Design of Experiments

- 1. Analysis of Variance
- 2. More about Single Factor Experiments
- 3. Randomized Blocks, Latin Squares
- 4. Factorial Designs
- 5. 2^k Factorial Designs
- 6. Blocking and Confounding

Montgomery, D.C. (1997): Design and Analysis of Experiments (4th ed.), Wiley.

1. Single Factor – Analysis of Variance

Example: Investigate **tensile strength** y of new synthetic fiber.

Known: y depends on the weight percent of cotton
(which should range within 10% – 40%).

Decision:
(a) test specimens at 5 levels of cotton weight: 15%, 20%, 25%, 30%, 35%.
(b) test 5 specimens at each level of cotton content.

Single Factor Experiment with $a = 5$ **levels** and $n = 5$ **Replicates**.

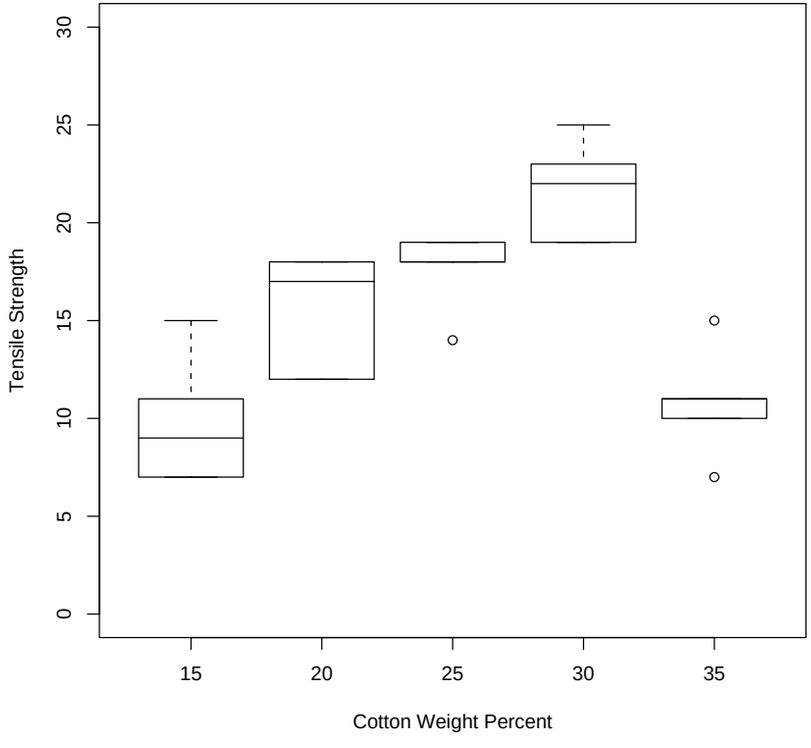
$\Longrightarrow$ 25 **runs**.

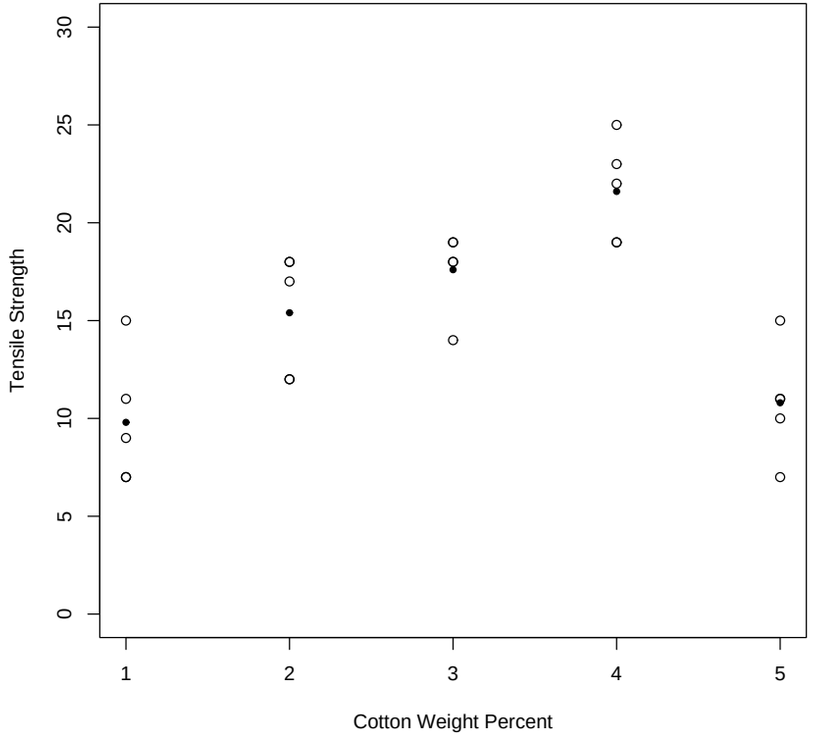
Runs should be in **Random Order** (prohibit warm up effects of machine ...)

Cotton	Observation						
Weight %	1	2	3	4	5	Total	Average
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						376	15.04

```
> y <- c( 7,  7, ... , 15, 11);  w <- gl(5, 5, labels=c(15, 20, 25, 30, 35))
> tapply(y, w, sum)  # total
 15  20  25  30  35
 49  77  88 108  54
> tapply(y, w, mean) # average
  15   20   25   30   35
 9.8 15.4 17.6 21.6 10.8
> mean(tapply(y, w, mean))  # mean average
 [1] 15.04
```

```
> boxplot(y~w);  plot(as.numeric(w), y); points(tapply(y, w, mean), pch=20)
```





We wish to test for differences between the mean strengths at all $a = 5$ levels of cotton weight percent $\Rightarrow$ **Analysis of Variance**.

Analysis of Variance (ANOVA)

Use the **Linear Regression Model**

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

for treatment $i = 1, \ldots, a$, and replication $j = 1, \ldots, n$.

Observation y_{ij} (ith treatment, jth replication)
Parameter μ is common to all treatments (**Overall Mean**)
Parameter τ_i is unique to the ith treatment (i**th Treatment Effect**)
Random variable ϵ_{ij} is the **Random Error** component.

Further assumption: $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

Our interest is in the treatment effects.

Treatment Effects τ_i:

Fixed: the a treatments are chosen by the experimenter.
(tests and conclusions will only apply to the factor levels considered)
Fixed Effects Model

Random: the a treatments are a random sample from a population of treatments.
(we are able to extend conclusions to all treatments in the population)
Random Effects Model / Components of Variance Model

Fixed Effects Model

Treatment effects τ_i are usually defined as the deviations from the overall mean

$$\mu := \frac{1}{a}\sum_{i=1}^{a} \mu_i = \frac{1}{a}\sum_{i=1}^{a} (\mu + \tau_i) = \mu + \frac{1}{a}\sum_{i=1}^{a} \tau_i \,,$$

Thus, we have a restriction on these effects, namely

$$\sum_{i=1}^{a} \tau_i = 0 \,.$$

Here, $\mu_i = \mathsf{E}(y_{ij})$ is the mean of all observations y_{ij} in the ith treatment (row).

ANOVA Decomposition

We are interested in testing the equality of the a treatment means

$$H_0\text{: } \mu_1 = \mu_2 = \cdots = \mu_a \quad \Longleftrightarrow \quad H_0\text{: } \tau_1 = \tau_2 = \cdots = \tau_a$$

which is equivalent to testing the equality of all treatment effects.

The Sum of Squares decomposition in Regression is valid

$$SST = SSR + SSE$$

where SSR, the **S**um of **S**quares due to the **R**egression model, is only related to the treatment effects τ_i. Hence, we have

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu}_i)^2$$

$\hat{\mu}$ estimates the overall mean μ, where we assume that all the y_{ij} are from the same population. Thus, this estimate is given as

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} =: \overline{y}_{..}$$

where $N = an$ is the total number of observations.

$\hat{\mu}_i$ estimates the mean of the y_{ij} coming only from the ith row (treatment). This gives the estimate

$$\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij} =: \overline{y}_{i.}$$

Together this gives

$$\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{..})^2 = n \sum_{i=1}^{a} (\overline{y}_{i.} - \overline{y}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{i.})^2$$

Therefore, the total variability in the data can be partitioned into a sum of squares of the differences **between** the treatment averages and the grand average, plus a sum of squares of the differences of observations **within** treatments from the treatment average.

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatments	SSR	$a-1$	MSR	MSR/MSE
Error (within Treatments)	SSE	$N-a$	MSE	
Total	SST	$N-1$		

Tensile Strength Data: Test

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ against H_1: some means are different

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_{4,20}$	p-value
Cotton Weight Percent	475.76	4	118.94	14.76	< 0.001
Error (within Treatments)	161.20	20	8.06		
Total	639.96	24			

Thus, we reject H_0 and conclude that the treatment means differ!

```
> summary(aov(y~w))
            Df Sum Sq Mean Sq F value    Pr(>F)
w            4 475.76  118.94  14.757 9.128e-06 ***
Residuals   20 161.20    8.06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Estimation of the Model Parameters

Remember the model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

with overall mean μ, treatment means $\mu_i = \mu + \tau_i$, and treatment effects τ_i. Their estimates are

$$\hat{\mu} = \overline{y}_{..} \qquad \hat{\mu}_i = \overline{y}_{i.} \qquad \Longrightarrow \qquad \hat{\tau}_i = \overline{y}_{i.} - \overline{y}_{..}$$

Because of $y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma^2)$

$$\overline{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij} \sim N\left(\mu_i, \frac{1}{n}\sigma^2\right)$$

Moreover, MSE estimates σ^2 and the $(1-\alpha)$ confidence interval for the ith treatment mean μ_i is

$$\left[\overline{y}_{i\cdot} \pm t_{1-\alpha/2,N-a}\sqrt{MSE/n}\right]$$

```
> W <- C(w, treatment); coefficients(aov(y~W)) # default contrast for w
 (Intercept)         W20         W25         W30         W35
         9.8         5.6         7.8        11.8         1.0
> W <- C(w, sum); coefficients(aov(y~W))
 (Intercept)          W1          W2          W3          W4
       15.04       -5.24        0.36        2.56        6.56
> options(contrasts=c("contr.sum", "contr.poly")) # for all factors
```

Bartlett's Test for Equality of Variances: H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$

K^2 is based on the (pooled) sample variances and approximately χ^2_{a-1}.

```
> bartlett.test(y~W)

        Bartlett test for homogeneity of variances

data:  y by W
Bartlett's K-squared = 0.9331, df = 4, p-value = 0.9198
```

$\Longrightarrow$ Conclude that all 5 variances are the same!

This test is very sensitive to the normality assumption!

Variance-Stabilizing Transformation:

Let $\mathrm{E}(y) = \mu$ be the mean of y and suppose that the standard deviation is proportional to a power of the mean

$$\sigma_y \propto \mu^{\alpha}$$

Task: find a transformation of y that yields a constant variance. Suppose this is

$$y^* = y^{\lambda}$$

where $\lambda = 0$ implies the log transformation. Then

$$\sigma_{y^*} \propto \mu^{\lambda-(1-\alpha)}$$

Setting $\lambda = 1 - \alpha$, then the variance of the transformed data is constant.

Relationship b/w σ_y and μ	α	$\lambda = 1 - \alpha$	Transformation
$\sigma_y \propto$ const	0	1	no transformation
$\sigma_y \propto \mu^{1/2}$	$1/2$	$1/2$	Square Root
$\sigma_y \propto \mu$	1	0	Log
$\sigma_y \propto \mu^{3/2}$	$3/2$	$-1/2$	Reciprocal Square Root
$\sigma_y \propto \mu^2$	2	-1	Reciprocal

Selection of the Power: If $\sigma_{y_i} \propto \mu_i^\alpha = \theta \mu_i^\alpha$ then

$$\log \sigma_{y_i} = \log \theta + \alpha \log \mu_i$$

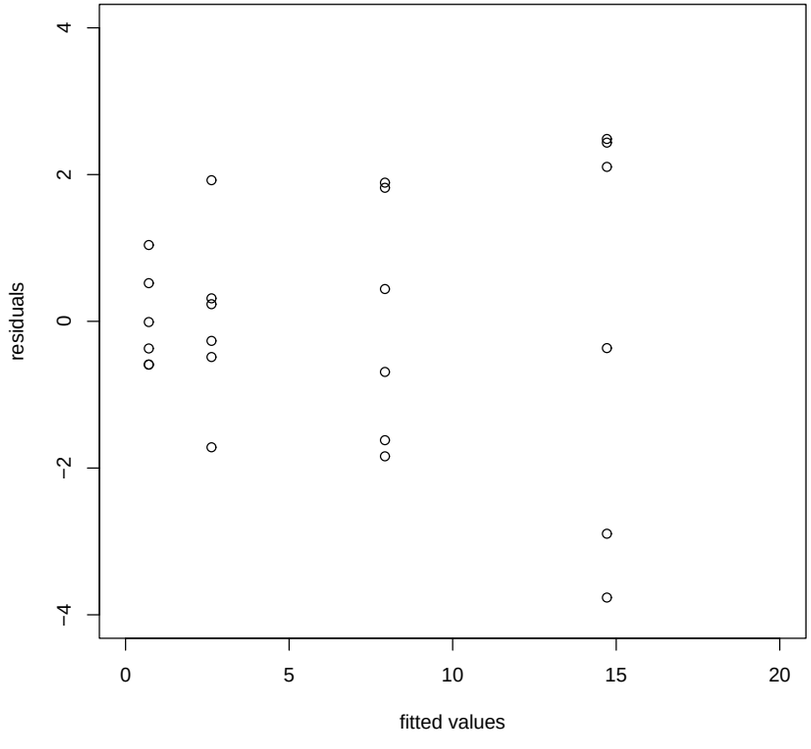
A plot of $\log \sigma_{y_i}$ versus $\log \mu_i$ is a straight line with slope α. Substitute σ_{y_i} and μ_i by their estimates S_i and $\overline{y}_{i\cdot}$ and guess the value of α from the plot.

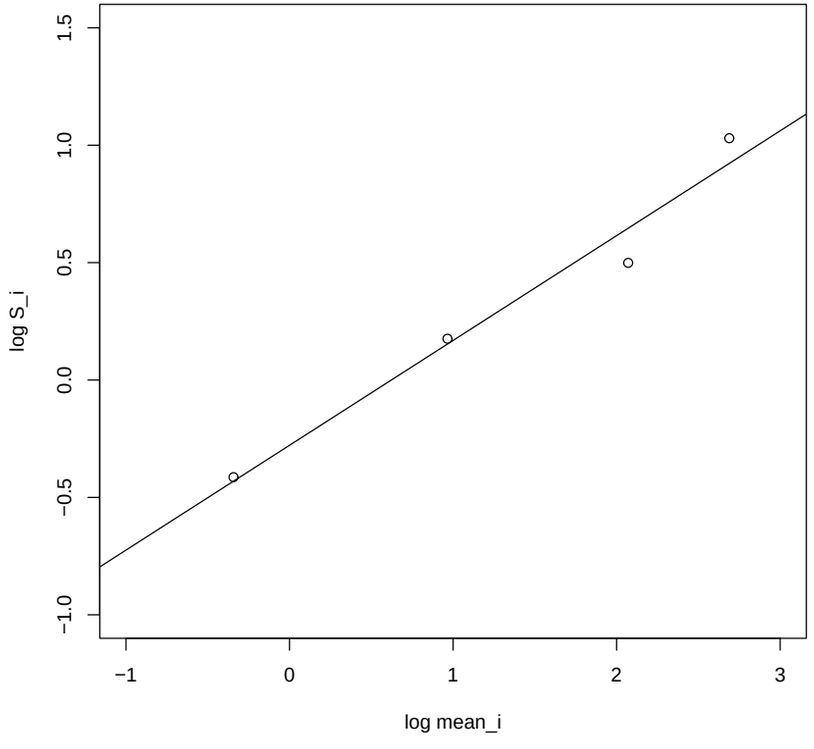
Example: 4 different estimation methods of the peak discharge applied to the same watershed.

Method	discharge (cubic feet / second)						$\overline{y}_{i\cdot}$	S_i
1	0.34	0.12	1.23	0.70	1.75	0.12	0.71	0.66
2	0.91	2.94	2.14	2.36	2.86	4.55	2.63	1.09
3	6.31	8.37	9.75	6.09	9.82	7.24	7.93	1.66
4	17.15	11.82	10.95	17.20	14.35	16.82	14.72	2.77

```
> y <- c(0.34, 0.12, ..., 16.82);  m <- gl(4, 6, labels=c(1, 2, 3, 4))
> tapply(y, m, mean);  tapply(y, m, sd)
       1        2        3        4
 0.710000 2.626667 7.930000 14.715000
       1        2        3        4
 0.661090 1.192202 1.647070  2.800891
> summary(aov(y~m))
             Df Sum Sq Mean Sq F value    Pr(>F)
 m            3 708.35  236.12  76.067 4.111e-11 ***
 Residuals   20  62.08    3.10
```

```
> r <- residuals(aov(y~m));  f <- fitted(aov(y~m));  plot(f, r)
> ls <- log(tapply(y, m, sd));  lm <- log(tapply(y, m, mean))
> plot(lm, ls);  abline(lm(ls~lm)) # gives slope = 0.45
```





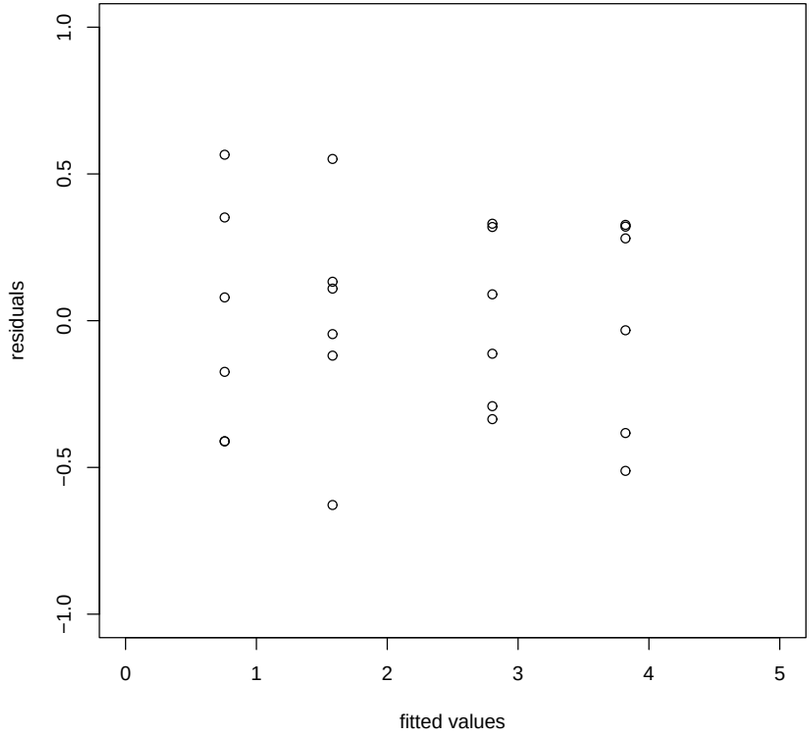
```
> bartlett.test(y~m)

        Bartlett test for homogeneity of variances

 data:  y by m
 Bartlett's K-squared = 8.9958, df = 3, p-value = 0.02935
```

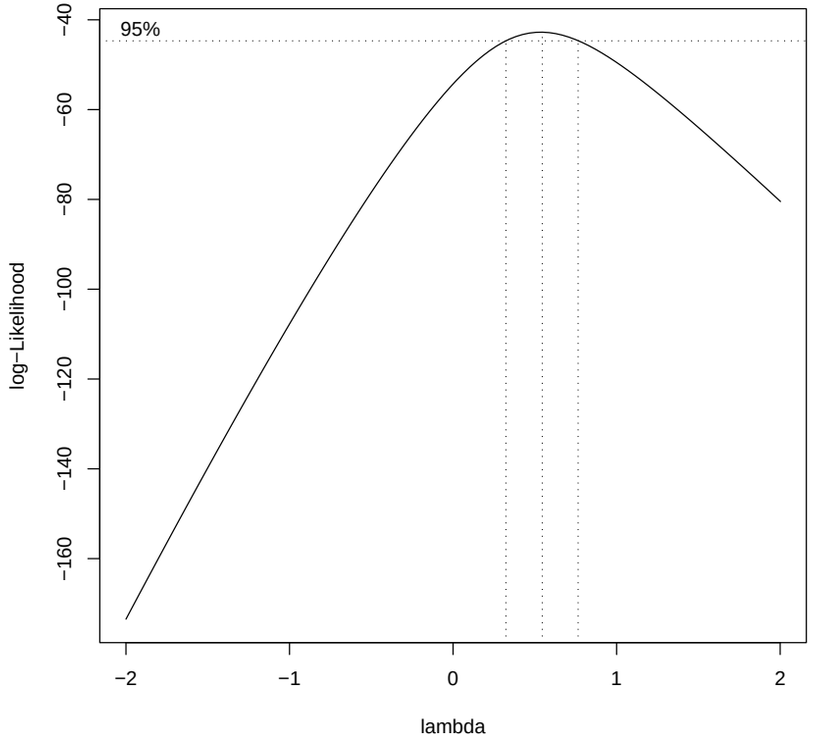
The Bartlett Test rejects Equality of Variances. Thus we analyze $y^* = \sqrt{y}$.

```
> ry <- sqrt(y);  tapply(ry, m, sd)
         1         2         3         4
 0.4044534 0.3857295 0.2929908 0.3734610
> summary(aov(ry~m))
             Df Sum Sq Mean Sq F value    Pr(>F)
 m            3 32.684  10.895  81.049 2.296e-11 ***
 Residuals   20  2.688   0.134
```

To account for the use of the data to estimate α we reduce the error degrees of freedom by one. This gives $F = 76.99$ again with p-value < 0.001.

```
> r <- residuals(aov(ry~m));  f <- fitted(aov(ry~m));  plot(f, r)
> library(mass);  boxcox(y~m)
```





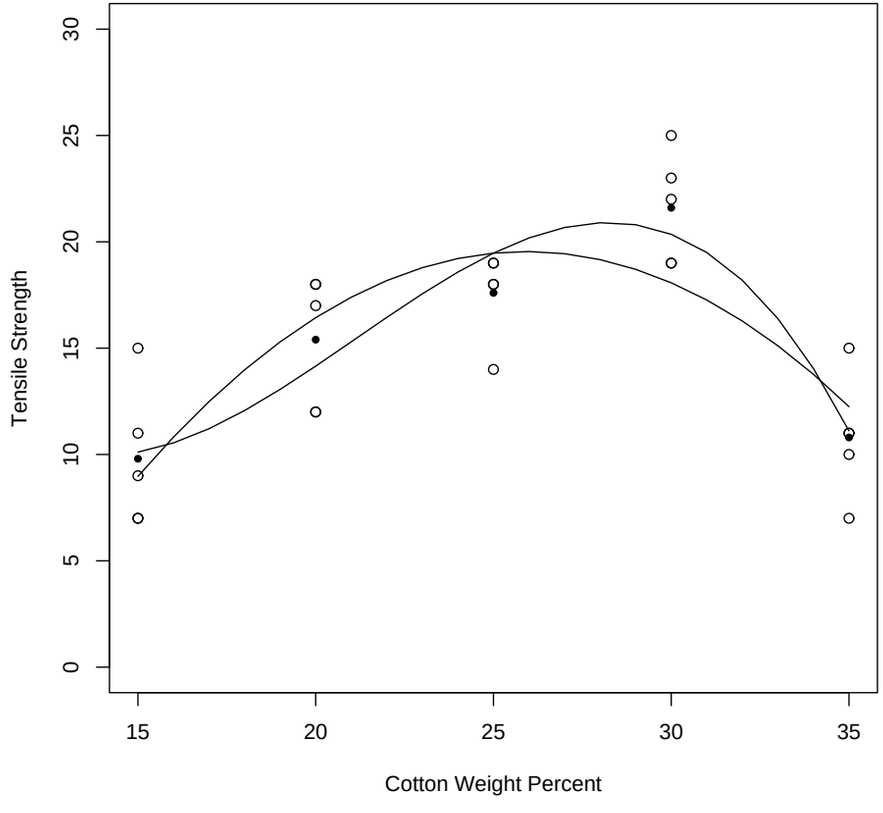
Practical Interpretation of Results:

So far we assumed that the factor (treatment) involved in the experiment is either **quantitative** or **qualitative**. With a quantitative factor we are usually interested in the entire range of values (regression analysis).

Example: For the Tensile Strength response y we either assume a quadratic or cubic model in Cotton Weight Percent x. Previous analysis showed that the maximal strength is produced for $x \approx 30\%$ (**process optimization**).

```
> x <- as.numeric(levels(w)[w])
> m2 <- lm(y ~ x + I(x^2));  m2
 Coefficients:
 (Intercept)            x       I(x^2)
   -39.98857      4.59257     -0.08857
> m3 <- lm(y ~ x + I(x^2) + I(x^3));  m3
 Coefficients:
 (Intercept)            x       I(x^2)       I(x^3)
    62.6114      -9.0114       0.4814      -0.0076
```

```
> p2 <- predict(m2, data.frame(x=seq(15,35)))
> p3 <- predict(m3, data.frame(x=seq(15,35)))
> plot(x, y);  points(seq(15,35,5), tapply(y, w, mean), pch=20)
> lines(15:35, p2);  lines(15:35, p3)
```



Random Effects Model

We are interested in a factor that has a large number of possible levels. If the experimenter randomly selects a of these levels from the population of factor levels, then we say that the factor is **random**.

Example: A textile company weaves fabric on a large number of looms. The looms should be homogeneous so that the fabric is of uniform strength. They select 4 looms at random and make 4 strength determinations.

	observations				
Loom	1	2	3	4	$y_{i\cdot}$
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

Again the model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

but both, τ_i and ϵ_{ij} are random variables here. If they are independent and $\text{Var}(\tau_i) = \sigma_\tau^2$ and $\text{Var}(\epsilon_{ij}) = \sigma^2$, then the variance of any observation is

$$\text{Var}(y_{ij}) = \sigma_\tau^2 + \sigma^2 \,.$$

σ_τ^2 and σ^2 are called **variance components**. To test hypotheses we also need

$$\tau_i \overset{iid}{\sim} N(0, \sigma_\tau^2) \qquad \text{and} \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \,.$$

Hypotheses on individual treatment effects are meaningless. Instead we test

$$H_0\text{: } \sigma_\tau^2 = 0 \text{ versus } H_1\text{: } \sigma_\tau^2 > 0.$$

$\sigma_\tau^2 = 0$: all treatments are identical; $\sigma_\tau^2 > 0$: variability exists between treatments.

The ANOVA decomposition $SST = SSR + SSE$ is still valid. Thus, under the null hypothesis where $\sigma_\tau^2 = 0$ and hence $\tau_1 = \tau_2 = \cdots = \tau_a = 0$, the ratio

$$F = \frac{SSR/(a-1)}{SSE/(N-a)} = \frac{MSR}{MSE}$$

is distributed as F with $a-1$ and $N-a$ degrees of freedom.

Further calculus results in

$$\mathsf{E}(MSR) = \sigma^2 + n\sigma_\tau^2 \quad \text{and} \quad \mathsf{E}(MSE) = \sigma^2\,.$$

Thus under H_0 both are unbiased estimators of σ^2. But under H_1 the expected numerator is larger than the expected denominator. Thus we reject H_0 for values of F which are too large (if $F > F_{1-\alpha;a-1,N-a}$).

How to find estimators of the variance components?

AoV Method: Equating observed and expected mean squares gives

$$MSR = \hat{\sigma}^2 + n\hat{\sigma}_\tau^2 \quad \text{and} \quad MSE = \hat{\sigma}^2$$

$$\hat{\sigma}^2 = MSE \quad \text{and} \quad \hat{\sigma}_\tau^2 = \frac{1}{n}(MSR - MSE)\,.$$

Notice that $\hat{\sigma}_\tau^2$ might be negative!!

Example: Are the looms homogeneous?

```
> y <- c(98, 97, ..., 98);  l <- gl(4, 4, labels=c(1, 2, 3, 4))
> tapply(y, l, sd)  # loom-specific standard deviations
       1         2         3         4
1.2909944 1.2909944 0.9574271 1.8257419
> summary(aov(y~l))
             Df Sum Sq Mean Sq F value    Pr(>F)
 l            3 89.188  29.729  15.681 0.0001878 ***
 Residuals   12 22.750   1.896
```

Hence, we reject H_0 and conclude that there is variability between the looms.

We also get the estimate $\hat{\sigma}^2 = MSE = 1.90$ and $\hat{\sigma}^2_\tau = (MSR - MSE)/4 = 6.96$.

The variance of any observation on strength is estimated by $\hat{\sigma}^2 + \hat{\sigma}^2_\tau = 8.86$. Most of this variability is attributable to differences **between** looms.

The process engineer must now try to reduce the differences in loom performance (possibly caused by faulty set-up, poor maintenance, ...).

If these sources of between-loom variability could be identified and eliminated, then the variance of the process output (strength of fabric) could be reduced, perhaps as low as $\hat{\sigma}^2 = 1.90$. This would greatly increase the quality of the fiber product.

More About Single-Factor Experiments

Fitting Response Curves:

Polynomial regression model for the tensile Strength experiment:

```
> m4 <- lm(y ~ x + I(x^2) + I(x^3) + I(x^4))
> anova(m4)
 Analysis of Variance Table

 Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
 x          1  33.62   33.62  4.1712   0.05452 .
 I(x^2)     1 343.21  343.21 42.5824 2.326e-06 ***
 I(x^3)     1  64.98   64.98  8.0620   0.01013 *
 I(x^4)     1  33.95   33.95  4.2116   0.05347 .
 Residuals 20 161.20    8.06
```

ANOVA and equivalent Linear Regression Model:

Suppose we have a single-factor ANOVA model with $a = 3$ treatments, so

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

The equivalent LME is

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_{ij}$$

with the indicators (reference category is treatment 3)

$$x_{1j} = \begin{cases} 1 & \text{if } y_{ij} \in \text{treatment 1} \\ 0 & \text{otherwise} \end{cases} \qquad x_{2j} = \begin{cases} 1 & \text{if } y_{ij} \in \text{treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

How do the parameters $(\beta_0, \beta_1, \beta_2)$ compare to $(\mu, \tau_1, \tau_2, \tau_3)$ where $\sum_{i=1}^{a} \tau_i = 0$?

Treatment		ANOVA	LRM
1	$\mu_1 =$	$\mu + \tau_1$	$\beta_0 + \beta_1$
2	$\mu_2 =$	$\mu + \tau_2$	$\beta_0 + \beta_2$
3	$\mu_3 =$	$\mu - \tau_1 - \tau_2$	β_0

Thus $\beta_0 = \mu_3$, $\beta_1 = \mu_1 - \mu_3$, $\beta_2 = \mu_2 - \mu_3$.

Now let us test H_0: $\tau_1 = \tau_2 = \tau_3 = 0$, or equivalently H_0: $\mu_1 = \mu_2 = \mu_3$.

If H_0 is true, then the respective LRM parameters has $\beta_0 = \mu$, $\beta_1 = 0$, $\beta_2 = 0$.

In general, if there are a treatments, the LRM will have $a - 1$ variables

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_{a-1} x_{a-1,j} + \epsilon_{ij}$$

with the indicators (reference category is treatment a)

$$x_{ij} = \begin{cases} 1 & \text{if } y_{ij} \in \text{treatment i} \\ 0 & \text{otherwise} \end{cases}$$

Kruskal-Wallis rank sum test:

If the normality assumption is unjustified, a nonparametric alternative to the ANOVA F test should be used to check on differences in a treatment means μ_i.

The Kruskal-Wallis test tests H_0: $\mu_1 = \cdots = \mu_a$.

For the tensile data we get

```
> kruskal.test(y~w)

        Kruskal-Wallis rank sum test

data:  y by w
Kruskal-Wallis chi-squared = 19.0637, df = 4, p-value = 0.0007636
```

We again reject the null hypothesis and conclude that the treatments differ.

This is the same conclusion as from the usual ANOVA F test.

Repeated Measures:

Experimental units are often people. Because of differences in their experience, the responses of different people to the same treatment may be different. Unless it is controlled, this variability becomes part of the experimental error.

To control it, we use a design in which each of the a treatments is used on each person (or subject). Such a design is called **repeated measures design**.

An experiment involves a treatments and every treatment is used exactly once on each of n subjects. Let y_{ij} be the response of subject j to treatment i.

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \,,$$

where τ_i is the effect of the ith treatment, and β_j is the parameter associated with the jth subject. We assume that treatments are fixed (so $\sum_i \tau_i = 0$) but the subjects employed are a random sample from a large population. Thus we assume $\mathsf{E}(\beta_j) = 0$ and $\mathsf{Var}(\beta_j) = \sigma^2_\beta$.

	Subject				
Treatment	1	2	$\cdots$	n	Totals
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	$\cdots$	$y_{\cdot n}$	$y_{\cdot\cdot}$

Consider ANOVA partition:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{\cdot\cdot})^2 = a\sum_{j=1}^{n}(\overline{y}_{\cdot j}-\overline{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{\cdot j})^2$$

Total Sum of Squares is separated into a sum of squares from variation **between subjects** and a sum of squares from variation **within subjects**.

We write

$$SS_{total} = SS_{between} + SS_{within}$$

with degrees of freedom

$$an - 1 = (n - 1) + n(a - 1)\,.$$

Differences within subjects depend on both, differences in treatment effects and uncontrolled variability (noise or error). Thus, we further decompose SS_{within} as

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{.j})^2 = n\sum_{i=1}^{a}(\overline{y}_{i.} - \overline{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2$$

First term measures the contribution of the difference between treatment means to SS_{within}, the second term is the residual variation due to error.

Thus

$$SS_{within} = SS_{treatments} + SS_E$$

with degrees of freedom

$$n(a-1) = (a-1) + (a-1)(n-1)\,.$$

To test the hypotheses of no treatment effect, that is

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \text{at least one } \tau_1 \neq 0$$

use the ratio

$$F = \frac{SS_{Treatments}/(a-1)}{SS_E/(a-1)(n-1)} = \frac{MS_{Treatments}}{MS_E}$$

Analysis of Covariance:

Consider a study performed to determine if there is a difference in the breaking strength (y, response) of a monofilament fiber produced by three different machines (discrete factor). This possibly also depends on the diameter (thickness) of the sample (x, co-variable). A thicker fiber will generally be stronger than a thinner one.

Machine 1		Machine 2		Machine 3	
y	x	y	x	y	x
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15
207	126	216	130	180	106

Procedure: We have a single factor experiment with one **covariate**. An appropriate statistical model is

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \overline{x}..) + \epsilon_{ij} \, ,$$

y_{ij} is the jth observation taken under the ith treatment (machine),

x_{ij} is the measurement on the covariate corresponding to y_{ij},

$\overline{x}..$ is its mean,

μ is the overall mean parameter,

τ_i is the fixed effect of the ith treatment ($\sum_i \tau_i = 0$),

β describes the linear dependency of y_{ij} on x_{ij}.

Notice: the covariate is centered and expressed as $(x_{ij} - \overline{x}..)$ instead of x_{ij} so that the parameter μ is preserved as the overall mean.

```
> y <- c(36, 41, ..., 32);  x <- c(20, 25, ..., 15);  machine <-  gl(3, 5)

> mean(y)
[1] 40.2
> options(contrasts=c("contr.treatment", "contr.poly"))  # default
> lm(y ~ machine + x))
 Coefficients:
  (Intercept)     machine1     machine2            x
       17.360        1.037       -1.584        0.954
> lm(y ~ machine + I(x-mean(x)))
 Coefficients:
  (Intercept)     machine1     machine2   I(x-mean(x))
       40.382        1.037       -1.584        0.954
> options(contrasts=c("contr.sum", "contr.poly"))
> lm(y ~ machine + I(x-mean(x)))
 Coefficients:
  (Intercept)     machine1     machine2   I(x-mean(x))
       40.200        0.182        1.219        0.954
```

To test on the machine effect, machine has to enter the model last.

```
> anova(lm(y ~ I(x-mean(x)) + machine))
 Analysis of Variance Table

 Response: y
                Df  Sum Sq Mean Sq  F value   Pr(>F)
 I(x - mean(x))  1 305.130 305.130 119.9330 2.96e-07 ***
 machine         2  13.284   6.642   2.6106   0.1181
 Residuals      11  27.986   2.544
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Thus, we cannot reject the *no machine effect* hypotheses!

How to test if there is a *diameter effect*?

```
> summary(lm(y ~ I(x-mean(x)) + machine))

 Coefficients:
                Estimate Std. Error t value Pr(>|t|)
 (Intercept)     40.2000     0.4118  97.611  < 2e-16 ***
 I(x - mean(x))   0.9540     0.1140   8.365 4.26e-06 ***
 machine1         0.1824     0.5950   0.307    0.765
 machine2         1.2192     0.6201   1.966    0.075 .
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Residual standard error: 1.595 on 11 degrees of freedom
 Multiple R-Squared: 0.9192,     Adjusted R-squared: 0.8972
 F-statistic: 41.72 on 3 and 11 DF,  p-value: 2.665e-06
```

We reject H_0:$\beta = 0$. There is a linear relationship between breaking strength and diameter. Thus, the adjustment provided by the ANCOVA was necessary.

Ignoring a covariate will sometimes cause an incorrect analysis!

```
> anova(lm(y ~ machine))  # ignoring diameter
 Analysis of Variance Table

 Response: y
           Df  Sum Sq Mean Sq F value  Pr(>F)
 machine    2 140.400  70.200  4.0893 0.04423 *
 Residuals 12 206.000  17.167
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

This would gives evidence that there is an significant machine effect.

With $\hat{\beta} = 0.954$ we can compute **adjusted treatment means** as

$$(\hat{\mu} + \hat{\tau}_i) = \overline{y}_{i\cdot} - \hat{\beta}(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})\,, \qquad i = 1, \ldots, a\,.$$

These are much closer together ($\Rightarrow$ ANCOVA was necessary!)

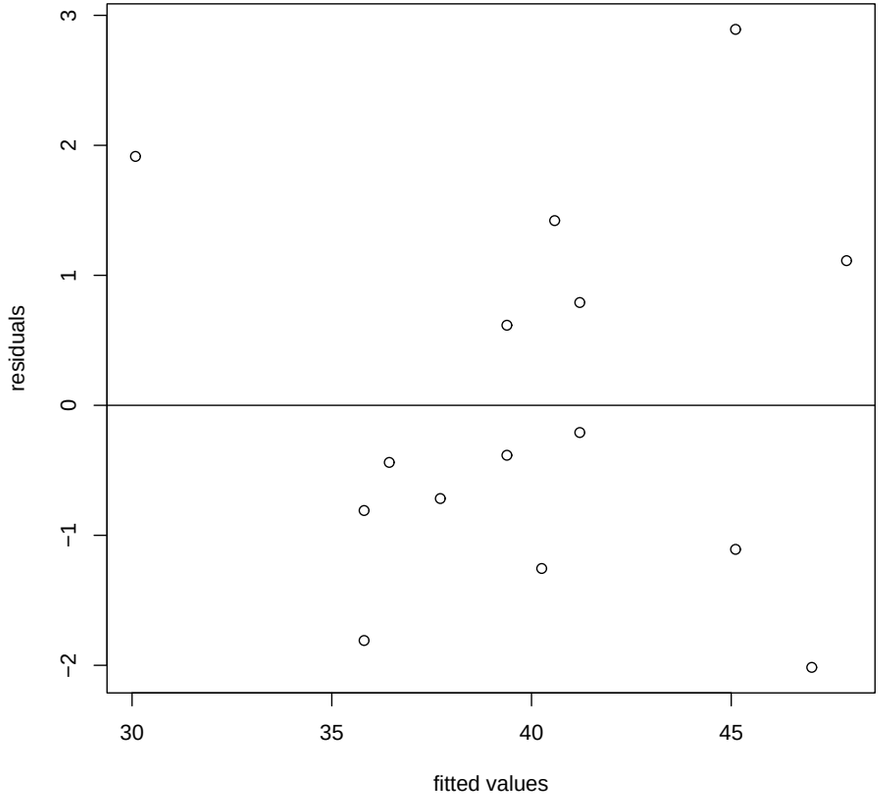
$$\begin{aligned} adjusted(\overline{y}_{1\cdot}) &= 41.40 - 0.954(25.2 - 24.13) = 40.38 \\ adjusted(\overline{y}_{2\cdot}) &= 43.20 - 0.954(26.0 - 24.13) = 41.42 \\ adjusted(\overline{y}_{3\cdot}) &= 36.00 - 0.954(21.2 - 24.13) = 38.80 \end{aligned}$$

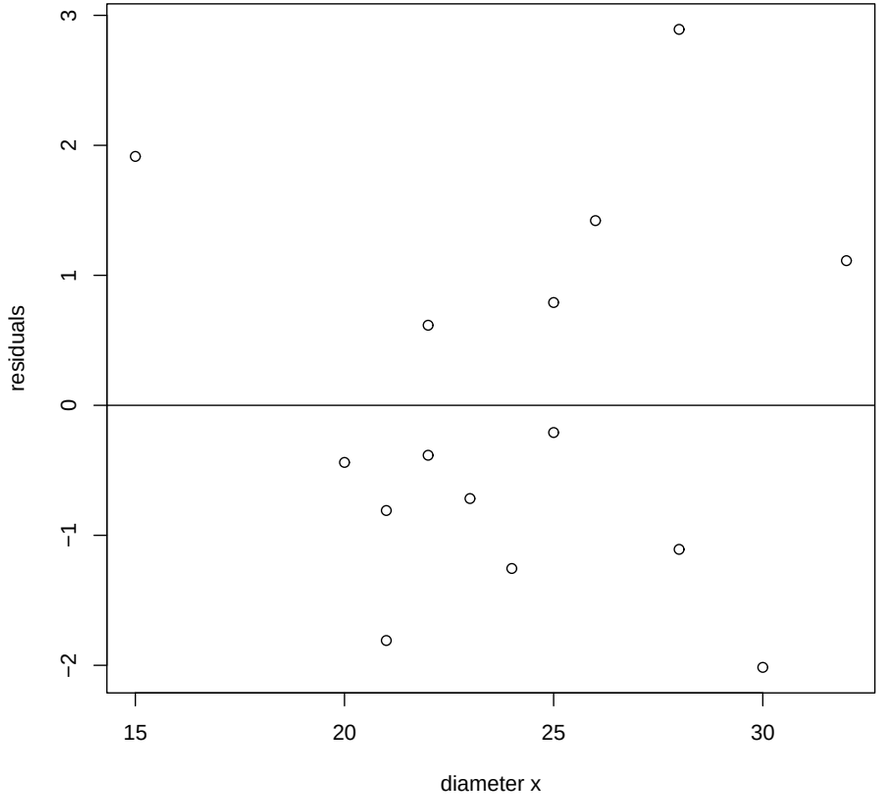
Checking the model is based on **residuals** $e_{ij} = y_{ij} - \hat{y}_{ij}$, with **fitted values**

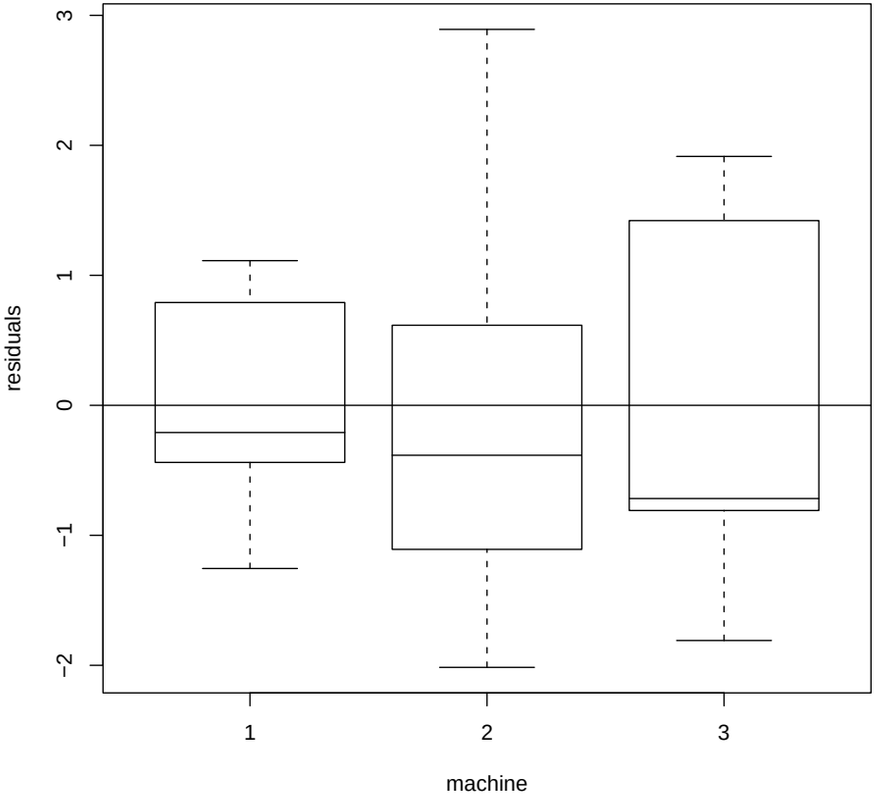
$$\begin{aligned}
\hat{y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}(x_{ij} - \overline{x}_{..}) \\
&= \overline{y}_{..} + \left[\overline{y}_{i.} - \overline{y}_{..} - \hat{\beta}(x_{i.} - \overline{x}_{..})\right] + \hat{\beta}(x_{ij} - \overline{x}_{..}) \\
&= \overline{y}_{i.} + \hat{\beta}(x_{ij} - \overline{x}_{i.})
\end{aligned}$$

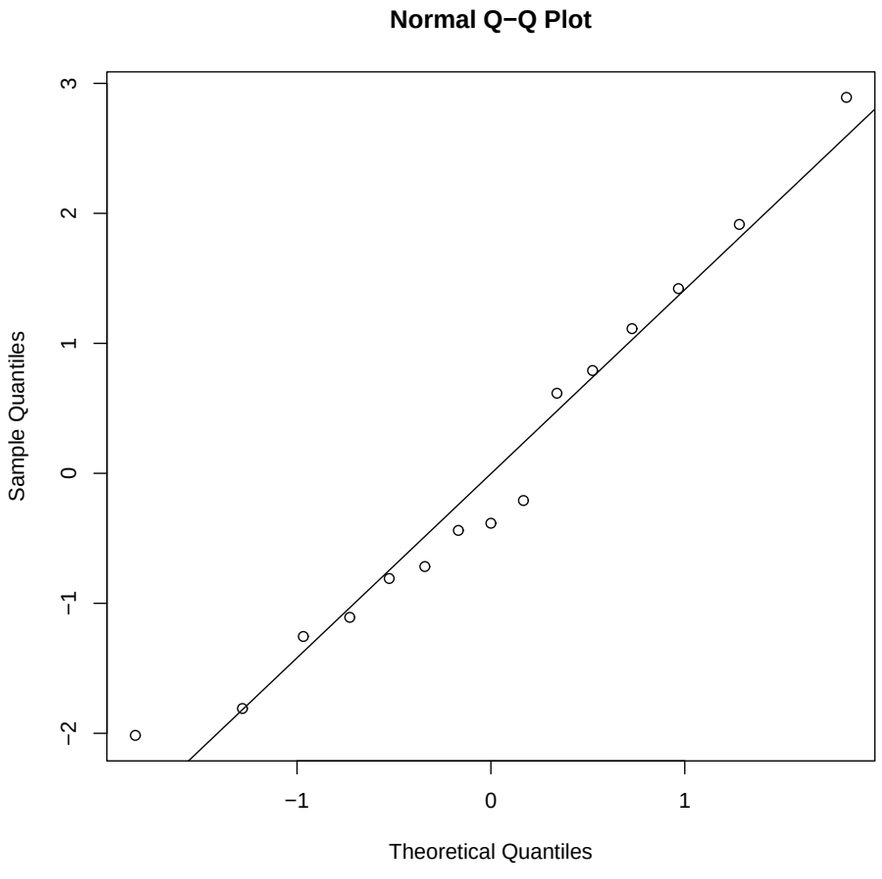
We plot the residuals versus the fitted values, versus the covariate, and versus the machines. Produce also a normal probability plot of the residuals.

```
> e <- my.mod$residuals
> f <- my.mod$fitted
> plot(f, e);  abline(h=0)        # plot residuals vs fitted
> plot(x, e);  abline(h=0)        # plot residuals vs x
> plot(machine, e);  abline(h=0) # plot residuals vs machine
> qqnorm(e);  qqline(e)          # QQ-plot with reference line
```


residuals
3
2
1
0
−1
−2
30
35
40
45
fitted values


residuals
3
2
1
0
−1
−2
15
20
25
30
diameter x





No major departures from the assumptions are indicated !!

3. Randomized Blocks & Latin Squares Designs

3.1 The Randomized Complete Block Design

Define a **nuisance factor** as a design factor that probably has an effect on the response, but we are not interested in that effect.

• If a nuisance factor is **unknown** and, hence, **uncontrolled**, we don't know that it exists and it may even change levels during the experiments. **Randomization** is the design technique used to guard against such a *lurking* nuisance factor.

• Often, it is **known** but **uncontrolled**. If we are able to observe its value (yarn thickness), then we compensate for it by using the ANCOVA model.

• When the nuisance factor is **known** and **controllable**, then **blocking** can be used to systematically eliminate its effect on the statistical comparisons among treatments.

Example: Suppose we wish to determine whether or not 4 different tips produce different readings on a hardness testing machine. The machine operates by pressing the tip into a metal test coupon (from the depth of the resulting depressing, the hardness of the coupon is determined). We've decided to obtain 4 observations for each tip.

There is only 1 factor (tip type) and a **completely randomized single-factor design** would consist of **randomly** assigning each one of the 4×4 runs to an **experimental unit** (metal coupon) and observing the resulting hardness. Thus, 16 different test coupons would be required, one for each run in the design.

Potentially serious problem: if the coupons differs slightly in their hardness, then they will contribute to the variability observed in the hardness data.

$\Rightarrow$ experimental error will reflect random error and variability between coupons.

We would like to remove this variability from the experimental error. Such a design would require to test each tip once on each of the 4 coupons. This design is called a **randomized complete block design**. $Complete$ indicates that each block (coupon) contains all the treatments (tips). In this design, the blocks form a more homogeneous experimental unit on which to compare the tips (eliminates the variability among the blocks). Within a block, the order in which the 4 tips are tested is randomly determined.

	Test Coupon			
Tip	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

Statistical Analysis:
We have a treatments that are to be compared and b blocks. There is 1 observation per treatment in each block, and the order in which the treatments are run within each block is determined randomly (blocks represent a restriction on randomization).

Thus, we apply the model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \,,$$

μ is the overall mean,

τ_i is the effect of the ith treatment,

β_j is the effect of the jth block.

Treatments and blocks are fixed factors with $\sum_i \tau_i = 0$ and $\sum_j \beta_j = 0$.

Test **equality of treatment means $\Longleftrightarrow$ no treatment effects**

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \quad \Longleftrightarrow \quad H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0\,.$$

Partition the total sum of squares as

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2 &= b\sum_{i=1}^{a}(\overline{y}_{i.} - \overline{y}_{..})^2 + a\sum_{j=1}^{b}(\overline{y}_{.j} - \overline{y}_{..})^2 \\ &\quad + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2\end{aligned}$$

Thus we have

$$SS_{total} = SS_{treatments} + SS_{blocks} + SS_E$$

with associated degrees of freedom, df, $(N = ab)$

$$(N-1) = (a-1) + (b-1) + (a-1)(b-1)\,.$$

SS divided by df is a mean square. The expected value of the mean squares are

$$\begin{aligned}
\mathsf{E}(MS_{treatment}) &= \sigma^2 + \frac{b}{a-1}\sum_{i=1}^{a}\tau_i^2\,, \\
\mathsf{E}(MS_{blocks}) &= \sigma^2 + \frac{a}{b-1}\sum_{j=1}^{b}\beta_j^2\,, \\
\mathsf{E}(MS_E) &= \sigma^2
\end{aligned}$$

To test equal treatment means, we use the test statistic $F = MS_{treatments}/MS_E$.

```
> hard <- c(9.3, 9.4, 9.6, ..., 10.2);  tip <- gl(4, 4);  coupon <- gl(4,1, 16)
> anova(lm((hard-9.5)*10 ~ tip + coupon))
 Analysis of Variance Table

 Response: (hard - 9.5) * 10
           Df Sum Sq Mean Sq F value    Pr(>F)
 tip        3 38.500  12.833  14.438 0.0008713 ***
 coupon     3 82.500  27.500  30.938 4.523e-05 ***
 Residuals  9  8.000   0.889
```

We conclude that the type of tip affects the mean hardness reading.

Also the coupons (blocks) seem to differ significantly. But since the blocks represent a restriction on randomization, $F = MS_{blocks}/MS_E$ is no longer an exact F test statistic. However, we can use it at least approximately, indicating that blocking is necessary also in future experiments.

What happens, if we ignore the randomized block design?

Suppose we used 4 coupons, randomly assigned the tips to each, and (by chance) the same design results. The incorrect analysis as a completely randomized single-factor design is:

```
> anova(lm((hard-9.5)*10 ~ tip))
Analysis of Variance Table

Response: (hard - 9.5) * 10
          Df Sum Sq Mean Sq F value Pr(>F)
tip        3 38.500  12.833  1.7017 0.2196
Residuals 12 90.500   7.542
```

The Hypothesis of equal mean hardness from the 4 tips cannot be rejected!

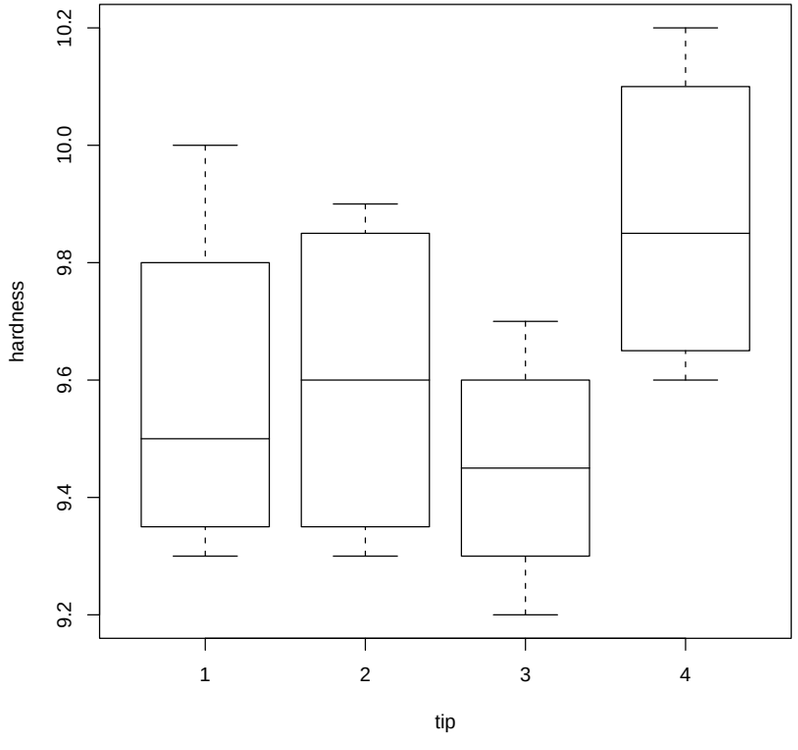
Thus, the randomized block design reduces the amount of noise sufficiently.

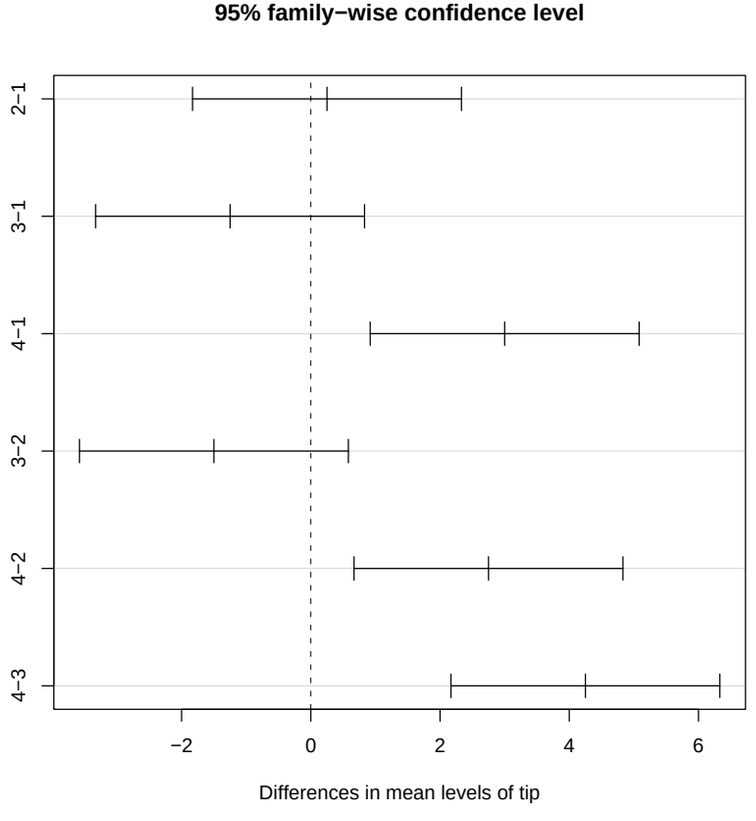
Multiple Comparisons: The analysis indicates a significant difference in treatment means. Now we are interested which treatment means differ.

We create a set of confidence intervals on the differences between the means of the levels of `tip`. The intervals are based on the Studentized range statistic, *Tukey's Honest Significant Difference* method.

```
> hardness.aov <- aov((hard-9.5)*10 ~ tip + coupon)
> TukeyHSD(hardness.aov, which="tip", ordered=FALSE, conf.level = 0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level
$tip
     diff     lwr    upr
2-1  0.25 -1.8312 2.3312
3-1 -1.25 -3.3312 0.8312
4-1  3.00  0.9188 5.0812
3-2 -1.50 -3.5812 0.5812
4-2  2.75  0.6688 4.8312
4-3  4.25  2.1688 6.3312
```

```
> plot(tip, hard);  plot(TukeyHSD(hardness.aov, "tip"))
```





Thus, tip type 4 produce a mean hardness reading that is significantly higher than the means from the other type of tips.

3.2 The Latin Square Design

The randomized complete block design was introduced to reduce the residual error by removing variability due to a known and controllable nuisance parameter.

There are several other designs that utilize the blocking principle.

Suppose that an experimenter is studying the effects of 5 different formulations of a rocket propellant on the observed burning rate. Each formulation is mixed from a batch of raw material that is only large enough for 5 formulations to be tested. Furthermore, the formulations are prepared by several operators, and there may be substantial differences in the skills and experience of the operators. Thus, it seems that there are 2 nuisance factors to be $averaged\ out$ in the design: batches of raw material and operators.

The appropriate design for this problem consists of testing each formulation exactly once in each batch of raw material and for each formulation to be prepared exactly once by each of 5 operators (**Latin Square Design**).

Batches of	Operators				
Raw Material	1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

Design is a square arrangement and the 5 formulations (treatments) are denoted by Latin letters (A, B, C, D, E).

The Latin square design is used to eliminate 2 nuisance sources of variability: it systematically allows blocking in 2 directions
$\Longrightarrow$ rows and columns represent $2\ restrictions\ on\ randomization$.

In general, a $p \times p$ Latin square design contains p rows and p columns. Each of the p^2 cells contains one of the p letters, and each letter occurs once and only once in each row and column.

Some examples of Latin squares:

4×4				5×5					6×6					
A	B	C	D	A	D	B	E	C	A	D	C	E	B	F
B	C	D	A	D	A	C	B	E	B	A	E	C	F	D
C	D	A	B	C	B	E	D	A	C	E	D	F	A	B
D	A	B	C	B	E	A	C	D	D	C	F	B	E	A
				E	C	D	A	B	F	B	A	D	C	E
									E	F	B	A	D	C

The statistical model for a Latin Square is:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk}\,,$$

where y_{ijk} is the observation in the ith row and kth column for the jth treatment

μ is the overall mean,

α_i is the ith row effect,

τ_j is the jth treatment effect,

β_k ist the kth column effect,

ϵ_{ijk} is the random error.

The model is completely **additiv**. There are no interactions between rows, columns, and treatments.

Since there is only 1 observation in each cell, only 2 of the 3 subscripts i, j, k are needed. E.g., if $i = 2$ and $k = 3$ we automatically find $j = 4$ (formulation D) (Because each treatment appears exactly once in each row and column.)

ANOVA decomposition:

$$SS_{Total} = SS_{Rows} + SS_{Columns} + SS_{Treatments} + SS_E$$

with respective degrees of freedom:

$$p^2 - 1 = (p-1) + (p-1) + (p-1) + (p-2)(p-1)$$

The appropriate statistic for testing for no differences in treatment means is

$$F = \frac{MS_{Treatments}}{MS_E}$$

```
> y <- c(24, 20, ..., 29, 31)
> oper  <- gl(5, 1, 25);  batch <- gl(5, 5)
> form <- as.factor(c("A","B","C","D","E", "B","C" ...,"D"))
> summary(aov(y ~ oper + batch + form))
            Df Sum Sq Mean Sq F value   Pr(>F)
 oper        4 150.00   37.50  3.5156 0.040373 *
 batch       4  68.00   17.00  1.5937 0.239059
 form        4 330.00   82.50  7.7344 0.002537 **
 Residuals  12 128.00   10.67
```

We conclude, that there is a significant difference in the mean burning rate generated by the different formulations.

There is also an indication that there are differences between operators, so blocking on this factor was a good precaution.

There is no strong evidence of a difference between batches of raw material, so it seems that in this particular experiment we were unnecessarily concerned about this source of variability.

A Latin square in which the first row and column consists of the letters in alphabetical order is called a **standard Latin square** (as in the example).

As with any experimental design, the observations in the Latin square should be taken in random order. E.g, if $p = 3$ there exist a total number of 12 Latin square designs. For our example with $p = 5$ we could already select out of 161,280 suitable Latin square designs.
Usual procedure: select a Latin square from a table of such designs, and then arrange the order of rows and columns, and letters at random.

With Latin squares we can investigate 3 factors (rows, columns, and letters), each at p levels, in only p^2 runs. This design assumes that there are no interactions between the factors.

Disadvantage of small Latin squares: they provide relatively small number of error df. E.g., a 3×3 design has only 2 error df, a 4×4 design has only 6 error df.
Solution: **replicate** them n times to increase error df! (There are several ways to do that.)

3.3 The Graeco-Latin Square Design

Consider a $p \times p$ Latin square, and superimpose on it a second $p \times p$ Latin square in which the treatments are denoted by Greek letters. If the two squares when superimposed have the property that each Greek letter appears once and only once with each Latin letter, the design obtained is called a **Graeco-Latin square**.

Example of a 4×4 Graeco-Latin square:

	Column			
Row	1	2	3	4
1	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
2	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
3	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
4	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

Such a design can be used to control systematically 3 sources of extraneous variability, that is, to block in 3 directions. The design allows investigation of 4 factors (rows, clomns, Latin and Greek letters), each at p levels in only p^2 runs.

Statistical model:

$$y_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \psi_l + \epsilon_{ijkl}\,,$$

where y_{ijkl} is the observation in row i and column l for Latin letter j and Greek letter k,

μ is the overall mean,

θ_i is the ith row effect,

τ_j is the effect of Latin letter j treatment,

ω_k is the effect of Greek letter k treatment,

ψ_l is the lth column effect,

ϵ_{ijkl} is the random error, assumed to be $N(0, \sigma^2)$.

Only 2 of the 4 subscripts are necessary to completely identify an observation.

ANOVA very similar to that of a Latin square.

$$SS_{Total} = SS_{Rows} + SS_{Columns} + SS_L + SS_G + SS_E$$

with respective degrees of freedom:

$$p^2 - 1 = (p-1) + (p-1) + (p-1) + (p-1) + (p-3)(p-1)$$

The appropriate F statistic for testing for no differences in rows, columns, Latin letters, and Greek letters is the respective mean square divided by the mean square error.

Example: Suppose that in the rocket propellant experiment an additional factor, $test\ assemblies$, could be of importance. Let there be 5 test assemblies denoted by the Greek letters $\alpha, \beta, \gamma, \delta$, and ϵ.

Here is the resulting 5×5 Graeco-Latin square design:

Batches of	Operators				
Raw Material	1	2	3	4	5
1	$A\alpha = 24$	$B\gamma = 20$	$C\epsilon = 19$	$D\beta = 24$	$E\delta = 24$
2	$B\beta = 17$	$C\delta = 24$	$D\alpha = 30$	$E\gamma = 27$	$A\epsilon = 36$
3	$C\gamma = 18$	$D\epsilon = 38$	$E\beta = 26$	$A\delta = 27$	$B\alpha = 21$
4	$D\delta = 26$	$E\alpha = 31$	$A\gamma = 26$	$B\epsilon = 23$	$C\beta = 22$
5	$E\epsilon = 22$	$A\beta = 30$	$B\delta = 20$	$C\alpha = 29$	$D\gamma = 31$

Notice that, since the totals for batches of raw material (rows), operators (columns), and formulations (Latin letters) are identical to those before, we have

$$SS_{Batches} = 68.0, \quad SS_{Operators} = 150.0, \quad SS_{Formulations} = 330.0.$$

```
> assem <- as.factor(c("a","c","e","b","d", "b","d", ...,"c"))
> summary(aov(y ~ oper + batch + form + assem))

            Df Sum Sq Mean Sq F value   Pr(>F)
oper         4 150.00   37.50  4.5455 0.032930 *
batch        4  68.00   17.00  2.0606 0.178311
form         4 330.00   82.50 10.0000 0.003344 **
assem        4  62.00   15.50  1.8788 0.207641
Residuals    8  66.00    8.25
```

Formulations are significantly different at 1%. Compared to the previous result, we see that removing the variability due to test assemblies has decreased the experimental error. However, we have also reduced the error df from 12 to 8. Thus, our estimate of error has fewer df, and the test may be less sensitive.

3.4 Balanced Incomplete Block Design

In some randomized block designs, it may not be possible to apply all treatments in every block. For example, in the hardness testing experiment, suppose that because of their size each coupon can be used only for testing 3 tips.

The question is: Which tips are to be tested on the first coupon, which on the second and so on if information is desired in all four tips?

A solution to this problem is to use a (**balanced incomplete block design**).

An incomplete block design is simply one in which there are more treatments than can be put in a single block.

A **balanced incomplete block design** is an incomplete block design in which every pair of treatments occurs the same number of times in the experiment.

The number of blocks necessary for balancing will depend on the number of treatments that can be run in a single block.

Example: Does time of reaction for a chemical process depend on the type of 4 catalyst employed? The experimental procedure consists of: select a batch of raw material, apply each catalyst in a separate run, observe reaction time. Since batches may affect the performance of the catalysts, we use batches as blocks.

However, each batch is only large enough to permit 3 catalysts to be run.
The order in which the catalysts are run in each block is randomized.

Treatment	Block (Material Batch)				
(Catalyst)	1	2	3	4	$y_{i.}$
1	73	74	—	71	218
2	—	75	67	72	214
3	73	75	68	—	216
4	75	—	72	75	222
$y_{.j}$	221	224	207	218	$870 = y_{..}$

Note that each pair of catalysts such as $(1, 2)$, occurs together twice in the experiment.

Assume that there are a treatments $(a = 4)$ and b blocks $(b = 4)$. Each block contains k treatments $(k = 3)$, each treatment occurs r times in the design $(r = 3)$, there are $N = ar = bk$ total observations $(N = 12)$.

The number of times each pair of treatments appears together throughout the experiment is $\lambda = r(k-1)/(a-1)$ $(\lambda = 2)$.

If $a = b$, the design is symmetric. λ must be an integer.

Statistical model (BIBD):

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}\,,$$

y_{ij} is the ith observation in the jth block, μ is the overall mean,

τ_i is the effect of the ith treatment,

β_j is the effect of the jth block,

ϵ_{ij} is the random error, assumed to be $N(0, \sigma^2)$.

Partition the total variability as

$$SS_{total} = SS_{treatments(adjusted)} + SS_{blocks} + SS_E$$

Because each treatment is represented in a different set of r blocks, the adjustment is necessary to extract treatment effect from blocks. The BIBD is **not orthogonal**.

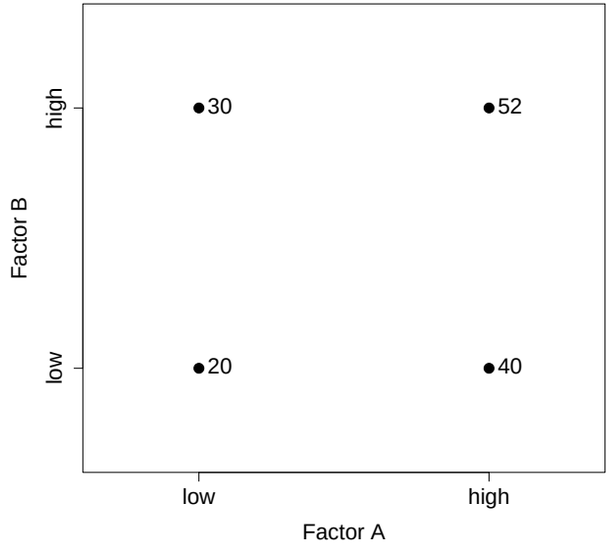
```
> y <- c(73,74,NA,71,...,75);  cat <- gl(4,4);   batch <- gl(4,1,16)
> summary(aov(y ~ cat + batch))  # yields unadjusted SS's
            Df Sum Sq Mean Sq F value    Pr(>F)
 cat         3 11.667   3.889  5.9829 0.0414634 *
 batch       3 66.083  22.028 33.8889 0.0009528 ***
 Residuals   5  3.250   0.650

> summary(aov(y ~ batch + cat))  # yields adjusted SS's
            Df Sum Sq Mean Sq F value   Pr(>F)
 batch       3 55.000  18.333  28.205 0.001468 **
 cat         3 22.750   7.583  11.667 0.010739 *
 Residuals   5  3.250   0.650
```

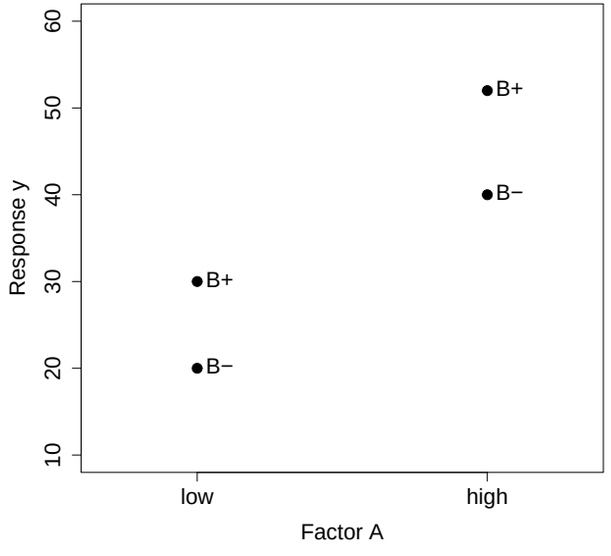
4. Introduction to Factorial Designs

4.1 Basic Definitions and Principles

Suppose there are now 2 factors of interest to the experimenter. For simplicity, let both factors have only 2 levels `low` and `high`, and denote them by $(-, +)$.

```
> A <- as.factor(c("low","low","high","high"))
> B <- as.factor(c("low","high","low","high"))
> y1 <- c(20, 30, 40, 52)
> y2 <- c(20, 40, 50, 12)
```

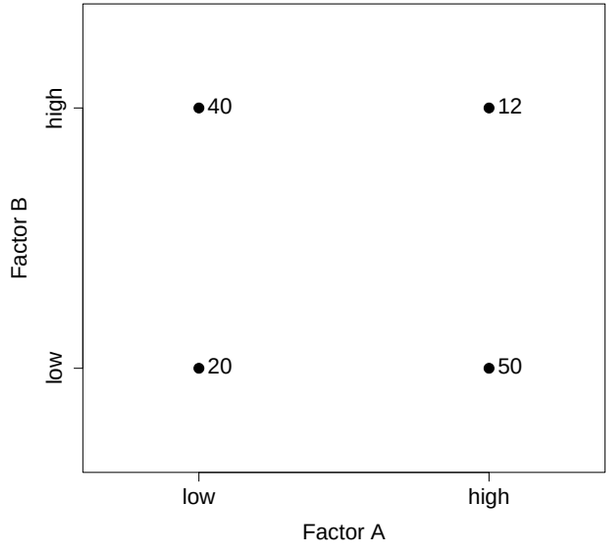


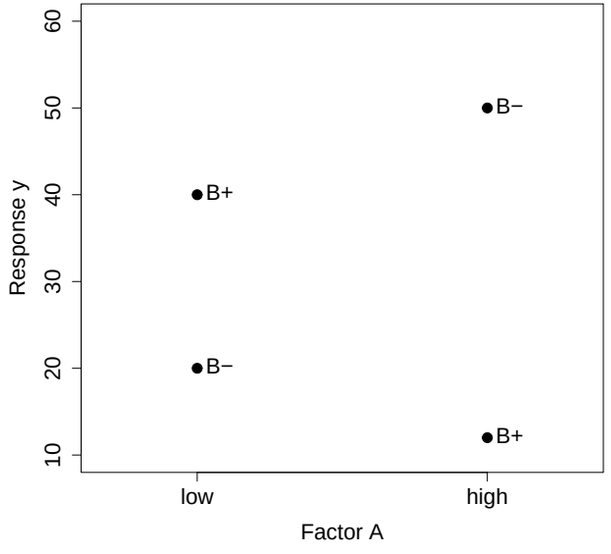


Definition of a factor effect: The change in the mean response when the factor changed from `low` to `high`.

$$A = \overline{y}_{A+} - \overline{y}_{A-} = \frac{40+52}{2} - \frac{20+30}{2} = 21$$

$$B = \overline{y}_{B+} - \overline{y}_{B-} = \frac{30+52}{2} - \frac{20+40}{2} = 11$$





In case of interaction:

$$
\begin{aligned}
A &= \overline{y}_{A+} - \overline{y}_{A-} = \frac{50+12}{2} - \frac{20+40}{2} = 1 \\
B &= \overline{y}_{B+} - \overline{y}_{B-} = \frac{40+12}{2} - \frac{20+50}{2} = -9 \\
AB &= \frac{20+12}{2} - \frac{40+50}{2} = -29
\end{aligned}
$$

The advantage of a factorial experiment:

1. More efficiency than on-factor-at-a-time experiments,

2. All data are used in computing both effects. (Note that all 4 observ's are used in determining the average effect of factor A and the average of factor B.),

3. Some information is provided on possible interaction between the 2 factors.

4.2 The Two-Factor Factorial Design

There are a levels of factor A and b levels of factor B. In addition, there are n replications for all ab treatment combinations.

The order in which the abn observations are taken is selected at random so that this design is a **completely randomized design**.

Example (Battery Design Experiment): Effective life time (in hours) of a battery possibly depend on the plate material of the battery, and the temperature (^{o}F) of the device for which the battery is used. $n = 4$ batteries are tested at each combination of material and temperature. All 36 tests are run in random order.

The engineer wants to answer the following questions.

- What effect do material type and temperature have on battery life?
- Is there a material that give *uniformly long life regardless of temperature*?

Life (in hours) data for the battery Design Example:

Material Type	Temperature (^{o}F)					
	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

The statistical (effects) model is:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}\,,$$

μ is the overall mean,
τ_i is the effect of the ith level of the row factor A, $\sum_i \tau_i = 0$,
β_j is the effect of the jth level of the column factor B, $\sum_j \beta_j = 0$,
$(\tau\beta)_{ij}$ is the interaction effect between τ_i and β_j, $\sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$,
ϵ_{ij} is the random error, assumed to be $N(0, \sigma^2)$.

The statistical (means) model is:

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}\,,$$

where $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$.

We are interested in testing the following hypotheses

1. The equality of row treatment effects

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$

2. The equality of column treatment effects

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$$

3. The exist of interaction

$$H_0 : (\tau\beta)_{ij} = 0 \quad \text{for all } i, j$$

The total variability can be expressed as (**two-factor ANOVA**)

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\overline{y}_{...})^2 &= bn\sum_{i=1}^{a}(\overline{y}_{i..}-\overline{y}_{...})^2+an\sum_{j=1}^{b}(\overline{y}_{.j.}-\overline{y}_{...})^2\\ &\quad +n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{ij.}-\overline{y}_{i..}-\overline{y}_{.j.}+\overline{y}_{...})^2\\ &\quad +\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\overline{y}_{ij.})^2\end{aligned}$$

Thus

$$SS_{total}=SS_A+SS_B+SS_{AB}+SS_E$$

with associated df

$$abn-1=(a-1)+(b-1)+(a-1)(b-1)+ab(n-1)$$

The expected values of the mean squares are

$$\begin{aligned}
\mathsf{E}(MS_A) &= \sigma^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\tau_i^2 \\
\mathsf{E}(MS_B) &= \sigma^2 + \frac{an}{b-1}\sum_{j=1}^{b}\beta_j^2 \\
\mathsf{E}(MS_{AB}) &= \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2 \\
\mathsf{E}(MS_E) &= \sigma^2
\end{aligned}$$

Under the three H_0's, MS_A, MS_B, MS_{AB}, and MS_E all estimates σ^2.

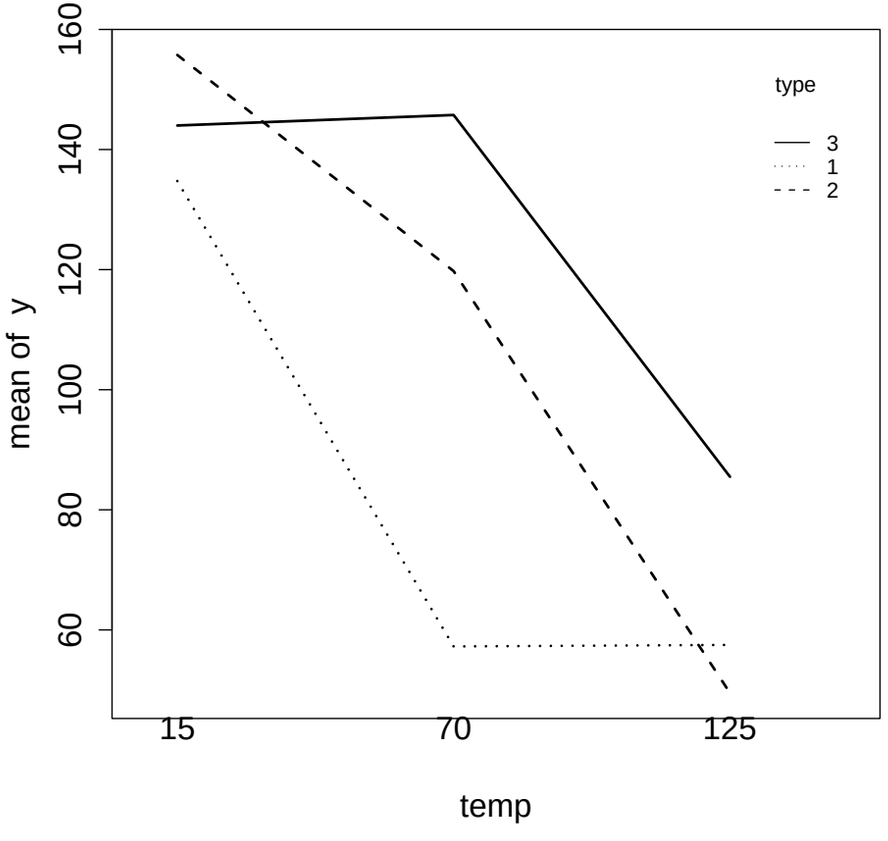
Under the three H_1's, $MS_A > MS_E$, $MS_B > MS_E$, $MS_{AB} > MS_E$, that is large values of the ratios imply that the data do not support the null hypotheses.

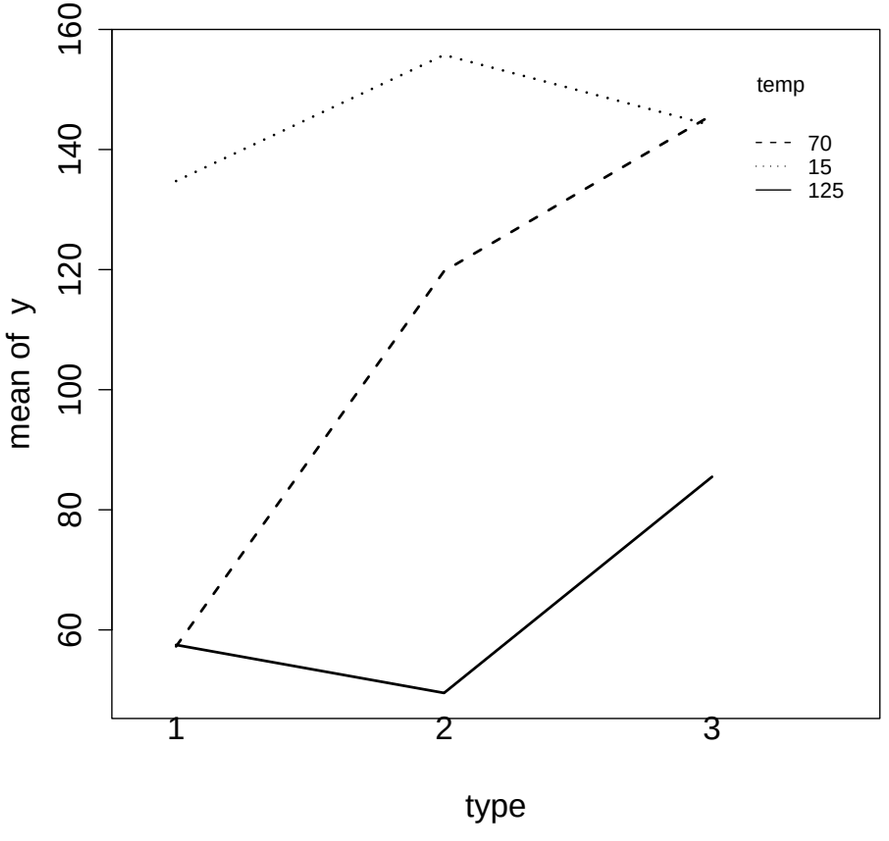
```
> y <- c(130,155,74,180, 34,40,80,75, ..., 96,104,82,60)
> type <- gl(3, 12)
> temp <- gl(3, 4, 36, levels = c("15", "70", "125"))
> life <- lm(y ~ type*temp)
> anova(life)
 Analysis of Variance Table

 Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
 type       2  10684    5342  7.9114  0.001976 **
 temp       2  39119   19559 28.9677 1.909e-07 ***
 type:temp  4   9614    2403  3.5595  0.018611 *
 Residuals 27  18231     675
```

Conclude that there is a significant interaction between material type and temperature. Both main effects are also significant.

Construct a graph of the average response at each treatment combination. The significant interaction is indicated by the lack of parallelism of the lines.

```
> # compute sample means (= fitted means) of each cell and plot it
> interaction.plot(temp, type, y)
> interaction.plot(type, temp, y)
```





Multiple Comparisons:

Once we fail to reject $H_0 : (\tau\beta)_{ij} = 0$ for all i, j, we can test the main effects.

Suppose that we reject $H_0 : \tau_i = 0$ or $H_0 : \beta_j = 0$. We then need to do multiple comparisons to discover specific differences between row or column means.

If interaction is significant, we could compare all ab cell means to determine which ones differ. This gives 36 comparisons between all possible pairs of the 9 means.

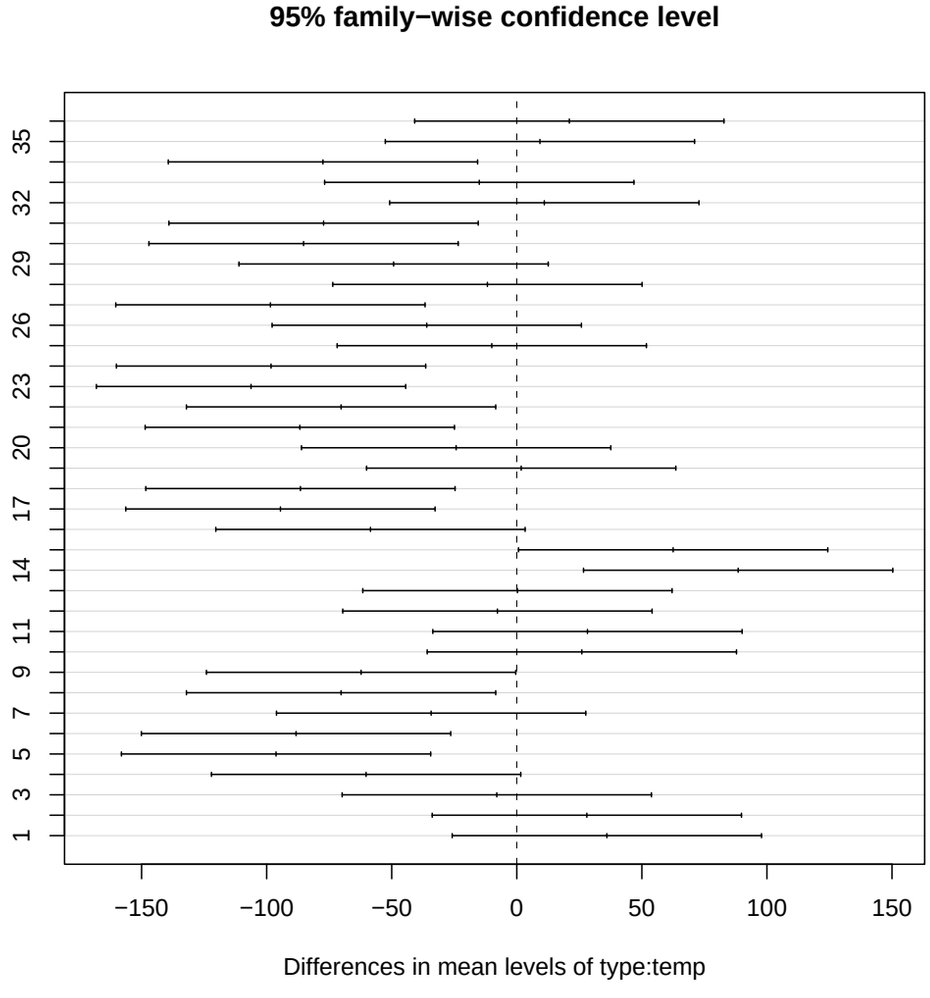
```
> tapply(fitted(life), list(type, temp), mean)
       15     70  125
 1 134.75  57.25 57.5
 2 155.75 119.75 49.5
 3 144.00 145.75 85.5
 > tapply(y, list(type, temp), mean)
       15     70  125
 1 134.75  57.25 57.5
 2 155.75 119.75 49.5
 3 144.00 145.75 85.5
```

```
> # we could compare pairs of row and/or column means (not appropriate here)
> # as also all pairs of cell means by:

> life.aov <- aov(y ~ type*temp)
> TukeyHSD(life.aov)
  Tukey multiple comparisons of means
     95% family-wise confidence level

 $type
         diff    lwr   upr
  2-1   25.17  -1.14 51.47
  3-1   41.92  15.61 68.22
  3-2   16.75  -9.55 43.05

 $temp
           diff      lwr     upr
  70-15  -37.25   -63.55  -10.95
 125-15  -80.67  -106.97  -54.36
 125-70  -43.42   -69.72  -17.11
```

```
$"type:temp"
         diff     lwr    upr
 [1,]   21.00  -40.82  82.82    #  (2, 15) - (1, 15)
 [2,]    9.25  -52.57  71.07    #  (3, 15) - (1, 15)
 [3,]  -77.50 -139.32 -15.68    #  (1, 70) - (1, 15)
 [4,]  -15.00  -76.82  46.82    #  (2, 70) - (1, 15)
 [5,]   11.00  -50.82  72.82    #  (3, 70) - (1, 15)
 [6,]  -77.25 -139.07 -15.43    #  (1,125) - (1, 15)
 ...
[22,]   62.50    0.68 124.32    #  (2, 70) - (1, 70)
[23,]   88.50   26.68 150.32    #  (3, 70) - (1, 70)
 ...
[27,]   26.00  -35.82  87.82    #  (3, 70) - (2, 70)
 ...
[36,]   36.00  -25.82  97.82    #  (3,125) - (2,125)
```

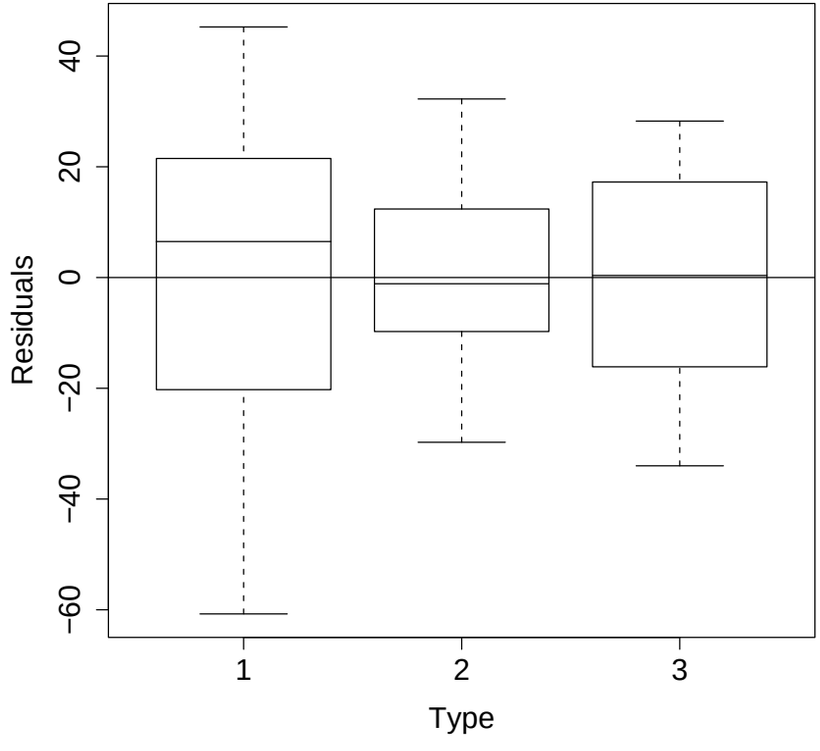
E.g., fix `temp=70` and test if mean battery life is the same for material types. Mean life is equal for material 2 and 3, but both of these materials are significantly better than material 1.

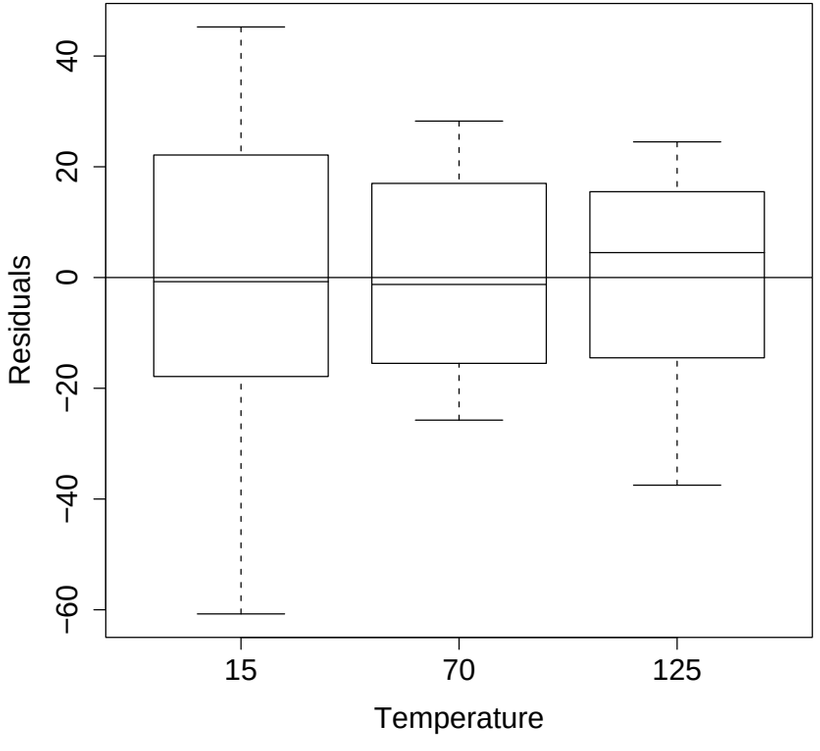
```
> plot(TukeyHSD(life.aov))  # notice: (22,23,27) corresponds to (15,14,10)
```

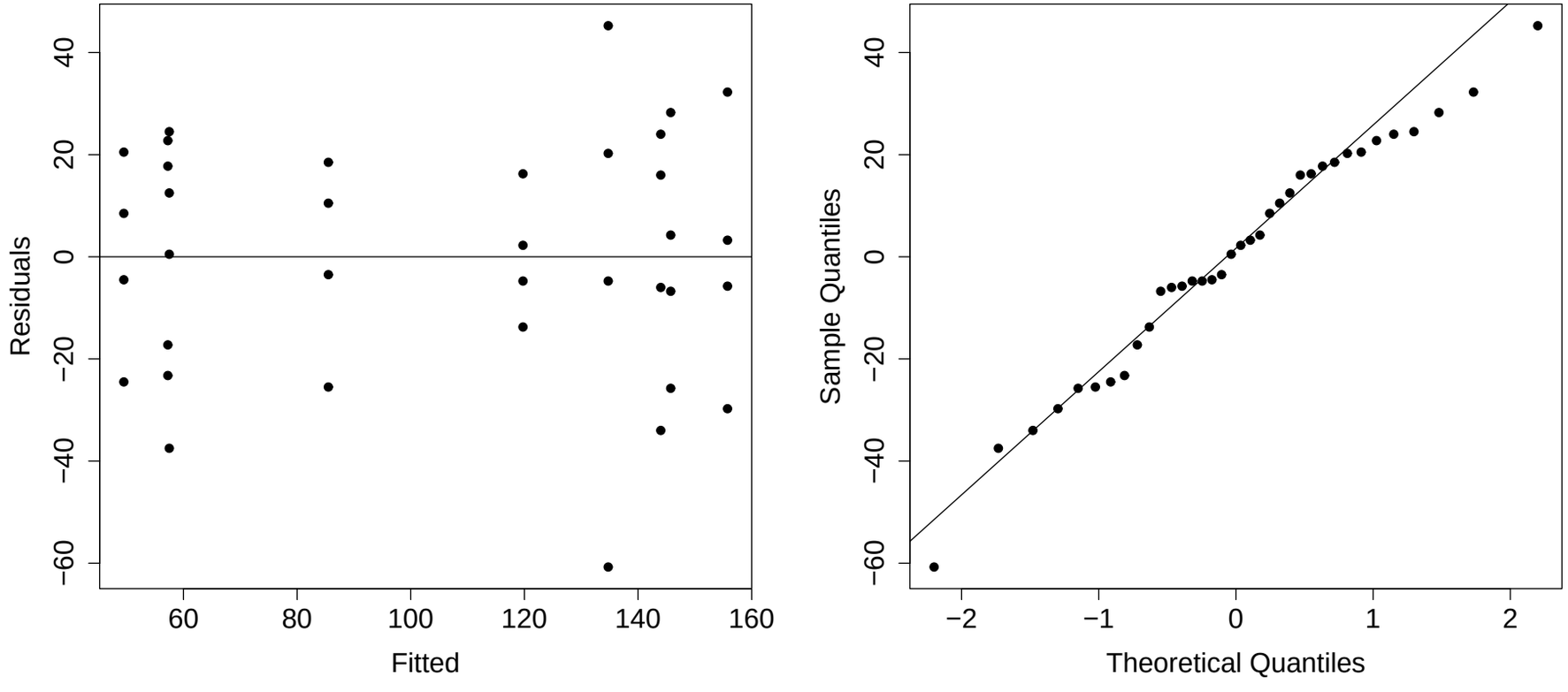


Model Diagnostics:

```
> e <- residuals(life);  f <- fitted(life)
> plot(type, e);  plot(temp, e)
> plot(f, e);  qqnorm(e);  qqline(e)
```







No major departures can be detected (variances only slightly increase as life increases). Since $\hat{\sigma} = 26$, only 1 residual (-60.75 from material 1, 15^o) is larger than $2\hat{\sigma}$. Notice that the second largest residual (45.26) is from the same cell.

One Observation per Cell

Two-factor experiment with only a single replicate ($n = 1$). The model is:

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij} \,,$$

with both factors again assumed to be fixed.

Under this model the error variance σ^2 is **not estimable**. The model is said to be **saturated** and results in $SS_E = 0$. If there is no interaction effect present, then $(\tau\beta)_{ij} = 0$ for all i, j, and we consider the main effects model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \,.$$

If this model is appropriate, then $\mathsf{E}(MS_E) = \sigma^2$, and the main effects A and B may be tested by comparing MS_A and MS_B to MS_E, respectively.

• How to test whether or not 2 factors interact when $n = 1$?

Instead of assuming the interaction model (no main effects can be tested) or the main effects model (which is too simple), Tukey considered the two-factor model

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{ij} \, ,$$

where γ is an unknown constant. By defining the interaction term this way, we may use a regression approach to test on $H_0 : \gamma = 0$.

The test partition the residual sum of squares $SS_{Residual}$ into a single-degree-of-freedom component (SS_N describing the non-additivity sum of squares related to γ and hence to the interaction) and SS_E, a component for error with $df_E = (a-1)(b-1) - 1$. That is

$$SS_E = SS_{Residual} - SS_N$$

$F = SS_N/(SS_E/df_E)$ is used to test on interaction.

The sum of squares for non-additivity is computed as

$$SS_N = \frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{i.}-\overline{y}_{..})(\overline{y}_{.j}-\overline{y}_{..})y_{ij}\right]^2}{\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2\sum_{j=1}^{b}(\overline{y}_{.j}-\overline{y}_{..})^2},$$

giving the ANOVA decomposition

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Rows (A)	SS_A	$a-1$	MS_A	MS_A/MS_E
Columns (B)	SS_B	$b-1$	MS_B	MS_B/MS_E
Non-additivity	SS_N	1	MS_N	MS_N/MS_E
Error	SS_E	$(a-1)(b-1)-1$	MS_E	
Total	SS_T	$ab-1$		

Example:

The impurity present in a chemical product is affected by two factors: Pressure and temperature. We have data from a single replicate of a factorial experiment.

Temperature	Pressure					
(^{o}F)	25	30	35	40	45	$y_{i.}$
100	5	4	6	3	5	23
125	3	1	4	2	3	13
150	1	1	3	1	2	8
$y_{.j}$	9	6	13	6	10	44

```
> y <- c(5, 4, 6, 3, ..., 2);  temp <- gl(3, 5, labels=c("100","125","150"))
> press <- gl(5, 1, 15, labels=c("25","30","35","40","45"))
> anova(lm(y ~ temp * press))  # saturated model => SSE=0
           Df Sum Sq Mean Sq F value Pr(>F)
 temp       2 23.333  11.667
 press      4 11.600   2.900
 temp:press 8  2.000   0.250
 Residuals  0  0.000
```

```
> a <- anova(lm(y ~ temp + press));  a
 Analysis of Variance Table
 Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
 temp       2 23.333  11.667  46.667 3.885e-05 ***
 press      4 11.600   2.900  11.600  0.002063 **
 Residuals  8  2.000   0.250

> SStemp  <- a[1,2];  dft <- a[1,1]
> SSpress <- a[2,2];  dfp <- a[2,1]
> SSresid <- a[3,2];  dfr <- a[3,1]

> # Now use the function tukey.1df() to calculate SSN
> source("tukey.1df.R")
> data <- matrix(c(as.numeric(temp), as.numeric(press), y), nrow=length(y))
> colnames(data) <- c("temp","press","y")
> SSN <- tukey.1df(data); SSN
[1] 0.09852217
```

```
> SSE <- SSresid - SSN;  dfE <- dfr-1;  MSE <- SSE/dfE
> Ftemp <- (SStemp/dft)/MSE;  Fpress <- (SSpress/dfp)/MSE;  FN <- SSN/MSE
> 1-pf(Ftemp, dft, dfE);  1-pf(Fpress, dfp, dfE);  1-pf(FN, 1, dfE)
```

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P-Value
Temperature	23.33	2	11.67	42.95	0.0001
Pressure	11.60	4	2.90	10.68	0.0042
Non-additivity	0.098	1	0.098	0.36	0.5660
Error	1.902	7	0.272		
Total	36.93	14			

From the test statistic for non-additivity $F = 0.36$ (with p-value 0.566) we conclude that there is no evidence of interaction in this data. The main effects of temperature and pressure are significant.

4.3 The General Factorial Design

The results for the two-factor factorial design can be extended to the general case with a levels of factor A, b levels of factor B, c levels of factor C, and so on. We assume again that there are $n \geq 2$ replicates of the complete experiment.

For example consider the three-factor analysis of variance model

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}\,,$$

with all factors A, B, and C fixed, and $\epsilon_{ijkl} \sim N(0, \sigma^2)$.

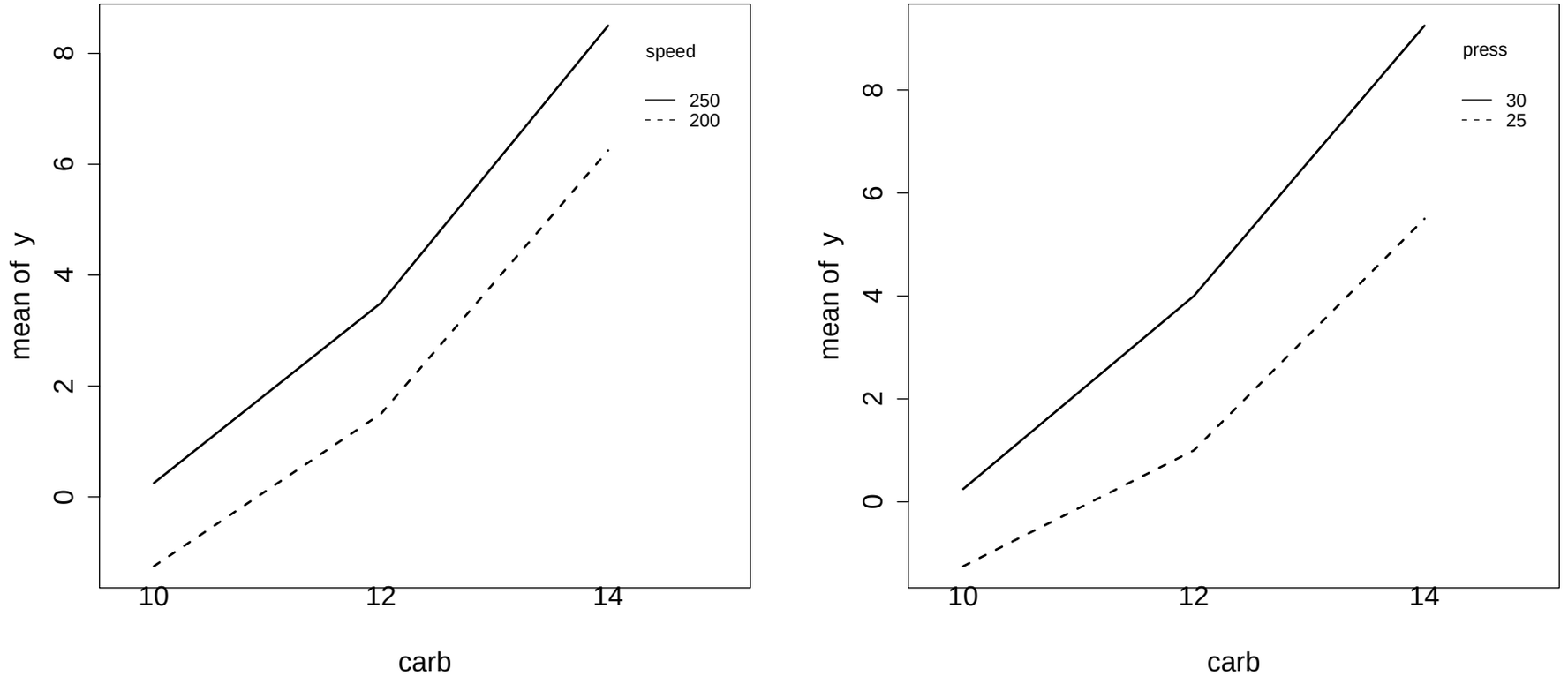
Example: A soft drink bottler is interested in obtaining more uniform fill heights in the bottles. The engineer can control 3 variables during the filling process: the percent carbonation (A), the operating pressure in the filler (B), and the bottles produced per minute (C, line speed). The response observed is the deviation from the target fill height.

Percent Carbonation	Operating Pressure: 25 psi, Line Speed 200	25 psi, Line Speed 250	30 psi, Line Speed 200	30 psi, Line Speed 250	$y_{i...}$
10	-3	-1	-1	1	-4
	-1	0	0	1	
12	0	2	2	6	20
	1	1	3	5	
14	5	7	7	10	59
	4	6	9	11	
$y_{.jk.}$	6	15	20	34	$y_{....} = 75$
$y_{.j..}$	21		54		

```
> y <- c(-3,-1,-1,0,-1,0,1,1,0,1,2,1,2,3,6,5,5,4,7,6,7,9,10,11)
> carb <- gl(3, 8, labels=c("10", "12", "14"))
> press <- gl(2, 4, 24, labels=c("25", "30"))
> speed <- gl(2, 2, 24, labels=c("200", "250"))
```

```
> anova(lm(y ~ carb*press*speed))
 Analysis of Variance Table

 Response: y
                  Df  Sum Sq Mean Sq  F value    Pr(>F)
 carb              2 252.750 126.375 178.4118 1.186e-09 ***
 press             1  45.375  45.375  64.0588 3.742e-06 ***
 speed             1  22.042  22.042  31.1176 0.0001202 ***
 carb:press        2   5.250   2.625   3.7059 0.0558081 .
 carb:speed        2   0.583   0.292   0.4118 0.6714939
 press:speed       1   1.042   1.042   1.4706 0.2485867
 carb:press:speed  2   1.083   0.542   0.7647 0.4868711
 Residuals        12   8.500   0.708
```

We see that carbonation, pressure, and speed significantly affect the fill volume. The carbonation/pressure interaction F ratio has a p-value of 0.0558, indicating some interaction between these two factors.



So we decide to recommend the low level of pressure (25 psi) and the high level of line speed (250 bpm, which will maximize the production rate). The carbonation rate, which is difficult to control, should be also kept low.

5. The 2^k Factorial Design

We consider k factors, each at only 2 levels (they could be either quantitative or qualitative and are usually denoted by `low` and `high`, or $-, +$). A complete replicate of such a design requires $2 \times 2 \times \cdots \times 2 = 2^k$ observations and is called 2^k **factorial design**. This class of designs is very widely used in industrial experimentation.

Throughout this chapter we assume that the factors are fixed, the designs are completely randomized, and the usual normality assumptions are satisfied.

5.1 The 2^2 Design

Only two factors A and B, each run at two levels. Typical for chemical process data, where A denotes reactant concentration (15 and 25%), and B is the amount of catalyst used (`low=1pound` and `high=2pounds`). The experiment is replicated three times.

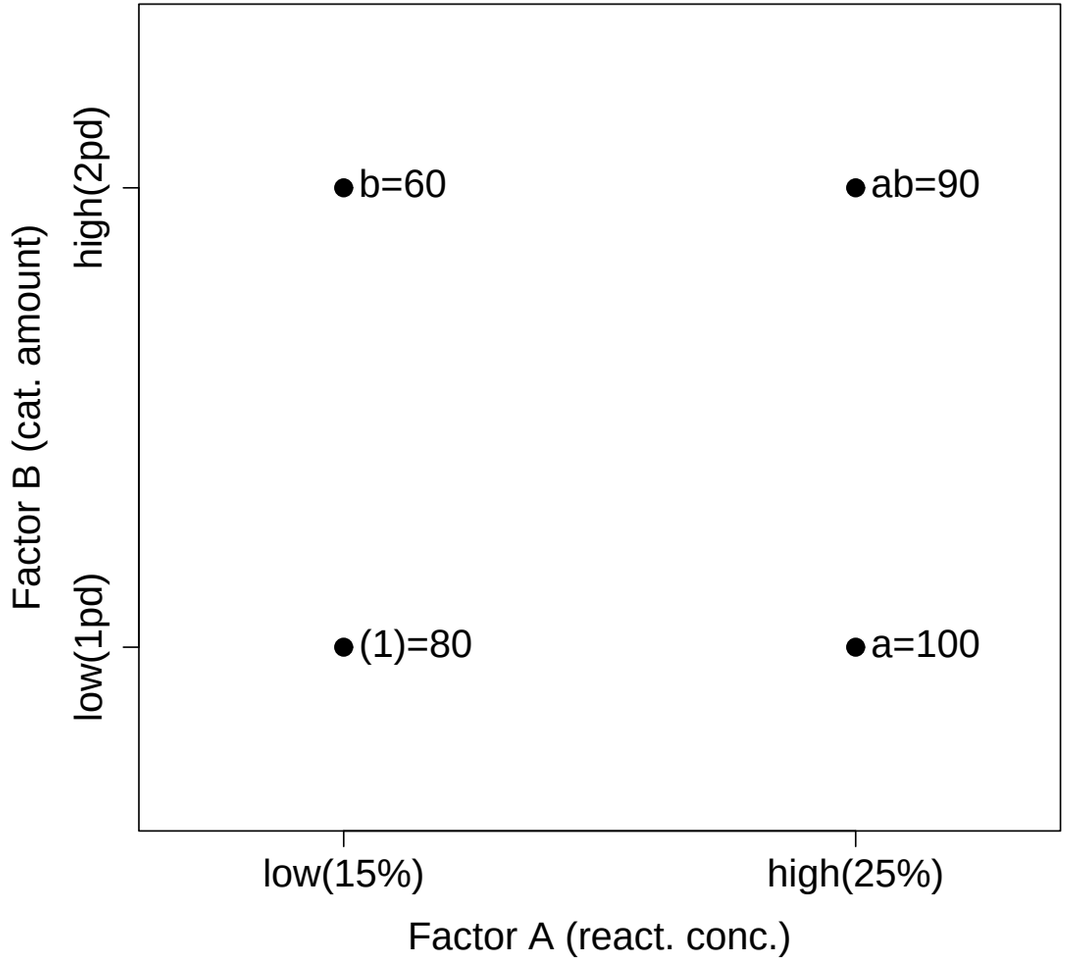
Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
−	−	A low, B low	28	25	27	80
+	−	A high, B low	36	32	32	100
−	+	A low, B high	18	19	23	60
+	+	A high, B high	31	30	29	90

By convention we denote the effect of a factor by a capital Latin letter.

The high level of any factor in the treatment combination is denoted by the corresponding lowercase letter.

The low level of any factor in the treatment combination is denoted by the absence of the corresponding letter.

Thus, a represents the treatment combination of A at high level and B at the low level. ab represents both factors at the high level, and (1) is used to denote both factors at the low level.


b=60
ab=90
(1)=80
a=100
high(2pd)
low(1pd)
Factor B (cat. amount)
low(15%)
high(25%)
Factor A (react. conc.)

Analysis procedure for a factorial design:

- Estimate factor effects; main effects A and B, and interaction AB
- Statistical testing (ANOVA); compute SS terms according to A, B, AB, and error; build ANOVA table and test
- Analyze residuals; check normality assumption and constant variance

Compute main effects and interaction effect

Treatment Combination	Effect of Factor I	A	B	AB
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

In a two-level factorial design, we define the average effect of a factor as the change in response produced by a change in the level of that factor averaged over the levels of the other factor.

The effect of A at the low level of B is $[a-(1)]/n$ and at the high level of B it is $[ab-b]/n$. Averaging these quantities yields the **main effect** of A. Applying this principle also onto B and AB gives

$$\begin{aligned} A &= \frac{1}{2n}\{[ab-b]+[a-(1)]\} = \frac{1}{2n}\{[ab+a]-[b+(1)]\} \\ B &= \frac{1}{2n}\{[ab-a]+[b-(1)]\} = \frac{1}{2n}\{[ab+b]-[a+(1)]\} \\ AB &= \frac{1}{2n}\{[ab-b]-[a-(1)]\} = \frac{1}{2n}\{[ab+(1)]-[a+b]\} \end{aligned}$$

For the chemical experiment, we get $A = 8.33$, $B = -5.00$, and $AB = 1.67$.

The effect of A is positive; increasing reactant conc. from low to high will increase the yield. The effect of B is negative; increasing the amount of catalyst will decrease the yield. The interaction effect appears to be relatively small.

Both main effects and the interaction effect were estimated by means of **contrasts**. These are linear combinations of the treatment totals, e.g. $C = \sum_{i=1}^{a} c_i y_{i.}$ with the restriction $\sum_{i=1}^{a} c_i = 0$. The sum of squares due to a contrast C is

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{n \sum_{i=1}^{a} c_i^2} .$$

We define contrasts in 2^2 designs as

$$\begin{aligned} \text{contrast}_A &= ab + a - b - (1) \\ \text{contrast}_B &= ab - a + b - (1) \\ \text{contrast}_{AB} &= ab - a - b + (1) . \end{aligned}$$

These 3 contrasts are **orthogonal**. The sum of squares due to contrasts are

$$\begin{aligned} SS_A &= \frac{(\text{contrast}_A)^2}{4n} \\ SS_B &= \frac{(\text{contrast}_B)^2}{4n} \\ SS_{AB} &= \frac{(\text{contrast}_{AB})^2}{4n}. \end{aligned}$$

In the example $n = 3$, giving sum of squares $SS_A = 50^2/12 = 208.33$, $SS_B = (-30)^2/12 = 75.00$, and $SS_{AB} = 10^2/12 = 8.33$. SS_T and SS_E are computed in the usual way giving $SS_T = 323.00$ and $SS_E = SS_T - SS_A - SS_B - SS_{AB} = 31.33$. We summarized these results again in an ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
A	SS_A	1	MS_A	MS_A/MS_E
B	SS_B	1	MS_B	MS_B/MS_E
AB	SS_{AB}	1	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	$2^2(n-1)$	MS_E	
Total	SS_T	$n2^2-1$		

```
> y <- c(28, 25, 27, 36, ..., 29);  rep <- gl(3,1,12)
> A <- gl(2, 3, 12, labels=c("-","+"));  B <- gl(2, 6, 12, labels=c("-","+"))
> anova(lm(y ~ A*B))
 Analysis of Variance Table
 Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
 A          1 208.333 208.333 53.1915 8.444e-05 ***
 B          1  75.000  75.000 19.1489  0.002362 **
 A:B        1   8.333   8.333  2.1277  0.182776
 Residuals  8  31.333   3.917
```

It is often convenient to to write down the treatment combinations in the order $(1), a, b, ab$, This is referred to as **standard order**. Using this standard order, we see that the contrast coefficients are:

Effects	(1)	a	b	ab
A	-1	$+1$	-1	$+1$
B	-1	-1	$+1$	$+1$
AB	$+1$	-1	-1	$+1$

Regression Approach: For the chemical process example the model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \,,$$

where x_1 and x_2 are **coded variables** representing the **natural variables**, reactant concentration and the amount of catalyst used.

The relationship between the natural variable and the coded variable is

$$x_1 = \frac{\text{conc} - (\text{conc}_l + \text{conc}_h)/2}{(\text{conc}_h - \text{conc}_l)/2}, \qquad x_2 = \frac{\text{cata} - (\text{cata}_l + \text{cata}_h)/2}{(\text{cata}_h - \text{cata}_l)/2}$$

When the natural variables have only 2 levels, this coding will produce the familiar ± 1 notation. In the example, this gives

$$\begin{aligned} x_1 &= \frac{\text{conc} - (15+25)/2}{(25-15)/2} = \frac{\text{conc} - 20}{5} \\ x_2 &= \frac{\text{cata} - (1+2)/2}{(2-1)/2} = \frac{\text{cata} - 3/2}{1/2} \end{aligned}$$

If concentration is at the high level 25%, then $x_1 = +1$ (low level 15% results in $x_1 = -1$). If catalyst is at high level 2 pounds, then $x_2 = 1$ (low level 1 pound results in $x_2 = -1$).

```
> mod <- lm(y ~ A+B, x=TRUE)  # provides also the design matrix X
> mod$x                       # 'low' is coded as +1 and 'high' as -1
   (Intercept) A1 B1
1            1  1  1
2            1  1  1
3            1  1  1
4            1 -1  1
5            1 -1  1
6            1 -1  1
7            1  1 -1
8            1  1 -1
9            1  1 -1
10           1 -1 -1
11           1 -1 -1
12           1 -1 -1
```

The factor level appearing first is always coded as $+1$ in R. Arranging the data appropriately is the only chance in order not to get $wrong$ signs of the estimates.

```
> summary(mod)
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  27.5000      0.606  45.377 6.13e-12 ***
 A1           -4.1667      0.606  -6.875 7.27e-05 ***
 B1            2.5000      0.606   4.125  0.00258 **
```

The intercept is the grand average of all 12 observations, and the regression coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ are one-half the corresponding factor effect estimates. (Because the regression coefficient measures the effect of a unit change in x on the mean of y, and the factor effect is based on a two-unit change from -1 to $+1$.)

The fitted mean model is ($x_1 = x_2 = +1$ if concentration and catalyst are low)

$$\hat{\mu} = 27.5 + \left(\frac{-8.33}{2}\right) x_1 + \left(\frac{+5.00}{2}\right) x_2$$

and can be compared to the estimated factor effects $A = +8.33$ and $B = -5.00$.

Notice, a regression model with interaction effect results the same main effect estimates as the main effects only model (but slightly different standard errors).

```
> summary(lm(y ~ A*B))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  27.5000     0.5713  48.135 3.84e-11 ***
 A1           -4.1667     0.5713  -7.293 8.44e-05 ***
 B1            2.5000     0.5713   4.376  0.00236 **
 A1:B1         0.8333     0.5713   1.459  0.18278
```

Fitted values: Using the model w/o interaction effect we get as fitted cell means

Reactant Concentration	Amount of Catalyst: 1 Pound ($x_2 = 1$)	Amount of Catalyst: 2 Pounds ($x_2 = -1$)
15% ($x_1 = 1$)	$27.5 - 4.167 + 2.5 = 25.83$	$27.5 - 4.167 - 2.5 = 20.83$
25% ($x_1 = -1$)	$27.5 + 4.167 + 2.5 = 34.17$	$27.5 + 4.167 - 2.5 = 29.17$

The model with interaction gives the observed cell means as fitted values.

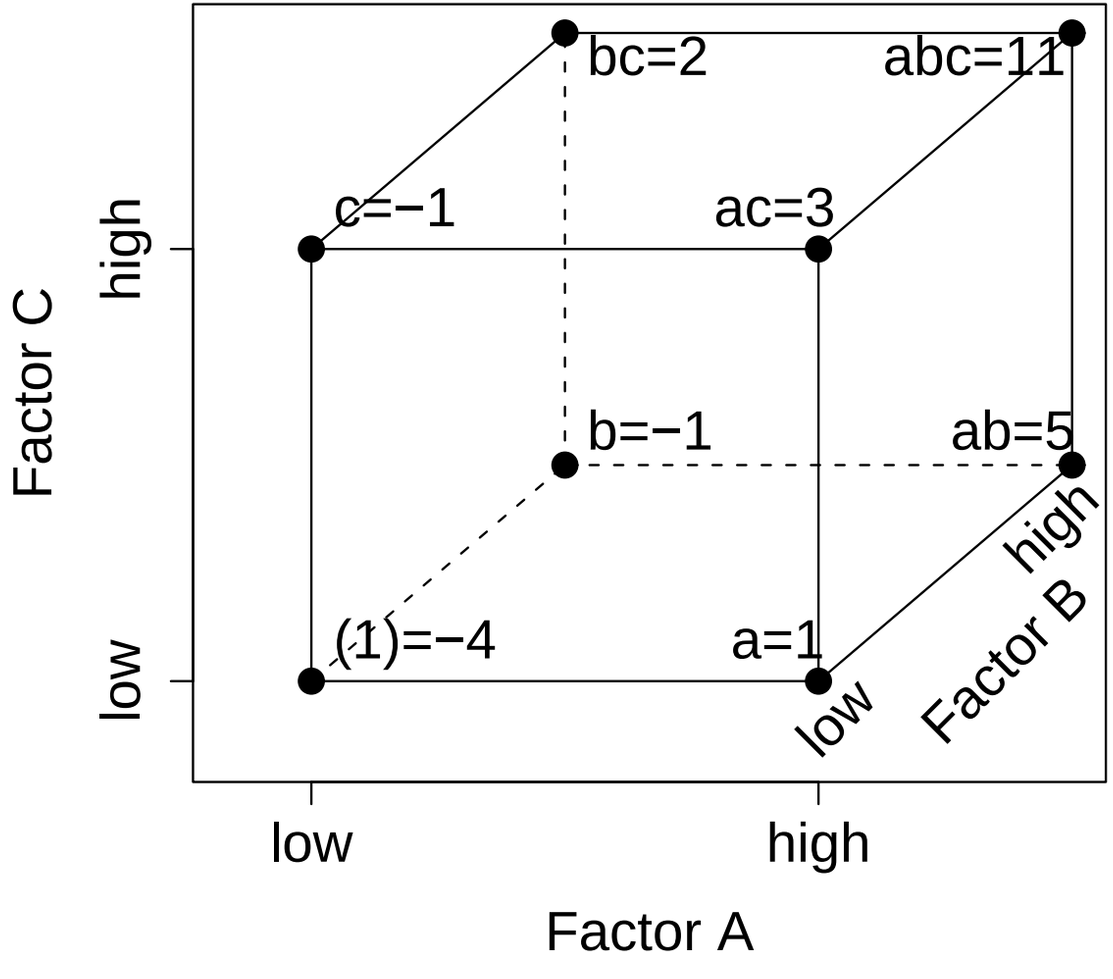
5.2 The 2^3 Design

Suppose that three two-level factors A, B, and C are of interest.

Example: Recall the previous bottle filling example, where 3 levels of percent carbonation (10, 12, and 14%) were used. Suppose that only the first two are studied. Then the data can be described as a 2^3 factorial experiment.

	Operating Pressure (B)			
	25 psi		30 psi	
Percent	Line Speed (C)		Line Speed (C)	
Carbonation (A)	200	250	200	250
10	-3	-1	-1	1
	-1	0	0	1
12	0	2	2	6
	1	1	3	5

The 8 treatment combinations can now be displayed geometrically as a cube.


bc=2
abc=11
c=−1
ac=3
b=−1
ab=5
(1)=−4
a=1
high
low
Factor C
high
low
Factor B
low
high
Factor A

The effect of A when B and C are at the low level is $[a-(1)]/n$. The effect of A when B is high and C is low is $[ab-b]/n$. The effect of A when C is high and B is low is $[ac-c]/n$. When B and C are high, the effect of A is $[abc-bc]/n$. Thus the average effect of A is just the average of these 4 effects, i.e.

$$A = \frac{1}{4n}\left[(a-(1)) + (ab-b) + (ac-c) + (abc-bc)\right]$$

After arranging these term we get

$$\begin{aligned} A &= \left[a+ab+ac+abc-(1)-b-c-ac\right]/4n \\ B &= \left[b+ab+bc+abc-(1)-a-c-ac\right]/4n \\ C &= \left[c+ac+bc+abc-(1)-a-b-ac\right]/4n \end{aligned}$$

Similar expressions could be found for the interaction effects. Sum of Squares can be computed by $SS = (\text{contrast})^2/8n$.

```
> y <- c(-3, -1, -1, 0, -1, 0, ..., 6, 5)
> carb <- gl(2, 8, labels=c("10", "12"))
> press <- gl(2, 4, 16, labels=c("25", "30"))
> speed <- gl(2, 2, 16, labels=c("200", "250"))
> anova(lm(y ~ carb*press*speed))
 Analysis of Variance Table

 Response: y
                  Df Sum Sq Mean Sq F value    Pr(>F)
 carb              1 36.000  36.000    57.6 6.368e-05 ***
 press             1 20.250  20.250    32.4 0.0004585 ***
 speed             1 12.250  12.250    19.6 0.0022053 **
 carb:press        1  2.250   2.250     3.6 0.0943498 .
 carb:speed        1  0.250   0.250     0.4 0.5447373
 press:speed       1  1.000   1.000     1.6 0.2415040
 carb:press:speed  1  1.000   1.000     1.6 0.2415040
 Residuals         8  5.000   0.625
```

The main effects are very strong. Only the AB interaction is slightly significant.

Regression Model Approach: We use a model with all main effects and carbonation/pressure interaction only to predict fill height deviation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_{12} + \epsilon$$

```
> summary(lm(y ~ carb+press+speed+carb:press))
              Estimate Std. Error t value Pr(>|t|)
 (Intercept)     1.000      0.203   4.927 0.000452 ***
 carb1          -1.500      0.203  -7.391 1.38e-05 ***
 press1         -1.125      0.203  -5.543 0.000175 ***
 speed1         -0.875      0.203  -4.311 0.001233 **
 carb1:press1    0.375      0.203   1.848 0.091700 .
```

Remember, that a factor at its low (high) level is again coded as $x = -1$ ($x = +1$).

$$\hat{\mu} = 1.00 + \left(\frac{3.00}{2}\right) x_1 + \left(\frac{2.25}{2}\right) x_2 + \left(\frac{1.75}{2}\right) x_3 + \left(\frac{-0.75}{2}\right) x_{12}$$

5.3 A Single Replicate of the 2^k Design

For even a moderate number of factors, the total number of treatment combination in a 2^k design is large (a 2^6 design has 64 treatment combinations). Frequently, available resources only allow a **single replicate** of the design to be run.

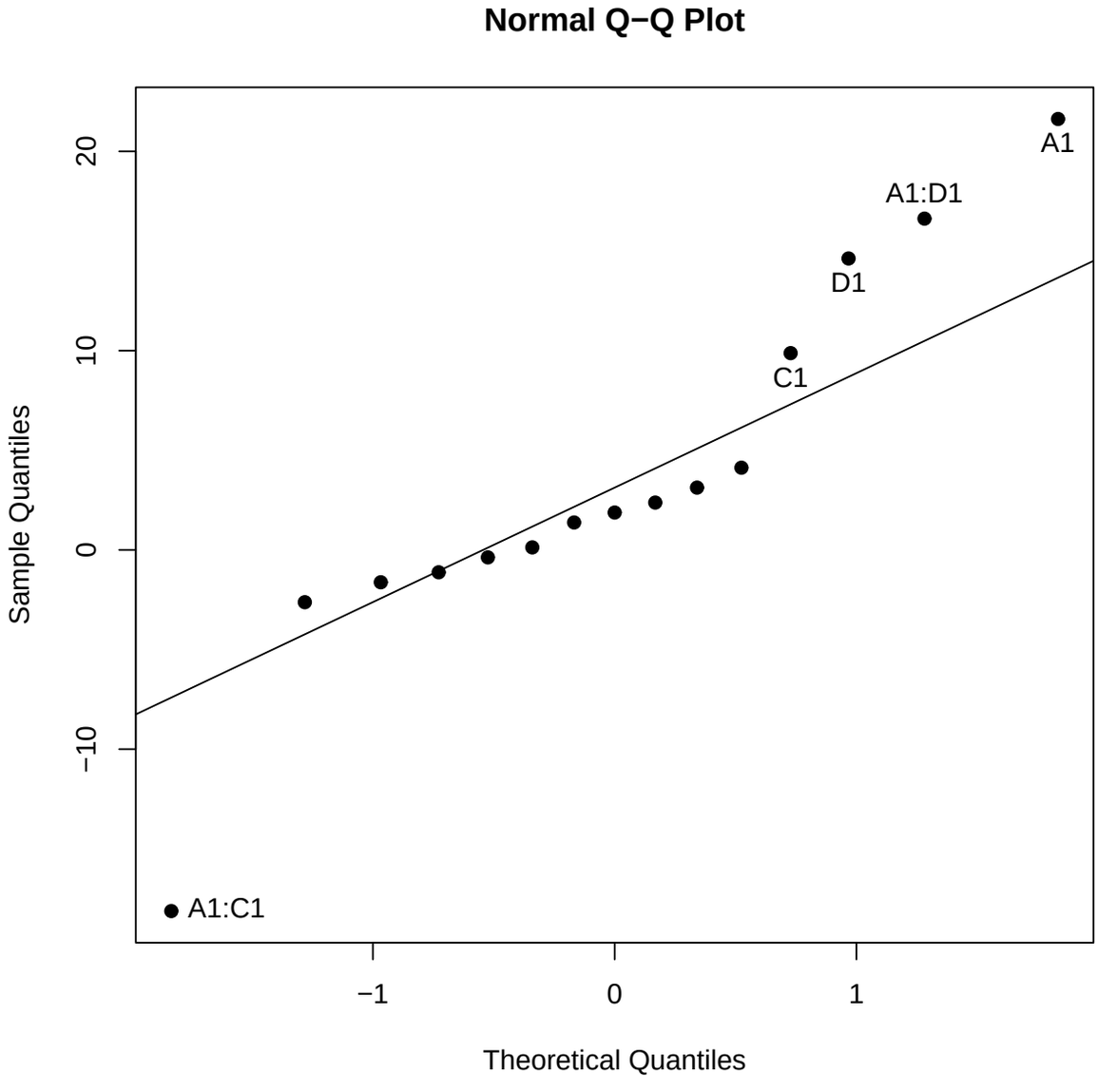
Example: A chemical product is produced in a pressure vessel. The four factors temperature (A), pressure (B), concentration of formaldehyde (C), and stirring rate (D) are possible influencing the mean filtration rate y. $2^4 = 16$ runs are made in random order.

The process engineer is interested in maximizing the filtration rate. The engineer also would like to reduce the formaldehyde concentration as much as possible. Currently, the process uses the concentration at the high level (low level always results in lower filtration rates).

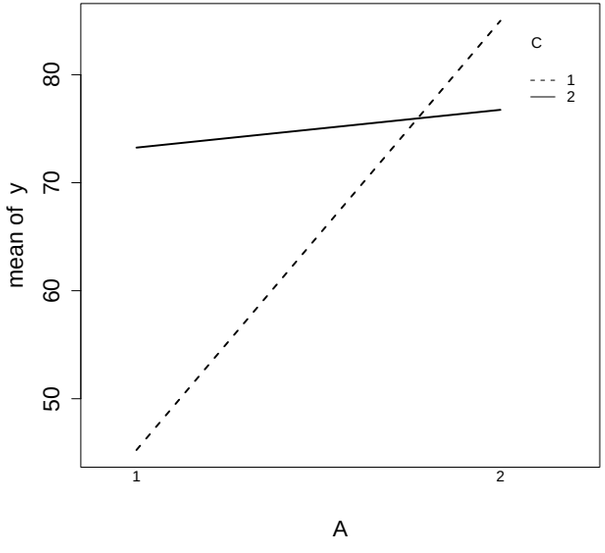
Run Number	Factor A	B	C	D	Run Label	Filtration Rate
1	−	−	−	−	(1)	45
2	+	−	−	−	a	71
3	−	+	−	−	b	48
4	+	+	−	−	ab	65
5	−	−	+	−	c	68
6	+	−	+	−	ac	60
7	−	+	+	−	bc	80
8	+	+	+	−	abc	65
9	−	−	−	+	d	43
10	+	−	−	+	ad	100
11	−	+	−	+	bd	45
12	+	+	−	+	abd	104
13	−	−	+	+	cd	75
14	+	−	+	+	acd	86
15	−	+	+	+	bcd	70
16	+	+	+	+	$abcd$	96

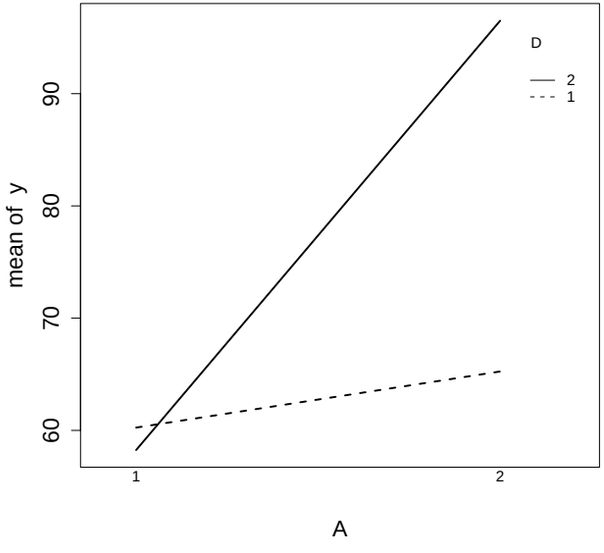
```
> y <- c(45,71,48,65,68,60,80,65,43,100,45,104,75,86,70,96)
> A <- gl(2,1,16);  B <- gl(2,2,16);  C <- gl(2,4,16);  D <- gl(2,8,16)
> mod <- lm(y ~ A*B*C*D)
> fac.effects <- mod$coeff[2:16] * c(-2,-2,-2,-2,2,2,2,2,2,2,-2,-2,-2,-2,2)
> fac.effects
        A1          B1          C1          D1
    21.625       3.125       9.875      14.625
     A1:B1       A1:C1       A1:D1       B1:C1
     0.125     -18.125      16.625       2.375
     B1:D1       C1:D1    A1:B1:C1    A1:B1:D1
    -0.375      -1.125       1.875       4.125
  A1:C1:D1    B1:C1:D1 A1:B1:C1:D1
    -1.625      -2.625       1.375
> qqnorm(fac.effects);  qqline(fac.effects)
```

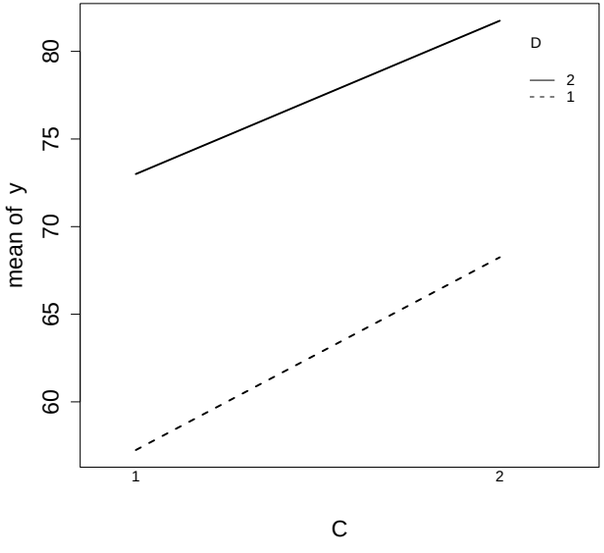
All of these effects that lie along the straight line in the probability plot are negligible, whereas the large effects are far from the line. Thus, the important effects are the main effects of A, C, D, and the AC and AD interaction. The model considered is saturated, thus the ANOVA table does not give F tests.


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
20
10
0
−10
−1
0
1
A1
A1:D1
D1
C1
A1:C1

```
> interaction.plot(A,C,y);  interaction.plot(A,D,y);  interaction.plot(C,D,y)
```







We study a regression model with the main factors A, C, D, and all of their interactions included.

```
> anova(lm(y ~ A * C * D ))
 Analysis of Variance Table

 Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
 A          1 1870.56 1870.56 83.3677 1.667e-05 ***
 C          1  390.06  390.06 17.3844 0.0031244 **
 D          1  855.56  855.56 38.1309 0.0002666 ***
 A:C        1 1314.06 1314.06 58.5655 6.001e-05 ***
 A:D        1 1105.56 1105.56 49.2730 0.0001105 ***
 C:D        1    5.06    5.06  0.2256 0.6474830
 A:C:D      1   10.56   10.56  0.4708 0.5120321
 Residuals  8  179.50   22.44
```

The same conclusions can be drawn as from the probability plot. The interactions CD and ACD are not significant.

```
> summary(lm(y ~ A * C * D - C:D - A:C:D))   # F1 stands for  F=low
 Coefficients:                               # add nothing if F=high
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)   70.062      1.104  63.444 2.30e-14 ***
 A1           -10.812      1.104  -9.791 1.93e-06 ***
 C1            -4.938      1.104  -4.471  0.00120 **
 D1            -7.313      1.104  -6.622 5.92e-05 ***
 A1:C1         -9.062      1.104  -8.206 9.41e-06 ***
 A1:D1          8.312      1.104   7.527 2.00e-05 ***
```

This gives fitted values (remember: if $x = +1$ for $high$, and $x = -1$ for low)

$$\begin{aligned}\hat{\mu} &= 70.06 + \left(\frac{21.625}{2}\right) x_1 + \left(\frac{9.875}{2}\right) x_3 + \left(\frac{14.625}{2}\right) x_4 \\ &\quad + \left(\frac{-18.125}{2}\right) x_{13} + \left(\frac{16.625}{2}\right) x_{14}\end{aligned}$$

The Addition of Center Points to the 2^k Design

A potential concern in the use of two-level factorial experiments is the assumption of **linearity** in the factor effects. If the k factors are **quantitative**, a more appropriate model in some situations is the **second-order response surface model**

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \epsilon \,,$$

where β_{jj} represent **pure quadratic effects**. We also add **center points** to the 2^k design. These consist of n replicates at the points $x_i = 0$, $i = 1, \ldots, k$.

Consider a 2^2 design with 1 observation at each factorial point $(-,-)$, $(+,-)$, $(-,+)$, and $(+,+)$ and with n_C observations at the center point $(0,0)$. Let $\overline{y}_F$ be the average of the 4 runs at the 4 factorial points, and let $\overline{y}_C$ be the average of the n_C runs at the center point. If the difference $\overline{y}_F - \overline{y}_C$ is small, then the center points lie on or near the plane passing through the factorial points (no quadratic effect).

A single degree-of-freedom **sum of squares for pure quadratic curvature** is

$$SS_{Pure\ Quadratic} = \frac{n_F n_C (\overline{y}_F - \overline{y}_C)^2}{n_F + n_C},$$

where n_F is the number of factorial design points. $F = SS_{Pure\ Quadratic}/MS_E$ actually tests $H_0 : \sum_j \beta_{jj} = 0$. Furthermore, if the factorial points are unreplicated, one may use the n_C center points to construct an estimate of error

$$SS_E = \sum_{center\ points} (y_i - \overline{y}_C)^2$$

with $n_C - 1$ df.

Example: The yield of a chemical process depend on reaction time (A: low is 30, high is 40 min) and reaction temperature (B: low is 150, high is 160 degrees). Because we are uncertain about the linearity, we conduct a 2^2 factorial experiment (with a single replicate) augmented with 5 center points run at 35 minutes, 155 degrees.

Run	A	B	Yield
1	low	low	39.3
2	low	high	40.0
3	high	low	40.9
4	high	high	41.5
5	center	center	40.3
6	center	center	40.5
7	center	center	40.7
8	center	center	40.2
9	center	center	40.6

```
> y <- c(39.3,40.0,40.9,41.5,40.3,40.5,40.7,40.2,40.6)
> A <- as.factor(c("-1","-1","1","1","0","0","0","0","0"))
> B <- as.factor(c("-1","1","-1","1","0","0","0","0","0"))
> m.f <- mean(y[1:4]);  m.f          # mean of obs at factorial points
[1] 40.425
> m.c <- mean(y[A==0 & B==0]);  m.c  # mean of obs at center points
[1] 40.46
```

```
> MSE <- var(y[A==0 & B==0]);  MSE   # MS from center points (df=4)
[1] 0.043
> SSPQ <- 4*5*(m.f-m.c)^2/9;  SSPQ   # SS Pure Quadratic (df=1)
[1] 0.002722222
> SSPQ/MSE  # Test statistic on no-curvature hypothesis
[1] 0.06330749

> anova(lm(y ~ A*B))  # wrong ANOVA !!
 Analysis of Variance Table

 Response: y
           Df Sum Sq Mean Sq F value   Pr(>F)
 A          2 2.4052  1.2026 27.9677 0.004454 **   # also includes SSPQ
 B          1 0.4225  0.4225  9.8256 0.035030 *    # correct
 A:B        1 0.0025  0.0025  0.0581 0.821316      # correct
 Residuals  4 0.1720  0.0430                       # correct
```

To get $df = 1$ for both three-level factors we define A and B as variables:

```
> A <- c(-1,-1,1,1,0,0,0,0,0);  B <- c(-1,1,-1,1,0,0,0,0,0)
> anova(lm(y ~ A*B))
           Df  Sum Sq Mean Sq F value    Pr(>F)
 A          1 2.40250 2.40250 68.7520 0.0004166 ***
 B          1 0.42250 0.42250 12.0906 0.0177127 *
 A:B        1 0.00250 0.00250  0.0715 0.7997870
 Residuals  5 0.17472 0.03494                      # also includes SSPQ

> anova(lm(y ~ A*B + I(A^2) + I(B^2)))
           Df  Sum Sq Mean Sq F value   Pr(>F)
 A          1 2.40250 2.40250 55.8721 0.001713 **
 B          1 0.42250 0.42250  9.8256 0.035030 *
 I(A^2)     1 0.00272 0.00272  0.0633 0.813741    # SSPQ is now separated
 A:B        1 0.00250 0.00250  0.0581 0.821316
 Residuals  4 0.17200 0.04300
```

Conclusion: significant main effects, no interaction and no second-order curvature.

Remember our regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon$$

with 6 parameters included. But the 2^2 design plus center points only has 5 independent runs. Thus, 1 parameter is not estimable.

```
> summary(lm(y ~ A*B + I(A^2) + I(B^2)))
 Coefficients: (1 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) 40.46000    0.09274 436.291 1.66e-10 ***
 A            0.77500    0.10368   7.475  0.00171 **
 B            0.32500    0.10368   3.135  0.03503 *
 I(A^2)      -0.03500    0.13910  -0.252  0.81374
 A:B         -0.02500    0.10368  -0.241  0.82132
```

The quadratic B effect cannot be estimated. In a **central composite design** the 2^k design is augmented with central points and some further **axial points** like $(\sqrt{2}, 0)$, $(-\sqrt{2}, 0)$, $(0, -\sqrt{2})$, and $(0, \sqrt{2})$ for a 2^2 design (very effective!).

6. Blocking and Confounding in the 2^k Factorial Design

Blocking a replicated Design

- Blocking is a technique for dealing with controllable nuisance variables
- If there are n replicates of the design, then each replicate is a block
- Each replicate is a run of the blocks (time periods, batches of raw materials, etc.)
- Runs within the block are randomized

Example: The chemical process experiment with A (concentration) and B (catalyst) from the previous chapter.

Suppose that only 4 experimental trials can be made from a single batch of raw material. Therefore, 3 batches of raw material will be required to run all three replicates of this design.

Block 1	Block 2	Block 3
$(1) = 28$	$a = 32$	$ab = 29$
$a = 36$	$ab = 30$	$(1) = 27$
$b = 18$	$(1) = 25$	$b = 23$
$ab = 31$	$b = 19$	$a = 32$
$B_1 = 113$	$B_2 = 106$	$B_3 = 111$

```
> A <- as.factor(c("low","high","low","high", ..., "high"))
> B <- as.factor(c("low","low","high","high", ...,  "low"))
> Block <- as.factor(c(1, 1, 1, 1, 2, ..., 3))
> y <- c(28, 36, 18, 31, 32, 30, 25, 19, 29, 27, 23, 32)
```

```
> anova(lm(y ~ Block + A*B))
 Analysis of Variance Table

 Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
 Block      2   6.500   3.250  0.7852 0.4978348
 A          1 208.333 208.333 50.3356 0.0003937 ***
 B          1  75.000  75.000 18.1208 0.0053397 **
 A:B        1   8.333   8.333  2.0134 0.2057101
 Residuals  6  24.833   4.139
```

The conclusions from this analysis, had the design been run in blocks, are identical to those before (relatively small block effect).

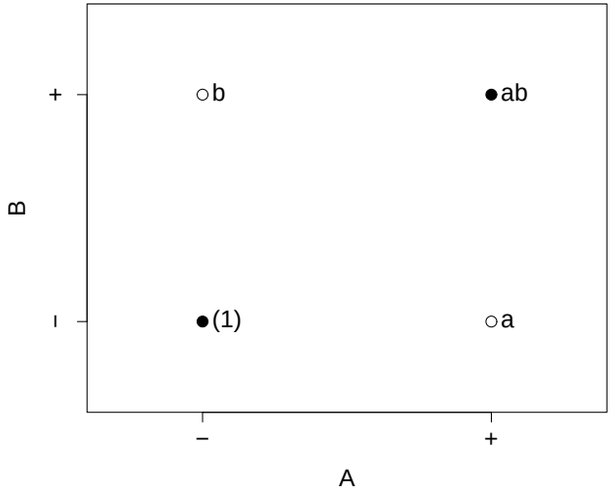
Confounding in the 2^k factorial design

When the block size is smaller than the number of treatment combinations in one replicate, confounding is a design technique for arranging a complete factorial experiments in blocks. Usually, higher order interactions are confounded with blocks.

Even though the designs presented are incomplete block designs, because each block does not contain all the treatments or treatment combinations, the special structure of the 2^k factorial system allows a simplified method of analysis.

Simple Confounding: Run a single replicate of a 2^2 design. Each batch of raw material is only large enough for 2 treatment combinations. Thus, 2 batches are required and we consider batches as blocks. One possible design is

Block 1	Block 2
(1)	a
ab	b



Treatment combinations on opposing diagonals are assigned to different blocks. The order in which the treatment combinations are run within a block is randomly determined. We also randomly decide which block to run first.

Suppose we estimate the main effects of A and B as if no blocking had occurred:

$$A = \frac{1}{2}[ab + a - b - (1)] \qquad B = \frac{1}{2}[ab + b - a - (1)]$$

Note that both A and B are unaffected by blocking since in each estimate there is 1 plus and 1 minus treatment combination from each block. That is, any difference between block 1 and block 2 will cancel out.

Now consider the AB interaction

$$AB = \frac{1}{2}[ab + (1) - a - b]$$

Since the 2 treatment combinations with plus sign, ab and (1), are in block 1 and the 2 with a minus sign, a and b, are in block 2, the block effect and the AB interaction are identical. That is, AB is is indistinguishable from, or **confounded with blocks**.

The is apparent from the table of plus and minus signs:

Treatment	Factorial Effect			
Combination	I	A	B	AB
(1)	$+$	$-$	$-$	$+$
a	$+$	$+$	$-$	$-$
b	$+$	$-$	$+$	$-$
ab	$+$	$+$	$+$	$+$
				$\Uparrow$

All treatment combinations that have a plus sign on AB are assigned to block 1, whereas all treatment combinations that have a minus sign on AB are assigned to block 2. This approach can be used to confound any effect (A, B, or AB) with blocks.

This approach can be used to confound any effect (A, B, AB) with blocks. E.g., for

Block 1	Block 2
(1)	a
b	ab

Treatment Combination	Factorial Effect I	A	B	AB
(1)	$+$	$-$	$-$	$+$
a	$+$	$+$	$-$	$-$
b	$+$	$-$	$+$	$-$
ab	$+$	$+$	$+$	$+$
		$\Uparrow$		

the main effect A would have been confounded with blocks.

This scheme can be used to confound any 2^k design in two blocks. As a second example, consider the 2^3 design run in 2 blocks. Suppose we wish to confound the 3-factor interaction ABC with blocks. From the table of plus/minus signs we assign the treatment combinations that are minus on ABC to block 1 and those that are plus on ABC to block 2.

Treatment Combination	Factorial Effect I	A	B	AB	C	AC	BC	ABC
(1)	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$
a	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$+$
b	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
ab	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$
c	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$
ac	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$-$
bc	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$
abc	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

Block 1	Block 2
(1)	a
ab	b
ac	c
bc	abc

Run the treatment combinations within a block in random order!

Example: Recall the example in which temperature (A), pressure (B), concentration of formaldehyde (C), and stirring rate (D) are studied to determine their effect on filtration rate.

We make 2 modifications:

• suppose the 16 treatment combinations cannot all be run using 1 batch of raw material. We can run 8 from a single batch, so a 2^4 design confounded in 2 blocks seems appropriate. Of course, we confound the highest-order interaction $ABCD$ with blocks.

• We introduce a block effect, so that the utility of blocking can be demonstrated. suppose that one of the 2 batches of raw material is of much poorer quality (batch 1), and, as a result, all responses will be 20 units lower in this batch than in the other. Now all the tests in block 1 are performed first (in random order)

Block 1	Block 2
$(1) = 25$	$a = 71$
$ab = 45$	$b = 48$
$ac = 40$	$c = 68$
$bc = 60$	$d = 43$
$ad = 80$	$abc = 65$
$bd = 25$	$bcd = 70$
$cd = 55$	$acd = 86$
$abcd = 76$	$abd = 104$

```
> y <- c(45,71,48,65,68,60,80,65,43,100,45,104,75,86,70,96) # original data
> A <- gl(2,1,16);  B <- gl(2,2,16);  C <- gl(2,4,16);  D <- gl(2,8,16)
```

```
> block <- as.factor(c(1,2,2,1,2,1,1,2,2,1,1,2,1,2,2,1))
> y <- y - 20*(block=="1")

> options(contrasts=c("contr.sum", "contr.poly"))
> mod <- lm(y ~ A*B*C*D) # as if no blocking had occurred
> summary(mod)
 Residuals:
 ALL 16 residuals are 0: no residual degrees of freedom!

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  60.0625         NA      NA       NA
 A1          -10.8125         NA      NA       NA
 :
 A1:B1:C1:D1  -9.3125         NA      NA       NA

 Residual standard error: NaN on 0 degrees of freedom
 Multiple R-Squared:     1,      Adjusted R-squared:   NaN
 F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

The estimates of all 4 main effects, 6 two-factor interactions, and the 4 three-factor interactions are identical to the effect estimates obtained previously, where there was no block effect.

What about the $ABCD$ interaction effect? The estimate in the original experiment was $ABCD = 1.375$. Now it is $ABCD = -18.625$. Since $ABCD$ is confounded with blocks, the $ABCD$ interaction estimates the *original interaction effect* plus the *block effect*, which is -20.

```
> anova(lm(y ~ A + C + D +A:C + A:D + block))

           Df  Sum Sq Mean Sq F value    Pr(>F)
A           1 1870.56 1870.56  89.757 5.600e-06 ***
C           1  390.06  390.06  18.717 0.0019155 **
D           1  855.56  855.56  41.053 0.0001242 ***
block       1 1387.56 1387.56  66.581 1.889e-05 ***
A:C         1 1314.06 1314.06  63.054 2.349e-05 ***
A:D         1 1105.56 1105.56  53.049 4.646e-05 ***
Residuals   9  187.56   20.84
```

Experiments with Random Factors

1. Introduction:

- Previous chapters have considered fixed factors.

  1. A specific set of factor levels is chosen for the experiment
  2. Inference confined to those levels.
  3. Often quantitative factors are fixed (why?)

- When factor levels are chosen at random from a larger population of potential levels, the factor is random

  1. Inference is about the entire population of levels.
  2. Industrial applications include measurements system studies.

1. Introduction:

- Example