

## Medical Statistics (BME) / Medical Biometry and Clinical Epidemiology (AF Medizin)

---

### Exercise - Repetition

---

Deadline: -

Beside the exercises try to answer the following questions on the given topics as a preparation for the exam. Some of them are considered in the repetition exercise, in general all topics have been dealt with in the exercise or the lecture. All information can be found in the lectures and exercise slides.

#### (1) Study Design and Bias

- What kinds of study designs are named in the lecture besides parallel design and what is their main feature?
- What is the difference between cross-sectional and cohort design?
- Name problems which can influence the outcome and/or interpretation of you data when you conduct a clinical trial.
- Name a procedure which helps to reduce/prevent certain forms of bias.

#### (2) Diagnostic Tests

- Repeat the definition of the important medical terms used in diagnostic tests, such as gold standard, prevalence, sensitivity, specificity, predictive values, likelihood quotient, odds, cut-off.
- Describe which of the previously mentioned values are influenced by the test decision and how a cut-off relates to that if a high value indicates an illness.
- How can the quality of a diagnostic test be evaluated?
- How can the influence of a cut-off on the test decision be illustrated and how do you chose an appropriate cut-off.
- How does a decision rule influence sensitivity and specificity of a combined test if tests are conducted in a serial or parallel manner.

#### (3) Confidence Intervals

- Repeat the formulas for confidence intervals for two-and one-sided continuous parameters which are assumed to be normally distributed. How do they look like if we are interested in the difference of the parameter between two unpaired groups.
- What does the exact and approximate two-sided confidence interval for a probability/frequency look like? How does the approximate CI look like if we are interested in the difference of the probabilities between two unpaired groups.

- How can you find the quantils for the confidence intervals in the corresponding quantil tables.

#### (4) Statistical Tests

- Determine under which conditions which test/analysis is used: binomial test, paired/unpaired t-test,  $\chi^2$ -test, ANOVA, Agreement (Bland-Altman), Agreement (Kappa).
- Repeat the procedures for the paired/unpaired t-test and  $\chi^2$ -test (Formulation of the Hypothesis, calculation of the test statistic, determination of the critical value, test decision, interpretation)
- What do the type-1 and type-2 error describe
- How can the paired/unpaired t-test and  $\chi^2$ -test be conducted with the help of confidence intervals?
- How is the test performed if you use a p-value instead of the test statistic?

#### (5) ANOVA

- Repeat the procedure for the ANOVA test (Formulation of the Hypothesis, calculation of the test statistic, determination of the critical value, test decision, interpretation)
- What do you have to consider when you analyse the residuals?
- How do you evaluate the model fit of the ANOVA?
- Why is it important to consider adjusting the significance level for post-hoc tests if the ANOVA is significant.
- Name the methods from the lecture used for multiple testing adjustment and describe their features, how the adjustment is performed and how the test decision is performed for the adjusted values.

#### (6) Agreement

- What do you analyze if you investigate the agreement between two methods with a continuous parameter?
- Which parameters do you have to calculate to create a Bland-Altman Plot and how do you draw a Bland-Altman Plot?
- Which characteristics do you need to analyse in a Bland-Altman plot and how?
- What is the difference between association and agreement?
- Which relationships and problems are taken into account by Lin's Concordance Correlation Coefficient and how is Lin's CC evaluated?
- How can you measure agreement if your parameter is categorical? An how is Kappa evaluated?
- How do you calculate the corresponding confidence interval for the categorical agreement parameter in  $SE(\text{kappa})$  is given?
- Describe the paradox of Cohen's Kappa.

#### (6) SAS

- How can you calculate the mean and standard deviation of a parameter in SAS. Which option has to be added if you want to consider your parameter for different groups?
- How can you calculate covariances and (Pearson) correlations between two (or more) variables in SAS?
- How do you create  $n \times m$  tables in SAS?
- Which procedures are needed to conduct the paired/unpaired t-test,  $\chi^2$ -test and ANOVA in SAS and how is the general structure of the code to perform the corresponding test. How can you find the corresponding values for the test decision in the output?
- How can you calculate Kappa (Agreement) in SAS?

## Overview Answers

Keine Garantie für Vollständigkeit oder Richtigkeit

### 1. Study Design and Bias

1. On study designs we got

1. parallel design
2. Cross Over Design
3. Faktorial Design

Their main features are

1. looking at progression of disease/disorder, no carryover effects, issue Variability in response occurring within the same patient
  2. statistically efficient, confounding covariates effects are reduced, several bias as well
  3. studying effect of certain value by using combinations
2. Problems influencing outcome/interpretation of data when conducting clinical trial are different forms of bias(selection attrition detection ascertainment confounding) selection means eg non selective population so not randomized(external)
- attrition means losing track of patients
  - detection means wrong measurement method
  - ascertainment eg non selective population so not randomized(internal)
  - confounding means relating unrelated things
- ! Procedure to reduce/prevent certain forms of bias is 1.randomization 2. controlling
- ! Thereby removing with 1 selection and concealment bias and with 2 Ascertainment and Performance Bias

### 2. Diagnostic Tests

1. Definition of important medical terms used in diagnostic tests

1. gold standard means best performing test available
2. prevalence/Prävalenz  $Pr = P(K^+) = \frac{\text{Number of people suffering illness}}{\text{size of people}}$  and independent of test decision
3. sensitivity/Sensitivität  $Se = P(T^+|K^+) = \frac{\text{Number of diseased patients with positive test}}{\text{Number of diseased patients}}$  number of true positive cases correctly identified, prob. of positive test within diseased population
4. specificity/Spezifität  $Sp = P(T^-|K^-) = \frac{\text{Number of healthy patients with negative test result}}{\text{Number of healthy people}}$  number of true negative cases correctly identified by test, prob. of negative test result within diseased population
5. cut-off directly influences Specificity and Sensitivity, a good ratio has to be found with respect to clinical reasons, at or near shoulder of curve and according to Youden Index ( $Y=Se+Sp-1$ ), dot plot might help to identify, otherwise ROC-Curve helps
6. predictive values are
  - A. PPV pos.pred.value  $PV^+ = P(K^+|T^+) = \frac{\text{healthy persons with pos Test}}{\text{Persons with pos Test}} = \frac{Se \times Pre}{Se \times (Pre) + (1 - Sp) \times (1 - Pre)}$
  - B. NPV neg.pred.value  $PV^- = P(K^-|T^-) = \frac{\text{healthy persons with neg Test}}{\text{Persons with neg Test}} = \frac{Sp \times (1 - Pre)}{Sp \times (1 - Pre) + (1 - Se) \times Pre}$
  - C. Bayes theorem  $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$
7. odds a-priori odds  $\frac{Prevalence}{1 - Prevalence}$ , a-postiori  $\frac{PV^+}{1 - PV^+}$
8. likelihood ratio LR defines usefulness of test
  - $LR^+ = \frac{Se}{1 - Sp}$

- $LR^- = \frac{1-Se}{Sp}$
  - Accetable:  $LR^+ > 3, LR^- < 0.3$ , excellent:  $LR^+ > 10, LR^- < 0.1$
  - higher  $LR^+$  of test, but not if its a good indicator(depends on prevalence)
2. Which of previously mentioned values are influenced by test decision? We can decide our significance level and the appropriate test to choose in relation to the study who was evaluated. Depending on the data we can additionally choose a cut-off value
  3. How does a cut-off relates to that if high value  $\rightarrow$  illness? Cut-off only is an indicator for Sens and Spec on data, but can not change Prevalence and statistically significance for all the biases
  4. Quality of diagnostic tests can be evaluated by  $LR^+, LR^-$  and prevalence
  5. Illustration of cut-off on test decision and how to choose appropriate cut-off  
Checking for AUC by Youden Index we can approximate a cut off value, this means we evaluate approximately for all Spec and Sens the given data finding an optimum looking at costs for false tests as well
  6. How does decision rule influence sensitivity and specificity of combined test if conducted in serial or parallel manner? Serial for examination whenever fast judgement not needed, parallel conduct on fast judgment is needed Depending on statistically independency the probabilities can be calculated as well as Spec and Sens

### 3. Confidence Intervals

MasterFormula

(normal dist.) CI (unknown variance)

$$\left[ \bar{x} - t_{(1-\frac{\alpha}{2}; n-1)} SEM ; \bar{x} + t_{(1-\frac{\alpha}{2}; n-1)} SEM \right]$$

standard error of the mean (SEM)  $SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$  where  $s_x$  ist STD

$n > 30$  or variance  $\sigma^2$  known  $\rightarrow$  using standardised normal distribution for t

1. CI one-sided continuous parameter(normal dist.) using MasterFormula except with doubling  $\alpha$  since we know there are only values between 0 and 100 %
2. CI two-sided continuous parameter(normal dist.) using MasterFormula
3. difference of parameters between 2 unpaired groups

$$\left[ \bar{x}_1 - \bar{x}_2 - SE_{diff} * t_{(1-\frac{\alpha}{2}; n_1+n_2-1)} ; \bar{x}_1 - \bar{x}_2 + SE_{diff} * t_{(1-\frac{\alpha}{2}; n_1+n_2-1)} \right]$$

$$SE_{diff} = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1-1)s_{x_1}^2 + (n_2-1)s_{x_2}^2}{n_1+n_2-2}}$$

4. exact/approximate unpaired CI for proportions with  $\chi^2$ -Test  $p_{1,2} = \frac{m_1}{n_1} - \frac{m_2}{n_2} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{m_1}{n_1}(1-\frac{m_1}{n_1})}{n_1} + \frac{\frac{m_2}{n_2}(1-\frac{m_2}{n_2})}{n_2}}$   
m number of positive tests, n total test number or imagining abcd table with illness above:

$$T = n * \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

5. exact/approximate paired CI for proportions  $p_{1,2} = \frac{m}{n} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}$  m number of positive tests, n total test number
6. approximate paired CI for frequency  $\bar{d} \pm \frac{s}{\sqrt{n}} * t_{n-1, 1-\frac{\alpha}{2}}$
7. quantil tables checking

### 4. Statistical Tests

1. Test to use — Procedure — Formula

Condition For Binom, t-Test and Chi-Test

Stating  $H_0$  first(usually some equality of values), we calculate p-Value by formula and check under  $H_0$  (0-hypothesis) if the appropriate  $t/\chi$ /binom value is  $\geq 0.05$  or whatever significance level is given. If yes, we reject, otherwise we can assume  $H_0$  is true, the opposite goes for type 2 error  $\beta$  which has to be  $\leq$  significance value for  $H_1$  to be false

1. Binomial Test — only 2 events, calculation of possibilities due to hypothesis — assuming we have binomial distribution we can calculate the CI and check if we are biased or not or the p-Value —  $bin(k, n, p) = \binom{n}{k} * p^k * (1 - p)^{n-k}$
  2. paired t-Test — dependencies between groups —  $T = \sqrt{n} \frac{\bar{d}}{s_x}$
  3. unpaired t-Test — no dependencies between groups —  $T = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
  4.  $\chi^2$ -test — only for frequencies —  $\chi^2_T = \sum_{allCells} \frac{(freq_{obs} - freq_{exp})^2}{freq_{exp}}$
  5. ANOVA SEE BELOW
  6. Agreement(Bland-Altman) we plot mean difference of value and the difference of data points and paint asymptotic  $\pm 2s_d$  line over mean difference, for exact we plot  $\frac{s_d}{\pm} \sqrt{n} * t_{n-1, 1-\frac{\alpha}{2}}$  lines
  7. Agreement(Kappa) Cohens Kappa  $\kappa = \frac{p_o - p_e}{1 - p_e}$ , (zeile,spalte) $p_o = (T^+, K^+) * (T^-, K^-)$ ,  $p_e = (\frac{\sum T^+}{n} * \frac{\sum K^+}{n} + \frac{\sum T^-}{n} * \frac{\sum K^-}{n})$ ,  $> 0.81 \rightarrow$  very good,  $0.61 - 0.8 \rightarrow$  good,  $0.41 - 0.6 \rightarrow$  fair,  $0.21 - 0.4 \rightarrow$  slight,  $< 0.2 \rightarrow$  poor
2. type-1 errors is probability of error occuring on  $H_0$  hypothesis, type-2 errors is probability of error occuring on  $H_1$  hypothesis
  3. How can paired/unpaired t-test and  $\chi^2$ -test be conducted with help of CI ? Yes, checking the CI we can identify for significance. If 0 is including the CI area we know that there is no statistically significance. We can do this for all of the tests

## 5. ANOVA

1. procedure repeating for ANOVA
  1. Calculating Model  $SS_A = \sum_{i=1}^k n_i (y_{i\bullet} - \bar{y}_{\bullet\bullet})^2$
  2. Calculating corrected total  $SS_{CT} = \sum_{i,j} (y_{ij} - \bar{y}_{\bullet\bullet})^2$
  3. Calculating error  $SS_R = SS_{CT} - SS_A$
  4. Calculating Test Statistics  $F_0 = \frac{\frac{1}{k-1} SS_A}{\frac{1}{N-k} SS_R}$
  5. Calculating Model Fit(Residual)  $R^2 = \frac{SS_A}{SS_{CT}}$  has to be high(close to 1 is good)
2. Analazing residuals we have to consider  $R^2$ -Test and  $F_0$ -Test to check if our statement is correct
3. Evaluating model fit on ANOVA by SEE ABOVE
4. Considering adjusting the significance level for post-hoc tests, if ANOVA is significant we are doing, because ANOVA does not tell which group was affected and these are trying to do
5. Methods for multiple testing adjustment are Bonferroni and Hochberg method
6. Their features are simpleness and efficiency for small test sets for Bonferroni and accuracy and correctness also for large test sets for Hochberg
7. Adjustment is performed by taking into effect that positive test should not lower the p-Value in Hochberg
8. Test decision is performed for the adjusted values by calculating p-Values and afterwards comparing comparing test results via Hochberg or Bonferri where for bigger varity of test one shall use Hochberg

## 6. Agreement

1. Investigating the agreement between 2 methods with contiuous parameters we analyze location and scale shift from expected data to measured data by Bland-Altman Plot visually or by  $\kappa$  for linear correlation
2. In an Bland-Altman plot we need to analyze agreement by calculating the differences and the mean of them as well as expected variance over difference and looking for scale and location shift

3. Difference between Association and Agreement is Agreement means with low error probability the data fit a certain model while association means the correlation or dependency of two data set
4. Relationships and problems taken into account by Lin's Concordance Correlation Coefficient are location and scale shift by using Accuracy on Pearson linear coefficient
5. Lin's CC is evaluated by  $\frac{2s_{12}}{s_1^2 + s_2^2 + (\bar{x}_1 - \bar{x}_2)^2}$
6. Agreement if parameter is categorical can be measured by calculating the sum over appearance (to sample it to a group) and measure if this was correct
7. Kappa is evaluated by SEE ABOVE
8. For categorical agreement parameter in SE(kappa) we can calculate the corresponding confidence interval by
9. The paradox of Cohen's Kappa means if we have symmetric imbalance we get lower  $\kappa$ , because it distinguishes between agreement of findings in positive and negative ratings

## 7. SAS

1. Mean and standard deviation of a parameter in SAS can be calculated by

```
/*ii) calculate mean and standard deviation for height seperated by sex */
PROC MEANS data=WORK.workingdata MEAN STD;
  VAR HEIGHT;
  BY SEX;
RUN;
```

2. Option for different groups can be added by

```
PROC MEANS DATA=bodyfat MEAN STD MIN MAX CLM;
  VAR bodyfat height weight;
  CLASS gender;
RUN;
```

3. Calculating covariance and (pearson) correlation between two(or more) variables in SAS by

```
/*c pairwise Person and Spearman correlation*/
/*already sorted data by sex*/
PROC CORR data=WORK.workingdata PEARSON SPEARMAN;
  VAR HEIGHT WEIGHT1;
RUN;
```

4. Creating  $n \times m$  table in SAS by

```
data Ex11b;
input treatment $ outcome numbers;
datalines;
C 0 7
C 1 3
T 0 2
T 1 8
;
run;
```

5. Procedure to conduct paired/unpaired t-test,  $\chi^2$ -test and ANOVA

```
/* unpaired t-test SIDES=2 for twosided, h0=value for different hypothesis */
PROC TTest DATA=bodyfat;
  VAR bodyfat age;
  CLASS gender;
RUN;
/*PAIRED pulse1*pulse2; for paired t-test*/
/*$Chi-test */
PROC freq data=earinfection;
  tables location*sex / chisq;
RUN;
/* ANOVA */
PROC GLM DATA=WORK.MinutesData;
```

```

CLASS COUNTRY ;
MODEL TVTIME = COUNTRY ;
RUN;

```

6. General structure of code to perform corresponding test

```

PROC TESTNAME DATA=DATA_STORE;
  VAR variable_to_test;
  CLASS combined_variables;
  PAIRED pair1*pair2; /*optional*/
RUN;

```

7. How to find corresponding values for the test decision in output  
 paired/unpaired t-Test → t-Value, checking for significance level next to it  
 $\chi^2$ -Test → standing next to it and looking on p-Value next to it for checking significance level  
 ANOVA → F-Value checking, probability lower than test and higher next to it,  $R^2$  for model fit

8. Calculating Kappa(Agreement)

```

proc freq data=SkinCondition order=data;
  tables Derm1*Derm2 / agree noprint;
  test kappa;
  weight Count;
run;

```