Experimental Design and Data Analysis, Lecture 5

Eduard Belitser

VU Amsterdam

Lecture overview

1. two-way ANOVA
2. randomized block design
3. repeated measures
4. Friedman test

two way ANOVA
(completely randomized design)

Setting

An experiment with:

- a numerical outcome Y;
- two factors (categorical variables) that can be fixed at I and J levels (categories), respectively.

EXAMPLE Agricultural experiment with outcome total yield from a plot and factors type of fertilizer and crop variety.

EXAMPLE Quality of a genetic algorithm to determine the minimal value of a criterion function with outcome CPU time needed to find true minimum and factors mutation probability and population size.

EXAMPLE Outcome time to develop mold on bread and factors temperature and humidity.

Design

- Select NIJ experimental units randomly from the population of interest.
- Assign combined levels (i, j) of the factors to a random set of N units.
- Independently perform the NIJ experiments.

Randomization in R:

```
> I=4; J=2; N=3
> rbind(rep(1:I,each=N*J),rep(1:J,N*I),sample(1:(N*I*J)))
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
[1,]    1    1    1    1    1    1    2    2    2     2     2     2     3
[2,]    1    2    1    2    1    2    1    2    1     2     1     2     1
[3,]   20    1    3   14   17   24   19   12   22    13    16    15     4
     [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
[1,]     3     3     3     3     3     4     4     4     4     4     4
[2,]     2     1     2     1     2     1     2     1     2     1     2
[3,]    23     8    10     2     7    21     9     5     6    18    11
```

For unit 20 use levels (1,1) of (factor 1, factor 2); for unit 1 use levels (1,2); ...; for unit 11 use levels (4,2).

Two-way ANOVA model: data

Data

$$\text{sample } (i,j): \quad Y_{ij1}, Y_{ij2}, \ldots, Y_{ijn_{ij}}, \quad i = 1, \ldots, I; \;\; j = 1, \ldots, J.$$

Assume that these are sampled independently from IJ normal populations with (possibly different) population means μ_{ij}, and with equal population variances.

Commonly, one considers balanced design: $n_{ij} = N$ for all subgroups (i,j).

We want to test the following null hypotheses:

- no interaction between the two factors A and B,
- no main effect of the first factor A,
- no main effect of the second factor B.

The overall nullhypothesis $H_0 : \mu_{ij} = \mu_{kl}$ for every i, j, k, l is of modest interest.

We also estimate the corresponding parameters.

Two-way ANOVA model: assumptions

The two-way ANOVA model is:

$$Y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, n_{ij}.$$

Assumption: the indep. errors $e_{ijk} \sim N(0, \sigma^2)$, with unknown variance σ^2.

We decomposed the (i,j)-group means as $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where

- μ is the overall mean,
- α_i is the main effect of level i of the first factor A,
- β_j is the main effect of level j of the second factor B,
- γ_{ij} is the interaction effect of levels i and j of the first and second factors.

Now we can formalize the hypothesis to test:

- $H_{AB} : \gamma_{ij} = 0$ for every (i,j) (no interactions between factor A and B),
- $H_A : \alpha_i = 0$ for every i (no main effect of factor A),
- $H_B : \beta_j = 0$ for every j (no main effect of factor B).

For the parameters to be identifiable, we need to impose $I + J + 1$ linear restrictions, (done by command contrasts in R). The default in R is the `treatment` parametrization: $\alpha_1 = \beta_1 = \gamma_{1j} = \gamma_{i1} = 0$, $j = 1, \ldots, J$, $i = 1, \ldots, I$. Often one uses the `sum` parametrization: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i \gamma_{ij} = 0$ for all $j = 1, \ldots, J$, and $\sum_j \gamma_{ij} = 0$ for all $i = 1, \ldots, I$.

Tests in two-way ANOVA

Setting: a two-way ANOVA model: $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$.

Hypotheses: we want to test H_{AB}, H_A, H_B against their negations.

Test statistics: F_{AB} for testing H_{AB}, F_A for testing H_A, and F_B for testing H_B.

Distribution of F's under H_{AB}, H_A, H_B: $F_{AB} \sim F_{(I-1)(J-1),n-IJ}$, $F_A \sim F_{I-1,n-IJ}$, $F_B \sim F_{J-1,n-IJ}$). $F_{m,k}$ is the F-distribution with m and k degrees of freedom.

Test: larger values of $F_{AB} = f_{AB}$ give more evidence against H_{AB}, hence we reject H_{AB} if F_{AB} is large. The test is therefore always right-sided: compare the p-value $p_{right} = P(F > f_{AB})$ with a significance level α. Similarly for F_A, F_B.

In R: the p-value is in `anova(lm(y~f1*f2))`, with `f1` and `f2` the two factors.

Balanced design: equal group size $n_{ij} = N$ for each i and j, thus $n = NIJ$.

Formula `y~f1*f2` is the same as `y~f1+f2+f1:f2`, meaning that the model includes μ (μ is always included by default), and all α_i's, β_j's and γ_{ij}'s.
If H_{AB} is not rejected (i.e., we concluded that all $\gamma_{ij} = 0$), then it is proper practice to test for main effects A and B under the additive model $\mu_{ij} = \mu + \alpha_i + \beta_j$ (in R: `y~f1+f2`). Otherwise, we proceed to test for main effects using the full model.

F-statistics in two-way ANOVA

The idea of the F-statistics is $F = \dfrac{\text{explained variance}}{\text{unexplained variance}} = \dfrac{\text{between-groups SS}}{\text{within-groups SS}}$.

Denote the total mean $\bar{Y}_{...} = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}$, and

$$\bar{Y}_{ij\cdot} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}, \quad \bar{Y}_{i\cdot\cdot} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}, \quad \bar{Y}_{\cdot j\cdot} = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}.$$

The test statistics are

$$F_{AB} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{...})^2/((I-1)(J-1))}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2/(n-IJ)},$$

$$F_{A} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{...})^2/(I-1)}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2/(n-IJ)},$$

similarly for F_B.

General form of ANOVA tables

One-way ANOVA results are usually presented in an one-way ANOVA table:

Source	Df	Sum Sq	Mean Sq	F value	p-value
Factor A	$I-1$	SS_A	$SS_A/(I-1)$	$F_A = \frac{SS_A/(I-1)}{RSS/(n-I)}$	$P_A(F_A > f)$
Residuals	$n-I$	RSS	$RSS/(n-I)$		
Total	$n-1$	SS_T			

Two-way ANOVA results are usually presented in a two-way ANOVA table:

Source	Df	Sum Sq	Mean Sq	F value	p-value
Factor A	$I-1$	SS_A	$SS_A/(I-1)$	$F_A = \frac{SS_A/(I-1)}{RSS/(n-IJ)}$	$P_A(F_A > f)$
Factor B	$J-1$	SS_B	$SS_B/(J-1)$	$F_B = \frac{SS_B/(J-1)}{RSS/(n-IJ)}$	$P_B(F_B > f)$
Interaction	$(I-1)(J-1)$	SS_{AB}	$SS_{AB}/(I-1)(J-1)$	$F_{AB} = \frac{SS_{AB}/[(I-1)(J-1)]}{RSS/(n-IJ)}$	$P_{AB}(F_{AB} > f)$
Residuals	$n-IJ$	RSS	$RSS/(n-IJ)$		
Total	$n-1$	SS_T			

$$SS_T = SS_A + SS_B + SS_{AB} + RSS = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{\cdots})^2.$$

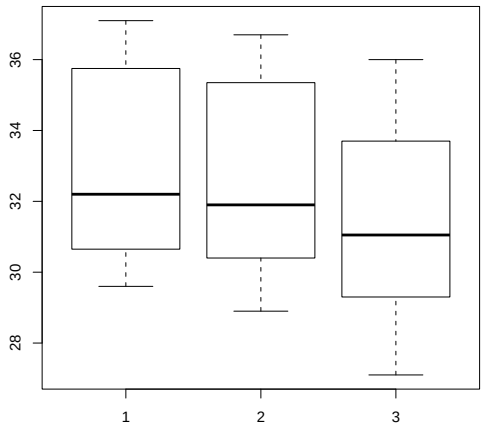
Example: pvc (1)

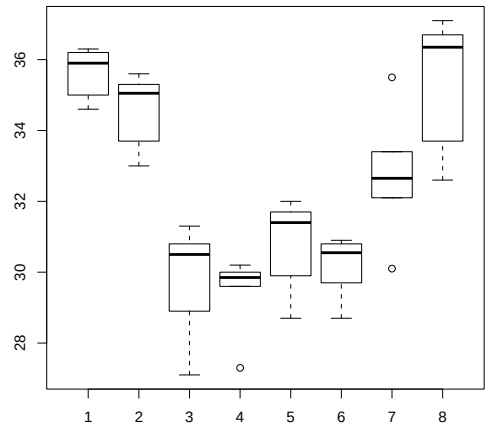
The following data is from an experiment to study factors affecting the production of the plastic PVC, 3 `operators` used 8 different devices called `resin` railcars to produce PVC, two samples for each of the 24 combinations.

```
> pvc=read.table(file="pvc.txt",header=TRUE)
> pvc[1:4,]
  psize operator resin
1  36.2        1     1
2  36.3        1     1
3  35.3        1     2
4  35.0        1     2
```

```
> attach(pvc)
> boxplot(psize~operator)
> boxplot(psize~resin)
```

These pictures give an idea of the main effects of the factors. Interactions are not visible.

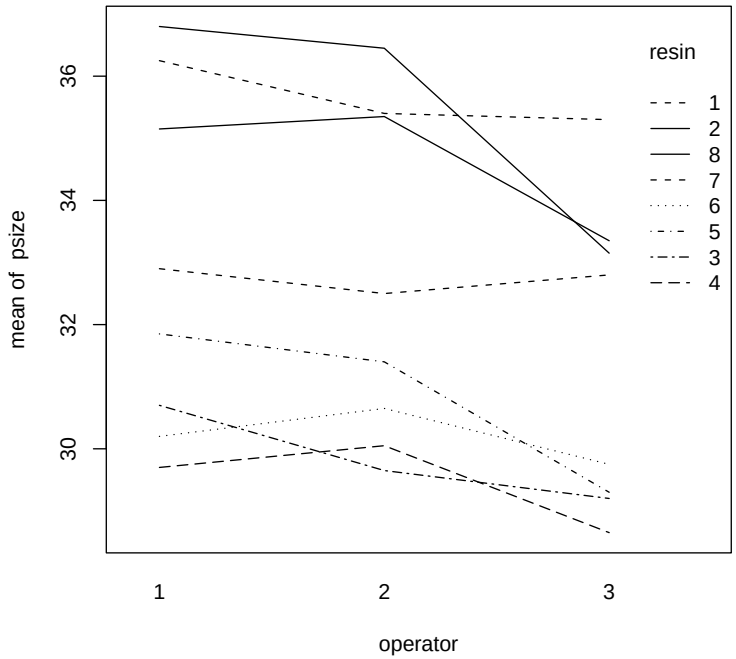


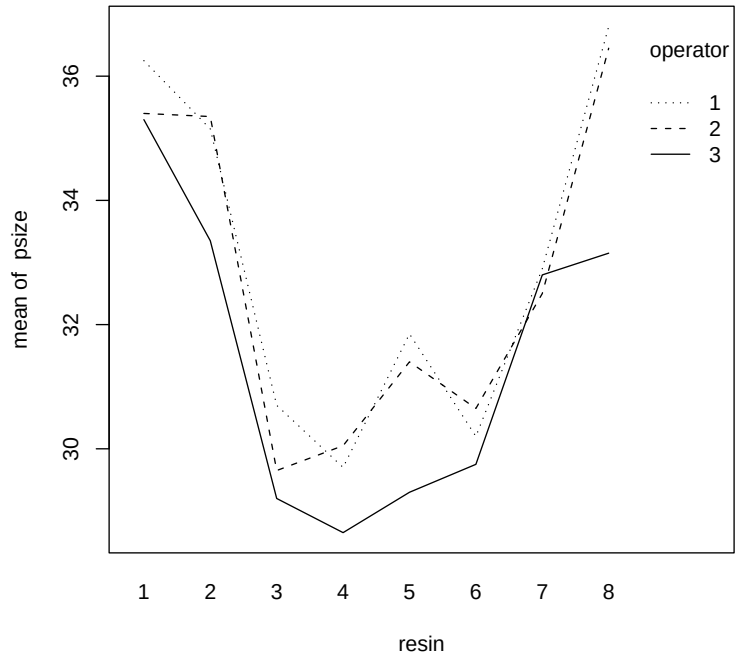


Example: pvc (2)

An interaction plot fixes one factor and plots the average outcome (vertical axis) against the levels of the other factor (horizontal axis). Interaction shows up as nonparallel curves.

```
> interaction.plot(operator,resin,psize)
> interaction.plot(resin,operator,psize)
```





Lines may be unparallel, because of interactions, but also because of noise in the data.

Example: pvc (3)

```
> pvc$operator=as.factor(pvc$operator); pvc$resin=as.factor(pvc$resin)
> pvcaov=lm(psize~operator*resin); anova(pvcaov)
[ some output deleted ]
Response: psize
              Df  Sum Sq Mean Sq F value    Pr(>F)
operator       2  20.718  10.359  7.0072   0.00401 **
resin          7 283.946  40.564 27.4388 5.661e-10 ***
operator:resin 14  14.335   1.024  0.6926   0.75987
Residuals      24  35.480   1.478
```

The p-value for testing $H_0 : \alpha_i = 0$ for all i is 0.00401; for $H_0 : \beta_j = 0$ for all j is `5.661e-10`; for $H_0 : \gamma_{i,j} = 0$ for all (i, j) is 0.75987. So, there is no evidence for interaction (both factors seems to have a main effect but one should not draw conclusions about the factors at this stage).

The command `as.factor` (or `factor`) is necessary, because the 2nd and 3rd columns of the data matrix were read in as numerical variables (with values 1, 2, 3, 4), but should be treated as factors in the analysis.

Example: pvc (4)

```
> summary(pvcaov) # estimates in the default treatment contrasts
[ some output deleted ]
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       36.2500     0.8598  42.164  < 2e-16 ***
operator2         -0.8500     1.2159  -0.699 0.491216
operator3         -0.9500     1.2159  -0.781 0.442245
resin2            -1.1000     1.2159  -0.905 0.374615
[ some output deleted ]
resin8             0.5500     1.2159   0.452 0.655078
operator2:resin2   1.0500     1.7195   0.611 0.547175
[ some output deleted ]
operator3:resin8  -2.7000     1.7195  -1.570 0.129454
```

The output of `summary(pvcaov)` shows estimates of $\mu, \alpha_2, \alpha_3, \beta_2, \beta_2, \ldots, \beta_8, \gamma_{22}, \ldots, \gamma_{38}$ in the default treatment parametrization: $\alpha_1 = \beta_1 = \gamma_{1j} = \gamma_{i1} = 0$, $i = 1, 2, 3$, $j = 1 \ldots, 8$. The corresponding estimates $\hat{\alpha}_1 = \hat{\beta}_1 = \hat{\gamma}_{11} = \ldots = \hat{\gamma}_{31} = 0$ are not shown. The p-values in column `Pr(>|t|)` are for testing the individual null hypothesis that the coefficient is 0. The test statistic, computed as $\texttt{t value} = \frac{\texttt{Estimate}}{\texttt{Std. Error}}$, has t_{n-IJ}-distribution under H_0.

Example: pvc (5)

The command `contrasts` overrules the default treatment parametrization (e.g., to sum parameterization), `lm` and anova have to be run again.

```
> contrasts(pvc$operator)=contr.sum; contrasts(pvc$resin)=contr.sum
> pvcaov2=lm(psize~operator*resin,data=pvc); summary(pvcaov2)
[ some output deleted ]
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      32.35417    0.17550 184.359  < 2e-16 ***
operator1         0.58958    0.24819   2.376 0.025855 *
operator2         0.32708    0.24819   1.318 0.199983
resin1            3.29583    0.46432   7.098 2.45e-07 ***
[ some output deleted ]
resin7            0.37917    0.46432   0.817 0.422183
operator1:resin1  0.01042    0.65664   0.016 0.987474
[ some output deleted ]
operator2:resin7 -0.56042    0.65664  -0.853 0.401844
```

The output shows estimates of $\mu, \alpha_1, \alpha_2, \beta_1, \ldots, \beta_7, \gamma_{11}, \gamma_{12}, \ldots$ in the sum parametrization. The estimates of α_3 (operator 3) and β_8 (resin 8) are not shown. These can be found from the restrictions $\sum_{i=1}^{3} \hat{\alpha}_i = 0$, $\sum_{j=1}^{8} \hat{\beta}_j = 0$; similarly for the interactions: $\sum_{i=1}^{3} \hat{\gamma}_{ij} = 0$ for $j = 1, \ldots, 8$ and $\sum_{j=1}^{8} \hat{\gamma}_{ij} = 0$ for $i = 1, 2, 3$. The p-values in `Pr(>|t|)` are for testing individual hypothesis H_0: `coefficient=0`.

Example: pvc (6)

As we see, the previous analysis says there are no interactions. Now we remove interaction term from the model and fit the additive model

$$\mu_{ij} = \mu + \alpha_i + \beta_j, \quad i = 1, \ldots, I, \;\; j = i \ldots, J.$$

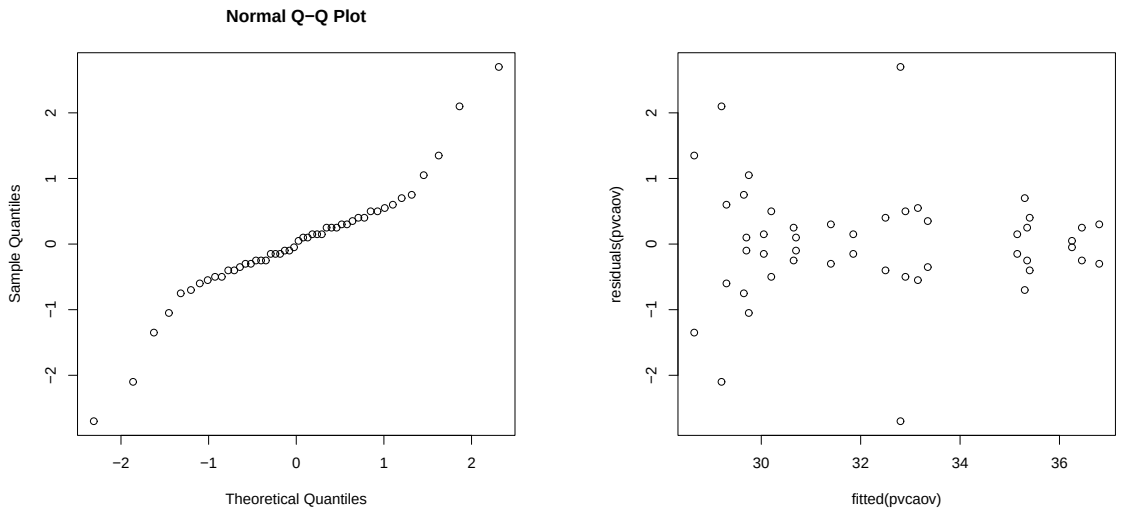
```
> pvc$operator=as.factor(pvc$operator)
> pvc$resin=as.factor(pvc$resin)
> pvcaov=lm(psize~operator+resin,data=pvc)
> anova(pvcaov)
[ some output deleted ]
Response: psize
                Df  Sum Sq   Mean Sq  F value    Pr(>F)
operator         2    20.718  10.359     7.902     0.00135 **
resin            7   283.946  40.564    30.943     8.111e-14 ***
Residuals       38    49.815   1.311
```

The p-value for testing $H_A : \alpha_i = 0$ for all i is 0.00135; for $H_B : \beta_j = 0$ for all j is $8.111e-14$. So both factors have a main effect in the additive model as well.

Example pvc (7)

We check the normality and the assumption of equal variances. The residuals $\hat{e}_{ijk} = Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}$ are the data corrected for the different population means and ought to look normal. The fitted value $\hat{Y}_{ijk}$ for Y_{ijk} is the estimated mean $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}$. The spread in the residuals should not change systematically with any variable, in particular not with the fitted values.

```
> qqnorm(residuals(pvcaov2)); plot(fitted(pvcaov2),residuals(pvcaov2))
```



Left plot: normality is doubtful. Right plot: the spread in the residuals seems to be bigger for smaller fitted values. Some data-points also seem extreme. Perhaps transform the data or consider outliers.

One observation per cell (1)

The following dataset contains the strength of a thermoplastic composite depending on power of a laser and speed of a tape.

```
> composite=read.table("composite.txt",head=T); composite
  strength laser   tape
1    25.66   40W   slow
2    29.15   50W   slow
3    35.73   60W   slow
4    28.00   40W medium
5    35.09   50W medium
6    39.56   60W medium
7    20.65   40W   fast
8    29.79   50W   fast
9    35.66   60W   fast
```

Notice that we have only one observation per cell (i.e., per each combination of levels of the two factors `laser` and `tape`). But then there is a problem in the test statistics F for intercations: since $n_{ij} = 1$, $n = IJ$ and the denominator $RSS/(n - IJ)$ is not well defined. To estimate and test interaction effects, it is necessary to have at least 2 observations per combination (i, j) of factor levels.

One observation per cell (2)

R produces a warning message if the data is not sufficient to fit the model, in this case it is impossible to estimate interactions with one observation per cell:

```
> attach(composite); anova(lm(strength~laser*tape))
           Df  Sum Sq Mean Sq F value Pr(>F)
laser       2 224.184 112.092
tape        2  48.919  24.459
laser:tape  4  10.503   2.626
Residuals   0   0.000
Warning message:
In anova.lm(lm(strength ~ laser * tape, data = composite)) :
  ANOVA F-tests on an essentially perfect fit are unreliable
```

If it can be assumed a priori that all interactions are 0, then it is possible to test and estimate main effects. (Interaction plots may help to justify this assumption.)

```
> anova(lm(strength~laser+tape,data=composite))
          Df  Sum Sq Mean Sq F value   Pr(>F)
laser      2 224.184 112.092 42.6893 0.002003 **
tape       2  48.919  24.459  9.3151 0.031242 *
Residuals  4  10.503   2.626
```

randomized block design

Setting

An experiment with:

- a numerical outcome Y ("dependent variable"),
- a factor of interest that can be fixed at I *levels* ("treatment"),
- a factor that is *not* of interest that can be fixed at B *levels* ("block").

The purpose is to understand the dependence of Y on the treatment factor. The block variable is thought (or known) to be of influence. It is used to create homogeneous groups of experimental units, in which the treatment effect is easier to see and not blurred by variation due to the block factor.

EXAMPLE Chemical production process with outcome total yield, treatment variable temperature fixed at levels low, medium and high and block blend of raw material.

EXAMPLE Study of web design with outcome total time of a user on webpage, treatment variable type of design and block user skill. Each user is tested with a single type of web design.

Design

Independently, for $b = 1, 2, \ldots, B$:

- select NI experimental units randomly from the population of units with block level b,
- assign level i of the factor to a random set of N units ($i = 1, 2, \ldots, I$),
- perform the experiment NI times, independently.

Randomization in R.

```
> I=4; B=5; N=1
> for (i in 1:B) print(sample(1:(N*I)))
[1] 3 1 2 4
[1] 4 3 2 1
[1] 1 4 2 3
[1] 3 4 1 2
[1] 2 4 3 1
```

For block 1 assign unit 3 to treatment 1, unit 1 to treatment 2, etc., for block 2 assign unit 4 to treatment 1, unit 3 to treatment 2, etc.

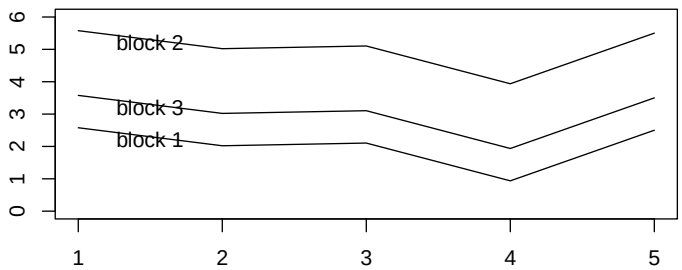
Given many blocks, it is typical to use one replicate per treatment level per block: $N = 1$.

Analysis

Data (Y_{ibk}) are assumed to follow the model

$$Y_{ibk} = \mu + \alpha_i + \beta_b + e_{ibk}, \qquad i = 1, \ldots, I; \quad b = 1, \ldots, B; \quad k = 1, \ldots, N,$$

where the "errors" (e_{ibk}) are a random sample from a normal population.



The pattern $(\alpha_1, \alpha_2, \ldots, \alpha_I)$ of treatment effects is assumed to be the same within every block.

We test the null hypothesis $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$.
We also estimate the treatment effects $\alpha_1, \alpha_2, \ldots, \alpha_I$.

The model is the same as in a two-way factorial experiment, with the block as a second factor, but with zero interactions.

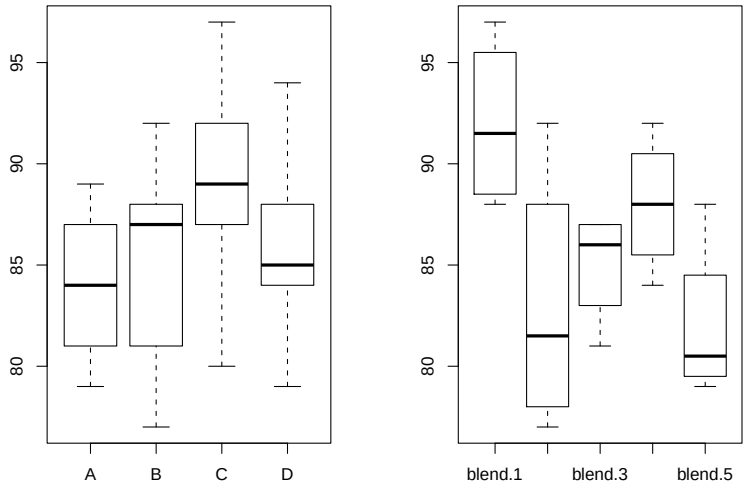
Analysis in R: data input

The following data frame contains the data about penicllin made by production processes A, B, C, D (treatment); with 5 different blends of raw material (blocks), as in a two-way factorial experiment.

```
> penicillin
   treat   blend yield
1      A blend.1    89
2      B blend.1    88
3      C blend.1    97
4      D blend.1    94
5      A blend.2    84
[ some output deleted ]
20     D blend.5    88
> xtabs(yield~treat+blend,data=penicillin)
     blend
treat blend.1 blend.2 blend.3 blend.4 blend.5
    A      89      84      81      87      79
    B      88      77      87      92      81
    C      97      92      87      89      80
    D      94      79      85      84      88
```

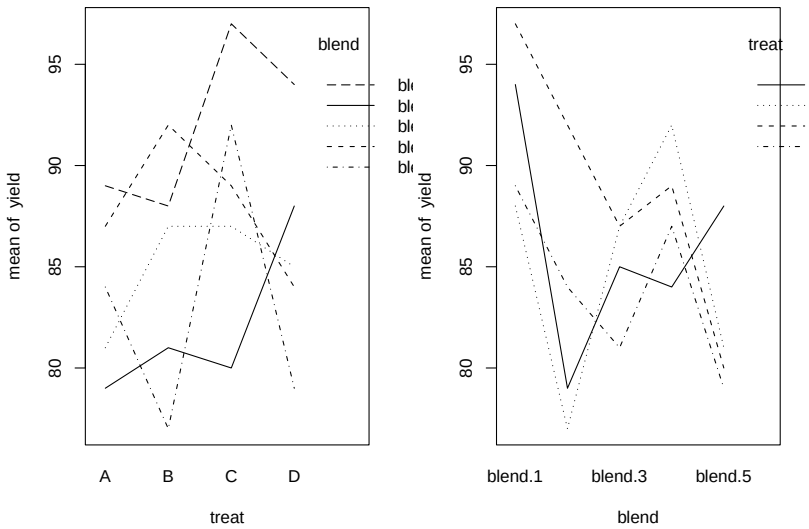
Analysis in R: graphics

```
> attach(penicillin)
> par(mfrow=c(1,2))
> boxplot(yield~treat)
> boxplot(yield~blend)
```




```
> par(mfrow=c(1,2))
> interaction.plot(treat,blend,yield)
> interaction.plot(blend,treat,yield)
```

The left plot gives estimates of the treatment patterns per block.



Analysis in R: testing and estimation

```
> aovpen=lm(yield~treat+blend)
> anova(aovpen)
Response: yield
          Df Sum Sq Mean Sq F value  Pr(>F)
treat      3     70  23.333  1.2389 0.33866
blend      4    264  66.000  3.5044 0.04075 *
Residuals 12    226  18.833
```

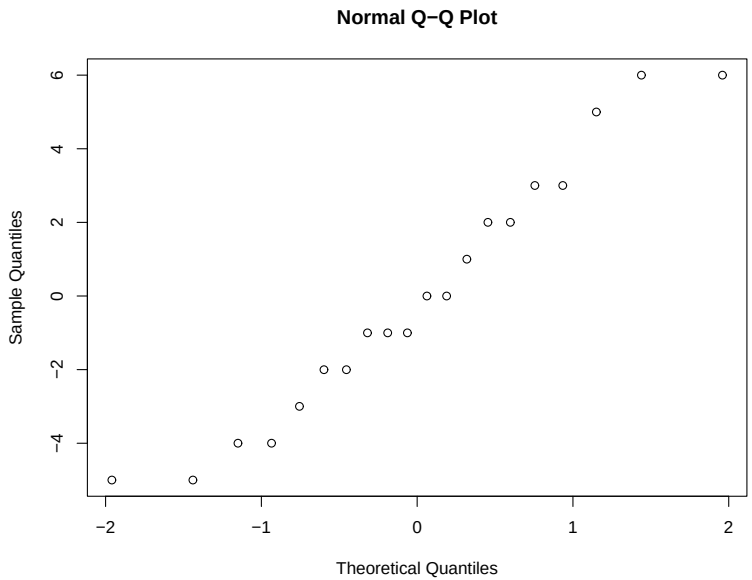
The treatment effects are not significantly different from 0. The blocks (`blend`) are, but this was not the research question.

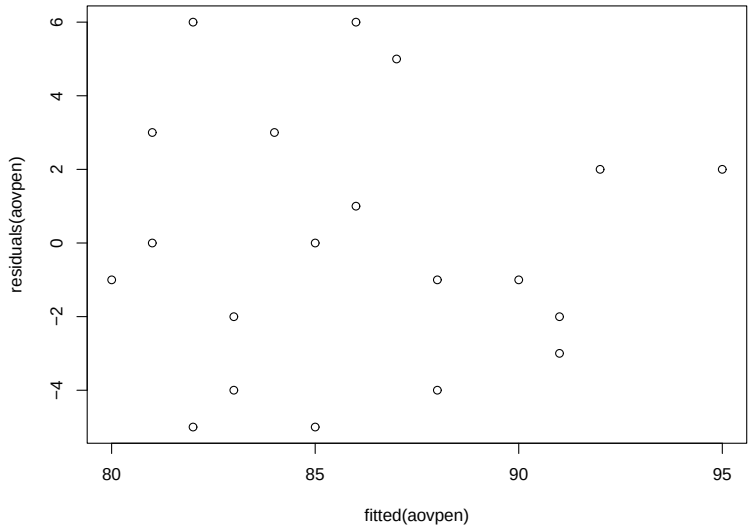
```
> summary(aovpen)
[ some output deleted ]
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    90.000      2.745  32.791  4.1e-13 ***
treatB          1.000      2.745   0.364  0.72194
treatC          5.000      2.745   1.822  0.09351 .
treatD          2.000      2.745   0.729  0.48018
blendblend.2   -9.000      3.069  -2.933  0.01254 *
blendblend.3   -7.000      3.069  -2.281  0.04159 *
blendblend.4   -4.000      3.069  -1.304  0.21686
blendblend.5  -10.000      3.069  -3.259  0.00684 **
```

The yield of treatment C is estimated 5 higher than that of treatment A, etc.

Analysis in R: diagnostics

```
> qqnorm(residuals(aovpen))
> plot(fitted(aovpen),residuals(aovpen))
```





Look OK; perhaps a slight curve in the qq-plot. The interaction plots (see some slides back) can also be considered diagnostic.

Discussion

- The advantage of the block design is that more precise conclusions can be obtained by removing variation, present due to block factor. The units must be similar within the blocks, and dissimilar between the blocks.
- Assuming that *the pattern of treatment effects is the same for each block* means assuming the absence of interaction between block and treatment. Without replications ($N = 1$), this cannot be tested, with $N > 1$ it can.
- If treatment and blocks do interact, the interpretation of the results of a factorial analysis is more subtle.
- Multiple treatment factors: a multi-way factorial experiment can be done within every block (rather than a one factor experiment).
- Multiple block factors: all combinations of levels of the block factors can be viewed as a new, single block factor, to which the block design applies.

repeated measures

Setting and design

Setting: an experiment with

- a numerical outcome Y ("dependent variable"),
- a factor of interest that can be fixed at I levels, ("treatment").
- experimental units that are measured at every treatment level.

The purpose is to understand the dependence of Y on the treatment factor.
The same experimental units are used for every treatment, because this is thought to reduce "extraneous variation": the units serve as blocks.
For $I = 2$ treatments, this is simply the paired sample design.

EXAMPLE Study of web design with outcome total time on webpage, treatment variable type of design. Each user is tested with every type of design.

EXAMPLE The velocity of a ball is measured for different types of tennis rackets for a number of players, where every player uses all types of rackets.

Design:

- Select B experimental units randomly from a population of units.
- Measure each unit at every treatment level, if possible in random order.

Exchangeable case

Data vectors $(Y_{1b}, Y_{2b}, \ldots, Y_{Ib})$ for B units are assumed to follow the model

$$Y_{ib} = \mu + \alpha_i + \beta_b + e_{ib}, \qquad i = 1, \ldots, I; \quad b = 1, \ldots, B,$$

- the "error vectors" $(e_{1b}, \ldots, e_{Ib})$ for the B units are a random sample from a (multivariate) normal distribution.
- the "errors" $e_{1b}, \ldots, e_{Ib}$ within a single unit are exchangeable (i.e., the ordering is irrelevant, in a way, generalizing the pared samples).
- the effects $\beta_1, \ldots, \beta_B$ of the units may be considered fixed or random.

The pattern $(\alpha_1, \ldots, \alpha_I)$ of treatments is assumed to be the same for each unit.

We want to test the null hypothesis $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$.
We also want to estimate the treatment effects $\alpha_1, \alpha_2, \ldots, \alpha_I$.

The model is the same as in a randomized block experiment, with the units as blocks, except for the assumption on the errors. These are allowed to be dependent within the units, even though still "exchangeable".

Analysis in R: data input

Data input is as in a block design, with columns for outcome, treatment level, and block level (=identification of unit).

```
> ashinalong
   pain id order treatment
1  -167  1    pa         a
2  -102  1    pa         p
3  -127  2    pa         a
4   -39  2    pa         p
5   -58  3    pa         a
6    32  3    pa         p
7  -103  4    pa         a
8    28  4    pa         p
[ some output deleted ]
31  -72 16    ap         a
32  -36 16    ap         p
```

The data frame `ashinalong` contains the same data as `ashina`, but every individual is represented by two lines, one for the treatment with the active drug, the other for the placebo. The extra column `id` shows the pairing of the measurements.

Analysis in R: exchangeable case

Analysis is as for a randomized block design, with every unit being a block.

```
> ashinalong$id=factor(ashinalong$id)
> aovashina=lm(pain~treatment+id,data=ashinalong); anova(aovashina)
Analysis of Variance Table
Response: pain
          Df Sum Sq Mean Sq F value   Pr(>F)
treatment  1  14706 14706.1  10.413 0.005644 **
id        15  51137  3409.2   2.414 0.049184 *
Residuals 15  21184  1412.3
```

Compare to the two sample t-test:

```
> t.test(ashina[,1],ashina[,2],paired=TRUE)
        Paired t-test
data:  ashina[, 1] and ashina[, 2]
t = -3.2269, df = 15, p-value = 0.005644
```

The p-value for treatment is identical to the one of the paired-sample t-test found previously (the order of the treatments was ignored). The p-value for `id` is not interesting. Note that R had to be told to treat `id` as labels, not as numbers.

Lack of exchangeability

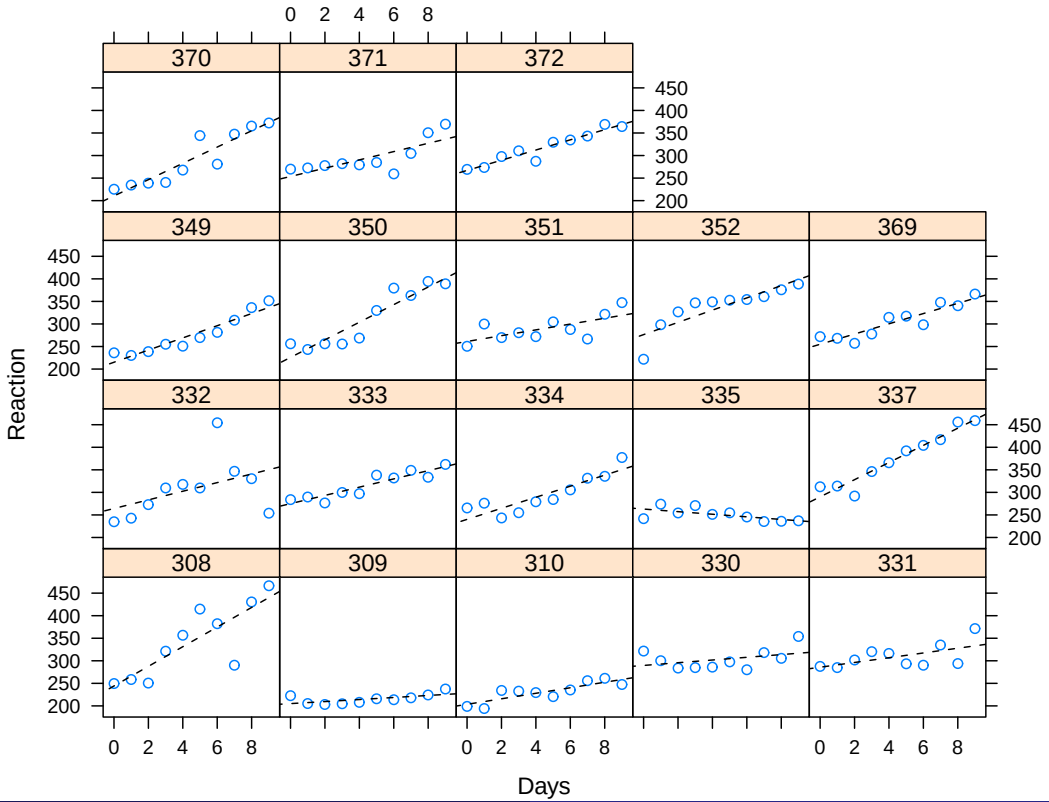
Repeated measures may not be exchangeable.

- Time effect: growth, increasing or decreasing variation.
- Learning effect: subject becomes better or bored at tasks (cf. crossover design).
- Dissimilar subjects: the pattern of response to treatment varies too much.

A block design analysis is then invalid.

Example Truck drivers were deprived of sleep and their reaction time was measured daily over a period of 10 days. Each panel gives the reaction times of one truck driver with best fitting line overlaid.

Drivers follow different lines, which contradicts the model assumptions.



Discussion

- The discussed repeated measures design corrects for some dependencies.
- Taking repeated measures is attractive, because fewer experimental units are needed and "extraneous" variation between units is reduced.
- However, in many studies, in particular most "longitudinal studies", where individuals are followed over time, the assumption of "exchangeability" fails. More complicated models are then necessary.
- Models with random effects (called mixed effects models) are a possibility.

Friedman test

Setting and design

Setting and design for the Friedman test are either as in a randomized block design with $N = 1$ or as in repeated measures. An experiment with:

- a numerical outcome Y ("dependent variable").
- a factor of interest that can be fixed at I *levels.* ("treatment").
- a number of blocks or units that are measured at *every* treatment level.

Data

	block1	block2	...	blockB
level 1:	Y_{11},	Y_{12},	...,	Y_{1B}
level 2:	Y_{21},	Y_{22},	...,	Y_{2B}
⋮				
level I:	Y_{I1},	Y_{I2},	...,	Y_{IB}

Data (Y_{ib}) are not assumed to come from a normal distribution.

We want to test the null hypothesis of no treatment effect taking the blocks into account, by using ranks.

The underlying idea of this test: the Friedman test computes the ranks of the i-th measurement within each block. Under H_0 the rank of Y_{ib} should lie randomly between 1 and I for each b. If the average rank of Y_{ib} (averaged over blocks) is lower/higher than expected, this indicates that H_0 might not be true.

The sign test (two-sided) is equivalent to a Friedman test on two groups.

Analysis in R: data input

The dataset `itch.tx` contains the numbers of hours subjects were itching after treatment with 7 different drugs (incl. `No_Drag` and `Placebo`) against itching.

```
> itch=read.table("itch.txt",header=TRUE,sep=","); itch
```

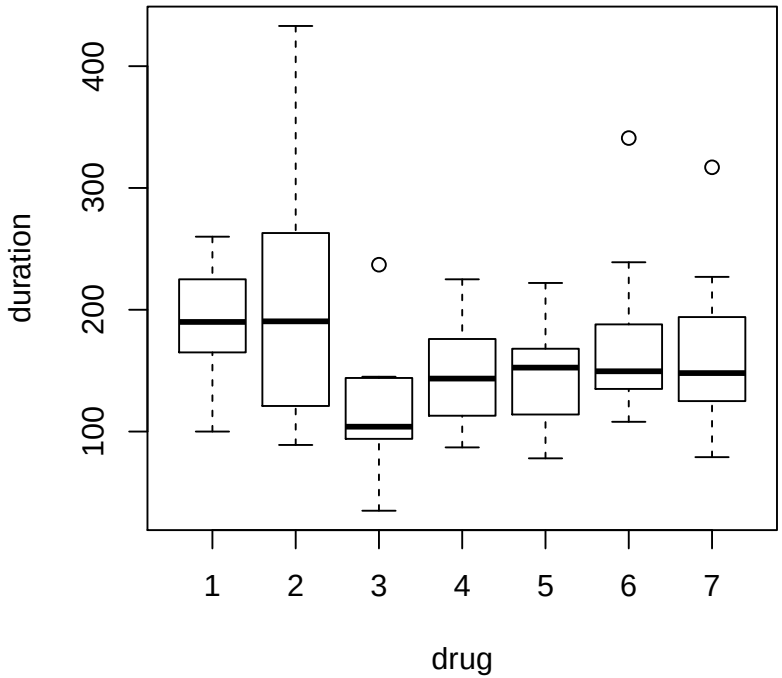
	Subject	No_Drug	Placebo	Papaverine	Morphine	Aminophylline	Pentabarbital	Tripelennamine
1	BG	174	263	105	199	141	108	141
2	JF	224	213	103	143	168	341	184
3	BS	260	231	145	113	78	159	125
4	SI	225	291	103	225	164	135	227
5	BW	165	168	144	176	127	239	194
6	TS	237	121	94	144	114	136	155
7	GM	191	137	35	87	96	140	121
8	SS	100	102	133	120	222	134	129
9	MU	115	89	83	100	165	185	79
10	OS	189	433	237	173	168	188	317

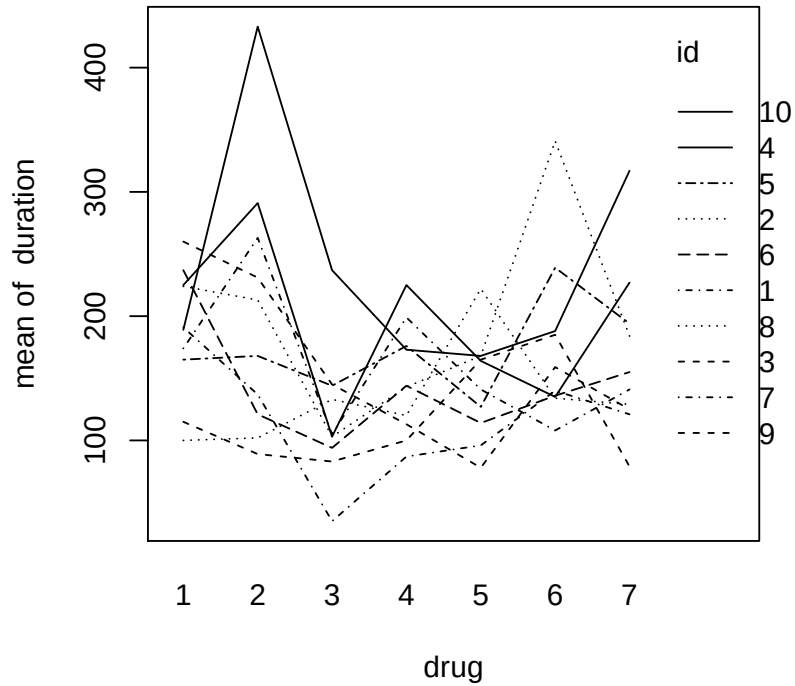
Create a data frame with `duration` as 1st, `id` as 2d, and `drug` as 3d columns.

```
> duration=as.vector(as.matrix(itch[,2:8]))
> id=as.factor(rep(1:10,7)); drug=as.factor(rep(1:7,each=10))
> itchdata=data.frame(cbind(duration,id,drug)); itchdata[1:3,]
   duration id drug
1       174  1    1
2       224  2    1
3       260  3    1
```

Analysis in R: graphics

```
> boxplot(duration~drug,xlab="drug",ylab="duration")
> interaction.plot(drug,id,duration)
```





Parallel lines in the interaction plot indicate that there is no significant interaction effect. But beware that we're dealing with $N = 1$.

Analysis in R: testing (1)

```
> friedman.test(duration,drug,id,data=itchdata)

        Friedman rank sum test

data:  duration, drug and subject
Friedman chi-squared = 14.2796, df = 6, p-value = 0.02666
```

Command `friedman.test(duration,drug,id,data=itchdata)` performs the Friedman test, testing the relevance of factor `drug` taking into account the blocking factor `id`. The p-value for testing (H_0 : no treatment effect) is 0.02666, so H_0 is rejected, there is a treatment effect.

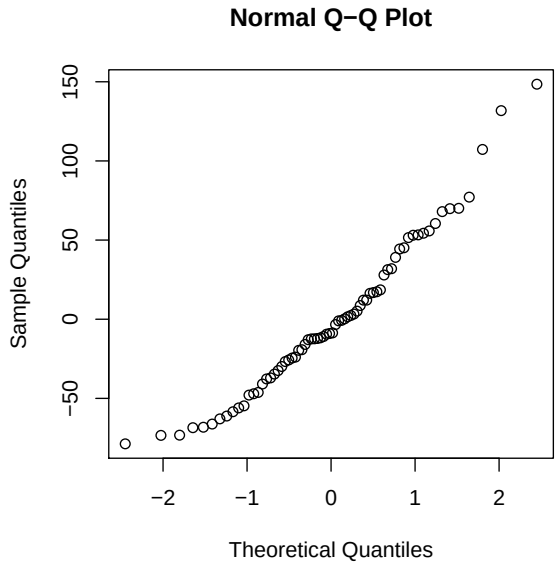
Analysis in R: testing (2)

Compare the Friedman test results to results for the repeated measures design:

```
> itchaov=lm(duration~drug+subject); anova(itchaov)
Analysis of Variance Table
Response: duration
          Df Sum Sq Mean Sq F value   Pr(>F)
drug       6  51487  8581.2  2.7893 0.019494 *
subject    9 101253 11250.3  3.6569 0.001261 **
Residuals 54 166127  3076.4
```

```
> qqnorm(itchaov$residuals)
```

In a randomized block design we also find a significant treatment effect. The QQ-plot looks ok, perhaps slightly bowed.



To finish

Today we discussed:

1. 2-way ANOVA
2. randomized block design
3. repeated measures
4. Friedman test

Next time: general factorial and incomplete block designs, random effects, more block designs.