Høgskolen i Telemark

EXAM

4301 – NATURAL SCIENCE METHODS

02.10.2015

Time:	*9-14*
Language:	*English*
Pages:	*11 (this page included)*
Help means:	*Calculator, statistical tables, English dictionary*
Notes:	*None*
Appendices:	*None*

The exam results will be made available on Studentweb.

EXAM IN NATURAL SCIENCES METHODS
02.10.2015, kl. 09.00 - 14.00

All main questions (1-6) count the same and within each main question all sub-questions count the same.

1. Explain briefly the following statistical concepts:
    a. continuous numerical variable,
    b. interval scale,
    c. mean,
    d. SD,
    e. Central Limit Theorem,
    f. Confidence interval,
    g. Type I and Type II error.

2. Systolic blood pressure was measured (in units of mm Hg) during preventive health examinations on people in Dallas, USA. Here are the measurements for a subset of these patients: 112, 128, 108, 129, 125, 153, 155, 132, 137.

    a. How many individuals are in the sample?
    b. What is the mean of this sample?
    c. What is the sum of the squares of the measurements?
    d. What is the standard deviation of this sample?
    e. What is the coefficient of variation of this sample?
    f. What is the standard error of the mean?
    g. Calculate an approximate confidence interval of the mean.

3. Assume a random sample. What effect does increasing the sample size have on:
    a. The probability of committing a Type I error?
    b. The probability of committing a Type II error?
    c. The power of a test?
    d. The effect size?
    e. The significance level?

4. In a test of Murphy's law, pieces of toast were buttered on one side and then dropped. Murphy's law predicts as you may know that they will fall butterside down. Out of 9821 slices of toast dropped, 6101 landed butter-side down. Do these data support Murphy's law, or not?

    a. Give the proper H_0 and H_A.
    b. Which test would you use, and why?
    c. What are the assumptions of the test?
    d. Do the test (see attached formulae and table sheets).

5. Below is a printout from an ANOVA performed in DataAnalysis in EXCEL.

    a) Which types of ANOVA have we learned, and which is used here?
    b) What are the assumptions for using ANOVA?
    c) Which are the explanatory and response variables here, and what types of variables are they?
    d) Explain what the different columns and rows in the ANOVA table (below) show, and fill in the table.

e) If you were asked to investigate how 3 different levels of fat and fibers, respectively, in the bun in of MacDonald's cheeseburger influence the weight of their customers, how would you design such a study?

Anova: Two-Factor With Replication

SUMMARY	A	B	C	D	E	Total	
1							
Count		2	2	2	2	2	10
Sum		826	788	832	958	824	4228
Average		413	394	416	479	412	422,8
Variance		2	50	2	2	18	951,955556
2							
Count		2	2	2	2	2	10
Sum		916	783	864	867	742	4172
Average		458	391,5	432	433,5	371	417,2
Variance		8	4,5	2	4,5	8	1101,73333
3							
Count		2	2	2	2	2	10
Sum		801	776	909	914	805	4205
Average		400,5	388	454,5	457	402,5	420,5
Variance		4,5	2	4,5	2	4,5	950,5
4							
Count		2	2	2	2	2	10
Sum		805	792	841	823	751	4012
Average		402,5	396	420,5	411,5	375,5	401,2
Variance		4,5	2	0,5	4,5	24,5	263,511111
5							
Count		2	2	2	2	2	10
Sum		828	813	921	973	770	4305
Average		414	406,5	460,5	486,5	385	430,5
Variance		2	24,5	4,5	0,5	32	1552,5
Total							
Count		10	10	10	10	10	
Sum		4176	3952	4367	4535	3892	
Average		417,6	395,2	436,7	453,5	389,2	
Variance		488,266667	52,6222222	356,233333	871,611111	283,733333	

ANOVA

urce of Variati	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	4674,52	4	1168,63	134,017202	1,6733E-16	2,75871047
Columns	29584,12	4	7396,03	848,168578	2,6915E-26	2,75871047
Interaction	13579,68	16	848,73	97,331422	1,7206E-18	2,06908764
Within	218	25	8,72			
Total	48056,32	49				

6. In a survey of effects of running on personality 231 male runners that each ran about 30 km per week, were given the Cattell Sixteen Personality Factor Questionnaire, a 187-questions multiple choice-test which is widely used by psychologists. A news article (New York Times, 15 February 1988) stated that "Researchers have found statistically significant differences in personality among runners in the 30-year old population as a whole." And further in the headline "Research shows that running may change the runners state of mind."

a) Explain what is meant by statistically significant.
b) Which test do you think have been used, and why?
c) Explain why the headline is misleading.
d) Design an experiment which will demonstrate if what the headline states is correct.

Formulae for Basic Statistics

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$s = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2}{n-1}} \qquad s = \sqrt{\frac{\Sigma\left(Y_i^2\right) - n\bar{Y}^2}{n-1}}$$

Standard error of the mean

$$s/\sqrt{n}$$

χ^2 test of goodness-of-fit

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

$$\ln(\hat{OR}) - Z_\alpha \, SE[\ln(\hat{OR})] \le \ln(OR) \le \ln(\hat{OR}) + Z_\alpha \, SE[\ln(\hat{OR})]$$

$$\hat{OR} = \frac{ad}{bc}$$

Poisson Probability Distribution

$$P[x] = \frac{\mu^x e^{-\mu}}{x!}$$

Binomial Probability Distribution

$$P[x] = \binom{N}{x} p^x (1-p)^{N-x}$$

Normal Probability Distribution

$$P[x] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

Confidence Interval for the mean of a normal distribution

$$\bar{Y} \pm SE_{\bar{Y}} t_{\alpha(2),df}$$

Confidence Interval for the variance of a normal distribution

$$\frac{df\, s^2}{\chi^2_{\frac{\alpha}{2},df}} \le \sigma^2 \le \frac{df\, s^2}{\chi^2_{1-\frac{\alpha}{2},df}}$$

Formulae for regression and correlation

$$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})=\left(\sum X_iY_i\right)-\frac{\sum X_i\sum Y_i}{n}$$

$$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum\left(X_i^2\right)-\frac{\left(\sum X_i\right)^2}{n}$$

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum\left(Y_i^2\right)-\frac{\left(\sum Y_i\right)^2}{n}$$

Formulae for regression and correlation

$$\sum(X-\overline{X})(Y-\overline{Y}) = \sum(XY) - \frac{(\sum X)(\sum Y)}{n}$$

$$b = \frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sum(X_i-\overline{X})^2}$$

$$a = \overline{Y} - b\overline{X}$$

$$SS_{Total} = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}$$

$$SS_{regression} = b\sum(X_i-\overline{X})(Y_i-\overline{Y})$$

$$SS_{residual} + SS_{regression} = SS_{Total}$$

$$MS_x = \frac{SS_x}{DF_x}$$

$$r^2 = \frac{SS_{regression}}{SS_{Total}}$$

$$SE_b = \sqrt{\frac{MS_{residual}}{\sum(X_i-\overline{X})^2}}$$

$$MS_{residual} = \frac{\sum(Y_i-\overline{Y})^2 - b\sum(X_i-\overline{X})(Y_i-\overline{Y})}{n-2}$$

$$b \pm t_{\alpha[2],\nu} SE_b$$

$$\hat{Y} \pm t_{\alpha[2],\nu} SE_{\hat{Y}}$$

$$t = \frac{b-\beta_0}{SE_b}$$

$$t = \frac{(b_1-b_2)-(\beta_1-\beta_2)}{SE_{b1-b2}}$$

$$(MS_{error})_p = \frac{(SS_{error})_1 + (SS_{error})_2}{(DF_{error})_1 + (DF_{error})_2}$$

$$SE_{b1-b2} = \sqrt{\frac{(MS_{error})_p}{\left(\sum(X-\overline{X})^2\right)_1} + \frac{(MS_{error})_p}{\left(\sum(X-\overline{X})^2\right)_2}}$$

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

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

$$z = 0.5 \ln\left(\frac{1 + r}{1 - r}\right)$$

$$\sigma_z = \sqrt{\frac{1}{n - 3}}$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

ANOVA etc.

$$F = \frac{MS_{\text{groups}}}{MS_{\text{error}}}$$

$$MS_{\text{error}} = s^2_{\text{pooled}} = \frac{\sum s_i^2 (n_i - 1)}{N - k}$$

$$MS_{\text{groups}} = \frac{\sum n_i (\overline{Y}_i - \overline{Y})^2}{k - 1}$$

$$\overline{Y} = \frac{\sum n_i (\overline{Y}_i)}{N}$$

$$R^2 = \frac{SS_{\text{groups}}}{SS_{\text{total}}}$$

Kruskal-Wallis

$$H = \frac{12}{N(N+1)} \left[\sum \frac{R_i^2}{n_i} \right] - 3(N+1)$$

Tukey-Kramer:

$$q = \frac{\overline{Y}_i - \overline{Y}_j}{SE} \qquad SE = \sqrt{s^2_{\text{pooled}} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}.$$

Mann-Whitney U

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U' = n_1 n_2 - U$$

$$Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/3}}$$

Critical value of F, $\alpha(1)$=0.05, $\alpha(2)$=0.10

	Numerator *df*									
den. *df*	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32

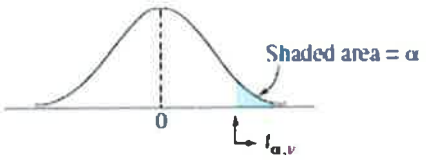

TABLE 2
Percentage points of Student's t distribution

df/α =	.40	.25	.10	.05	.025	.01	.005	.001	.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
35	0.255	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.255	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
inf.	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

TABLE 7
Percentage points of the chi-square distribution

$\alpha = .10$	.05	.025	.01	.005	.001	df
2.706	3.841	5.024	6.635	7.879	10.83	1
4.605	5.991	7.378	9.210	10.60	13.82	2
6.251	7.815	9.348	11.34	12.84	16.27	3
7.779	9.488	11.14	13.28	14.86	18.47	4
9.236	11.07	12.83	15.09	16.75	20.52	5
10.64	12.59	14.45	16.81	18.55	22.46	6
12.02	14.07	16.01	18.48	20.28	24.32	7
13.36	15.51	17.53	20.09	21.95	26.12	8
14.68	16.92	19.02	21.67	23.59	27.88	9
15.99	18.31	20.48	23.21	25.19	29.59	10
17.28	19.68	21.92	24.72	26.76	31.27	11
18.55	21.03	23.34	26.22	28.30	32.91	12
19.81	22.36	24.74	27.69	29.82	34.53	13
21.06	23.68	26.12	29.14	31.32	36.12	14
22.31	25.00	27.49	30.58	32.80	37.70	15
23.54	26.30	28.85	32.00	34.27	39.25	16
24.77	27.59	30.19	33.41	35.72	40.79	17
25.99	28.87	31.53	34.81	37.16	42.31	18
27.20	30.14	32.85	36.19	38.58	43.82	19
28.41	31.41	34.17	37.57	40.00	45.31	20
29.62	32.67	35.48	38.93	41.40	46.80	21
30.81	33.92	36.78	40.29	42.80	48.27	22
32.01	35.17	38.08	41.64	44.18	49.73	23
33.20	36.42	39.36	42.98	45.56	51.18	24
34.38	37.65	40.65	44.31	46.93	52.62	25
35.56	38.89	41.92	45.64	48.29	54.05	26
36.74	40.11	43.19	46.96	49.65	55.48	27
37.92	41.34	44.46	48.28	50.99	56.89	28
39.09	42.56	45.72	49.59	52.34	58.30	29
40.26	43.77	46.98	50.89	53.67	59.70	30
51.81	55.76	59.34	63.69	66.77	73.40	40
63.17	67.50	71.42	76.15	79.49	86.66	50
74.40	79.08	83.30	88.38	91.95	99.61	60
85.53	90.53	95.02	100.43	104.21	112.32	70
96.58	101.88	106.63	112.33	116.32	124.84	80
107.57	113.15	118.14	124.12	128.30	137.21	90
118.50	124.34	129.56	135.81	140.17	149.45	100
140.23	146.57	152.21	158.95	163.65	173.62	120
268.47	277.14	284.80	293.89	300.18	313.44	240