Vergleich eindimensionaler Stichproben

Zweistichproben-Problem: unabhängige oder **verbundene Stichproben?**

2 Populationen X und Y liegen vor.

Z.B. Aufschluss über $\theta = E(Y) - E(X)$. Geeigneter Schätzer $T = \bar{Y} - \bar{X}$.

Prinzipieller Unterschied in

$$\begin{aligned} \mathsf{var}(T) &= \mathsf{var}(\overline{Y}) + \mathsf{var}(\overline{X}) - 2\mathsf{cov}(\overline{Y}, \overline{X}) \\ &= \frac{\mathsf{var}(Y)}{n} + \frac{\mathsf{var}(X)}{n} - 2\rho\frac{\sqrt{\mathsf{var}(Y)\mathsf{var}(X)}}{n}, \end{aligned}$$

mit $\rho = \mathsf{cor}(X, Y)$. Für unabhängige Stichproben gilt $\rho = 0$.

Sind X und Y hoch positiv korreliert, verringert sich die Varianz von T.

Graphische Verfahren

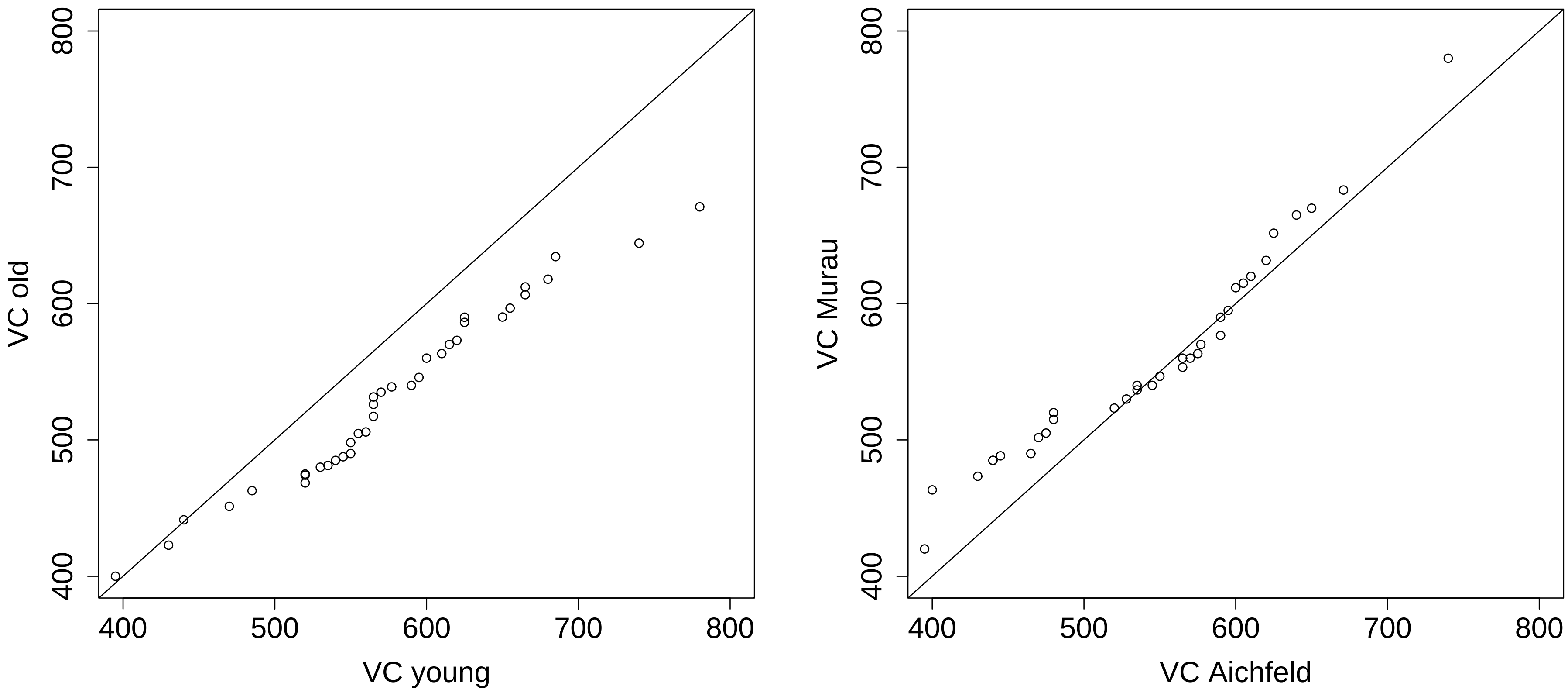
Empirischer Quantil-Quantil-Plot (EQQ-Plot):

$$q_Y(p) \Longleftrightarrow q_X(p)\,, \qquad 0 < p < 1$$

- $n = m$: wird der EQQ-Plot durch die Punkte $(x_{(i)}, y_{(i)})$ gebildet.
- $n \neq m$: Quantile der größeren Stichprobe werden durch Interpolation bestimmt.

Interpretation: Falls X und Y ident verteilt, so resultiert die Gerade $x = y$.

```
> qqplot(VC[age<30], VC[age>=30]);  abline(0, 1)
> qqplot(VC[region=="A"], VC[region=="M"]);  abline(0, 1)
```



EQQ-Plot von `VC` für den Vergleich jung/alt (links), sowie für die Bezirke Aichfeld/Murau (rechts).

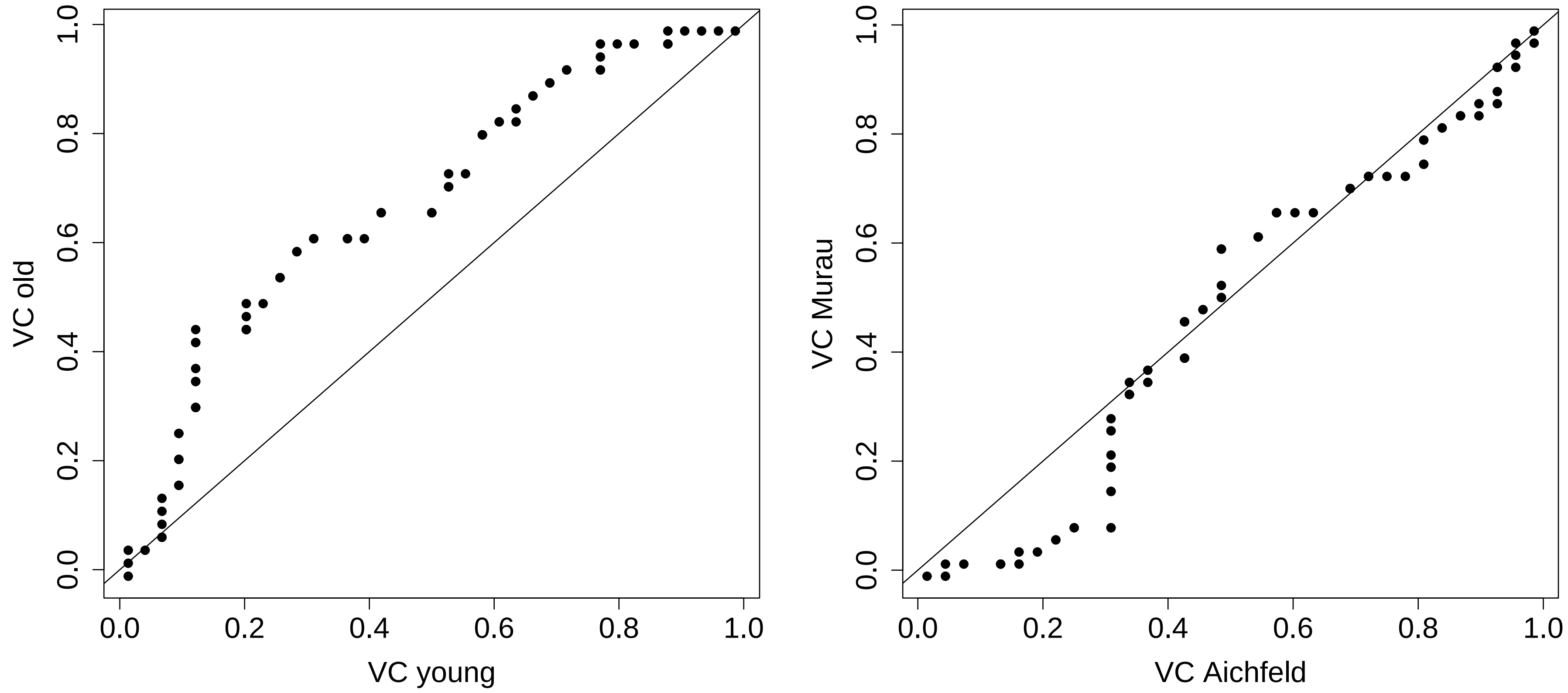
Empirischer Prozent-Prozent-Plot (EPP-Plot)

$$p_Y(q) \Longleftrightarrow p_X(q)$$

q durchläuft den Datenbereich: zB. die kombinierte geordnete Stichprobe.

Falls X, Y beide $U(0,1)$-verteilt, sind der EQQ-Plot und der EPP-Plot ident.

```
> s <- sort(VC);  n <- length(VC)
> sx <- VC[age< 30];  sy <- VC[age>=30]  # Altersgruppen
> px <- py <- 1:n
> for (i in 1:n) {
    px[i] <- (sum(sx <= s[i]) - 1/2)/lenx
    py[i] <- (sum(sy <= s[i]) - 1/2)/leny
  }
>  plot(px, py);  abline(0, 1)
```



EPP-Plot der Variablen `VC` für die beiden Altersgruppen (links) und für die beiden der Bezirke Aichfeld/Murau (rechts).

Vergleich mehrerer Gruppen

Boxplot Serien: mit **Modifikationen**

1. **Variable-Width Boxplot:** Breite der Box proportional zu Stichprobenumfang.

2. **Notched Boxplot:** hat Kerben der Form $\tilde{X} \pm cS_{\tilde{X}}$

Bemerkungen zu den Kerben im Fall zweier Gruppen:

Test auf Gleichheit von $\mathsf{E}(\tilde{X}) = m_X$ und $\mathsf{E}(\tilde{Y}) = m_Y$ derart konzipieren, dass bei Überlappung der Kerben $H_0 : m_X - m_Y = 0$ nicht verworfen werden kann.

Annahme: $\tilde{X} \sim N(m_X, \sigma^2)$ und $\tilde{Y} \sim N(m_Y, k^2\sigma^2)$. Unter H_0 folgt

$$\frac{(\tilde{X} - \tilde{Y}) - (m_X - m_Y)}{\sqrt{1 + k^2}\sigma} \overset{H_0}{=} \frac{\tilde{X} - \tilde{Y}}{\sqrt{1 + k^2}\sigma} \overset{H_0}{\sim} N(0, 1)\,.$$

H_0 nicht verwerfen, falls die Null im Intervall

$$\left[\tilde{X} - \tilde{Y} - z_{1-\alpha/2}\sqrt{1+k^2}\sigma, \tilde{X} - \tilde{Y} + z_{1-\alpha/2}\sqrt{1+k^2}\sigma\right]$$

enthalten ist. Die Null ist aber gerade dann in diesem Intervall, wenn

$$\begin{aligned}
\tilde{X} - \tilde{Y} - z_{1-\alpha/2}\sqrt{1+k^2}\sigma &\leq 0 \\
\tilde{X} - \tilde{Y} + z_{1-\alpha/2}\sqrt{1+k^2}\sigma &\geq 0\,.
\end{aligned}$$

Sei $z_{1-\alpha/2}\sqrt{1+k^2}\sigma = c\sigma + ck\sigma = c\sigma(1+k)$, so folgt $c = z_{1-\alpha/2}\sqrt{1+k^2}/(1+k)$ und die beiden obigen Bedingungen sind äquivalent mit

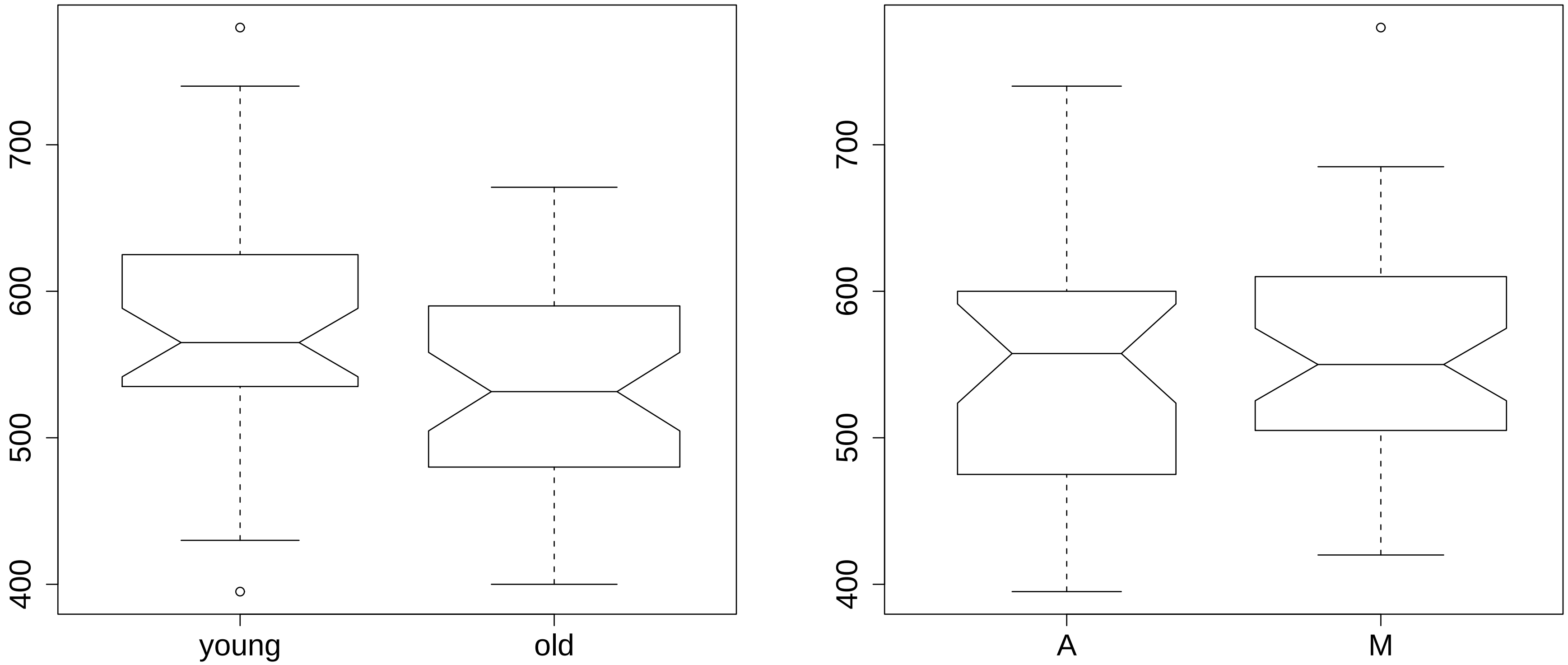
$$\begin{aligned}
\tilde{X} - z_{1-\alpha/2}\frac{\sqrt{1+k^2}}{1+k}\sigma &\leq \tilde{Y} + z_{1-\alpha/2}\frac{\sqrt{1+k^2}}{1+k}k\sigma \\
\tilde{X} + z_{1-\alpha/2}\frac{\sqrt{1+k^2}}{1+k}\sigma &\geq \tilde{Y} - z_{1-\alpha/2}\frac{\sqrt{1+k^2}}{1+k}k\sigma\,.
\end{aligned}$$

Bei $k = 1$, wähle für $\alpha = 0.05$ in jedem Boxplot $c = z_{1-\alpha/2}\sqrt{2}/2 = 1.386$.
Bei $k = 2$ resultiert $c = z_{1-\alpha/2}\sqrt{5}/3 = 1.461$.
Wähle $c = 1.7$, so ist man oft auf der sicheren Seite.

$$\tilde{X} \pm 1.7 S_{\tilde{X}} = \tilde{X} \pm 1.7 \frac{1.25 \mathsf{IQR}}{1.35\sqrt{n}} = \tilde{X} \pm 1.57 \frac{\mathsf{IQR}}{\sqrt{n}}\,.$$

Bei Überlappung kann die Gleichheitshypothese nicht verworfen werden.

```
> a <- as.factor(trunc(age/30))
> levels(a) <- c("young", "old")
> boxplot(VC ~ a, varwidth = TRUE, notch=TRUE)
```



Boxplots der Variablen `VC` mit Notches und proportionalen Breiten für die beiden Altersgruppen (links) und für die Regionen Aichfeld/Murau (rechts).

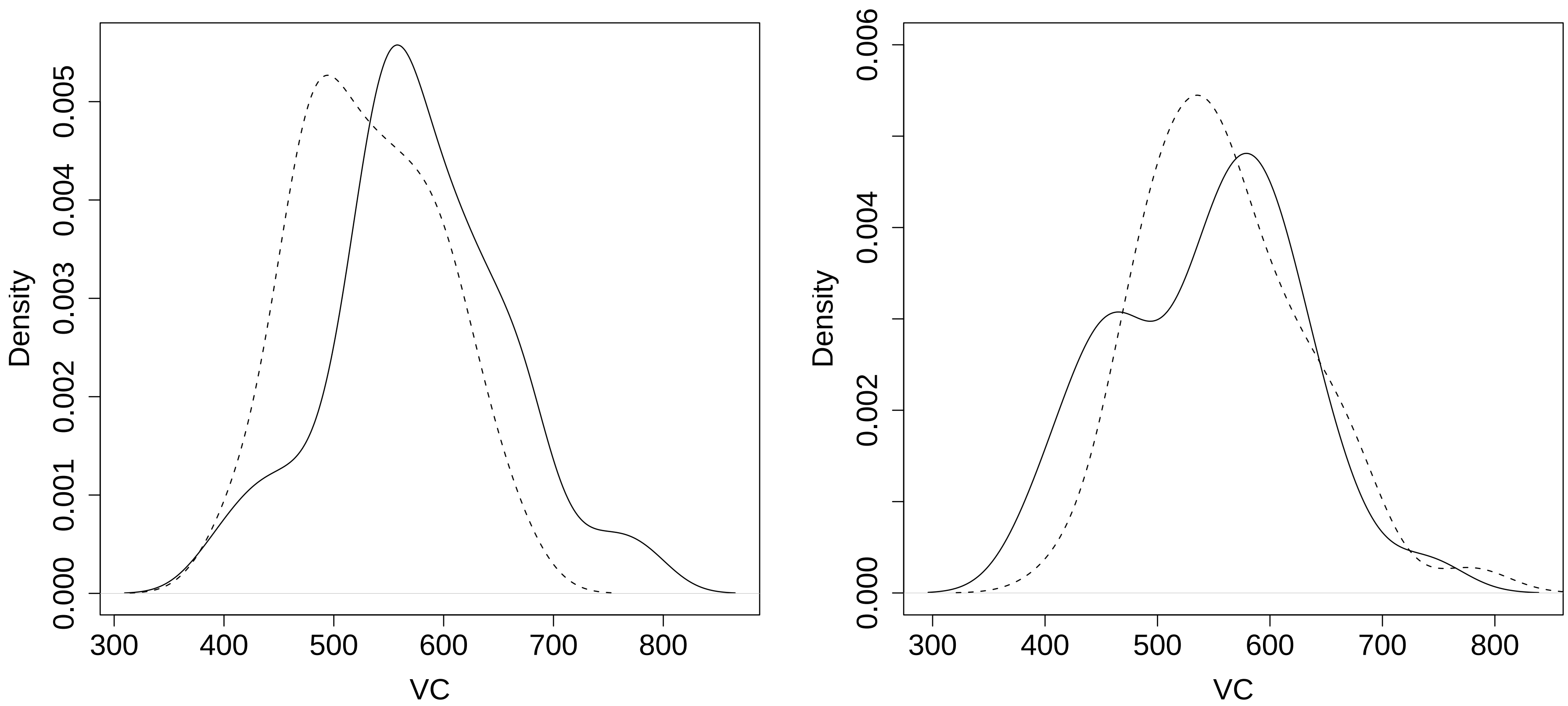
Kernschätzer: Beim Vergleich von Gruppen wähle h sowie Kern K einheitlich!

Z.B. Gauss-Kern und das Mittel beider optimalen Fensterbreiten

Alter: $h = 29.36$ (junge) und $h = 27.66$ (ältere), also $h = 28.5$ verwenden.

Regionen: $h = 36.47$ (Aichfeld) sowie $h = 30.20$ (Murau), also $h = 33.00$.

```
> plot(density(VC[a=="young"], bw=28.5)
> lines(density(VC[a=="old"], bw=28.5), lty=2)

> plot(density(VC[region=="A"], bw=33.0))
> lines(density(VC[region=="M"], bw=33.0), lty=2)
```



Kernschätzer der beiden VC-Gruppierungen für (links: jung/solid, älter/dashed) und (rechts: Aichfeld/solid, Murau/dashed).

Lineare Rangstatistik

Nichtparametrische Verfahren – Rangtests

Verwendete Teststatistik: Funktion der Ränge und nicht der Beobachtungen

$X_i \overset{iid}{\sim} F$, $i = 1, \ldots, m$, unabhängig von $Y_j \overset{iid}{\sim} G$, $j = 1, \ldots, n$

F und G stetige Verteilungen

Wir testen generell die Gleichheithypothese $H_0 : F(z) = G(z),\ \forall z \in \mathbb{R}$ gegen

allgemeine Alternative: $H_1 : F(z) \neq G(z)$,

Lokationsalternative: $H_1 : F(z) = G(z + \theta)$, $\theta \neq 0$,

Variabilitätsalternative: $H_1 : F(z) = G(z\theta)$, $0 < \theta \neq 1$.

Definition 9. *In der* **kombinierten Stichprobe** $Z = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ *sind die Ränge* R_i *der* X_i *für* $i = 1, \ldots, m$ *bestimmt durch:*

$$R_i = R(X_i) = \sum_{j=1}^{m} T(X_i - X_j) + \sum_{k=1}^{n} T(X_i - Y_k)$$

mit

$$T(u) = \begin{cases} 0 & \textit{für} \quad u < 0\,, \\ 1 & \textit{für} \quad u \geq 0\,. \end{cases}$$

Für Rangtests erweist es sich als sinnvoll, die **kombinierte, geordnete Stichprobe** $Z_{(\cdot)} = (Z_{(1)}, \ldots, Z_{(N)})$ mit Umfang $N = m + n$ durch den Vektor V zu charakterisieren, der die Zugehörigkeit zur Gruppe X beschreibt:

$$V_i = \begin{cases} 1 & \text{falls } Z_{(i)} \text{ eine } X\text{-Variable}\,, \\ 0 & \text{falls } Z_{(i)} \text{ eine } Y\text{-Variable}\,. \end{cases}$$

Statistiken, die auf Ränge basieren, lassen sich oft linear in den V_i darstellen:

$$L_N = \sum_{i=1}^{N} g_i V_i$$

mit reellen Gewichtungsfaktoren g_i. L_N heißt **lineare Rangstatistik**.

Satz 8. *Unter $H_0 : F = G$ gilt für alle $i = 1, \dots, N$:*

1. *E*$(V_i) = \frac{m}{N}$, *var*$(V_i) = \frac{mn}{N^2}$, *cov*$(V_i, V_j) = -\frac{mn}{N^2(N-1)}$.

2. *E*$(L_N) = \frac{m}{N} \sum_{i=1}^{N} g_i$,
   var$(L_N) = \frac{mn}{N^2(N-1)} \left(N \sum_{i=1}^{N} g_i^2 - \left(\sum_{i=1}^{N} g_i \right)^2 \right)$.

3. $P(V_1 = v_1, \dots, V_N = v_N) = 1/\binom{N}{m}$.

4. $P(L_N = c) = a(c)/\binom{N}{m}$, *wobei* $a(c)$ *die Anzahl der Vektoren* $v = (v_1, \ldots, v_N)$ *ist, für die* $L_N = c$ *gilt.*

5. L_N *ist symmetrisch um* $\mathit{E}(L_N)$ *verteilt, falls* $g_i + g_{N-i+1} = k$, *konstant.*

6. *Für* $m, n \to \infty$, *mit* $m/n \to \lambda$, $0 < \lambda < \infty$, *strebt unter* H_0

$$Z_N = \frac{L_N - \mathit{E}(L_N)}{\sqrt{\mathit{var}(L_N)}}$$

gegen eine $N(0,1)$*-verteilte ZV.*

Tests der allgemeinen Alternative

Iterations- (Runs-) Test:

Entspricht im **Einstichprobenfall:** Test auf Zufälligkeit
und im **Zweistichprobenfall:** Test auf Gleichheit zweier Verteilungsfunktionen

Bei binäre Daten (Geschlecht, Erfolg usw.) bedeutet Zufälligkeit, dass alle möglichen Reihenfolgen des Auftretens die gleiche Wahrscheinlichkeit haben.

Definition 10. *Unter einer* ***Iteration*** *(Run) versteht man eine Folge von einem oder mehreren identischen Symbolen, denen entweder ein anderes oder kein Symbol unmittelbar vorangeht oder folgt. Die Statistik R zählt die Anzahl der Iterationen.*

Beispiel: $n = 20$ Schüler ($n_1 = 8$ Jungen und $n_2 = 12$ Mädchen) warten in einer Schlange vor dem Würstchenstand und zwar in folgender Reihenfolge:

J J M M M M J J J M M M M M J J M M M J

Totale Anzahl der beobachteten Iterationen

$$r = r_J + r_M = \#(\text{Iterationen vom Typ J und M}) = 4 + 3 = 7.$$

Testproblem

- H_0 : Die Reihenfolge ist zufällig

Mögliche Alternativen

- H_1: Die Reihenfolge ist nicht zufällig (Gruppierung)
- H_1: geschlechtshomoge Gruppierung (wenige Iterationen)
- H_1: geschlechtsinhomoge Gruppierung (viele Iterationen)

Im zweiseitigen Test wird H_0 abgelehnt, falls $r \leq r_{\alpha/2}$ oder $r \geq r_{1-\alpha/2}$.

Richtung der Abweichung von der Zufälligkeit (einseitiger Test):

1. zu wenig Iterationen, d.h. H_0 wird abgelehnt, wenn $r \leq r_\alpha$ ist, oder

2. zu viele Iterationen, d.h. H_0 wird abgelehnt, wenn $r \geq r_{1-\alpha}$ ist.

Beispiel: für $\alpha = 0.1$ folgt für Alternative $Geschlechtsinhomogenität$ kritischer Wert $r_{0.9} = 14$, was wegen $r = 7 < r_{1-\alpha}$ nicht zur Ablehnung von H_0 führt. Bzgl. Alternative $Geschlechtshomogenität$ resultiert $r_{0.1} = 8$ was zur Ablehnung von H_0 führen würde.

Achtung!

Sei $n = 28$ mit $n_1 = n_2 = 14$. Für den zweiseitigen Iterationstest und $\alpha = 0.05$ resultiert als Annahmebereich $[10, 20]$.

Gewisse systematische Anordnungen werden nicht erkannt.

$$\begin{array}{l|c|ll}
0\ldots0|1\ldots1| & r = 2 & H_0 & \text{Ablehnung,} \\
0|1|0|1|\ldots|0|1| & r = 28 & H_0 & \text{Ablehnung,} \\
00|11|00|\ldots|11| & r = 14 & H_0 & \text{Annahme!!,} \\
000|111|000|\ldots|00|11| & r = 10 & H_0 & \text{Annahme!!.}
\end{array}$$

Wald-Wolfowitz Test

Iterationstest bei unabhängigen Stichproben:

- $H_0 : F(z) = G(z)$
- $H_1 : F(z) \neq G(z)$

$X_i \overset{iid}{\sim} F$ und $Y_i \overset{iid}{\sim} G$, stetig.

Prozedur:

Bilde kombinierte, geordnete Stichprobe

Zähle Iterationen von x und y Beobachtungen

Lehne H_0 ab, falls Iterationszahl klein.

Falls Iterationszahl sehr groß, ist dies ein starker Hinweis auf Gültigkeit von H_0

Beispiel: Körpergröße

$x_{(i)}$	117	120	122	124	126	126	128	132		
$y_{(j)}$	110	113	114	116	116	118	119	119	123	125

Liegt dieselbe Verteilung vor?

Als kombinierte, geordnete Stichprobe resultiert

110	113	114	116	116	117	118	119	119	120	122	123	124	125	126	126	128	132
y	y	y	y	y	x	y	y	y	x	x	y	x	y	x	x	x	x

$r = 8$ Iterationen.

Für $\alpha = 0.05$ folgt $r_{0.05} = 6$. Wegen $r > r_{0.05}$ kann H_0 nicht verworfen werden.

Kolmogorov-Smirnov Test

Anpassungstest für den Vergleich zweier Verteilungen.

Annahmen: $X_i \stackrel{iid}{\sim} F$, $i = 1, \ldots, m$, und $Y_j \stackrel{iid}{\sim} G$, $j = 1, \ldots, n$, stetig verteilt und X, Y unabhängig.

Hypothesen:

- Test A: $H_0 : F(z) = G(z), \quad H_1 : F(z) \neq G(z)$,
- Test B: $H_0 : F(z) \leq G(z), \quad H_1 : F(z) > G(z)$,
- Test C: $H_0 : F(z) \geq G(z), \quad H_1 : F(z) < G(z)$.

KS-Teststatistik ist definiert durch

- Test A: $K_{m,n} = \max_{z\in\mathbb{R}} |F_m(z) - G_n(z)|$,
- Test B: $K^+_{m,n} = \max_{z\in\mathbb{R}}(F_m(z) - G_n(z))$,
- Test C: $K^-_{m,n} = \max_{z\in\mathbb{R}}(G_n(z) - F_m(z))$.

Entscheidung: Ablehnung, falls

- Test A: $k_{m,n} > k_{1-\alpha}$; $P(K_{m,n} > k_{1-\alpha}) = \alpha$,
- Test B: $k^+_{m,n} > k^+_{1-\alpha}$; $P(K^+_{m,n} > k^+_{1-\alpha}) = \alpha$,
- Test C: $k^-_{m,n} > k^-_{1-\alpha}$; $P(K^-_{m,n} > k^-_{1-\alpha}) = \alpha$.

Verteilung von $K_{m,n}$ unter H_0: anhand eines Beispiels:

Sei $m = 2$ und $n = 3$ dann gibt es insgesamt

$$\binom{m+n}{n} = \binom{5}{2} = 10$$

Stichproben, welche alle unter H_0 gleichwahrscheinlich sind.

kombinierte, geordnete Stichproben	k	$P(K_{2,3} = k)$
$(xxyyy)$, $(yyyxx)$	1	$P(K_{2,3} = 1) = 2/10$
$(yxxyy)$, $(yyxxy)$, $(yyxyx)$, $(xyxyy)$	2/3	$P(K_{2,3} = 2/3) = 4/10$
$(xyyxy)$, $(xyyyx)$, $(yxyyx)$	1/2	$P(K_{2,3} = 1/2) = 3/10$
$(yxyxy)$	1/3	$P(K_{2,3} = 1/3) = 1/10$

Beispiel: Körpergröße

Intervall	$\lvert F_m(z) - G_n(z)\rvert$	Intervall	$\lvert F_m(z) - G_n(z)\rvert$
$-\infty < z < 110$	0	$120 \leq z < 122$	0.550
$110 \leq z < 113$	0.100	$122 \leq z < 123$	0.425
$113 \leq z < 114$	0.200	$123 \leq z < 124$	0.525
$114 \leq z < 116$	0.300	$124 \leq z < 125$	0.400
$116 \leq z < 117$	0.500	$125 \leq z < 126$	0.500
$117 \leq z < 118$	0.375	$126 \leq z < 128$	0.250
$118 \leq z < 119$	0.475	$128 \leq z < 132$	0.125
$119 \leq z < 120$	*0.675*	$132 \leq z < \infty$	0

Liegt dieselbe Verteilung vor?

Zweiseitiger KS-Test (Test A)

Mit $\alpha = 0.05$ folgt $k_{8,10} = 0.675 > k_{0.95} = 46/80 = 0.575$

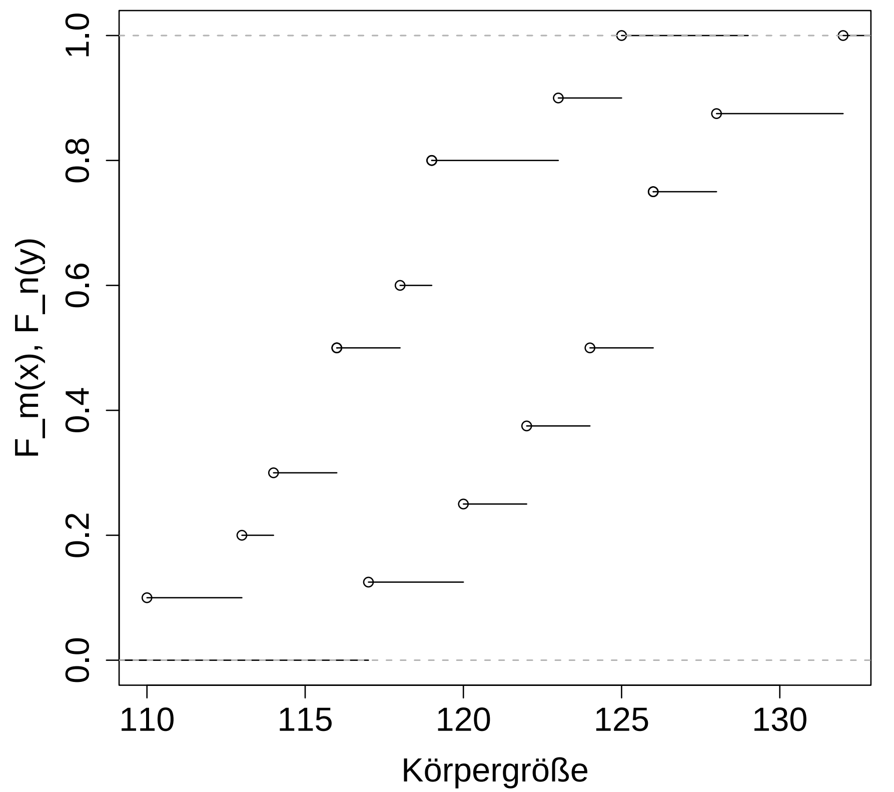
Also ist hierfür H_0 abzulehnen.

```
> x <- c(117,120,122,124,126,126,128,132)          # m=8 Mädchen
> y <- c(110,113,114,116,116,118,119,119,123,125)  # n=10 Knaben
> ks.test(x, y)

        Two-sample Kolmogorov-Smirnov test

 data:  x and y
 D = 0.675, p-value = 0.03484
 alternative hypothesis: two.sided

 Warning message:
 cannot compute correct p-values with ties in: ks.test(x, y)
```

Warnung wegen Bindungen (stetige Verteilungen würden diese nicht generieren).

Vergleich der beiden empirischen Verteilungsfunktionen: maximaler Abstand der beiden Treppen (KS-Abstand) im Intervall $[119, 129)$.

```
> library(stepfun)
> plot(ecdf(x)); lines(ecdf(y))
```



F_m(x), F_n(y)
0.0
0.2
0.4
0.6
0.8
1.0
110
115
120
125
130
Körpergröße

Tests bezüglich Lokationsalternativen

Zwei Stichproben sind bis auf Lage– (Lokations–) Parameter identisch verteilt.

Test auf Lageunterschiede:

- $H_0 : G(z) = F(z) \quad \forall z \in \mathbb{R}$,
- $H_1 : G(z) = F(z - \theta) \quad \forall z \in \mathbb{R}, \theta \neq 0$.

Alternativ–Hypothesen:

- (A) $\theta \neq 0$ (zweiseitig); d.h. $F \neq G$,
- (B) $\theta > 0$ (einseitig); d.h. $F \geq G$,
- (C) $\theta < 0$ (einseitig); d.h. $F \leq G$.

Parametrischer Test bei Normalverteilung

$X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$ und $Y_j \stackrel{iid}{\sim} N(\mu_Y, \sigma_Y^2)$:

Test auf Gleichheit der Erwartungswerte (t–Test)

H_0	H_1	Entscheidung gegen H_0, falls	kritische Werte
$\mu_Y - \mu_X = \theta = 0$	$\theta \neq 0$	$T < c_3$ oder $T > c_4$	$c_3 = t_{\alpha/2}$ $c_4 = t_{1-\alpha/2}$
$\mu_Y - \mu_X = \theta = 0$	$\theta > 0$	$T < c_1$	$c_1 = t_\alpha$
$\mu_Y - \mu_X = \theta = 0$	$\theta < 0$	$T > c_2$	$c_2 = t_{1-\alpha}$

1. $\sigma_X^2 = \sigma_Y^2$ (unbekannt) oder σ_X^2/σ_Y^2 bekannt.

$$T = \frac{\left(\overline{X} - \overline{Y}\right)\sqrt{\frac{nm}{n+m}}}{\underbrace{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}}_{S_P}} \sim t_{n+m-2}.$$

Diese vereinfacht sich bei $n = m$ zu

$$T = \frac{\left(\overline{X} - \overline{Y}\right)\sqrt{n}}{\sqrt{S_X^2 + S_Y^2}}.$$

S_P^2 bezeichnet man als *gepoolte Varianz*.

2. $\sigma_X^2 \neq \sigma_Y^2$.

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}} \overset{ap}{\sim} t_\nu \quad \text{mit} \quad \nu = \frac{\left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2}.$$

Entspricht Satterthwait's Approximation für die Freiheitsgrade einer Summe gewichteter unabhängiger χ^2-Größen. Man findet diesen Test auch unter der Bezeichnung Welch-Test.

```
> t.test(x, y, paired=FALSE, var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = 3.2357, df = 16, p-value = 0.005174
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
  2.439771 11.710229
sample estimates: mean of x mean of y
  124.375   117.300

> t.test(x, y, paired=FALSE, var.equal=FALSE)

        Welch Two Sample t-test

data:  x and y
t = 3.2196, df = 14.837, p-value = 0.005797
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.386632 11.763368
sample estimates: mean of x mean of y
  124.375   117.300
```

Der Wilcoxon-Rangsummentest (1945)

Annahme: $X_i \overset{iid}{\sim} F$, $Y_j \overset{iid}{\sim} G$, stetig verteilt, zumindest **ordinales** Meßniveau

Hypothesen:

- (A) $H_0 : G(z) = F(z); \quad H_1 : G(z) = F(z-\theta), \theta \neq 0$,
- (B) $H_0 : G(z) = F(z), \quad H_1 : G(z) = F(z-\theta), \theta > 0$,
- (C) $H_0 : G(z) = F(z), \quad H_1 : G(z) = F(z-\theta), \theta < 0$.

Teststatistik W_N:

$$W_N = \sum_{i=1}^{N} iV_i = \sum_{i=1}^{m} R_i \quad \text{mit} \quad N = m + n,$$

R_i Ränge der X_i in der kombinierten, geordneten Stichprobe.

Entscheidung: Verwerfe H_0, falls

- (A) $w_N \geq w_{1-\alpha/2}$ oder $w_N \leq w_{\alpha/2}$,
- (B) $w_N \leq w_\alpha$,
- (C) $w_N \geq w_{1-\alpha}$.

Verteilung von W_N unter H_0?

W_N ist eine **lineare Rangstatistik** L_N mit Gewichten $g_i = i$

- W_N ist symmetrisch verteilt wegen $g_i + g_{N-i+1} = i + N - i + 1 = N + 1 =$konstant
- $m(m+1)/2 \leq W_N \leq m(m+1)/2 + mn$
- Es gilt

$$\mathrm{E}(W_N) = \frac{m(N+1)}{2} \quad \text{und} \quad \mathrm{var}(W_N) = \frac{mn(N+1)}{12}.$$

- Zentraler Grenzwertsatz

$$Z = \frac{W_N - m(N+1)/2}{\sqrt{mn(N+1)/12}} \overset{as}{\sim} N(0,1)$$

für $m/n \to \lambda =$ konstant

- **Exakte Verteilung** von W_N unter H_0 anhand Beispiel:

Sei $m = 3$ und $n = 5$. Es gibt insgesamt

$$\binom{m+n}{m} = \binom{N}{m} = \binom{8}{3} = 56$$

verschiedene Vektoren $(v_1, \ldots, v_8)$, die unter H_0 gleich wahrscheinlich sind (1/56)

w	Ränge der X_i	$P(W_N = w)$
21	(6,7,8)	$1/56$
20	(5,7,8)	$1/56$
19	(4,7,8); (5,6,8)	$2/56$
18	(3,7,8); (4,6,8); (5,6,7)	$3/56$
17	(2,7,8); (3,6,8); (4,6,7); (4,5,8)	$4/56$
16	(1,7,8); (2,6,8); (3,5,8); (3,6,7); (4,5,7)	$5/56$
15	(1,6,8); (2,5,8); (2,6,7); (3,5,7); (3,4,8); (4,5,6)	$6/56$
14	(1,6,7); (1,5,8); (2,5,7); (2,4,8); (3,4,7); (3,5,6)	$6/56$

Z.B. ist 19 das $(1 - \alpha)$-Quantil für $\alpha = 4/56 \approx 0.071$.

Beispiel: Körpergrößen von $m = 8$ Knaben und $n = 10$ Mädchen.

	y_1	y_2	y_3	y_4	y_5	x_1	y_6	y_7	y_8	x_2	x_3	y_9	x_4	y_{10}	x_5	x_6	x_7	x_8
$z_{(i)}$	110	113	114	116	116	117	118	119	119	120	122	123	124	125	126	126	128	132
v_i	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	1
g_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

$w_{18} = \sum_{i=1}^{18} iv_i = 106.$

Für Test (A) mit $\alpha = 0.05$ ergibt Tabelle J $w_{\alpha/2} = 53.$

Damit ist $w_{1-\alpha/2} = 2\mathsf{E}(W_N) - w_{\alpha/2} = 152 - 53 = 99 < w_N$.

Wie schon beim KS-Test: Ablehnung von H_0.

Mann-Whitney-U Test

Anstelle des Wilcoxon-Rangsummen Tests.

$$U_N = \sum_{i=1}^{m} \sum_{j=1}^{n} W_{ij}$$

mit

$$W_{ij} = \begin{cases} 1 & \text{für} \quad Y_j < X_i \quad i = 1, \ldots, m \\ 0 & \text{für} \quad Y_j > X_i \quad j = 1, \ldots, n. \end{cases}$$

Es gilt

$$U_N = W_N - \frac{m}{2}(m+1).$$

Beispiel: Körpergrößen

	y_1	y_2	y_3	y_4	y_5	x_1	y_6	y_7	y_8	x_2	x_3	y_9	x_4	y_{10}	x_5	x_6	x_7	x_8
w	8	8	8	8	8	0	7	7	7	0	0	5	0	4	0	0	0	0

w gibt an, wie viele x einem y in der kombinierten, geordneten Stichprobe folgen. Summe der w ist $u_N = 70$ (entspricht $w_N - m(m+1)/2 = 106 - 36 = 70$).

```
> wilcox.test(x, y)

        Wilcoxon rank sum test with continuity correction

 data:  x and y
 W = 70, p-value = 0.00866
 alternative hypothesis: true mu is not equal to 0

 Warning message: Cannot compute exact p-value with ties
```

Van der Waerden X_N Test:

$$X_N = \sum_{i=1}^{N} \Phi^{-1}\left(\frac{i}{N+1}\right) V_i = \sum_{i=1}^{m} \Phi^{-1}\left(\frac{R_i}{N+1}\right)$$

Verteilung von X_N unter H_0: Wegen

$$\Phi^{-1}\left(\frac{i}{N+1}\right) + \Phi^{-1}\left(\frac{N-i+1}{N+1}\right) = g_i + g_{N-i+1} = 0$$

1. $\sum_{i=1}^{N} g_i = \sum_{i=1}^{N} \Phi^{-1}\left(\frac{i}{N+1}\right) = 0$,

2. $\mathsf{E}(X_N) = 0$, $\mathsf{var}(X_N) = \frac{mn}{N(N-1)} \sum_{i=1}^{N} \left(\Phi^{-1}\left(\frac{i}{N+1}\right)\right)^2$,

3. X_N ist symmetrisch um $\mathsf{E}(X_N) = 0$ verteilt.

Verteilung von X_n unter H_0 anhand des Beispiels mit $m = 3$ und $n = 5$:

i	1	2	3	4	5	6	7	8
g_i	-1.2206	-0.7647	-0.4307	-0.1397	0.1397	0.4307	0.7647	1.2206

(r_1, r_2, r_3)	x_8	P	(r_1, r_2, r_3)	x_8	P	(r_1, r_2, r_3)	x_8	P
(6,7,8)	2.416	1/56	(4,6,7)	1.056	1/56	(4,5,6)	0.431	
(5,7,8)	2.125	1/56	(3,5,8)	0.930	1/56	(1,6,8)	0.431	3/56
(4,7,8)	1.846	1/56	(2,6,8)	0.887	1/56	(2,6,7)	0.431	
(5,6,8)	1.791	1/56	(1,7,8)	0.765		(2,4,8)	0.316	1/56
(3,7,8)	1.555	1/56	(3,6,7)	0.765	3/56	(3,4,7)	0.194	1/56
(4,6,8)	1.512	1/56	(4,5,7)	0.765		(1,5,8)	0.140	
(5,6,7)	1.335	1/56	(3,4,8)	0.650	1/56	(2,5,7)	0.140	3/56
(2,7,8)	1.221		(2,5,8)	0.596	1/56	(3,5,6)	0.140	
(3,6,8)	1.221	3/56	(3,5,7)	0.474	1/56	(2,3,8)	0.025	1/56
(4,5,8)	1.221							

Für die restlichen 28 Möglichkeiten gilt: ist $X_N = x$ für (r_1, r_2, r_3), so ist $X_N = -x$ für $(N + 1 - r_1, N + 1 - r_2, N + 1 - r_3)$.

Beispiel: Körpergrößen

Gewichte $g_i = \Phi^{-1}\left(\frac{i}{19}\right)$. Man erhält $x_{18} = 4.9449$. Tabelle K liefert für $\alpha = 0.05$ den kritischen Wert $x_{1-\alpha/2} = 3.616$. Wegen $x_{18} > x_{1-\alpha/2}$ wird H_0 abgelehnt.

```
> x <- c(117,120,122,124,126,126,128,132)          # m=8 Mädchen
> y <- c(110,113,114,116,116,118,119,119,123,125)  # n=10 Knaben
> m <- length(x);  n <- length(y);  N <- m+n
> group <- c(rep("x", m), rep("y", n))
> V <- 1*(group[order(c(x, y))]=="x")  # Indikator(x) in komb-geord-StPr
> g <- qnorm((1:N)/(N+1))  # Gewichte
> X <- sum(g*V);  X       # Van der Waerden Statistik
[1] 4.944933
> var.X <- m*n/(N*(N-1))*sum(g^2)  # Varianz(X)
> var.X
[1] 3.468656
> p.value <- 2*(1 - pnorm(abs(X)/sqrt(var.X)))  # two-sided p.value
[1] 0.007928642
```

Als approximativen p-Wert liefert dies 0.008 und somit die gleiche Aussage.

Weitere Rangtests für Lagealternativen

1. **Fisher-Yates-Terry-Hoeffding:**

$$g_i = \mathsf{E}\left(Z_{(i)}\right)$$

g_i ist der Erwartungswert der i–ten geordneten Statistik $Z_{(i)}$ einer Stichprobe aus einer $N(0,1)$–verteilten Grundgesamtheit.

2. **Moods Median Test:**

$$g_i = \begin{cases} 0 & \text{für} \quad i \leq \frac{N+1}{2} \\ 1 & \text{für} \quad i > \frac{N+1}{2}. \end{cases}$$

Tests bezüglich Variabilitätsalternativen

F_X und G_Y seien stetig. X und θY haben dieselbe Verteilung, d.h.

$$\begin{aligned} F_X(z) &= G_Y\left(\frac{z}{\theta}\right) \\ \mu_X = \mathsf{E}(X) &= \theta \mathsf{E}(Y) = \theta\mu_Y \\ \sigma_X^2 = \mathsf{var}(X) &= \theta^2 \mathsf{var}(Y) = \theta^2\sigma_Y^2 \end{aligned}$$

Variabilitätsalternativen: Lage- und Streuungsunterschiede. Nur für $\mu_X = \mu_Y = 0$ sind Tests auf Variabilität Tests auf Streuung.

- $H_0 : G(z) = F(z)$,
- $H_1 : G(z) = F(\theta z)$, mit $\theta \neq 1$, $\theta > 1$ oder $\theta < 1$.

Parametrischer Test bei Normalverteilung

Test auf Gleichheit 2er Varianzen (F-Test): $X_i \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$, $i = 1, \ldots, m$, und $Y_j \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$, $j = 1, \ldots, n$, mit μ_X, μ_Y unbekannt und X_i, Y_j unabhängig.

H_0	H_1	Entscheidung gegen H_0, falls	kritische Werte
$\sigma_X = \sigma_Y$	$\sigma_X \neq \sigma_Y$ $(\theta \neq 1)$	$T < c_3$ oder $T > c_4$	$c_3 = f_{\alpha/2}$ $c_4 = f_{1-\alpha/2}$
$\sigma_X = \sigma_Y$	$\sigma_X > \sigma_Y$ $(\theta > 1)$	$T > c_1$	$c_1 = f_{1-\alpha}$
$\sigma_X = \sigma_Y$	$\sigma_X < \sigma_Y$ $(\theta < 1)$	$T < c_2$	$c_2 = f_{\alpha}$

Unter H_0 gilt:

$$T = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

F-Test ist sehr empfindlich gegenüber Abweichungen von der Normalverteilung.

```
> x <- c(117,120,122,124,126,126,128,132)          # m=8 Mädchen
> y <- c(110,113,114,116,116,118,119,119,123,125)  # n=10 Knaben
> var.test(x, y, ratio = 1)  # ratio=1 default

        F test to compare two variances

 data:  x and y
 F = 1.0886, num df = 7, denom df = 9, p-value = 0.8841
 alternative hypothesis: true ratio of variances is not equal to 1
 95 percent confidence interval:
 0.2593722 5.2505398
 sample estimates:
 ratio of variances
          1.088597
```

Hierbei überdeckt das angeführte Konfidenzintervall den wahren Varianzquotienten σ_X^2/σ_Y^2.

Siegel-Tukey Test (1960)

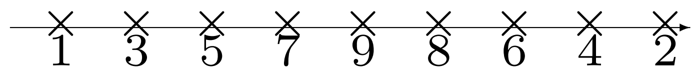
Zusätzliche Annahme gleichen Medians.

Testprobleme:

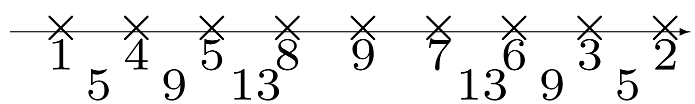
- Test A: $H_0 : G(z) = F(z)$, $H_1 : G(z) = F(\theta z)\,, \theta \neq 1, \theta > 0$,
- Test B: $H_0 : G(z) = F(z)$, $H_1 : G(z) = F(\theta z)\,, \theta > 1$,
- Test C: $H_0 : G(z) = F(z)$, $H_1 : G(z) = F(\theta z)\,, 0 < \theta < 1$.

Test auf Variabilität extrem große und den extrem kleine Meßwerte niedrige Rangzahlen, den mittleren Meßwerten hohe.

Z.B.



Symmetrie in der Summe benachbarter Gewichte für N gerade:



Ist N ungerade, dann wird die 'mittlere' Beobachtung aus der kombinierten, geordneten Stichprobe gestrichen und g_i für $N^* = N - 1$ berechnet.

Siegel-Tukey Teststatistik für gerades N

$$S_N = \sum_{i=1}^{N} g_i V_i \quad \text{mit} \quad g_i = \begin{cases} 2i & \text{für } i \text{ gerade und } 1 \leq i \leq N/2, \\ 2(N-i)+2 & \text{für } i \text{ gerade und } N/2 < i \leq N, \\ 2i-1 & \text{für } i \text{ ungerade und } 1 \leq i \leq N/2, \\ 2(N-i)+1 & \text{für } i \text{ ungerade und } N/2 < i \leq N. \end{cases}$$

Unter $H_0 : F = G$ hat S_N dieselbe Verteilung wie die Wilcoxon-Statistik W_N.

$$\mathsf{E}(S_N) = \frac{m(N+1)}{2} \quad \text{und} \quad \mathsf{var}(S_N) = \frac{mn(N+1)}{12}.$$

H_0 wird abgelehnt, falls:

- Test A: $S_N \geq w_{1-\alpha/2}$ oder $S_N \leq w_{\alpha/2}$,
- Test B: $S_N \leq w_{\alpha}$,
- Test C: $S_N \geq w_{1-\alpha}$.

Beispiel: Körpergrößen von $m = 8$ Mädchen und $n = 10$ Knaben.

	y_1	y_2	y_3	y_4	y_5	x_1	y_6	y_7	y_8	x_2	x_3	y_9	x_4	y_{10}	x_5	x_6	x_7	x_8
$z_{(i)}$	110	113	114	116	116	117	118	119	119	120	122	123	124	125	126	126	128	132
v_i	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	1
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
g_i	1	4	5	8	9	12	13	16	17	18	15	14	11	10	7	6	3	2

$s_{18} = \sum_i g_i v_i = 74$. Für $\alpha = 0.05$ entnehme man $w_{\alpha/2} = 53$ aus der Tabelle J.

Damit ist $w_{1-\alpha/2} = 2\mathsf{E}(W_N) - w_{\alpha/2} = 152 - 53 = 99$.

Wegen $53 < 74 < 99$ kann H_0 nicht verworfen werden.

```
> g <- rep(1, N);  g[N] <- 2
> odd <- 1-(even <- (trunc(1:N/2)==(1:N/2)))
> for (i in 2:(N/2)) g[i] <- g[i-1] + 1*odd[i] + 3*even[i]
> for (i in (N-1):(N/2+1)) g[i] <- g[i+1] + 1*odd[i] + 3*even[i]
> S <- sum(g*V)          # [1] 74
```

```
> E.S <- m*(N+1)/2       # [1] 76
> var.S <- m*n*(N+1)/12  # [1] 126.6667
> p.value <- 2*(1-pnorm(abs(X-E.S)/sqrt(var.X)))  # two-sided
> p.value
[1] 0.858955
```

Beachte!

Anwendung des S_N Test: F und G vom selben Verteilungstyp mit gleichem Median. Allgemeine Variabilitätsalternativen, wie Lage- und Streuungsunterschieder werden nicht erfasst.

Anschauliches Beispiel: Kombinierte, geordnete Stichprobe $xxxxxyyyyy$.

Man erhält dafür $S_N = 1 + 4 + 5 + 8 + 9 = 27$. Selbst für $\alpha = 0.5$ führt ein zweiseitiger S_N–Test nicht zur Ablehnung von $H_0 : F = G$.

Mood Test (1954)

Abweichungen der Ränge der x_i von der mittleren Rangzahl $(N+1)/2$

$$M_N = \sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right)^2 V_i$$

Große Werte von $M_N \quad \Rightarrow \quad x_i$ streuen stärker als y_j

Verteilung von M_N unter H_0:

$$\mathsf{E}(M_N) = \frac{m(N^2-1)}{12} \quad \text{und} \quad \text{var}(M_N) = \frac{mn(N+1)(N^2-4)}{180}$$

Nur für $m = n$ symmetrisch um den Erwartungswert.

Exakte Verteilung von M_N durch Auszählen: sei $m = 3$ und $n = 4$, $\binom{7}{3} = 35$ verschiedene Möglichkeiten der Rangzuordnung. Gewichte $g_i = (i-4)^2$. Somit

(r_1,r_2,r_3)	m_7	P	(r_1,r_2,r_3)	m_7	P	(r_1,r_2,r_3)	m_7	P
(1,2,7)	22	$2/35$	(1,2,4)	13	$4/35$	(2,3,4)	5	$4/35$
(1,6,7)	22		(1,4,6)	13		(2,4,5)	5	
(1,3,7)	19	$2/35$	(2,4,7)	13		(3,4,6)	5	
(1,5,7)	19		(4,6,7)	13		(4,5,6)	5	
(1,4,7)	18	$1/35$	(1,3,5)	11	$2/35$	(3,4,5)	2	$1/35$
(1,2,6)	17	$2/35$	(3,5,7)	11				
(2,6,7)	17		(1,3,4)	10	$4/35$			
(1,2,3)	14	$8/35$	(1,4,5)	10				
(1,2,5)	14		(3,4,7)	10				
(1,3,6)	14		(4,5,7)	10				
(1,5,6)	14		(2,3,6)	9	$2/35$			
(2,3,7)	14		(2,5,6)	9				
(2,5,7)	14		(2,4,6)	8	$1/35$			
(3,6,7)	14		(2,3,5)	6	$2/35$			
(5,6,7)	14		(3,5,6)	6				

Beispiel: Körpergrößen von $m = 8$ Knaben und $n = 10$ Mädchen.

Kombinierte, geordnete Stichprobe: $yyyyxyyyxxyxyxxxx$

Es ist $\frac{N+1}{2} = 9.5$ und damit $m_{18} = (6-9.5)^2 + (10-9.5)^2 + (11-9.5)^2 + (13-9.5)^2 + (15-9.5)^2 + (16-9.5)^2 + (17-9.5)^2 + (18-9.5)^2 = 228$.

Für $\alpha = 0.2$ ist $m_{\alpha/2} = 146$ und $m_{1-\alpha/2} = 284$, d.h. H_0 wird auf diesem Testniveau nicht abgelehnt.

```
> mood.test(x, y)

        Mood two-sample test of scale

 data:  x and y
 Z = 0.2341, p-value = 0.815
 alternative hypothesis: two.sided

> E.M <- m * (N^2-1)/12             # [1] 215.3333
> var.M <- m*n * (N+1)*(N^2-4)/180  # [1] 2702.222
```

```
> mood.test(x, y)$statistic*sqrt(var.M) + E.M + 1/2
    Z
 228.0
```

Die von R ausgegebene Statistik Z entspricht also der standardisieren Form von M_N (mit zusätzlicher Stetigkeitskorrektur $1/2$).

Weitere Rangtests für Variabilitätsalternativen:

Ansary-Bradley Test (1960) Lineare Rangstatistik mit Gewichten

$$g_i = \left(\frac{N+1}{2} - \left| i - \frac{N+1}{2} \right| \right) .$$

Sind die Abweichungen $|i - (N+1)/2|$ groß, so wird dadurch A_N klein. Dies ist ein Hinweis für stärker streuende x_i-Werte.

```
> ansari.test(x,y)

        Ansari-Bradley test

data:  x and y
AB = 39, p-value = 0.8574
alternative hypothesis: true ratio of scales is not equal to 1
```

Klotz (1962): verwendet das Quadrat von g_i in der X_N-Statistik als Gewichte.

Capon (1961): Erwartungswert des Quadrats von $Z_{(i)}$ (vgl. mit Fisher-Yates-Terry-Hoeffding Test für Lagealternativen).

Konkret:

$$K_N = \sum_{i=1}^{N} \left[\Phi^{-1}\left(\frac{i}{N+1}\right)\right]^2 V_i\,, \qquad C_N = \sum_{i=1}^{N} E\left(Z_{(i)}^2\right) V_i\,.$$

Der K_N-Test und der C_N-Test sind asymptotisch äquivalent.

Fligner-Killeen-Test (1976)

$$F_N = \sum_{i=1}^{N} \Phi^{-1}\left(\frac{1}{2} + \frac{i}{2(N+1)}\right) V_i\,.$$

```
> z <- c(x, y)                          # kombinierte Stichprobe
> sex <- c(rep("F", m), rep("M", n))  # "F"=Female, "M"=Male
> fligner.test(z ~ as.factor(sex))

        Fligner-Killeen test for homogeneity of variances

 data:  z by as.factor(sex)
 Fligner-Killeen:med chi-squared = 0.0081, df = 1, p-value = 0.9284
```