altersterm. Der Bestimmtheitsgrad ist in beiden Fällen im Vergleich zu den Niemalsrauchern(innen) kleiner und bei den Männern die Modellstreuung größer.

Die empirischen Mediane variieren besonders bei den Frauen aufgrund der wenigen Werte sehr stark. Bei den Frauen nehmen die MEF_{75}-Werte in etwa konstant ab, während sich die Abnahme der Werte bei den Männern mit zunehmendem Alter beschleunigt.

Die MEF_{75}-Werte der Frauen nehmen zwischen 20 und 60 im Schnitt um 3,25 (ml/s)/Jahr ab. Die Werte der Männer sinken zwischen 20 und 50 Jahren um 4,67 (ml/s)/Jahr und zwischen 50 und 70 Jahren um 10 (ml/s)/Jahr (6,8 (ml/s)/Jahr von 20 bis 70).

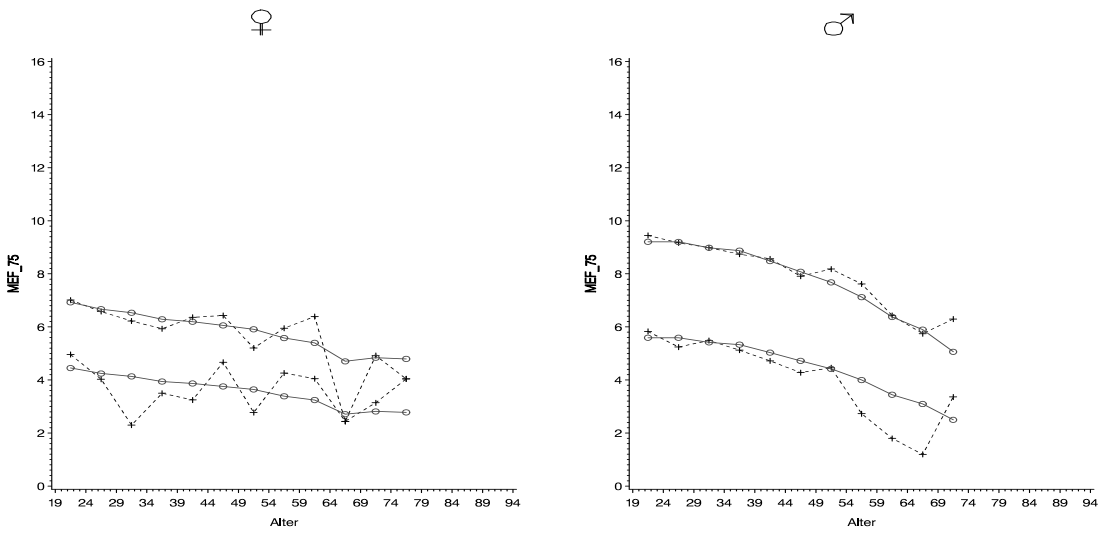


Abbildung 10.20: Medianplot

Während die MEF_{75}-Werte bei den Niemalsrauchern bei 30 Jahren ihr Maximum haben und bis 50 nur langsam sinken, nehmen die Werte der starken Raucher bereits ab 20 Jahren kontinuierlich ab.

Modellgleichung Kovarianzanalyse: $n = 7578$

$$\begin{aligned}\sqrt{MEF_{75}} &= 1,60 + 1,54\ln(H) + 0,014A - 0,000223A^2 \\ &\quad +0,23\ln(H)SEX + 0,009ASEX - 0,000087A^2SEX\end{aligned}$$

$$R^2 = 0,518 \quad s_e = 0,278$$

Das Modell für Frauen und Männer gemeinsam erklärt die Unterschiede zwischen den Geschlechtern durch Wechselwirkungen in allen drei erklärenden Variablen.

10.8 Raucher_schwer: MEF_{50}

Bei den Männern ist im Scatterplot deutlich die Abnahme der Werte zu erkennen.

Modellgleichungen

Frauen: $n = 158$

$$\sqrt{MEF_{50}} = 2,24 - 0,000130A^2$$

$$R^2 = 0,175 \quad s_e = 0,285$$

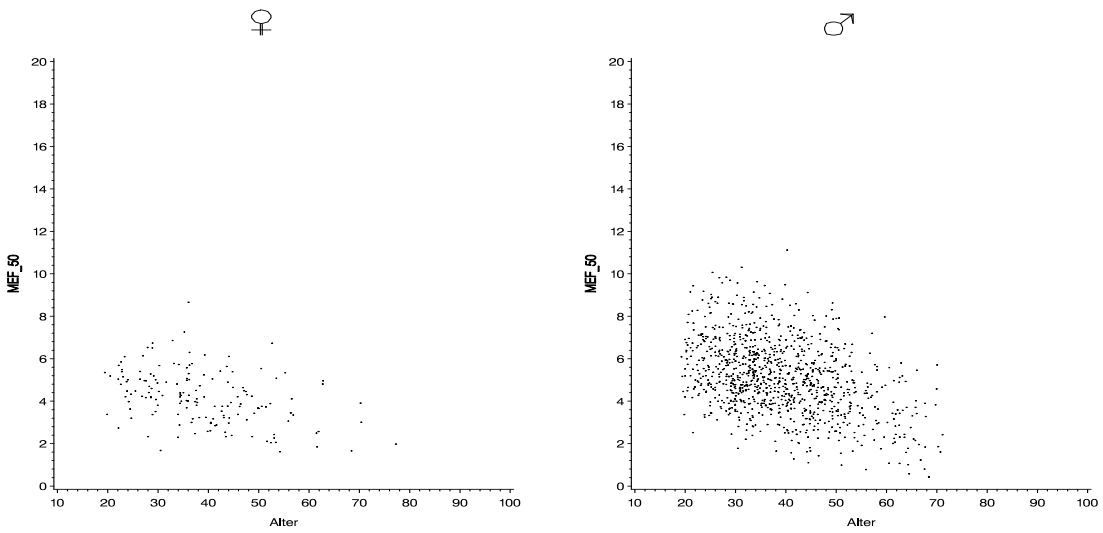


Abbildung 10.21: Scatterplot

Männer: $n = 1037$

$$\sqrt{MEF_{50}} = 1,84 + 1,18\ln(H) - 0,000177A^2$$

$$R^2 = 0,204 \quad s_e = 0,350$$

Bei den Frauen ist hier überhaupt nur mehr der quadratische Altersterm signifikant an der Modellerstellung beteiligt. Bei den Männern fällt dagegen die Variable Alter aus dem Modell heraus.

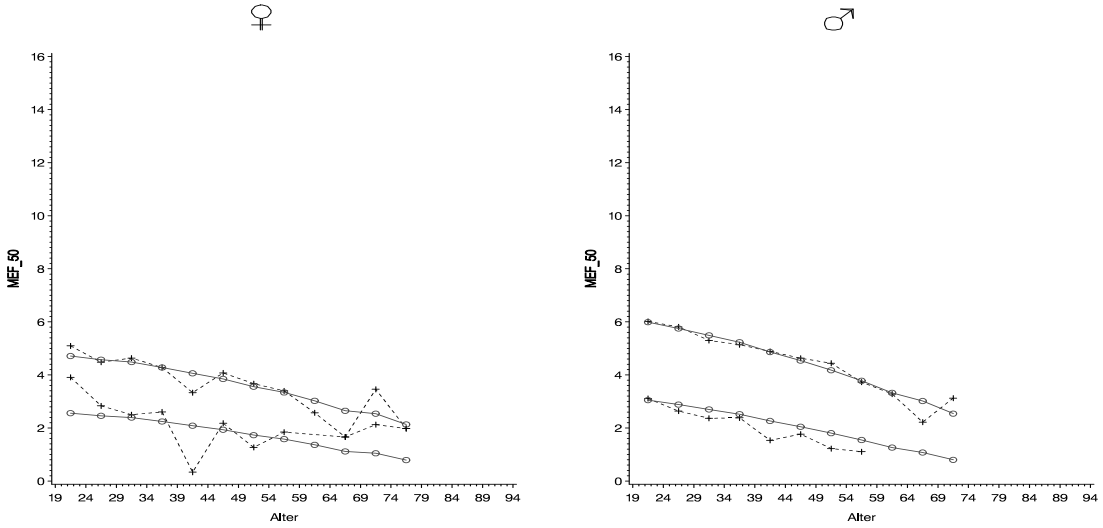


Abbildung 10.22: Medianplot

Bei Frauen und Männern nehmen die MEF_{50}-Werte ab 20 nahezu konstant ab. Mit dem Alter ist eine leicht beschleunigte Abnahme zu erkennen.

Die MEF_{50}-Werte der Frauen nehmen zwischen 20 und 60 in Schnitt um 5 (ml/s)/Jahr ab. Jene der Männer zwischen 20 und 70 im Schnitt um 6,4 (ml/s)/Jahr.

Modellgleichung Kovarianzanalyse: $n = 7581$

$$\begin{aligned}\sqrt{MEF_{50}} &= 1,52 + 1,14\ln(H) + 0,007A - 0,000165A^2 \\ &\quad +0,008ASEX - 0,000084A^2SEX\end{aligned}$$

$$R^2 = 0,372 \quad s_e = 0,299$$

Die Unterschied zwischen den Geschlechtern lassen sich durch Wechselwirkungen in den Alterstermen erklären.

10.9 Raucher_schwer: MEF_{25}

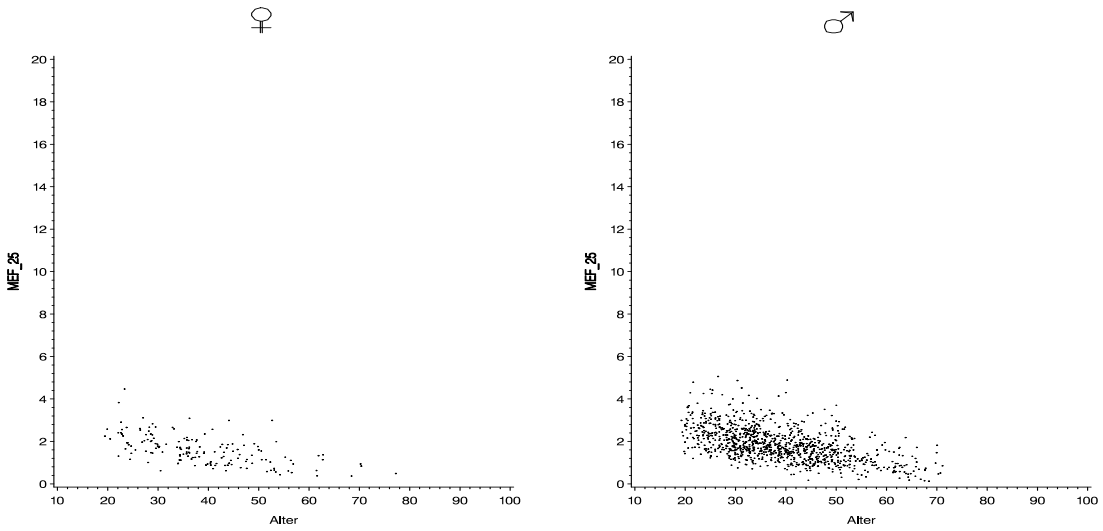


Abbildung 10.23: Scatterplot

Die Werte der Frauen und Männer liegen schon sehr nahe der unteren Schranke von $MEF_{25} = 0$.

Modellgleichungen

Frauen: $n = 158$

$$\sqrt{MEF_{25}} = 1,04 + 1,45\ln(H) - 0,014A$$

$$R^2 = 0,418 \quad s_e = 0,212$$

Männer: $n = 1037$

$$\sqrt{MEF_{25}} = 1,32 + 1,02\ln(H) - 0,015A$$

$$R^2 = 0,348 \quad s_e = 0,235$$

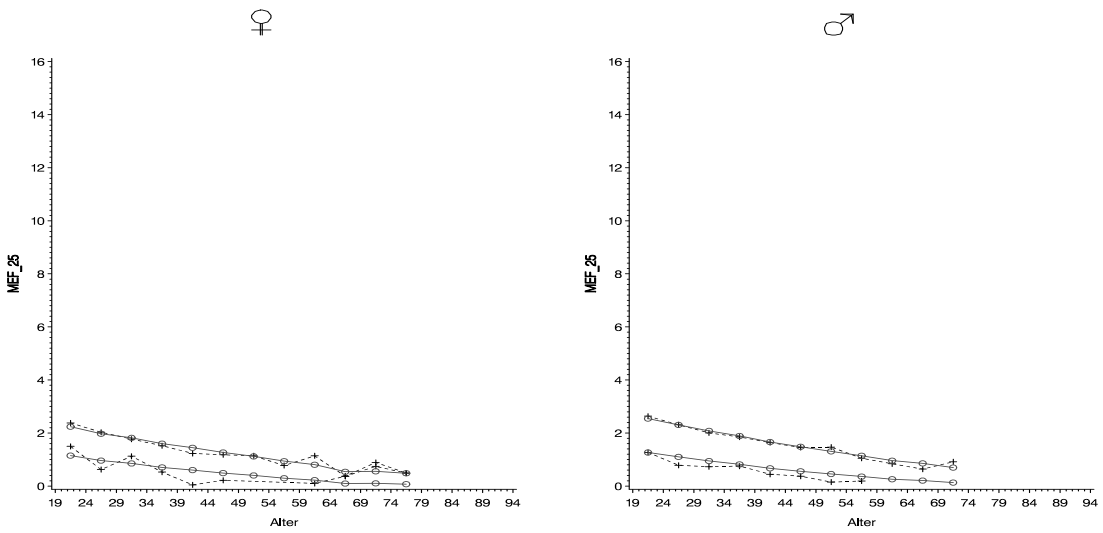


Abbildung 10.24: Medianplot

Die Modelle der Frauen und Männer stimmen in den Termen diesmal überein. Der quadratische Altersterm wird nicht in das Modell mit aufgenommen. Im Vergleich zu MEF_{75} und MEF_{50} besitzen die Modelle für MEF_{25} wieder einen höheren Bestimmtheitsgrad.

In beiden Medianplots ist die stetige Abnahme der Werte ab dem 20. Lebensjahr klar ersichtlich. So nehmen die MEF_{25}-Werte der Frauen zwischen 20 und 60 Jahren im Schnitt um 3,75 (ml/s)/Jahr ab. Jene der Männer zwischen 20 und 70 Jahren im Schnitt um 4 (ml/s)/Jahr.

Modellgleichung Kovarianzanalyse: n=7581

$$\begin{aligned}\sqrt{MEF_{25}} &= 1,43 + 0,83\ln(H) - 0,015A + 0,000036A^2 \\ &\quad -0,16SEX + 0,010ASEX - 0,000089A^2SEX\end{aligned}$$

$$R^2 = 0,483 \quad s_e = 0,222$$

Im Gegensatz zu den Modellen für Frauen und Männer getrennt, liefert der quadratische Altersterm hier wieder einen signifikanten Beitrag. Die Unterschiede zwischen den Geschlechtern beruhen auf einer konstanten Parallelverschiebung und Wechselwirkungen in den Alterstermen.

Kapitel 11

Statistische Grundlagen III

Im letzten Kapitel werden detaillierte Vergleiche der Niemals- mit den schweren Rauchern durchgeführt. Neben den Vergleichen durch Medianplots wird noch eine nichtparametrische Methode verwendet, um den mittleren Verlauf der Abnahme der Werte der Lungenfunktionsparameter mit zunehmendem Alter darzustellen.

11.1 Die Glättung von Scatterplots

Bezeichnen wir die abhängige Variable mit Y_i und die erklärende Variable mit x_i und betrachten diese als Realisationen der Zufallsvariable Y und der Kovariate X. Außerdem sei ein sogenannter *smoother* eine Prozedur, welche auf den bivariaten Daten $(x_1, y_1), \ldots, (x_n, y_n)$ beruht und folgende Zerlegung erzeugt:

$$Y_i = s(x_i) + r_i \qquad i = 1, \ldots, n$$

wobei $s(\cdot)$ die *smooth function* (Glättungfunktion) bezeichnet und r_i die Residuen.

Wir betrachten diese Glättungsfunktionen als Kurvenschätzer unter der Annahme, daß die Responsevariable Y_i erzeugt werden kann, indem zur Glättungsfunktion f ein nicht beobachtbarer Fehler ϵ hinzugefügt wird.

$$Y_i = f(x_i) + \epsilon_i \tag{11.1}$$

Um die Glättungsfunktion f zu schätzen verwenden wir den sogenannten *super − smoother*. Die diesen Glättungsfunktionen zugrundeliegende Idee ist die der lokalen Durchschnittsbildung. Diese Idee kann folgendermaßen begründet werden: ist die gemeinsame Verteilung von Y und X bekannt, so ist eine Möglichkeit $s(\cdot)$ zu finden, den Erwartungswert $E[Y - s(X)]^2$ zu minimieren. Unter der Bedingung $X = x$ ergibt sich die Lösung $s(x) = E[Y \mid X = x]$ für jedes x. Unser Ziel ist es nun $E[Y \mid X = x]$ aus den zur Verfügung stehenden Daten zu schätzen. Dies führt zur Klasse der *local average estimates*.

$$s(x_i) = Ave_{j \in N_i}\{y_i\} \tag{11.2}$$

wobei *Ave* ein Durchschnittsopererator wie z.B. das arithmetische Mittel oder der Median ist. N_i beschreibt eine gewisse Umgebung von X_i. In unserem Fall ist diese Umgebung folgendermaßen definiert:

$$N_i = \{max[i-(k-1)/2, 1], \ldots, i-1, i, i+1, \ldots, min[i+(k-1)/2, n]\}\,, \quad i = 1, \ldots, n$$

Der Parameter k heißt *span* (Spannweite) der Glättungsfunktion und bestimmt den Grad der Glättung des resultierenden Schätzers. Eine wünschenswerte Eigenschaft dieses $super - smoother$ wäre, daß die Spannweite k von der Streuung und der Nichtlinearität der Daten abhängig gemacht wird.

Der Einfachheit halber nehmen wir an, daß der Duchschnittoperater in (11.2) das arithmetische Mittel sei. Weiters nehmen wir an, daß die ϵ_i in 11.1 den Mittelwert 0 und konstante Varianz σ^2 besitzen. Daraus folgt, daß der *expected squared error* an der Stelle x_i folgendes Aussehen hat:

$$ESE(x_i \mid k) = \left[f(x_i) - \frac{1}{k} \sum_{i-\frac{k}{2}}^{i+\frac{k}{2}} f(x_j) \right]^2 + \frac{\sigma^2}{k}$$

Das ist eine Zerlegung in das Quadrat des Bias und der Varianz. Eine Erhöhung von k vergrößert den Bias und verringert die Varianz und umgekehrt.

Um nun ein Kriterium zur Bestimmung der Spannweite k zu erhalten, betrachten wir die kreuz-validierte Residuenquadratsumme ($CV(k, x_i)$) in einer bestimmten lokalen Umgebung von x_i.

$$CV(k, x_i) = \frac{1}{J} \sum_{j=i-\frac{J}{2}}^{i+\frac{J}{2}} [y_j - s_{(j)}(x_j \mid k)]^2$$

wobei $s_{(j)}(\cdot)$ die Glättungsfunktion, berechnet an der Stelle x_j ohne den Punkt (x_j, y_j) ist. Der Parameter J wird als zweite Spannweite neben k benutzt. J wird mit einem Wert aus dem Intervall $[0.2 \times n, 0.3 \times n]$ belegt. Die Spannweite wird auf diese Weise durch Minimierung von $CV(k, x_i)$ an jeder Stelle x_j neu berechnet.

Innerhalb dieser, durch die variable Spannweite k bestimmten Umgebung, wird der Glättungswert $s(x_i)$ als lokale Kleinste-Quadrate Regressionsschätzung (lok. KQRS) an der Stelle x_i bestimmt. Die lok. KQRS wird als Durchschnittsoperator *Ave* (siehe (11.2)) benutzt. Die Vorteile dieser lok. KQRS sind, daß die resultierende Glättungsfunktion:

1. durchgehende Kurven für aufeinanderfolgende x-Werte erzeugt, auch wenn die Differenzen $x_i - x_{i-1}$ nicht konstant sind;

2. an den Grenzen keinen großen Bias produziert, und
3. leicht zu progammieren ist und einfache Formeln zur Verfügung stehen, um die Spannweite k beim Schritt von x_i nach x_{i+1} neu zu berechnen.

Wir haben jetzt im Gegensatz zum multiplen linearen Regressionsmodell der Form

$$Y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i$$

ein Generaliziertes Additives Modell der Form

$$Y_i = \sum_{j=1}^{p} s_j(x_{ij}) + \epsilon_i$$

wobei $s_j(\cdot)$ die Glättungsfunktion ist.

Ein Algorithmus zur Lösung dieses Problems ist der $back - fitting\ \ algorithm.$

Für weitere Erläuterungen zu diesem Thema siehe Hastie-Tibshirani [7] und [8] sowie Segal et al. [17].

Unsere Auswertungen wurden mit dem Programmsystem $S - PLUS$ durchgeführt.

Kapitel 12

Vergleich Niemals- schwere Raucher(innen)

Um den negativen Einfluß des Rauchens auf die Lungenfunktionswerte zu charakterisieren, wird nun versucht diesen Einfluß mit verschiedenen Analysemethoden nachzuweisen. In diesem Kapitel werden die Niemals- mit den schweren Raucher(innen) verglichen.

12.1 Voranalyse der schweren Raucherinnen

In der Gruppe der schweren Raucherinnen sind mit nur 158 Beobachtungen wesentlich weniger Probandinnen als in der Gruppe der Niemalsraucherinnen mit 4124 Probandinnen. Bei den Männern ist die Aufteilung mit 3459 Niemalsrauchern und 1037 schweren Rauchern ausgewogener. Wobei es vor allem wichtig ist, daß in der Gruppe der schweren Raucher genügend Probanden vorhanden sind, um aussagekräftige Vergleichsuntersuchungen durchführen zu können.

Die einzige Möglichkeit mittels der verfügbaren Daten die Gruppe der schweren Raucherinnen zu vergrößern besteht darin, eine neue Gruppe zu generieren, welche aus den schweren Raucherinnen und einem Teil der mittleren Raucherinnen besteht. Bei genauerer Betrachtung der Gruppe der mittleren Raucherinnen erkennt man, daß 715 dieser 760 Raucherinnen angaben, mehr als 15 Zigaretten pro Tag und immerhin mehr als die Hälfte (434) sogar angaben 19 oder 20 Zigaretten pro Tag zu rauchen. Daraus könnte man den Schluß ziehen, daß Frauen nicht gerne zugeben mehr als eine Packung Zigaretten (= 20 Stück) zu rauchen.

Eine weitere Motivation ist der Vergleich der adjustierten Mittelwerte der mittleren Raucherinnen, welche 15 Zigaretten oder mehr pro Tag rauchen und der schweren Raucherinnen. Wie in der Abbildung 12.1 ersichtlich, ist der adjustierte Mittelwert des Parameters FEV_1 der mittleren Raucherinnen (≤ 15) nur unwesentlich und *nicht signifikant höher* als jener der schweren Raucherinnen.

General Linear Models Procedure
Least Squares Means

Z	FEV1 LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0	Pr > |T| H0: LSMEAN1=LSMEAN2
F1: 15-19	3.19951771	0.01528474	0.0001	0.0925
F2: >20	3.13889137	0.03256450	0.0001	

Abbildung 12.1: FEV_1

Alle nun folgenden Untersuchungen werden mit den Niemalsraucherinnen und den Raucherinnen, welche 15 oder mehr Zigaretten pro Tag rauchen durchgeführt. In diesem Kapitel werden jetzt jene als schwere Raucherinnen bezeichnet, welche mindestens 15 Zigaretten pro Tag rauchen.

In den folgenden drei Histogrammen wird nun die Alters- Größen- und Gewichtsverteilung der neu zusammengestellten Raucherinnengrupppe wiedergegeben.

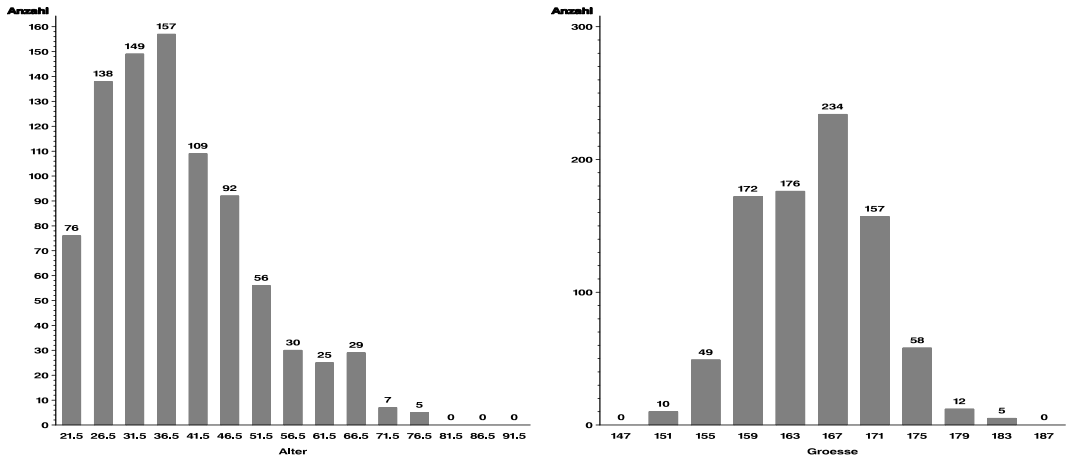


Abbildung 12.2: Alters- und Größenverteilung

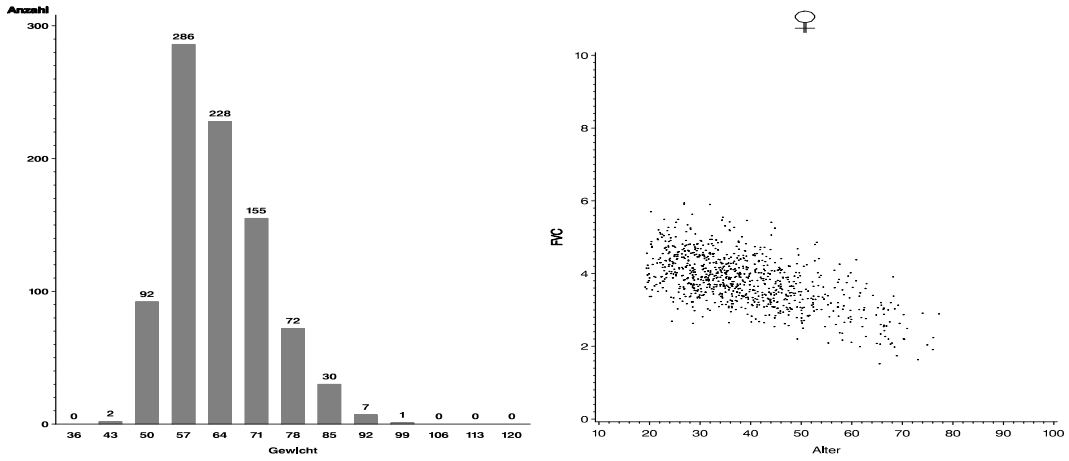


Abbildung 12.3: Gewichtsverteilung und FVC-Scatterplot

Im Altershistogramm und im Scatterplot kann man ersehen, daß bis etwa 70 Jahre ausreichend Daten vorhanden sind.

Die Kovarianzanalyse in Kapitel 4 hat gezeigt, daß sowohl bei den Frauen als auch bei den Männern die kleinsten adjustierten Mittelwerte jeweils in der Gruppe der schweren Raucher(innen) zu finden sind, während die höchsten adjustierten Mittelwerte bei den Frauen in den Gruppen der Niemals- und Ex-gel. Raucherinnen und bei den Männern vorwiegend in der Gruppe der Niemalsraucher zu finden sind. Nun beruhen die adjustierten Mittelwerte auf der gesamten Datenmenge der 19-94 jährigen, ist also sozusagen ein globaler Mittelwert. In diesem Kapitel wird nun untersucht, wie sich dieser vorhandene Unterschied zwischen den Niemals- und schweren Raucher(innen) bei Betrachtung des gesamten Altersbereiches entwickelt. Dabei ist interessant, ab wann und wie markant sich in den Graphiken ein Unterschied feststellen lßt. In weiterer Folge wird auch noch geprüft, ab welchen Altersklassen sich dieser Unterschied beim Vergleich der adjustierten Mittelwerte als signifikant erweist.

Im folgenden wird ein graphischer Vergleich der Mediane und Glättungskurven der Niemals- mit den schweren Rauchern(innen) durchgeführt. Wobei jeweils in der linken Graphik, dem Medianplot, die Mediane und ihr Verlauf der Niemals- und der schweren Raucher(innen) miteinander verglichen werden. Dabei sind die Mediane der Niemalsraucher(innen) durch $(- \circ -)$ und die durchgehend gezeichnete Kurve, die Mediane der schweren Raucher(innen) durch $-(+)-$ und die gestrichelte Kurve gekennzeichnet. In der rechten Graphik sind die beiden Glättungskurven wiedergegeben. Die durchgehende stärkere Kurve repräsentiert die Niemalsraucher(innen), die gestrichelte dünnere Kurve die schweren Raucher(innen).

12.2 Frauen: Vergleiche

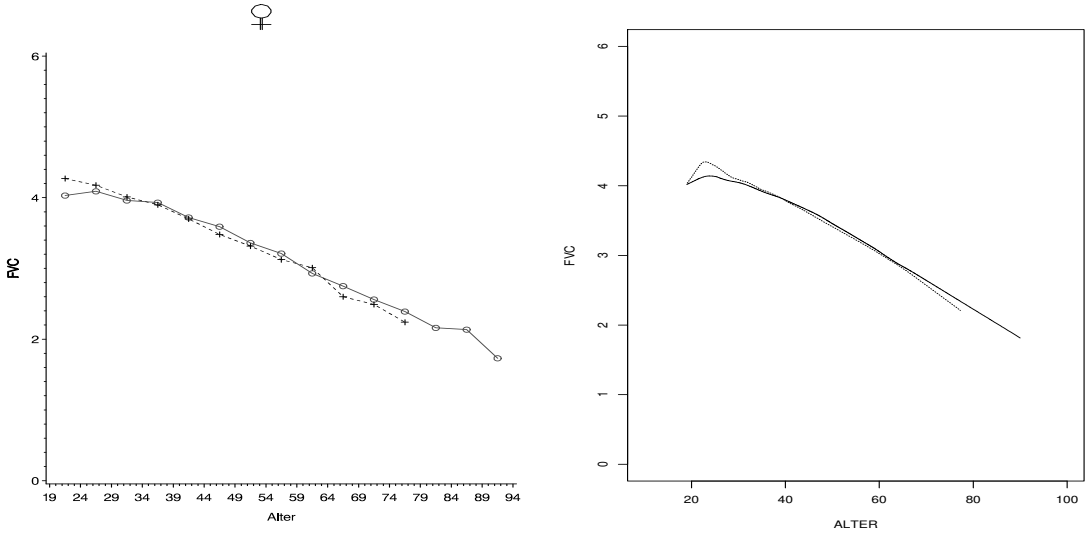


Abbildung 12.4: Frauen: FVC

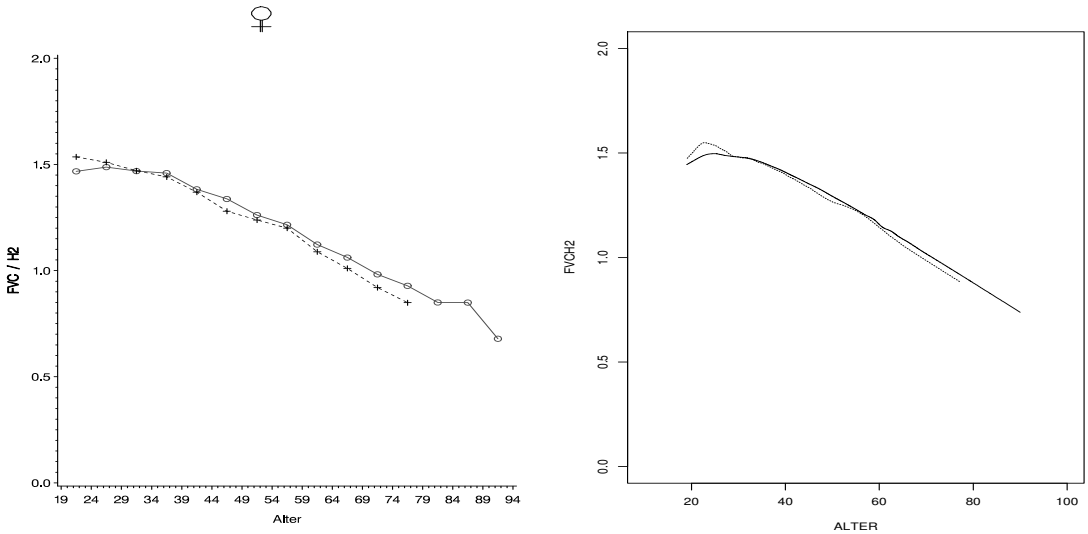


Abbildung 12.5: Frauen: FVC/H^2

Anhand der Parameter FVC und FVC/H^2 sind in den Abbildungen 12.4 und 12.5 keine wesentlichen Unterschiede zwischen den beiden Gruppen zu erkennen. Mit zunehmendem Alter ist eine leicht stärkere Abnahme der Mediane der schweren Raucherinnen festzustellen. Interessant sind die deutlich höheren Werte bei den schweren Raucherinnen im Altersbereich von etwa 20 - 30 Jahren was besonders deutlich an den Glättungskurven zu sehen ist.

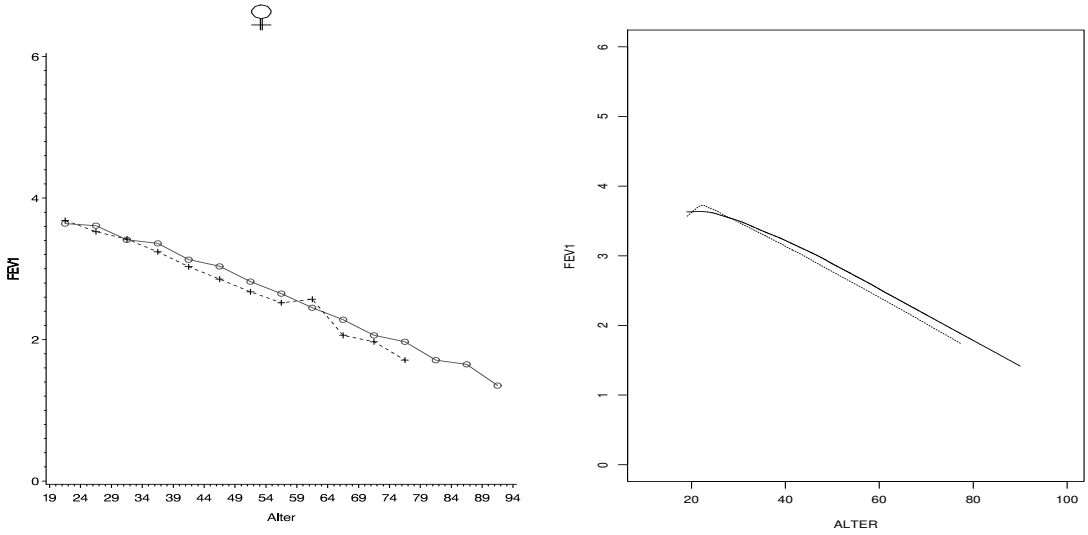


Abbildung 12.6: Frauen: FEV_1

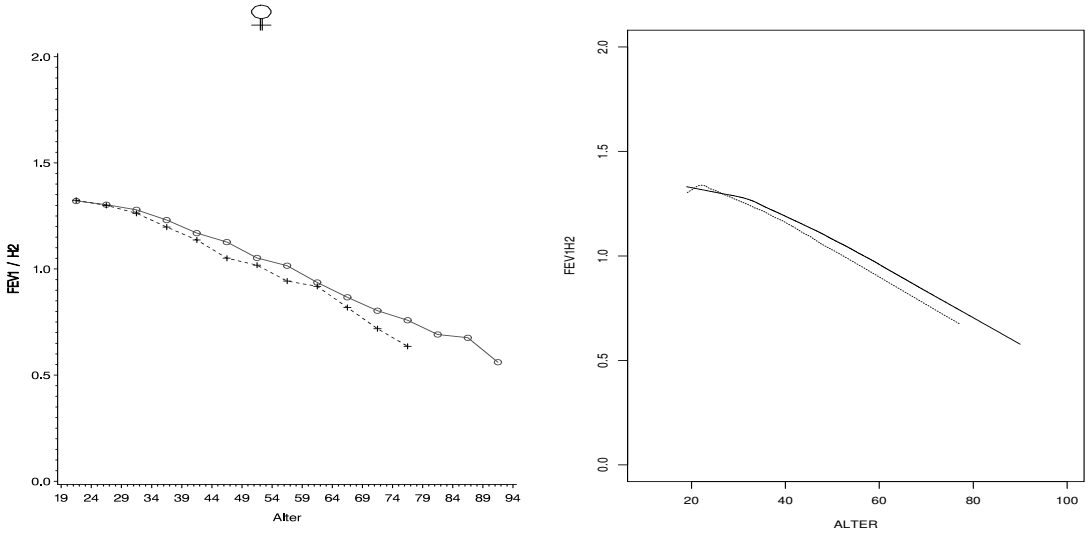


Abbildung 12.7: Frauen: FEV_1/H^2

Der Parameter FEV_1 zeigt im Vergleich zu FVC eine klare Unterscheidung der Gruppen. Hier sind besonders anhand der Glättungskurven die etwa ab 30 Jahren niedrigeren Werte der schweren Raucherinnen ersichtlich. Die zuvor deutlich höheren Werte der schweren Raucherinnen im Bereich von 20 - 30 Jahren treten nicht mehr in diesem Ausmaß auf.

Während die Glättungskurven einen kontinuierlichen Verlauf zeigen und die Kurve der schweren Raucherinnen konstant unter jener der Niemalsraucherinnen liegt, ist bei den Medianplots, besonders um 60 Jahre herum, sogar eine Überschneidung der Mediane festzustellen.

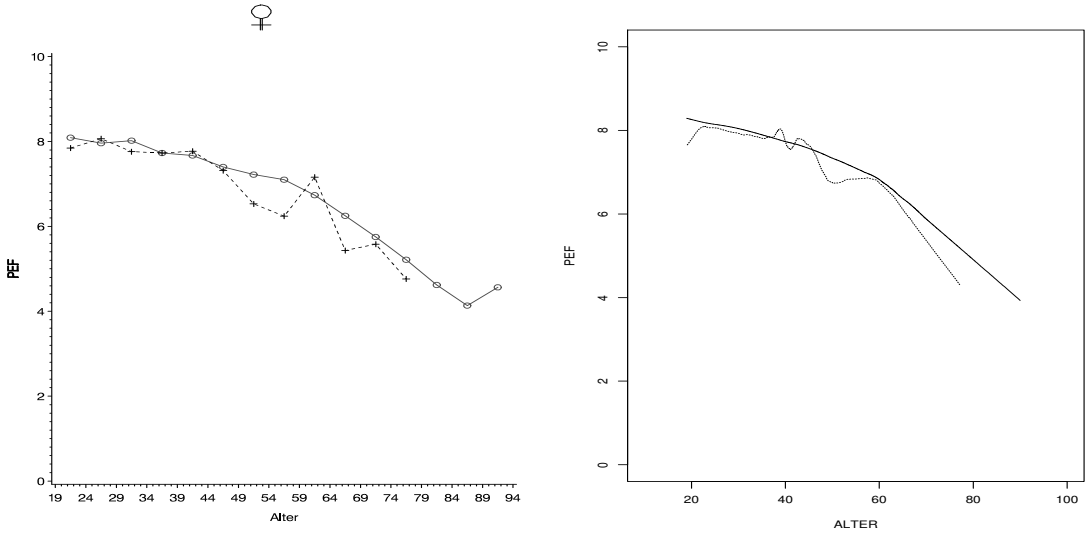


Abbildung 12.8: Frauen: PEF

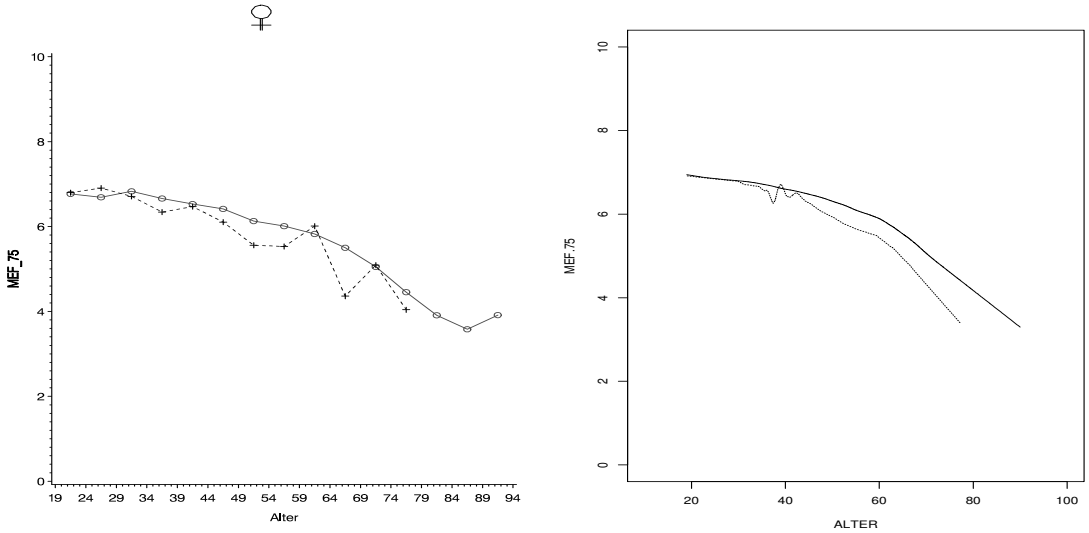


Abbildung 12.9: Frauen: MEF_{75}

Die Werte der Gruppe der schweren Raucherinnen unterliegen einer deutlich stärkeren Variabilität als die Werte der Niemalsraucherinnen. Beim Vergleich der Mediane ist wiederum der im Vergleich zu den vorhergehenden und nachfolgenden Altersklassen deutlich höhere Median der Klasse der 59-64 jährigen markant. Trotzdem sind, abgesehen von dieser einen Altersklasse, alle Mediane der schweren Raucherinnen niedriger als jene der Niemalsraucherinnen.

Bei den Glättungskurven zeigt vor allem der Parameter MEF_{75} eine deutlich stärkere Abnahme der Werte der schweren Raucherinnen.

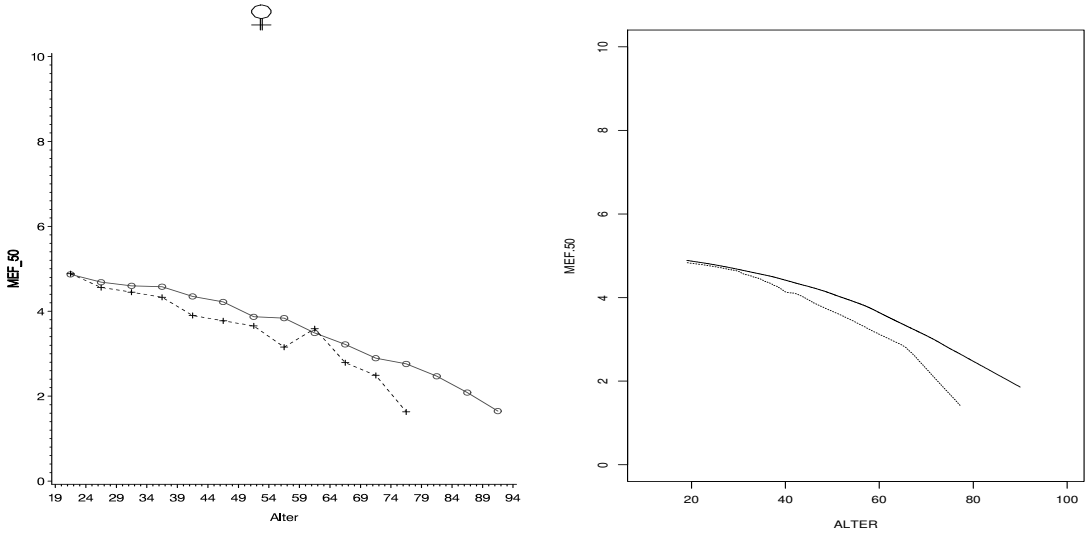


Abbildung 12.10: Frauen: MEF_{50}

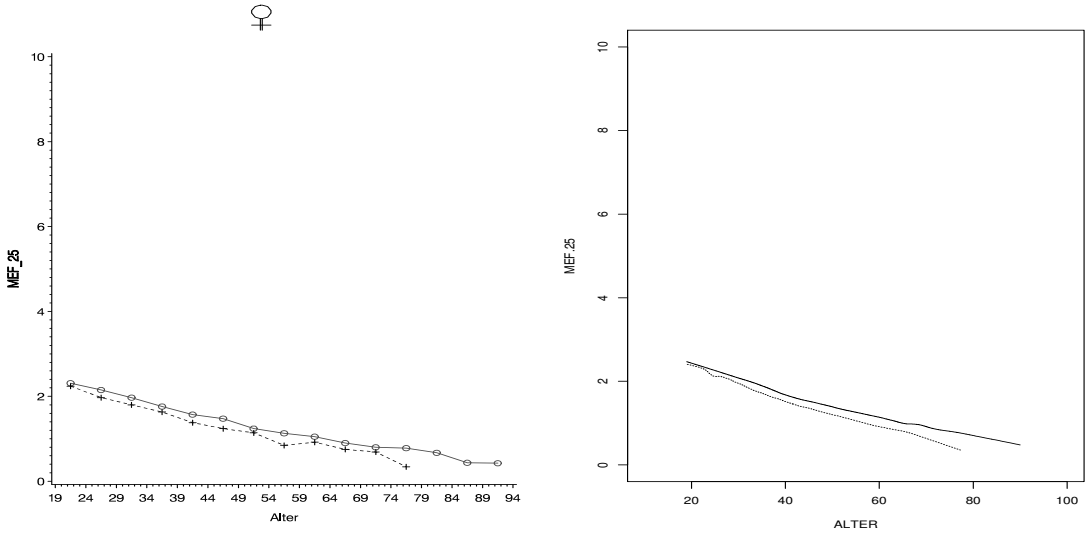


Abbildung 12.11: Frauen: MEF_{25}

Bei MEF_{50} ist der Median der 59-64 jährigen wiederum abweichend. Abgesehen von diesem 'Ausreißer' zeigt sich in beiden Graphiken, daß die Werte der schweren Raucherinnen stetig unter denen der Niemalsraucherinnen liegen und der Abstand mit zunehmendem Alter zusehends größer wird.

Für den Parameter MEF_{25} gilt in etwa die gleiche Charakteristik wie für MEF_{75} und MEF_{50}, wobei der absolute Unterschied aufgrund der niedrigen Werte geringer ist. Noch deutlicher sind hier schon ab 20 Jahren die niedrigen Werte der schweren Raucherinnen anhand der Kurven zu erkennen.

12.3 Männer: Vergleiche

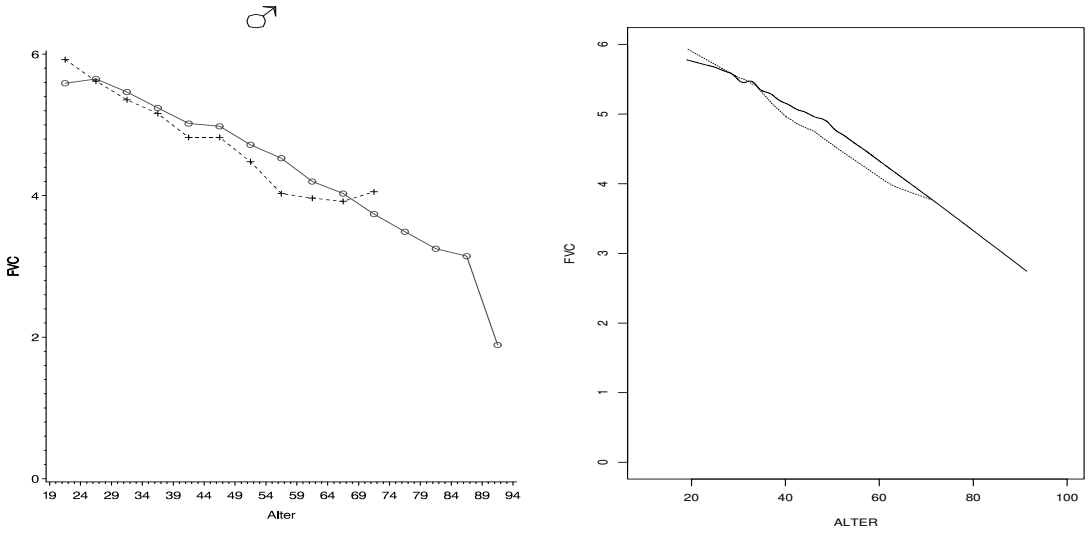


Abbildung 12.12: Männer: FVC

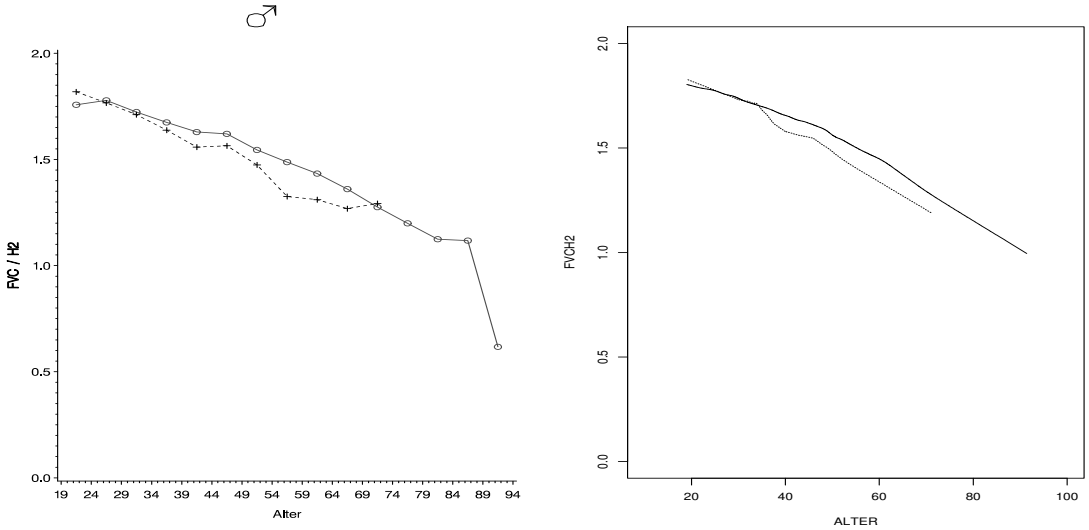


Abbildung 12.13: Männer: FVC/H^2

Bei den Männern ist im Gegensatz zu den Frauen bereits beim Parameter FVC ein deutlicher Unterschied der Gruppen ab etwa 30 Jahren zu erkennen. Die Überschneidung der Kurven in den letzten Altersklassen der schweren Raucher ist aufgrund der wenigen Werte in diesen Klassen nicht mehr als charakteristisch zu bewerten.

Bei den nach der Größe adjustierten Werten zeigen besonders die Glättungskurven die deutlich niedrigeren Werte der schweren Raucher.

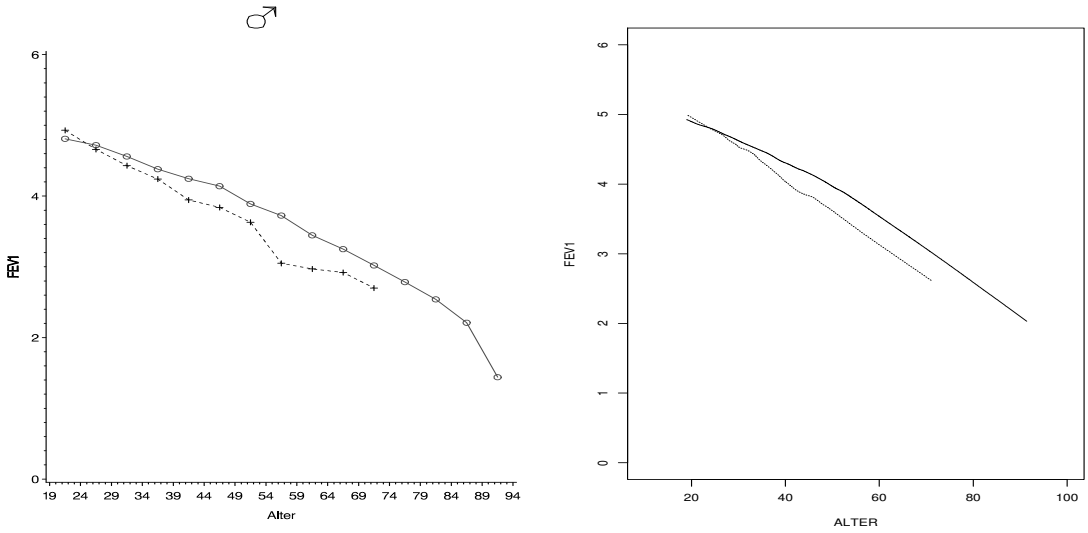


Abbildung 12.14: Männer: FEV_1

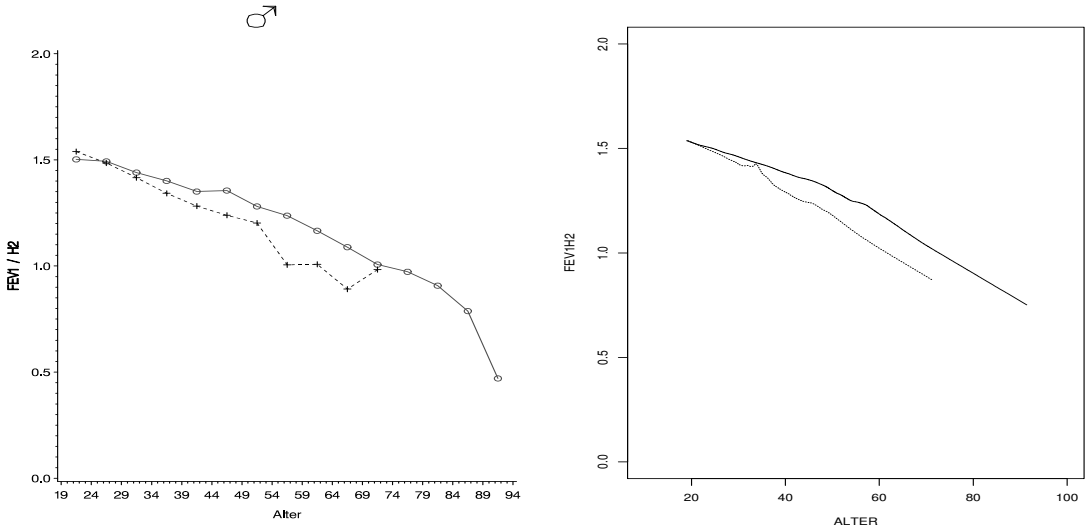


Abbildung 12.15: Männer: FEV_1/H^2

Hier ist der Unterschied zwischen beiden Gruppen besonders markant. Sowohl bei FEV_1 als auch bei FEV_1/H^2 liegen die Werte der schweren Raucher deutlich unter jenen der Niemalsraucher und der Abstand vergrößert sich noch mit zunehmendem Alter.

Medianplots und Glättungskurven zeigen bis auf die Gruppe der 69-74 jährigen ähnliche Tendenzen auf.

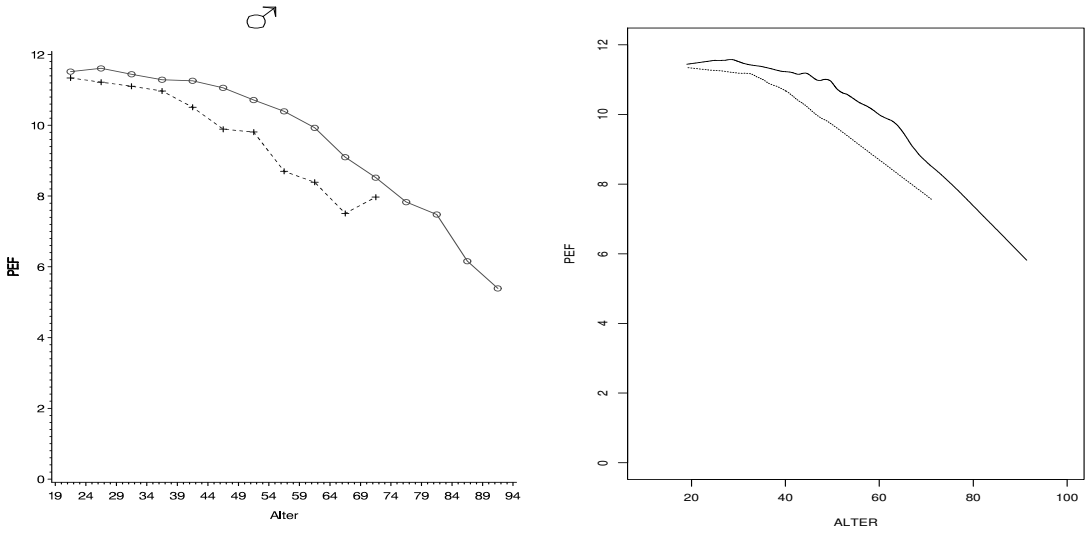


Abbildung 12.16: Männer: PEF

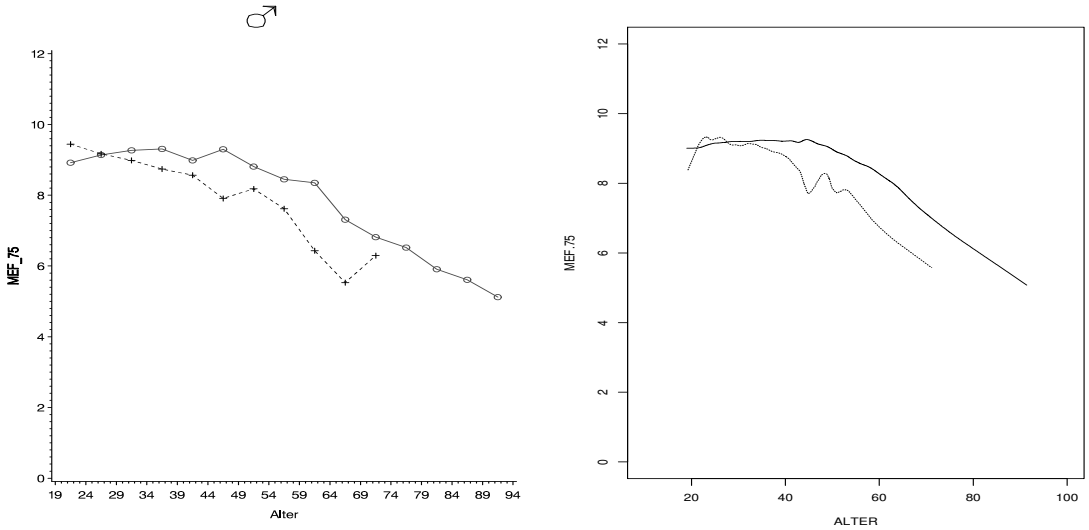


Abbildung 12.17: Männer: MEF_{75}

Bei PEF liegen sowohl die Mediane als auch die Glättungskurven der schweren Raucher durchgehend unter jenen der Niemalsraucher. Bereits ab 20 Jahren ist ein deutlicher Unterschied festzustellen.

Die Werte bei MEF_{75} der schweren Raucher variieren etwas stärker, wie an der Glättungskurve zu sehen ist. In Übereinstimmung mit den PEF-Werten zeichnen sich auch hier die schweren Raucher durch deutlich niedrigere Werte aus.

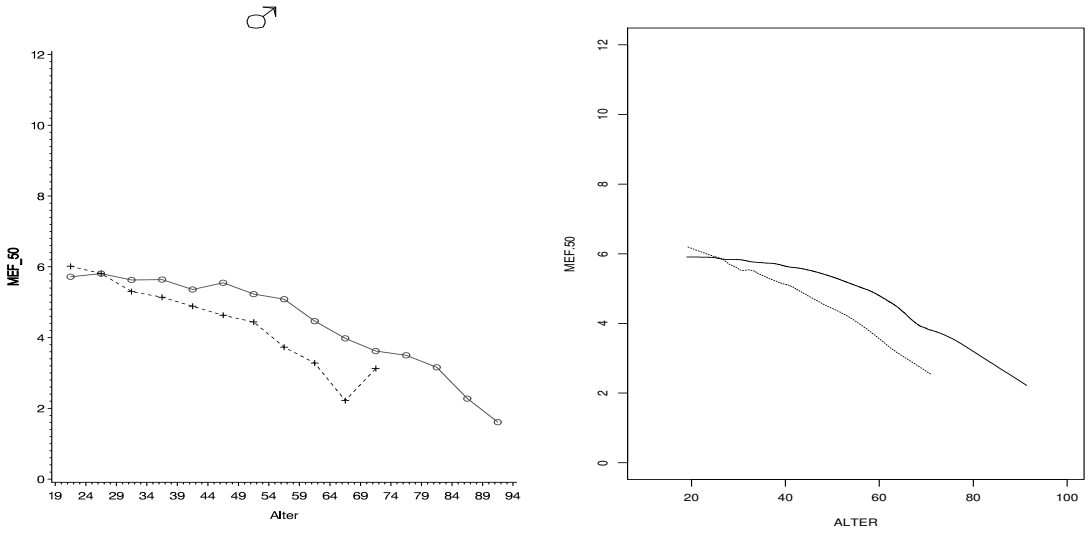


Abbildung 12.18: Männer: MEF_{50}

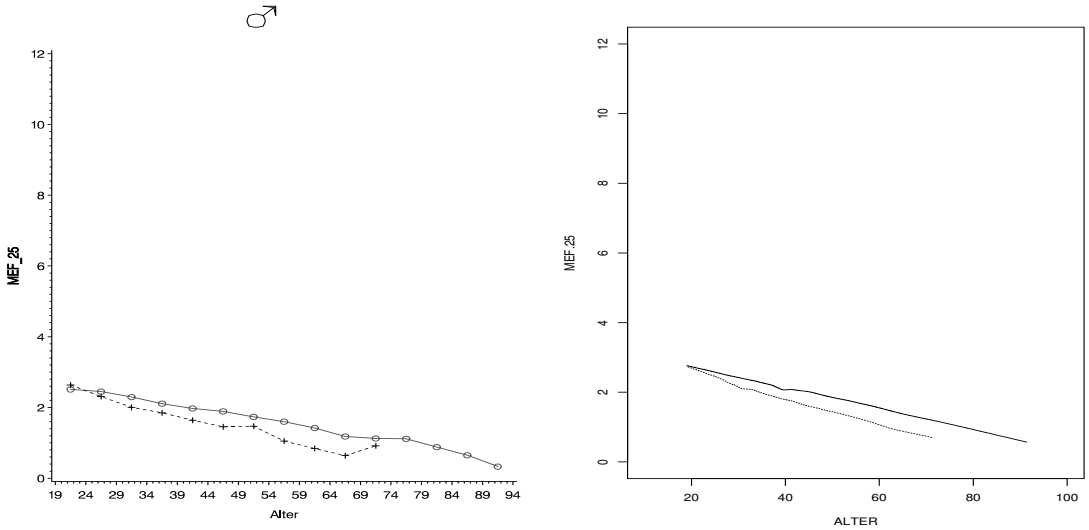


Abbildung 12.19: Männer: MEF_{25}

Bei beiden Parametern MEF_{50} und MEF_{25} haben die schweren Raucher niedrigere Werte als die Niemalsraucher. Ausgehend von einem etwa gleich hohen Niveau nehmen die Werte der schweren Raucher mit zunehmendem Alter stärker ab als die der Niemalsraucher.

Vergleicht man die Resultate der Frauen und der Männer miteinander, so kann gesagt werden, daß bei den Männern der negative Einfluß des Rauchens auf die Lungenfunktionsparameter deutlicher zu erkennen ist. Wobei allerdings zu beachten ist, daß bei den Frauen alle jene zu den schweren Rauchern gezählt werden, die 15 oder mehr Zigaretten/Tag rauchen.

12.4 Abschliessende Bemerkungen

Anhand der Graphiken ist, bei den Männern deutlicher als bei den Frauen, die schnellere Abnahme der Lungenfunktionswerte bei den schweren Raucher(innen) im Vergleich zu den Niemalsraucher(innen) zu erkennen. Wobei an dieser Stelle noch einmal darauf verwiesen sei, daß bei den Frauen zur Vergrößerung der Gruppe der schweren Raucherinnen jene aus der Gruppe der mittleren Raucherinnen hinzugenommen wurden, welche mindestens 15 Zigaretten/Tag rauchen.

Bei den Frauen ist diese schnellere Abnahme besonders bei den Parametern FEV_1 und MEF_{50} sowie etwas weniger deutlich an MEF_{75} und MEF_{25} zu sehen. Bei den Männern sind die Unterschiede bei allen Parametern deutlich zu sehen. Was die beiden Volumsparameter FVC und FEV_1 betrifft, ist zu sagen, daß jeweils in den Graphiken der nach der Größe adjustierten Parameter FVC/H^2 und FEV_1/H^2 der Unterschied markanter erscheint, wobei allerdings darauf zu achten ist, daß die y-Achse eine andere Skaleneinteilung besitzt. Der Vergleich der Graphiken der adjustierten mit den nicht adjustierten Volumsparameter ist noch einmal ein Hinweis dafür, in welchem Maße diese Parameter von der Größe abhängig sind.

Des weiteren wurden, ausgehend von der in den Medianplots verwendeten Altersklasseneinteilung, die adjustierten Mittelwerte berechnet. Die Adjustierung erfolgt wiederum nach dem Alter, der Größe sowie dem Gewicht (siehe 4.3). Folgende Tabelle gibt die p-Werte der Vergleiche bei den Frauen wieder.

	$H_0 : \mu_{NR} = \mu_R$, p-Werte : $P(\mid T \mid > t)$					
	[19-24)	[24-29)	[29-34)	[34-39)	[39-44)	[44-49)
FVC	0,0001	0,0517	0,9495	0,0001	0,0051	0,0002
FEV_1	0,0004	0,6601	0,0060	0,0001	0,0001	0,0001
PEF	0,4590	0,5020	0,0022	0,0129	0,0004	0,0004
MEF_{75}	0,0559	0,3505	0,0379	0,0015	0,0001	0,0001
MEF_{50}	0,1728	0,7037	0,0373	0,0001	0,0002	0,0001
MEF_{25}	0,0230	0,0001	0,0001	0,0001	0,0252	0,0001

Wie in der Tabelle schön zu sehen, läßt sich statistisch ein Unterschied zum Signifikanzniveau $p = 5\%$ zwischen den adjustierten Mittelwerten für alle Parameter erst ab der Altersklasse der 34-39 jährigen eindeutig nachweisen. Bei den Männern sind die adjustierten Mittelwerte der Niemalsraucher für alle Parameter bereits ab der Altersklasse der 19-24 jährigen *hoch signifikant höher* als jene der schweren Raucher.

Es kann gesagt werden, daß unsere Analysen die Schlußfolgerung erlauben, daß schwere Raucher(innen) im Vergleich zu den Niemalsraucher(innen) sowohl bei den Frauen und insbesondere bei den Männern deutlich schlechtere Lungenfunktionswerte besitzen. Besonders deutlich sind diese Unterschiede am Volumsparameter FEV_1 sowie an den MEF-Werten zu erkennen. Eine Interpretation dieses Ergebnisses ist, daß das Rauchen weniger das statische Lungenvolumen (FVC) als vielmehr das dynamische Verhalten der Lunge (FEV_1 und Flußwerte) negativ beeinflußt.

Literaturverzeichnis

[1] Bachmann K-D, Ewerbeck H, Kleihauer E, Rossi E, Stalder G: *Pädiatrie in Praxis und Klinik*, Band I, 2. neubearbeitete Auflage, Gustav Fischer Verlag Stuttgart - New York, Georg Thieme Verlag Stuttgart - New York.

[2] Dräger Medizintechnik: *Grundlagen der Physiologie*, http://www.draeger.com/-german/mt/mt-i/

[3] Falk M, Becker R, Mahron F: *Angewandte Statistik mit SAS*, Eine Einführung, Springer Verlag, Berlin, Heidelberg, 1995.

[4] Flury B, Riedwyl H: *Angewandte Multivariate Statistik - Computergestützte Analyse Mehrdimensionaler Daten*, Gustav Fischer Verlag, Stuttgart - New York, 1983.

[5] Friedl H: *Computeruntestützte Statistik*, Vorlesungsskriptum, Institut für Statistik, Technische Universität Graz, 1993.

[6] Hartung J, Elpelt B: *Statistik - Lehr- und Handbuch der angewandten Statistik* R. Oldenburg Verlag, München, 1995, 10.Auflage.

[7] Hastie TJ, Tibshirani RJ: *Generalized Additive Models, Monographs On Statistics And Applied Probability*, 43, Chapman & Hall, London, 1990.

[8] Hastie TJ, Tibshirani RJ: *Generalized additive models, Statistical Science*, 1, 297-318 (1986).

[9] Johns Hopkins School of Medicine's: *Interactive Respiratory Physiology*, http://infonet.welch.jhu.edu/∼omie/res...ForcedExpiration/

[10] Kleinbaum DG, Kupper LL, Muller KE: *Applied Regression Analysis and Other Multivariate Methodes*, 2^{nd} Edition, PWS-Kent, Boston, 1988.

[11] Kummer C: *Spirometrische Bezugswerte für Mädchen und Frauen*, Diplomarbeit, Institut für Statistik, Technische Universität Graz, 1995.

[12] Med Facts: *Spirometry Testing*, National Jewish Medical and Research Center, http://www.njc.org/MFhtml/SPI_MF.html

[13] Pierce R, Johns DP: *Spirometry; The Measurement And Interpretation Of Ventilatory Function In Clinical Practice*, The Thorac Society of Australia and New Zealand, http://hna.ffh.vic.gov.au/asthma/spiro

[14] Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yerault J-C: *Lung volumes and forced ventilatory flows*, European Respiratory Journal, 1993, 6, Suppl. 16, 5-40.

[15] Rapatz G: *Die Modellierung der Lungenfunktionsentwicklung bei Knaben und Männern*, Diplomarbeit, Institut für Statistik, Technische Universität Graz, 1995.

[16] SAS/STAT: *User's Guide*, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc., 1990.

[17] Segal MR, Weiss ST, Speizer FE: *Smoothing methods for epidemiologic analysis, Statistics in Medicine*, Vol. 7, 601-611 (1988).

[18] Shapiro SS, Wilk MB: *An analysis of variance test for normality (complete samples), Biometrika*, 1965, 52: 591-611.

[19] Shapiro SS, Wilk MB, Chen MJ: *A comparative study of various tests of normality, Journal of the American Statistical Association*, 1968, 63: 1343-1372.

[20] SPSS for Windows: *Base System Users's Guide*, Release 6.0, SPSS Inc., 1993.

[21] S-PLUS: *Guide to Statistical and Mathematical Analysis*, Version 3.3, StatSci Division, MathSoft, Inc., Seattle, Washington, September 1995.

[22] Stadlober E: *Wahrscheinlichkeitstheorie und Statistik für Telematiker*, Vorlesungsskriptum, 3. Auflage, Institut für Statistik, Technische Universität Graz, September 1992.