

EXAM

4301 – NATURAL SCIENCE METHODS

26.09.2014

Time: **9-12**

Language: ***English/bokmål***

#v of pages: ***10 (this page included)***

Help means: ***Calculator, statistical tables, English dictionary***

Notes: ***None***

Appendices: ***None***

The exam results will be made available on Studentweb.



Fakultet for allmennvitenskapelige fag

EXAM IN NATURAL SCIENCES METHODS
26.09.2014, kl. 08.30 - 12.30

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

1. Explain briefly the following statistical concepts: categorical variable

interval scale

response variable

SD

Poisson probability distribution

Central Limit Theorem

confidence interval

simple random sampling

cross-over design

replicates

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 sum of all the observations?
- c. What is the mean of this sample?
- d. What is the sum of the squares of the measurements?
- e. What is the variance of the sample?
- f. What is the standard deviation of this sample?
- g. What is the coefficient of variation of this sample?

3. What effect does reducing the value of the significance level from 0.05 to 0.01 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 sample size?

4. Some fish are famously able to develop into either sex, depending on social circumstances. One study of a coral reef fish, *Gobiodon erythrosilas*, placed juvenile fish with either an adult male or an adult female. Of the 12 juveniles placed with a male, 11 became female. Of the 10 juveniles placed with an adult female, six became male. What method can we use to test whether the social context of the juvenile fish affects which sex they became?

5. Male lizards in the species *Crotaphytus collaris* use their jaws as weapons during territorial interactions. Can weapons performance (bite force) predict territory size? The measurements for both variables are listed in the following table for 11 males:

Bite force (N)	Territory area (m ²)
28,2	437
33,9	589
29,5	871
39,8	977
41,7	1288
44,7	2138
46,8	2455
47,9	3548
36,3	2692
35,5	2042
33,9	3020
$\Sigma X =$	418,2
$\Sigma Y =$	20057
$\Sigma X^2 =$	16347,52
$\Sigma XY =$	803579,6
$\Sigma Y^2 =$	47633405
X mean =	38,01818182
Y mean =	1823,363636

- a. How rapidly does territory size increase with bite force? Calculate the equation for the regression line.

An ANOVA carried out to test the null hypothesis of zero slope for the regression yielded the following results

:

Source of variation	Sum of Squares	Df	Mean squares	F-ratio
Regression	3758539	1		
Residual	7303662	9		
Total				

- b. Complete the ANOVA table.
 c. Using the F-statistic, test the null hypothesis of zero slope at the significance level of $\alpha = 0,05$.
 d. What are your assumptions in part c?
 e. What does the MS_{residual} measure?
 f. Calculate the R^2 statistic. What does it measure?

6. A series of plot were placed in a large agricultural field in preparation for an experiment to investigate the effects of three fertilizers differing in their chemical composition. Before assigning treatments, it was noticed that plots differed along a moisture gradient. What strategy would you suggest the researcher implement to minimize the impact of this gradient on the ability to measure a treatment effect? Explain in words and with an illustration the experimental design you would recommend.

Formulae for Basic Statistics

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

$$s = \sqrt{\frac{\sum(Y_i - \bar{Y})^2}{n-1}} \quad s = \sqrt{\frac{\sum(Y_i^2) - n\bar{Y}^2}{n-1}}$$

Standard error of the mean

$$s / \sqrt{n}$$

χ^2 test of goodness-of-fit

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

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

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

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

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

Formulae for regression and correlation

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

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

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

Formulae for regression and correlation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$6 \sum d_i^2$$

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

ANOVA etc.

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

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

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

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

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

Kruskal-Wallis

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

Tukey-Kramer:

$$q = \frac{\bar{Y}_i - \bar{Y}_j}{SE} \quad SE = \sqrt{s_{\text{pooled}}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

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 , $\alpha(1)=0.05$, $\alpha(2)=0.10$

den. <i>df</i>	Numerator <i>df</i>									
	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	2.69	2.57	2.49	2.42	2.37	2.32

EKSAMEN I NATURVITENSKAPELIGE METODER
26.09.2014, kl. 08.30 - 12.30

Alle hovedspørsmål (1-6) teller likt og alle underspørsmål teller likt.

1. Forklar kort de følgende statistiske begreper: kategorisk variabel
interval skala
respons variabel
SD
Poisson probability distribution
Central Limit Theorem
konfidens interval
enkel tilfeldig sampling
cross-over design
replikater

2. Systolisk blodtrykk ble målt (I enhet mm Hg) i forbindelse med en forebyggende helseundersøkelse på innbyggere i Dallas, USA. Her er målingene for et lite utvalg av disse pasientene: 112, 128, 108, 129, 125, 153, 155, 132, 137.
 - a. Hvor mange individuer er det i samlet?
 - b. Hva er summen av alle observasjonene?
 - c. Hva er middelet i dette samlet?
 - d. Hva er kvadratsummen for målingene?
 - e. Hva er variansen i samlet?
 - f. Hva er standard avviket i dette samlet?
 - g. Hva er variasjonskoeffisienten i dette samlet?

3. Hvilken effekt har det å redusere verdien på signifikans nivået fra 0.05 til 0.01 på
 - a. Sannsynligheten for å gjøre en Type I feil?
 - b. Sannsynligheten for å gjøre en Type II feil?
 - c. Følsomheten til en test?
 - d. Sample størrelsen?

4. Noen fisk er kjent for evnen til enkeltindividet til å kunne utvikle seg til ulikt kjønn, avhengig av sosiale forhold. En undersøkelse av korallrev fisken *Gobiodon erythrosilas*, plasserte ung fisk sammen med enten en voksen hannfisk eller en voksen hunnfisk. Av de 12 unge som ble plassert sammen med en han, ble 11 til hunnfisk. Av de 10 unge som ble plassert sammen med en voksen hunn, utviklet seks seg til hannfisk. Hvilken metode kan vi bruke for å teste om sosiale forhold hos ungfisken påvirker hvilket kjønn de utvikler seg til?

5. Hanner av øgle arten *Crotaphytus collaris* bruker kjevene som våpen i territorielle kamper. Kan stryken på våpene (bitt kraft) forutsi størrelsen på territoriet? Målinger av begge variable er listen i følgende tabell for 11 hanner:

Bitt kraft (N)	Territorium (m ²)
28,2	437
33,9	589
29,5	871

	39,8	977
	41,7	1288
	44,7	2138
	46,8	2455
	47,9	3548
	36,3	2692
	35,5	2042
	33,9	3020
$\Sigma X =$		418,2
$\Sigma Y =$		20057
$\Sigma X^2 =$		16347,52
$\Sigma XY =$		803579,6
$\Sigma Y^2 =$		47633405
X mean =		38,01818182
Y mean =		1823,363636

- a. Hvor raskt øker territorium-størrelsen med litt kraften? Beregne ligningen for regresjons linjen.

En ANOVA gjennomført for å teste null hypotesen om ingen sammenheng, dvs. ingen linje I regresjonen , ga følgende resultater

Source of variation	Sum of Squares	Df	Mean squares	F-ratio
Regression	3758539	1		
Residual	7303662	9		
Total				

- b. Fyll inn ANOVA tabellen.
 c. Bruk F-statistic til å teste null hypotesen om ingen sammenheng med et signifikans nivå på $\alpha = 0,05$.
 d. Hva er forutsetningene for del c?
 e. Hva måler MS_{residual} ?
 f. Beregn R^2 statistic. Hva måler den?

6. En serie forsøksruter ble lag tut på et stort jorde for å gjøre et forsøk som skulle undersøke betydningen av tre forskjellige gjødselslag med ulik kjemisk sammensetning. Før de ulike behandlinger ble bestemt, la forskerne merke til at det var forskjeller mellom rutene pga. en fuktighetsgradient. Hva slags forsøksdesign vil du foreslå at forskerne bruker for å minimalisere betydningen av denne fuktighetsgradienten mht. å kunne måle betydningen av de ulike gjødselbehandlingene? Forklar i ord og med en illustrasjon den eksperimentelle design du vil anbefale.