

Tentamentsskrivning i Statistisk slutledning MVE155/MSG200, 7.5 hp.

Tid: Fredagen den 9 mars, 2012 kl 08.30-12.30

Examinator och jour: Serik Sagitov, tel. 772-5351, mob. 0736 907 613, rum H3026 i MV-huset.

Hjälpmedel: Chalmersgodkänd räknare, **egen** formelsamling (4 sidor på 2 blad A4) samt utdelade tabeller.

CTH: för "3" fordras 12 poäng, för "4" - 18 poäng, för "5" - 24 poäng.

GU: för "G" fordras 12 poäng, för "VG" - 20 poäng.

Important! For each problem do your best to

- describe and justify statistical models you apply,
 - state clearly hypotheses you test,
 - discuss different relevant approaches you have learned in the course.
-

1. (5 points) You have got a grant to measure the average weight of the hippopotamus at birth. You have seen in a previous publication by Stanley and Livingstone that for male calves the distribution of weights has a mean of roughly 70 kg and a standard deviation of 10 kg, while these numbers are 60 kg and 5 kg for females, but you are interested in a better remeasurement of the overall average.

The experimental procedure is simple: you wait for the herd of hippopotami to be sleeping, you approach a newborn, you put it quickly on the scales, and you pray for the mother not to wake up. You managed to weigh 13 female and 23 male newborns with the following results:

	Female	Male
Sample mean	62.8	69.7
Sample standard deviation	6.8	11.7

a. Test the null hypothesis of the equal sex ratio for the newborn hippopotami (meaning that the ratio of males to females at birth is 1 to 1).

b. Assuming the ratio of males to females at birth is 1 to 1, suggest two different unbiased point estimates for the overall average weight of the hippopotamus at birth.

c. Compute the standard errors for these point estimates.

d. What assumptions do you make for these calculations?

2. (5 points) Recognize important statistical terms hiding behind "bla-bla-bla" in the following extracts from the Internet (one term per item).

a. From a sample you can only get one value of a statistic like trimmed mean. You do not know the confidence interval of the trimmed mean or its distribution. Bla-bla-bla samples give more detail on the sampling distribution of the statistic of interest.

b. Bla-bla-bla is a measure of the "peakedness" of the probability distribution of a real-valued random variable, although some sources are insistent that heavy tails, and not peakedness, is what is really being measured by bla-bla-bla.

c. Note that bla-bla-bla is the probability of finding a difference that does exist, as opposed to the likelihood of declaring a difference that does not exist (which is known as a Type I error, or "false positive").

d. The bla-bla-bla is a measure of the tendency to fail; the greater the value of the bla-bla-bla, the greater the probability of impending failure... The bla-bla-bla is also known as the instantaneous failure rate.

e. Naive interpretation of statistics derived from data sets that include bla-bla-bla may be misleading. For example, if one is calculating the average temperature of 10 objects in a room, and most are between 20 and 25 degrees Celsius, but an oven is at 175 °C, the median of the data may be 23 °C but the mean temperature will be between 35.5 and 40 °C. In this case, the median better reflects the temperature of a randomly sampled object than the mean; however, naively interpreting the mean as "a typical sample", equivalent to the median, is incorrect. As illustrated in this case, bla-bla-bla may be indicative of data points that belong to a different population than the rest of the sample set.

3. (5 points) Officials of a small transit system with only five buses want to evaluate four types of tires with respect to wear. Applying a randomized block design, they decided to put one tire of each type on each of the five buses. The tires are run for 15,000 miles, after which the tread wear, in millimeters, is measured.

Bus	Tire 1	Tire 2	Tire 3	Tire 4	Mean
1	9.1	17.1	20.8	11.8	14.7
2	13.4	20.3	28.3	16.0	19.5
3	15.6	24.6	23.7	16.2	20.0
4	11.0	18.2	21.4	14.1	16.2
5	12.7	19.8	25.1	15.8	18.4
Mean	12.4	20.0	23.9	14.8	17.8

a. State the most appropriate null hypothesis by referring to a suitable parametric model. What are the main assumptions of the parametric model?

b. Using a non-parametric procedure test the null hypothesis of no difference between the four types of tires.

c. What kind of external effects are controlled by the suggested randomized block design? How the wheel positions for different tire types should be assigned for each of the five buses?

4. (5 points) A study is conducted of the association between the rate at which words are spoken and the ability of a "talking computer" to recognize commands that it is programmed to accept. A random sample of 50 commands is spoken first at a rate under 60 words per minute, and then the same commands are repeated at a rate over 60 words per minute. In the first case the computer recognized 42 out of 50 commands while in the second case it recognized only 35 commands. Is the observed difference statistically significant?

5. (5 points) . Suppose your prior beliefs about the probability p of success have mean $1/3$ and variance $1/32$. What is the posterior mean after having observed 8 successes in 20 trials?

6. (5 points) The data of the following table were gathered for an environmental impact study that examined the relationship between the depth of a stream and the rate of its flow

Depth	Flow rate
0.34	0.64
0.29	0.32
0.28	0.73
0.42	1.33
0.29	0.49
0.41	0.92
0.76	7.35
0.73	5.89
0.46	1.98
0.40	1.12

a. Draw the scatter plot for the given data using the x axis for depth. Fit by eye a regression line and plot the residuals against the depth. What does it say to you about the relevance of the simple linear regression model for this particular data?

b. The least square estimates for the parameters of the simple linear regression model are $b_0 = -3.98$, $b_1 = 13.83$. Given the standard deviations are $s_x = 0.17$ and $s_y = 2.46$ estimate the noise size (σ) and find the coefficient of determination.

c. The statistics for a quadratic model are given in the following table:

Coefficient	Estimate	Standard error	t value
β_0	1.68	1.06	1.59
β_1	-10.86	4.52	-2.40
β_2	23.54	4.27	5.51

Compute a 95 percent confidence interval for β_0 .

d. Is the quadratic term statistically significant? Carefully explain.

Statistical tables supplied:

1. Normal distribution table
2. Chi-square distribution table
3. t-distribution table
4. F-distribution table

Partial answers and solutions are also welcome. Good luck!

NUMERICAL ANSWERS

1a. Binomial model for the number of females $Y \sim \text{Bin}(36, p)$. Given $y_{\text{obs}} = 13$ we have to test $H_0 : p = 0.5$ against the two-sided alternative $H_1 : p \neq 0.5$. The approximate null distribution is $Y \sim N(18, 9)$, therefore, an approximate two-sided P-value becomes

$$2P(Y \leq 13) = 2P(Y < 14) \approx 2\Phi\left(\frac{13.5-18}{3}\right) = 2(1 - \Phi(1.5)) = 2 \times 0.067 = 13.4\%.$$

With such a high p-value we can not reject the null hypothesis of equal sex ratio.

1b. A simple sample mean

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n} = \frac{13 \times 62.8 + 23 \times 69.7}{36} = 67.2,$$

and a stratified sample mean

$$\bar{x}_s = \frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 = \frac{62.8 + 69.7}{2} = 66.3.$$

1c. The standard error of the stratified sample mean is

$$s_{\bar{x}_s} = \frac{1}{2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{1}{2}\sqrt{\frac{(6.8)^2}{13} + \frac{(11.7)^2}{23}} = 1.54.$$

To compute the sample variance from the simple random sample take some effort. First we observe that

$$\sum (x_{1i} - \bar{x}_1)^2 = \sum x_{1i}^2 - n\bar{x}_1^2.$$

It follows,

$$\begin{aligned} \sum x_{1i}^2 &= \sum (x_{1i} - \bar{x}_1)^2 + n_1\bar{x}_1^2 = (n_1 - 1)s_1^2 + n_1\bar{x}_1^2 \\ &= 12 \times (6.8)^2 + 13 \times (62.8)^2 = 51825, \\ \sum x_{2j}^2 &= (n_2 - 1)s_2^2 + n_2\bar{x}_2^2 \\ &= 22 \times (11.7)^2 + 23 \times (69.7)^2 = 114748, \\ s^2 &= \frac{\sum x_{1i}^2 + \sum x_{2j}^2}{35} - \frac{36}{35}\bar{x}^2 = 114.4. \end{aligned}$$

So that $s_{\bar{x}} = \sqrt{\frac{114.4}{36}} = 1.78$.

2a. Bootstrap.

2b. Kurtosis.

2c. Power of the test.

2d. Hazard function.

2e. Outliers.

3a. Under the two-way ANOVA model the most interesting is $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ the null hypothesis of no difference among different types of tires.

3b. The Friedman test

Bus	Tire 1	Tire 2	Tire 3	Tire 4
1	1	3	4	2
2	1	3	4	2
3	1	4	3	2
4	1	3	4	2
5	1	3	4	2
Mean rank	1.0	3.2	3.8	2.0

results in a test statistics $Q = 14.04$. The null distribution is approximated by a chi-square distribution with $df = 3$, whose table gives a p-value less than 0.5%. Reject H_0 .

4. Matched pairs design for 50 independent trials with four possible outcomes (correct, correct), (correct, wrong), (wrong, correct), (wrong, wrong). Assuming that in the slow regime the "talking computer" recognizes correctly all correct answers made in the fast regime we can summarize the results as follows

	Fast correct	Fast wrong	Totals
Slow correct	35	7	42
Slow wrong	0	8	8
Totals	35	15	50

McNemara's test statistics is $\frac{(7-0)^2}{7+0} = 7$. The null distribution is approximated by the χ_1^2 -distribution. Since the square root of 7 is 2.65, the standard normal distribution gives a (two-sided) P-value 0.8%. We conclude that the observed difference is statistically significant.

The conclusion will be different if our assumption is wrong. In the worst case the slow regime correct answers are totally different and the table of the outcomes looks as

	Fast correct	Fast wrong	Totals
Slow correct	27	15	42
Slow wrong	8	0	8
Totals	35	15	50

McNemara's test statistics is then $\frac{(7-0)^2}{8+15} = 2.13$. Since the square root of 2.13 is 1.46, the standard normal distribution gives a (two-sided) P-value 14%. We can not reject the null hypothesis in this case.

5. We use a Beta prior with parameters (a, b) satisfying

$$\frac{a}{a+b} = \frac{1}{3}, \quad \frac{\frac{1}{3}(1-\frac{1}{3})}{a+b+1} = \frac{1}{32}.$$

The prior pseudo-counts are well approximated by $a = 2$ and $b = 4$. Thus the posterior Beta distribution has parameters $(10, 16)$ giving the posterior mean estimate $\hat{p}_{\text{pme}} = 0.38$.

6b. First we find the sample correlation coefficient by $r = b_1 \frac{s_x}{s_y} = 0.96$. The coefficient of determination is $r^2 = 0.91$. Using formula

$$s^2 = \frac{n-1}{n-2} s_y^2 (1-r^2) = 0.589$$

the noise size is estimated as $s = \sqrt{0.589} = 0.77$.

6c. An exact 95% CI for β_0 is $b_0 \pm t_{n-3}(0.025)s_{b_0} = 1.68 \pm 2.365 \times 1.06 = [-0.83, 4.19]$.

6d. The observed test statistic $t = 5.51$ for the model utility test for $H_0 : \beta_2 = 0$ has an exact null distribution t_7 . After consulting the t_7 -distribution we reject this null hypothesis at 0.5% significance level. The quadratic term is therefore highly statistically significant.