university of groningen

name: .....

stud. nr.: S.....

First partial exam * * * SAMPLE * * *

Statistics II, PSBE2–07

October 2013

- For each of the following multiple-choice questions, write down the best answer to your knowledge on the separate pink answering sheet. There are **20 questions** [Partial exam 1a will consist of 25 items]. 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 scrap paper, the answer sheet, and the question set over to the proctor**. Your grade will only be released if your question set is returned, otherwise your score is invalid.
- For your own bookkeeping: You may transcribe your answers to each question to the grid on the last page of this question set. This last page will be handed over to you when you return the question set.
- The correct answer key will be uploaded to Nestor within two working days. Exam scores will be uploaded to Nestor within five working days.
- The exam is closed book. No other items are allowed on your desk 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's committee.

Good luck!!

1 What is meant in the following description?

> *The distribution of the values of the mean of variable y in all possible samples (of a certain size) from the same population.*

a. The sampling distribution of y.

b. The distribution of y in the population.

c. The sampling distribution of the mean of y.

d. The distribution of the mean of y in the population.

2 A sample is collected to calculate a confidence interval for the mean of the corresponding population. If the sample size increases then the confidence interval is narrower. Why?

a. Because the spread in the sampling distribution of the mean becomes smaller.

b. Because the number of degrees of freedom for the t-distribution becomes smaller.

c. Neither answer is correct.

d. Both answers are correct.

(Exam 2009-2010)

The following incomplete ANOVA table, the result of a Two-Way ANOVA with factors A and B, will be used for questions 3 to 5.

Tests of Between-Subjects Effects

Dependent Variable:score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
A	555,0	3	185,0	18,5	,000
B	190,0		95,0	9,5	,003
A * B	45,0			,8	
Error	120,0	12	10,0		
Corrected Total	910,0				

3 How many groups does factor B have, and what's the total sample size n?

a. Factor B has 2 groups; $n = 23$.

b. Factor B has 3 groups; $n = 23$.

c. Factor B has 2 groups; $n = 24$.

d. Factor B has 3 groups; $n = 24$.

4 Which of the following statements about the significance of the interaction $A \times B$ is true, when $\alpha = 5\%$?

a. The p-value $p_{A\times B}$ is always equal to the sum of the p-values p_A and p_B.

b. It is not significant, because $F_{A\times B} < 1$.

c. It is not significant, because SSAB is smaller than both SSA and SSB.

d. Since both Factor A and B are significant, so is the interaction.

5 Suppose that this data set was used to do a One-Way ANOVA using only factor A, where SSM would equal 555.0. What would be the mean square error in that case?

a. The MSE would then be 6.00.

b. The MSE would then be 17.75.

c. The MSE would then be 29.58.

d. The MSE would then be 112.50.

6 Which of the following null hypotheses can be tested with an ANOVA?

a. $H_0 : \beta_1 = 0$ in a regression model.

b. $H_0 : \mu_1 = \mu_2$ for comparing two groups.

c. Both null hypotheses can be tested with an ANOVA.

d. Neither null hypotheses can be tested with an ANOVA.

(Exam 2009-2010)

For 1404 test subjects it is measured how much they've spent on a supermarket visit. Also, it has been recorded whether they shopped for themselves, themselves and their spouse, or themselves and their family. The resulting SPSS output below is used in questions 7 to 10.

Amount spent

shopfor	Mean	N	Std. Deviation
Self	85,65	484	35,49
Self and spouse	95,44	560	43,29
Self and family	126,10	360	60,02

Amount spent

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	356624,45	2	178312,22	84,69	,000
Within Groups	2949620,91	1401	2105,36		
Total	3306245,36	1403			

7 Based on the SPSS output, which of the following claims can be drawn?

a. The assumption of equal variances does not hold.

b. There is a significant difference between the three means.

c. Factor 'amount spent' explains 11% of the variation in 'shopfor'.

d. The pooled variance is 2140.76.

8 The research team wants to study the contrast between people who shop for families and people who shop for themselves or themselves and their spouse. What contrast coefficients should one enter, and what is the value of the contrast?

a. $\frac{1}{2}$, $\frac{1}{2}$ and -1 and -35.55.

b. $\frac{1}{2}$, $\frac{1}{2}$ and -1 and +35.55.

c. -1, 0 and 1 and +40.46.

d. 1, 0 and -1 and +40.46.

9 Using the Bonferroni method to test differences between all means, what will be the value of the test statistic for the comparison between 'self' and 'self and spouse'?

a. $t = -12.49$.

b. $t = -9.79$.

c. $t = -3.44$.

d. $t = 0.29$.

10 Using the Bonferroni method to test two-sided differences between all means, what will be the significance level per test if the overall significance level is set to 10%?

a. $\alpha = 0.017$.

b. $\alpha = 0.033$.

c. $\alpha = 0.050$.

d. $\alpha = 0.100$.

The following five exercises (11 to 15) deal with the following SPSS output from a simple regression analysis:

Descriptive Statistics

	Mean	Std. Deviation	N
physfunc	65.56	29.35	28
emotwllb	83.21	14.97	28

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1			.149	27.07871

a. Predictors: (Constant), emotwllb

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression			4192.82		.024[a]
	Residual	19064.67				
	Total	23257.48				

a. Predictors: (Constant), emotwllb

b. Dependent Variable: physfunc

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.72	29.42		-.126	.900
	emotwllb	.83	.35	.425		

a. Dependent Variable: physfunc

The dependent variable is `physfunc` (physical functioning of subjects), the independent variable is `emotwllb` (questionnaire score that measures emotional well-being). For both variables, the higher the score the better the performance (physical/ emotional).

11 In one of the SPSS tables there is a value for the 'Standard Error of the Estimate'. What is meant by this term?

a. The estimated explained variance.

b. The estimated spread of the values of `physfunc` around the population regression line.

c. The estimated standard deviation of the predictions of `physfunc`.

d. The estimated standard error of the slope of the regression line.

12 What is the percentage of explained variance in the sample?

a. 15.3%.

b. 22.1%.

c. 18.0%.

d. 23.0%.

13 Test the null hypothesis $H_0 : \beta_1 = 0$. What applies to the p-value of this test?

a. p is greater than 0.10.

b. p is between 0.10 and 0.05.

c. p is between 0.05 and 0.01.

d. p is less than 0.01.

14 What is the 95% confidence interval for the slope β_1?

a. (0.07, 1.59).

b. (0.11, 1.55).

c. (0.15, 1.51).

d. (0.21, 1.45).

15 The Standard Error of the slope of the regression line is 0.35. What does this mean?

a. The spread in possible values of b_1 (in other samples from the same population) is equal to 0.35.

b. The population parameter β_1 equals 0.83 in 35% of the samples.

c. The dispersion of the points around the regression line equals 0.35, for each value of the independent variable `emotwllb`.

d. That 35% of the possible values of b_1 (in other samples from the same population) is equal to 0.83.

(Exam 2009-2010)

Questions 16 to 18 all deal within simple linear regression contexts.

16 Which is *NOT* an assumption or requirement of the simple linear regression model?

a. The residuals follow a normal distribution.

b. The variance of the subpopulations is assumed equal for each value of the independent variable.

c. The variance of the dependent variable is equal to the variance of the independent variable.

d. There is a linear relation between the independent variable and the mean of the dependent variable.

17 Which of the following claims, in the context of simple linear regression, is *true*?

a. The population regression line is given by $y = b_0 + b_1 x$.

b. For a given x-value, the 95% confidence interval for y is wider than the 95% prediction interval for y.

c. SSM, the Model Sum of Squares, is given by $\sum_i (y_i - \bar{y})^2$.

d. Testing H: $\beta_1 = 0$ vs. $\beta_1 \neq 0$ is equivalent to testing H: $R^2 = 0$ vs. $R^2 \neq 0$.

18 A simple linear regression analysis provided the following information: $b_0 = 4$, $\bar{x} = 6.25$, $\bar{y} = 104$. Compute b_1.

a. 16.00.

b. 17.28.

c. 24.44.

d. 101.50.

(Exam 2009-2010)

Questions 19 and 20 deal with a simple linear regression with $n = 15$, $r = 0.50$.

19 What can you say about the p-value of the two-sided test for H: $\rho = 0$ at $\alpha = 5\%$ level?

a. $p < 0.01$.

b. $0.01 < p < 0.05$.

c. $0.05 < p < 0.10$.

d. $p > 0.10$.

20 A second sample from the same population, also of size 15, provides $r = 0.80$. What does this imply?

a. In the second sample, the percentage of explained variance is 39% higher than in the first sample.

b. The value for the F-test in the ANOVA-output is larger in the first sample.

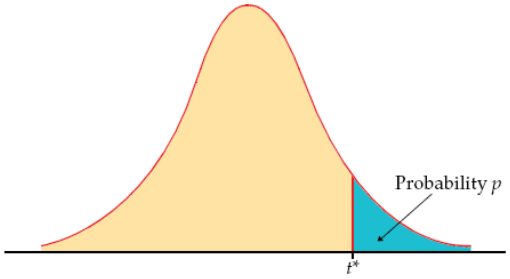
c. In the second sample, the percentage of explained variance is 30% higher than in the first sample.

d. The slope of the regression line is higher in the first sample.

End of exam

t distribution: critical values

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



	Upper-tail probability p								
df	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001
1	1.376	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3
2	1.061	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33
3	0.978	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21
4	0.941	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173
5	0.920	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893
6	0.906	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208
7	0.896	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785
8	0.889	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501
9	0.883	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297
10	0.879	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144
11	0.876	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025
12	0.873	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930
13	0.870	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852
14	0.868	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787
15	0.866	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733
16	0.865	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686
17	0.863	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646
18	0.862	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611
19	0.861	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579
20	0.860	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552
21	0.859	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527
22	0.858	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505
23	0.858	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485
24	0.857	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467
25	0.856	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450
26	0.856	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435
27	0.855	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421
28	0.855	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408
29	0.854	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396
30	0.854	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385
40	0.851	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307
50	0.849	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261
60	0.848	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232
80	0.846	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195
100	0.845	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174
1000	0.842	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098
z^*	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091
	60%	80%	90%	95%	96%	98%	99%	99.5%	99.8%
	Confidence level C								

Formulas

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:

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

Contrasts

Sample estimation:

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

Standard error:

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

Wilcoxon's rank sum test

Mean and SD of W:

$$\mu_W = \frac{n_1(N+1)}{2}, \quad \sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}.$$

Kruskal-Wallis test

Test Statistic:

$$H = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1).$$

Effect sizes

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

Form for marking answers, to take home

1. [a] [b] [c] [d]
2. [a] [b] [c] [d]
3. [a] [b] [c] [d]
4. [a] [b] [c] [d]
5. [a] [b] [c] [d]

6. [a] [b] [c] [d]
7. [a] [b] [c] [d]
8. [a] [b] [c] [d]
9. [a] [b] [c] [d]
10. [a] [b] [c] [d]

11. [a] [b] [c] [d]
12. [a] [b] [c] [d]
13. [a] [b] [c] [d]
14. [a] [b] [c] [d]
15. [a] [b] [c] [d]

16. [a] [b] [c] [d]
17. [a] [b] [c] [d]
18. [a] [b] [c] [d]
19. [a] [b] [c] [d]
20. [a] [b] [c] [d]

21. [a] [b] [c] [d]
22. [a] [b] [c] [d]
23. [a] [b] [c] [d]
24. [a] [b] [c] [d]
25. [a] [b] [c] [d]