

Some of the basic methods used in inferential statistics

LINEAR MODELS

A. Linear regression

- `lm()`
- basic case: one continuous independent variable (e.g. age) predicting the other continuous dependent variable (e.g. voice pitch)
- how much change is caused in the dependent variable by an amount of change in the predictor variable?
- BUT keep in mind: basically all the models below (*t*-test, ANOVA) which test the effect of one categorical dependent variable are linear regression models as well, in which you compare differences between group means (expressed in slopes) to 0 (different → difference between groups)
- for this reason each test may be executed in an ANOVA approach (the model is the group means compared to the grand mean) and in a linear regression approach (with intercept(s) and regression slopes)
- A general note to ANOVA models:
 - Type I sums of squares: the order of predictors entered into the model matters (effects evaluated after each other)
 - Type III sums of squares (default in `aov()`): the order of predictors does not matter (forced entry: all effects evaluated at the same time) → to do this: `model = aov(libido ~ partnerLibido + dose)` and then run `Anova(model, Type = "III")` (`Anova()` is in `car` package)
 - Type III sums of squares: main effects evaluated first (interactions second) – requires orthogonal contrasts!

B. Parametric tests for comparing groups (group means)

- for comparing data from normally distributed sampling frequencies (tested on distribution of scores) or more than minimally 30 items in a sample (from which we can assume normal distribution of sampling frequencies) – watch out! *not always groups scores must be normally distributed* (see e.g., paired sample *t*-test where the *differences* of scores must be normally distributed)
- in metrical data (which has a dimension, e.g. Hz, ms, °C), minimally interval scale
- **these all are actually linear regression models**, so they can be modeled as an ANOVA (e.g. `aov()`) or a linear regression model (`lm()`) with intercept being one group mean (the baseline) and the slopes (*b*) being the *differences* between group means

A guide to how ANOVAs (ANalysis Of VAriance) are named:

I.

<i>n</i> -way	<i>independent/repeated measures/mixed</i>	ANOVA
number of independent var.s	design (how variables are measured)	"ANOVA"
	between-group/within-group/mixed design	

II.

univariate	/	multivariate	ANOVA
ANOVA	/	MANOVA	
one measured outcome (dependent) var.	/	more dependent var.s(=MANOVA)	

Effect size measures – shows how much variance in the data can be accounted for by the independent variables we included in the model:

r^2 : see below at Pearson's r

η^2 : (basically r^2): the proportion of total variance explained by an effect (effect / total variance)

partial η^2 : it looks at the proportion of variance that a variable explains that is not explained by other variables in the analysis (effect / effect+ residual)

ω^2 : only if we have equal number of participants in each group

I. Independent samples

- 1 group and e.g., an assumed population mean (or zero) ("is the mean different from...?")
 - **one-sample t-test** (t, p)¹ : `t.test()`
- 2 groups: 1 predictor variable with 2 levels and 1 outcome variable:
 - **t-test** for comparing 2 group means (t, p) : `t.test()`
Assumptions:
 1. independence
 2. normal distribution of the group scores (actually normal distribution of sampling distribution!): **Shapiro-Wilk test** : `shapiro.test()`
 3. homogeneity of variances: **Levene's test** → only matters if you have unequal group sizes : car package : `levene.test()`
violation of the *assumption of homogeneity of variances* (significant result) → use **Welch's test** (parametric but without this assumption)
violation of the *assumption of normality* (significant result) → go to the non-parametric tests!
 - **one-way independent design analysis of variance (ANOVA or general linear model 1, a.k.a. GLM 1)**: for comparing 2 or more group means (actually generalized t-test, and both of them are actually cases of linear regression) ($F(df1, df2), p$)²
Assumptions:
 1. independence
 2. normal distribution of the group scores: **Shapiro-Wilk test**
 3. homogeneity of variances: **Levene's test**
violation of assumptions (significant result) → go to the non-parametric tests!
- 2 or more groups case #1: 1 predictor variable with 2 or more levels; and 1 outcome variable: **one-way independent design analysis of variance (ANOVA or GLM 1)** ($F(df1, df2), p$, preferably also effects size: ω^2)

¹ After each test name parameters to be reported in the publication are listed in brackets (in the order as they should be reported).

² df = degrees of freedom

Assumptions:

1. independence
2. normal distribution of the group scores: **Shapiro-Wilk test**
3. homogeneity of variances: **Levene's test**

violation of assumptions → go to the non-parametric tests!

- *2 or more groups case #2: 1 predictor variable with 2 or more levels; 1 outcome variable; and 1 covariate: **ANCOVA (or GLM 2)** ($F, (df01, df2), p, \text{partial } \eta^2$)*
 - to adjust eg., for the effect of age (covariate)
 - if the covariate and the independent variables are independent

Additional assumption (compared to ANOVA):

- homogeneity of regression slopes (i.e., no interaction between covariate and predictor)

- *2 or more groups case #3: more than 1 (n) predictor variable with 2 or more levels, and 1 outcome variable: **n-way or independent factorial design ANOVA (GLM 3)**: several independent variables or predictors and each has been measured using different entities (between groups) – outcome: main effects and interaction of predictors ($F(df1, df2), p, \omega^2$)*
`gogglesModel<-aov(attractiveness ~ gender + alcohol + gender:alcohol, data = gogglesData)`
OR
`gogglesModel<-aov(attractiveness ~ alcohol*gender, data = gogglesData)` # this includes interaction as well!

Assumptions:

1. independence
2. normal distribution of the group scores: **Shapiro-Wilk test**
3. homogeneity of variances: **Levene's test**

violation of assumptions (significant result) → go to the non-parametric tests!

Post hoc tests:

4. **Bonferroni**
5. **Tukey HSD**

- *2 or more groups case #4: 1 predictor variable with 2 or more levels; and more than 1 outcome variable: **one-way multivariate independent design MANOVA***
informs us of whether groups of participants can be distinguished by a combination of scores on several dependent measures
 - Words of warning: "In circumstances where there is a good theoretical basis for including some but not all of your dependent variables, you should run separate analyses: one for the variables being tested on a heuristic basis and one for the theoretically meaningful variables. The point to take on board here is not to include lots of dependent variables in a MANOVA just because you have measured them."

- e.g, we could compare a group of OCD sufferers after CBT and after behaviour therapy (BT) with a group of OCD sufferers who are still awaiting treatment (a no-treatment condition, NT).
- idea: we want to calculate underlying linear dimensions of the dependent variables. These linear combinations of the dependent variables are known as *variates* (or sometimes called *latent variables* or *factors*). In this context we wish to use these linear variates to predict which group a person belongs to (i.e., whether they were given CBT, BT or no treatment), so we are using them to discriminate groups of people. Therefore, these variates are called *discriminant functions* or **discriminant function variates**.

Assumptions

1. independence
 2. random sampling
 3. multivariate normality (we assume that the dependent variables (collectively) have multivariate normality within groups) **mshapiro.test()** in *mvnormtest* package
 1. homogeneity of covariance matrices (assumption of homogeneity of variances AND that the correlation between any two dependent variables is the same in all groups) *by(ocdData[, 2:3], ocdData\$Group, cov)*
 2. MANOVA is not robust for unequal sample sizes: if the larger group provides higher variance-covariance the *p* value should not be trusted!

violation of assumptions → go to the robust MANOVA!
- Main model:
 1. `ocdModel<-manova(outcome ~ Group, data = ocdData)` where
 2. `outcome<-cbind(ocdData$Actions, ocdData$Thoughts)`
 - Follow up analysis
 1. multiple univariate ANOVAs (with maybe **Bonferroni** correction) followed by
 2. **contrasts**
OR rather (also!) do
 3. **discriminant function analysis** `newModel<-lda(Group ~ Predictor(s), data = dataFrame, prior = prior probabilities, na.action = "na.omit")` where *Groups* is the groups you try to discriminate and *Predictors* are the continuous variable(s), e.g.: `ocdDFA<-lda(Group ~ Actions + Thoughts, data = ocdData)`

II. Dependent samples

- 2 groups: 1 predictor variable with 2 levels and 1 outcome variable
 - **dependent samples t-test**

Assumptions:

- a. normal distribution of the group score *differences* (actually normal distribution of sampling distribution *differences*!): **Shapiro-Wilk test** for score differences
- b. homogeneity of variances: **Levene's test** → only matters if you have unequal group sizes

violation of assumptions (significant result) → go to the non-parametric tests!

- **one-way repeated measures ANOVA (GLM 4)**

see below...

- 2 or more groups case #1: 1 predictor variable with 2 or more levels; and 1 outcome variable: **one-way repeated-measures ANOVA (GLM 4)** (ANOVA approach: $F(df1, df2)$, p , η^2 or r for planned comparisons, Mauchly's: (χ^2, p) corrections: (ϵ) ; multilevel approach: (χ^2, p))

ANOVA approach: `ezANOVA()` (similar output to SPSS, Mauchly included, corrections included) or `aov()`

linear regression approach: `lm()`, `lme()` → preferred to ANOVA approach, as these work with linear regression, and you don't have to be worried about Mauchly's!

reporting:

multilevel model: $(\chi^2(df), p)$

traditional ANOVA approach (e.g., `ezANOVA()`): $(F(df1, df2), p)$

1. linear regression (multilevel) approach:

GENERALLY: `newModel <- lme(outcome ~ predictor(s), random = random effects, data = dataframe, method = "ML")` where ML: maximum-likelihood

EXAMPLE:

```
baseline<-lme(Retch ~ 1, random = ~1 | Participant/Animal, data = longBush,
method = "ML")
```

```
bushModel<-lme(Retch ~ Animal, random = ~1 | Participant/Animal, data = long-
Bush, method = "ML")
```

and then:

```
anova(baseline, bushModel)
```

Further:

```
summary(bushModel)
```

2. ANOVA approach:

```
bushModel<-aov(Retch ~ Animal + Error(Participant/Animal), data = longBush)
```

Assumptions (if not treated as *lme* regression but an ANOVA as in `ezANOVA()`):

- a. instead of homogeneity of variances we have the assumption of *sphericity* (testable if we have minimum 3 groups): if the difference of the treatment levels is equal in variance **Mauchly's test** (non-significant → sphericity is met)

What to do if

- b. sphericity is violated:
 - i. when estimates of sphericity are greater than .75 the **Huynh–Feldt correction** should be used
 - ii. otherwise: **Greenhouse-Geisser correction** should be reported
- c. sphericity is NOT violated:
 - i. just check the main ANOVA results

Post hoc tests

not needed if contrast are used as in regression, but in the ANOVA approach:

- d. if sphericity is NOT violated: **Tukey**
- e. if sphericity is violated: **Bonferroni**

- *2 or more groups case #2: more than 1 predictor variable with 2 or more levels; and 1 outcome variable: **n-way (or factorial) repeated-measures ANOVA**: several independent variables or predictors and each has been measured using different entities (between groups) ($F(df1, df2)$, p , r for planned comparisons)*

baseline model:

```
baseline<-lme(attitude ~ 1, random = ~1 | participant/drink/imagery, data =  
longAttitude, method = "ML")
```

adding the first variable:

```
drinkModel<-lme(attitude ~ drink, random = ~1 | participant/drink/imagery, data =  
longAttitude, method = "ML")
```

OR:

```
drinkModel<-update(baseline, .~. + drink)
```

THEN add imagery:

```
imageryModel<-update(drinkModel, .~. + imagery)
```

THEN add interaction:

```
attitudeModel<-update(imageryModel, .~. + drink:imagery)
```

Comparison:

```
anova(baseline, drinkModel, imageryModel, attitudeModel)
```

Further:

```
summary(drinkModel)
```

III. Mixed designs (GLM 5)

several independent variables or predictors have been measured; some have been measured with different entities, whereas others were measured using the same entities

- these *minimally have 2 independent variables*: one measured in a within, and one in a between subject design – i.e., not every combination exists in one subject

- strong suggestion: **never use more than 3 independent variables if possible!** it yields too many interactions (v1:v2, v2:v3, v1:v3, v1:v2:v3) and contrasts (all of the levels combined yield minimally 96 pairs!), and you die analyzing them eventually
- Begin by interpreting the highest-order effect (i.e., the significant interaction that contains the most predictors). You should not interpret any lower order effects contained within that interaction (not even main effect).
- Basic method:
 - choosing contrasts
 - compute main model
 - compute contrasts or post hoc tests

1. Mixed design ANOVA approach ($F(df1, df2), p$) — *ezANOVA()*

Assumptions (if not treated as *lme* regression but e.g. *ezANOVA()*):

1. the assumption of *sphericity* must be met in repeated measures variables (and the corresponding interaction terms) if ANOVA approach is used: **Mauchly's test** (*non-significant*) → sphericity is met

2. Multilevel model (linear mixed model) approach (χ^2, p) for contrasts: ($p, t(df), p$)

- i. has the advantages that (1) we don't need to concern ourselves with sphericity, and (2) we can now break down these very complicated effects by looking at the model parameters (which reflect the contrasts that we used to code the predictor variables)
- ii. here you add variables and interactions to the model hierarchically and see if the model gets better
- iii. begin by interpreting the highest-order effect and go no further down

IV. Correlations and coefficients

to see if 2 variables change together, but causality is not assumed (no predictor variable is determined)

main idea: if the two variables “move together” (e.g., number of ads seen in telly are in a connection with number of chocolates eaten) the scores in the two variable have the same variance = their standardized covariance (covariance is not $(\text{score} - \text{mean})^2 / (n - 1)$ but $(\text{score}_a - \text{mean}_a) \times (\text{score}_b - \text{mean}_b) / (n - 1)$ is close to -1 or +1

1. Pearson's r (bivariate)

2. **Partial correlation:** if some of the variance explained in variable A by variable B is not unique to variable B (but may be accounted for also by variable C which we held constant in this comparison) → variable C is held constant, i.e., is *controlled for* in the comparison (see SPSS: “Analyze > Correlate > Partial > Controlling for:”)

Cohen's suggestion for the interpretation of r :

$r = 0,10$ → small effect (kis hatás): A hatás a teljes variancia ($0,1^2 \times 100 =$) 1%-át* magyarázza.

$r = 0,30$ → medium effect (közepes hatás): A hatás a teljes variancia ($0,3^2 \times 100 =$) 9%-át* magyarázza.

$r = 0,50 \rightarrow$ large effect (erős/nagy hatás): A hatás a teljes variancia ($0,5^2 \times 100 =$) 25%-át* magyarázza.

r^2 : determinációs együttható: Azt mutatja meg, hogy az egyik változó varianciája mennyit magyaráz meg a másik változó varianciájából.

Pl. $r = 0,30 \rightarrow$ közepes hatás esetén az egyik változó varianciája a másik változó varianciájának ($0,3^2 \times 100 =$) 9%-át magyarázza, míg a maradék 91%-ot más változók magyarázzák!
A hatás irányát nem mondja ki!

szignifikancia (p értelmezése):

A korrelációelemzésben azt is teszteljük (H_1), hogy a kapott érték szignifikánsan eltér-e 0-tól (azaz a „nincs összefüggés” (H_0)-tól).

$p < 0,05$ jelentése itt: a korreláció valóban létezik (5%-nál kevesebb az esélye, hogy ezt az értéket kaptam volna, ha a hatás nem létezik).

C. Non-parametric tests for comparing groups (group means)

- in ordinal data (e.g. Likert scale measures)
- when the criterion of normal distribution of sampling frequency (as tested on sample scores) is NOT met

I. Independent samples

1.2 groups

a. Mann-Whitney U test (Z, p)

b. Wilcoxon rank-sum test (W, p)

2. more than 2 groups

a. Kruskal-Wallis rank sum test ($H(df)$ or $\chi^2(df), p$)

post hoc

1. several Wilcoxon rank-sum tests (with correction?)
2. or sg that is very similar to `friedmanmc()` ??

II. Dependent samples

1.2 groups:

a. Wilcoxon signed rank-test (p (median is good and effects size is preferred))

2.2 or more groups

a. Friedmann's ANOVA (χ^2 or χ^2_F, p)

post hoc:

1. `friedmanmc()` (difference=)

III. Mixed design

1. ??

IV. Correlations and coefficients:

1. Spearman's rho (ρ) (bivariate)

2. **Kendall's tau (τ)** (bivariate) – should be used in smaller datasets (if number of tied ranks is large, i.e., many scores have the same rank) → no effect size (r^2) may be calculated here!

D. Categorical data

Relationship between two categorical variables (and no continuous outcome variables), e.g. testing if food or affection as reward (1st var: food/affection) are more likely to make cats to learn to or not to learn to dance (2nd variable: yes/no)

- 2 variables: **Pearson's chi square** (χ^2 , p)
 - Assumptions
 - i. independency (each person contributes to only one cell)
- for small samples: **Fisher's exact test**
- more than 2 variables: **loglinear analysis** (e.g. animal (cat/dog) x training (yes/no) x dance (yes/no))

LOGISTIC REGRESSION

- multiple regression
- *glm()* – *generalized* linear model → the basic linear model generalized to other sorts of situations
- Prediction of categorical outcome variables (e.g., *voiced/unvoiced*) based on measured categorical (e.g., %ofVoicing, duration) or continuous (“is final position”) predictor variables
- Idea: if the outcome is categorical, the assumption of linear relation between variables is violated → solution: transferring data using logarithmic transformation (it expresses the multiple linear regression equation in logarithmic terms.
- In logistic regression not the value but the **probability** of one of the two categories is **predicted** (S-curve).
- types:
 - **binary logistic regression**: When we are trying to predict membership of only two categorical outcomes
 - **multinomial (or polychotomous) logistic regression**: when we want to predict membership of more than two categories
- assessing the fit of the model: **log-likelihood** (we expect small numbers reflecting a small amount of unexplained variation, just as in residual sum of squares RSS in linear regression)
- **deviance** statistic: **-2LL**
- comparing models: **the likelihood ratio**
- assessing the contribution of predictors: **z-statistic** or (Wald statistic), which is basically identical to *t*-statistic ($z = b/SE$)
- Assumptions (for working examples see page 343 in Field):
 - **linearity**: a linear relationship between any continuous predictors and the logit of the outcome variable → test: by looking at whether the interaction term between the predictor and its log transformation is significant
 - **independence of errors**: the cases of data should not be related, for example, you cannot measure the same people at different points in time

- **multicollinearity:** predictors should not to be highly correlated → test with tolerance and VIF statistics, the eigenvalues of the scaled, uncentred cross-products matrix, the condition indices and the variance proportions
- packages in R: *car*, *mlogit*
- general form:

```
newModel<-glm(outcome ~ predictor(s), data = dataframe, family = name of a distribution, na.action = an action)
```

 where family is the name of a distribution (e.g., Gaussian, binomial, poisson, gamma), and binomial is used here, so: `eelModel.1 <- glm(Cured ~ Intervention, data = eelData, family = binomial())`
- start by adding 1 predictor at a time and see if it improves the fit of the model over the baseline using the
 - **deviance statistic** → it will decrease if the fit increases (and its associated χ^2 statistic shows if this improvement is significant)
 - and **z-statistic** → it tells if the *b* coefficient (i.e. slope) of the predictor is significantly different from 0
- or insert both (or all three...) predictors at the same time, and take **a look at the z-statistic and its significance**
- use the **odds ratio** for interpretation *exp(model\$coefficients)* → If the value is greater than 1 then as the predictor increases, the odds of the outcome occurring increase (and vice versa) (BUT for this interpretation to be reliable the confidence interval of the odds ratio should not cross 1!)

FACTOR ANALYSIS and PRINCIPAL COMPONENT ANALYSIS

MULTILEVEL LINEAR MODELS a.k.a. MIXED MODELS

(based here on Bodo Winter, see Field et al. 2012 in a separate txt)

- why “mixed”: it has a *fixed effect* (predictor) and a *random effect* as well, the latter giving structure to the error term (the part of variance that is not explained by our predictor variable), e.g. *subject* to account for idiosyncratic variation (in repeated measures designs) – why random? because we only have a *random sample* of possible conditions, not all levels are present (not all kinds of clinics, not all kinds of personalities, etc.).
- this actually means that there are *different levels* of variables which are *nested* in each other, e.g. tested children (i.e., the *cases* are level1) coming from different classrooms (level2) and the classes are also coming from different schools (level3) and the latter two are *contextual variables*, and in many cases *subjects* act as contextual variable (e.g., it depends on many internal factors how many items I remember, etc.)
- two options: *nlme* package → *lme()* function OR *lme4* package → *lmer()* function
- **Assumptions**
 - homoscedasticity (equality of variances across the levels of the predictor)
 - which I must be misunderstanding, because Field states that heterogeneity of regression slopes is OK here (this will be the random slope term in the model)..
hmm..?
 - normal distribution

- in *random intercept* case, the normal distribution of intercepts,
- in *random slopes* case the normal distribution of random slopes
- no independence is assumed → we can model this variance → by always including **subject as a random effect!**
- can deal with some missing data and unbalanced designs
- we must test all the levels of the fixed effect (cannot investigate e.g., “curiosity” by covering only the levels of “curious” and “somewhat curious”, we also have to include “indifferent”)
- **main model with random intercept** (relationships are expected to be the same):
 - $\text{pitch} \sim \text{politeness} + \text{sex} + (1|\text{subject}) + \epsilon$
where $1|\text{subject}$ stands for having different intercept for each subject, that is,
 - we tell the model that we have multiple responses per subject (and also epr other variables denoted as *random*) and
 - we expect that within subjects (and other effects denoted as *random*) there will be some similarites – and between subjects there may be some variation
 - R: package *lme4*, function: *lmer()* OR package *nlme* and function *lme*
 - `lmer(frequency ~ politeness + gender + (1|subject) + (1|scenario))`
- getting statistical significance: **likelihood ratio test**
 - needs two models to be compared – one NOT containing the factor I am interested in (baseline), and one that does
 - *baseline*: `politeness.null = lmer(frequency ~ gender + (1|subject) + (1|scenario), data=politeness, REML=FALSE)`
 - *model tested*: `politeness.model = lmer(frequency ~ attitude + gender + (1|subject) + (1|scenario), data=politeness, REML=FALSE)`
 - *likelihood ratio test*: `anova(politeness.null, politeness.model)`
- **main model with interaction term:**
 - `lmer(frequency ~ attitude*gender + (1|subject) + (1|scenario), data=politeness)`
- **likelihood ratio test with interaction**
 - `politeness.noint = lmer(frequency ~ attitude + gender + (1|subject) + (1|scenario), data=politeness, REML=FALSE)`
 - `politeness.int = lmer(frequency ~ attitude*gender + (1|subject) + (1|scenario), data=politeness, REML=FALSE)`
 - `anova(politeness.noint, politeness.int)`
- BUT what if we also want to include that the effect of politeness may differ according to gender or scenario? → **random slope model** → where people are allowed to have different slopes for the effect of politeness (e.g. some lower their voices but others raise)
- **random slope model**: this is what happens if we violate the homogeneity of regression slopes in ANCOVA – this assumes that regressions slopes are the same in all contexts (conditions) → if this is the case, we can use multilevel model to explicitly estimate that variability in slopes → this is unusual in itself, so:s
- **main model with random intercept and random slope:**
 - `politeness.model = lmer(frequency ~ attitude + gender + (1+attitude|subject) + (1+attitude|scenario), data=politeness, REML=FALSE)`

- `coef(politeness.model)` → shows different coefs for politeness, but all are e.g., negative
- **chi² likelihood ratio test:**
 - `politeness.null = lmer(frequency ~ gender + (1+attitude|subject) + (1+attitude|scenario), data=politeness, REML=FALSE)`
 - `anova(politeness.null, politeness.model)`
- **writing up the results:**
 - R citation
 - `lme4()` → found by *citation("lme")*
 - fixed effects of politeness and gender (with/without interaction term)
 - random effects: intercepts for subjects and items and random slopes for the effect of politeness
 - assumptions tested: homoscedasticity and normality → visual inspection
 - *p*-values obtained by likelihood ratio test of the full model with the effect in question against the model without the effect in question
- **growth curves** (see MULTILEVEL_models.txt)
 - Growth models are multilevel models in which changes in an outcome over time are modelled using potential growth patterns (i.e. they are an extension to logistic regression). Patterns may be quadratic, cubic, logarithmic, exponential, or anything you like really.
 - The hierarchy in the data is that time points are nested within people (or other entities). As such, it's a way of analysing repeated-measures data that have a hierarchical structure.
 - The *anova()* function can be used to compare the overall fit of hierarchical models. The resulting change in the log-likelihood and the significance of this change can be used to ascertain if the fit has been improved (a significant change equates to a significant improvement). The AIC and BIC can also be compared across models (but not significance tested).
 - The *intervals()* function can be used to get confidence intervals for model parameters. These intervals can tell us how much intercepts and slopes varied over our level 1 variable, and whether this variance is significant (if the interval does not cross zero, it is significant).
 - An autoregressive covariance structure, AR(1), is often assumed in time course data such as that in growth models.

DEALING WITH OUTLIERS

Keep in mind: „The consequences for the statistical model of applying the ‘wrong’ transformation could be worse than the consequences of analyzing the untransformed scores”

E. Options to detect outliers

- what you see in a boxplot as outlier: IQR x 1,5
- if z-transformed score is greater/smaller than approx. 3/-3
- if original score is greater than mean ± 3 SD (same as above)

F. Dealing with outliers and non-normal distributions

- Remove the case: only if you have very good reasons to do so!
- Transformation of the data (based on observed skewness)
 - i. correction for positive skew or unequal variances – more common problems
 - 1. log transform
 - 2. square root transform
 - 3. reciprocal transformation
 - ii. correction for negative skew
 - 1. reverse score transformation
- Changing the score
 - i. to the next highest score
 - ii. replace ≥ 3.29 z scores by the mean + 3 SD
 - iii. replace ≥ 3.29 z scores by the mean + 2 SD