university of
groningen

First Partial Exam (SAMPLE)

Statistics II, PSBE2–07

09–99–2099, 1 hour

- The exam for this course exists of two partial exams. This is the first partial exam.
- The **real** partial exam will consist of **18 multiple-choice questions**. This sample exam may have more or less than 18 questions.
- Do not infer that the weight or type of content of each topic in the real exam will be the same as in this sample exam. Simply, use this exam to *get an idea* about the real exam.
- For each multiple-choice question, write down the best answer to your knowledge on the separate pink answering sheet. Only one out of four answers is correct for each question.
- Write your name and student number on the answer sheet.
- At the end, **hand the questions set and your answers over to the proctor**.
- The exam is closed book. No items are allowed on your desk other than the papers provided, your student card, pens/pencils, and a calculator.
- It is not allowed to use a graphical calculator. It is also **not allowed to use a mobile phone**, also not as a calculator.
- At the end of the exam there is a table with critical values and a formula sheet. Formulas from the formula sheet may or may not be used to answering questions.
- Fraud (such as looking into other's work, allowing others to look into your work, any communication) is prohibited and will be reported to the Examination Committee.

Good luck!!

1 The confidence interval for a population mean becomes narrower if ...

a. ... α increases. $\surd$

b. ... β increases.

c. ... n decreases.

d. ... s increases.

Solution:
$\text{CI} = \bar{y} \pm t^* \frac{s}{\sqrt{n}}$,
t^* goes down if α goes up.

2 A sample of size 16 was randomly drawn from a normal population with known standard deviation $\sigma = 5$. The sample mean is equal to 20.5. The researcher wants to run the following z-test ($\alpha = 5\%$): $H_0 : \mu = 20$ versus $H_a : \mu < 20$.
What is the rejection region for this test?

a. $\overline{x} < 17.55$.

b. $\overline{x} < 17.94$. $\surd$

c. $\overline{x} < 18.05$.

d. $\overline{x} < 18.44$.

Solution:
Reject H_0 if:

$$\begin{aligned}
\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} &< z^*_{.05} = -1.645 \\
\frac{\bar{x} - 20}{5/4} &< -1.645 \\
\bar{x} - 20 &< -2.056 \\
\bar{x} &< 17.94.
\end{aligned}$$

3 What is a Type I error?

a. It is the probability of correctly not rejecting the null hypothesis.

b. It is the probability of correctly rejecting the null hypothesis.

c. It is the probability of incorrectly not rejecting the null hypothesis.

d. It is the probability of incorrectly rejecting the null hypothesis. $\surd$

Solution:
By definition.

Scores measuring racial prejudice towards foreigners in a European western country were regressed on an indicator measuring group identity. Some results are shown below. Use this information in the next two questions.

Descriptive Statistics

	Mean	Std. Dev.	N
RacPr	9.657	2.294	12
GrId		7.210	12

Coefficients

	B	S.E.	t	Sig.
(Constant)	5.147	7.483	0.688	0.507
RacPr	2.042	0.756	2.702	0.022

Dependent Variable: GrId
Alpha: 5%

4 What is the value of the correlation r between GrId and RacPr?

a. $r = 0.21$.

b. $r = 0.39$.

c. $r = 0.63$.

d. $r = 0.65$. $\surd$

Solution:
$r = B_{RacPr} \times s_{RacPr}/s_{GrId} = 2.042 \times 2.294/7.210 = 0.65$.

5 What is the 95% confidence interval of β_{RacPr}?

a. $(0.36, 3.73)$. $\surd$

b. $(0.40, 3.69)$.

c. $(0.56, 3.52)$.

d. $(0.67, 3.41)$.

Solution:
$95\%CI = B_1 \pm t^*_{df=12-2;.05} \times SE_{B_1} = 2.042 \pm 2.228 \times 0.756 = (0.36, 3.73)$.

For a bivariate sample of size $n = 84$, the correlation is $r = 0.7$. Use this for the next two questions.

6 Consider these two claims about the 95% confidence interval for the population correlation coefficient ρ:
Claim A: "The interval is symmetric around $r = 0.7$."
Claim B: "One needs the Fisher Z-transformation to obtain this interval".
What can be concluded?

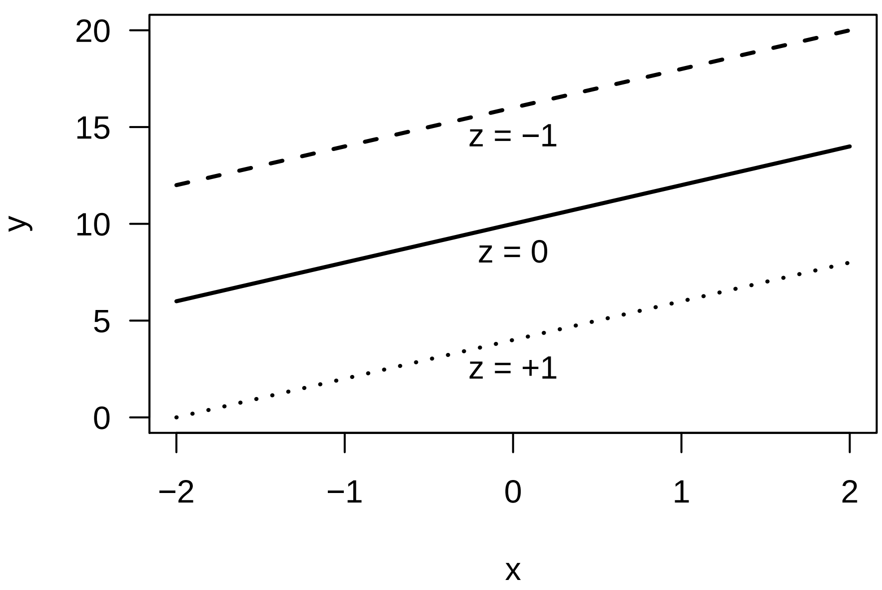
- **a.** Claim A is incorrect. Claim B is incorrect.
- **b.** Claim A is incorrect. Claim B is correct. $\surd$
- **c.** Claim A is correct. Claim B is incorrect.
- **d.** Claim A is correct. Claim B is correct.

Solution:
Claim A: False, symmetry is only expected when $\rho = 0$.
Claim B: True, since the Fisher Z-transformation takes the asymmetry of the sampling distribution into account.

7 Compute r_z.

- **a.** $r_z = 0.38$.
- **b.** $r_z = 1.04$.
- **c.** $r_z = 0.87$. $\surd$
- **d.** $r_z = 0.60$.

Solution:
$r_z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\ln\left(\frac{1+0.70}{1-0.70}\right) = 0.87$.



8 Consider the plot above corresponding to a regression analysis $Y = B_0 + B_1 x + B_2 z + B_3 xz$, where x and z are centered. What conclusion can be drawn?

a. The effect of x on Y is unrelated to z. $\surd$

b. The plot displays an example of a compromised interaction between x and z.

c. The effect of x on Y gets stronger as z increases (for the given values of the moderator).

d. The effect of x on Y gets weaker as z increases (for the given values of the moderator).

Solution:
The means plot shows parallel simple regression lines of Y on X for varying values of z. Parallel simple regression lines $\Rightarrow$ Equal simple slopes $\Rightarrow$ No $x \times z$ interaction (i.e., z does not moderate the $x \leftrightarrow Y$ relationship).

9 A regression has been performed on two continuous centered predictors. It is known that $n = 144$, $sd(x) = 2.73$, and $sd(z) = 2.01$. The estimated regression line is

$$\widehat{Y}_i = 286.27 + 17.88x_i + 28.18z_i + 7.53x_i z_i$$

Compute the simple regression equation of Y on x for z one standard deviation above mean.

a. $\widehat{Y}_i = 322.2 + 43.3x_i$

b. $\widehat{Y}_i = 314.5 + 25.4x_i$

c. $\widehat{Y}_i = 286.3 + 17.9x_i$

d. $\widehat{Y}_i = 342.9 + 33.0x_i$ $\surd$

Solution:
$\widehat{Y}_i = (286.27 + 28.18 \times 2.01) + (17.88 + 7.53 \times 2.01)x_i = 342.9 + 33.0x_i$.

10 Which of the alternatives below corresponds to the concept of ‘power’?

a. P(not rejecting $H_0 | H_0$ is false).

b. P(not rejecting $H_0 | H_0$ is true).

c. P(rejecting $H_0 | H_0$ is false). $\surd$

d. P(rejecting $H_0 | H_0$ is true).

Solution:
By definition.

11 In a sample, two variables, A and B, are positively correlated. What is **not** a possible explanation for this?

a. A and B both are common causes of C. $\surd$

b. A causes B through C.

c. Coincidence.

d. B causes A.

Solution:
Answer A is the wrong way round.

12 The population regression line in the multiple regression model with p predictors:

a. Is built on the principle DATA = FIT + RESIDUAL.

b. Is a straight line through the values of $x_1, \ldots, x_p$ and y.

c. Returns the value of the average of the dependent variable, for given values of the predictors. $\surd$

d. Provides p predictions for the mean dependent variable.

13 Three continuous variables, A, B and C, have been measured. The pairwise correlations between all three are positive and significant. In the multiple regression model $A = \beta_0 + \beta_1 B + \beta_2 C$, β_1 is non-significant whilst β_2 is significant. Which of the following describes this situation?

a. B is a mediator variable for the relation $A - C$.

b. C is a mediator variable for the relation $A - B$. $\surd$

c. B is a moderator variable for the relation $A - C$.

d. C is a moderator variable for the relation $A - B$.

14 A simple linear regression is carried out where X is used to predict Y. Both X and Y have been standardised. What is the value of the slope of the regression line?

a. 0

b. 1

c. $r_{x,y}$ $\surd$

d. s_y/s_x

15 Consider the following two claims about simple linear regression:
A: 'Prediction intervals for y at x-values close to $\bar{x}$ are smaller than those at x-values far from $\bar{x}$.'
B: 'The homogeneity assumption states that the variance of x is fixed.'

a. Claim A is correct, claim B is correct.

b. Claim A is correct, claim B is incorrect. $\surd$

c. Claim A is incorrect, claim B is correct.

d. Claim A is incorrect, claim B is incorrect.

A textiles company is opening new retail sale outlets and wants to staff their stores with personnel most likely to generate high income. The company studies the sales at several existing stores to determine whether extraversion and intelligence can be used to predict sales performance. If these turn out to be good predictors, the company shall use that information in the recruitment of staff for the new stores. For all current staff ($n = 50$), a psychological test yields scores on extraversion (continuous variable ext) and intelligence (continuous variable int). From the company's database, average weekly sales (in euro's, continuous variable sales) per staff member are extracted. The variables ext and int have been centered and you may assume that all relevant assumptions hold.

The following regression model is run:

$$\mathsf{sales}_i = B_0 + B_1\mathsf{int}_i + B_2\mathsf{ext}_i + B_3(\mathsf{int}_i \times \mathsf{ext}_i) + e_i.$$

The software provided the following table. Furthermore, $R^2 = 0.435$, $\text{sd}_{\mathsf{ext}} = 3.596$ and $\text{sd}_{\mathsf{int}} = 13.735$.

Use this setting to answer the next 4 questions.

	B	SE	p-value
(Constant)	869.693	32.860	0.000
int	9.871	2.461	0.000
ext	41.400	9.269	0.000
int×ext	0.276	0.838	0.743

16 Compute the value of the t-test statistic for testing H_0: $\beta_1 = 0$ versus one sided alternative H_a: $\beta_1 > 0$.

a. 2.01.

b. 4.01. √

c. 8.02.

d. 16.09.

Solution:
$t = \frac{9.871}{2.461}$.

17 Compute the value of the adjusted R^2.

a. 0.35.

b. 0.37.

c. 0.40. √

d. 0.42.

Solution:
$R^2_{\text{adj}} = 1 - \frac{50-1}{50-3-1}(1 - .435)$.

18 Compute the Simple Regression Equation of sales on ext, for intelligence at 1sd above mean.

a. $\widehat{\mathsf{sales}}_i = 1005.27 + 3.79\mathsf{ext}_i$

b. $\widehat{\mathsf{sales}}_i = 879.56 + 3.79\mathsf{ext}_i$

c. $\widehat{\mathsf{sales}}_i = 1005.27 + 45.19\mathsf{ext}_i$ $\surd$

d. $\widehat{\mathsf{sales}}_i = 879.56 + 45.19\mathsf{ext}_i$

Solution:
$\widehat{\mathsf{sales}}_i = (869.693 + 9.871 \times 13.735) + (41.400 + 0.276 \times 13.735)\mathsf{ext}_i$.

19 Provide the Simple Slope for sales on int, when ext $= 0$.

a. 10.15.

b. 9.87. $\surd$

c. 9.60.

d. 911.09.

t distribution: critical values[a]

ν	$\alpha_1 = .10$ $\alpha_2 = .20$	.05 .10	0.025 0.050	0.010 0.020	.005 .010
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
25	1.316	1.708	2.060	2.485	2.787
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
75	1.293	1.665	1.992	2.377	2.643
100	1.290	1.660	1.984	2.364	2.626
1000	1.282	1.646	1.962	2.330	2.581
∞	1.282	1.645	1.960	2.326	2.576

[a] α_1 holds the one-sided upper-tail value of the distribution with ν degrees of freedom; α_2 holds the corresponding two-sided value.

Formula sheet

At the exam, you will receive this formula sheet.

Pooled variance for i groups

$$s_p^2 = \frac{\sum_i (n_i - 1) s_i^2}{\sum_i (n_i - 1)}$$

Confidence interval for μ

$$\overline{y} \pm t^* \frac{s}{\sqrt{n}}.$$

t-test for H: $\mu_1 = \mu_2$

Test statistic:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Test for H: $\rho = 0$
Test statistic:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Contrasts

Sample estimation:

$$c = \sum_i a_i \bar{x}_i$$

Standard error:

$$\mathrm{SE}_c = s_p\sqrt{\sum_i \frac{a_i^2}{n_i}}.$$

Fisher Z-transformation

Transformation:

$$r_z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right).$$

Inverse transformation:

$$r = \frac{e^{2r_z} - 1}{e^{2r_z} + 1}.$$

(Semi-)partial correlations
Formula's valid when working with DV y and two predictors.

$$pr_1 = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}} = \sqrt{\frac{R^2 - r_{y2}^2}{1 - r_{y2}^2}}$$

$$sr_1 = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1-r_{12}^2}} = \sqrt{R^2 - r_{y2}^2}$$

Adjusted R^2

$$R^2_{\text{adj}} = R^2 - \frac{p}{n-p-1}\left(1 - R^2\right).$$

Effect sizes

$$\eta_p^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}}, \quad \omega^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \times MSE}{MSE + SS_{\text{total}}}$$

Binomial model

$$p(X = x|N,\theta) = \binom{N}{x}\theta^x(1-\theta)^{N-x}, \quad \binom{N}{x} = \frac{N!}{(N-x)!x!}$$