LINKÖPINGS UNIVERSITET
Matematiska institutionen
Matematisk statistik

Utbildningskod: TAMS24
Modul: TEN1
Datum: 2019-08-23
Institution: MAI

Exam in Statistics

TAMS24/TEN1 2019-08-23

You are permitted to bring:

- a calculator (no computer);
- Formel- och tabellsamling i matematisk statistik (from MAI);
- Formel- och tabellsamling i matematisk statistik, TAMS65;
- TAMS24: Notations and Formulas (by Xiangfeng Yang).

Grading (sufficient limits): 8-11 points giving grade 3; 11.5-14.5 points giving grade 4; 15-18 points giving grade 5. Your solutions need to be complete, well motivated, carefully written and concluded by a clear answer. Be careful to show what is random and what is not. Assumptions you make need to be explicit. Approximations are allowed if reasonable and clearly motivated. The exercises are in random order.

Solutions can be found on the homepage a couple of hours after the finished exam.

ONCE upon a time, in a land far far away, Rick and Gary started a flamingo farm that was called Exodus. When starting out, Exodus was filled up by a large amount of flamingos that were obtained from a woman living in the swamps of Louisiana. The farm was built in southern Florida in a suitable habitat for flamingos. Both Gary and Rick quickly became rather proficient in the art of breeding flamingos. Aiming to export flamingos both to the animal parks of the world and to private citizens, Rick and Gary proudly produced commercials exclaiming their competence in flamingo breeding. Before not too long, they got a call from their very first customer. Things did not turn out the way they expected...

1. A Danish doughnut company – with a secret recipe – wants to buy large amounts of flamingos each month for some reason. They are prepared to pay a certain amount per kilogram of flamingos, so heavier flamingos render more profit. Gary picks out a random sample of flamingos and measures their weight:

$$2.69 \quad 2.90 \quad 3.23 \quad 3.52 \quad 2.65 \quad 3.71 \quad 3.46 \quad 3.05$$

Assume that the samples are independent and from a normal distribution with variance 0.0625 and an unknown expectation μ.

   (a) Test the hypothesis $H_0 : \mu = 3.0$ against $H_1 : \mu \neq 3.0$ at the significance level 5%. (2p)

   (b) What is the power of this test at $\mu = 3.1$? (1p)

   (c) What is the highest level of confidence we can choose when using this sample and still reject H_0? Is it reasonable to use this calculation to choose the significance level of the test you want to perform? (2p)

2. When using the results from the previous exercise to decide what to charge the Danish company, things did not turn out exactly as calculated. To avoid a faster disaster, Gary and Rick decided to not assume that the variance is known. Using the same sample as in the previous exercise, answer the following questions.

   (a) Test the assumption that the variance actually is equal to 0.0625 against the alternate hypothesis that the variance is greater. Use the significance level 0.05. (2p)

   (b) Assume that the variance is unknown. Find a confidence interval for μ with 99% degree of confidence. (1p)

3. It turns out that in the contract with the Danish company, there was some fine print detailing that the flamingos were to be slaughtered prior to shipping. Obviously upset, Rick and Gary devised a plan for euthanizing the flamingos as humanly as possible by means of the FLAMINGO DECAPITATOR 2000™ (a shovel headed killing machine). The blades are very sharp, but need additional sharpening after a certain time to keep the efficiency of the strike of the beast.

   The distributor of the blades (a company called *metal command*) claim that the time until sharpening is necessary (assuming a certain prescribed use) is exponentially distributed with the expectation 1.0 days. Rick and Gary puts this to the test using 50 identical machines in parallel (and in exactly the same way) over the course of 2.5 days. They take note every 6 hours of how many machines that has been taken out for sharpening (these machines are then kept out of circulation to not interfer with the measurements).

Time (hours)	< 6	< 12	< 18	< 24	< 30	< 36	< 42	< 48	< 54	< 60
Frequency:	11	20	26	32	36	39	42	44	46	47

   Use a suitable test with significance level 10% to see if we can reject the hypothesis that the samples are from an Exp($\mu = 1.0$)-distribution. (2p)

4. A company called *pleasures of the flesh* got into contact and offered to sell a growth hormone specifically tailored to birds of a similar type as flamingos. The company claimed that the size of the flamingos was linearly dependent on the amount of hormone administered. An experiment to investigate a reasonable dosage was carried out, but when studying residual plots from a linear regression there were hints of something quadratic.

Size (kg)	2.1344	2.3870	2.5861	2.8209	3.0492	3.2521	3.6765	4.0815
Hormone (mg/kg)	0.1250	0.2500	0.3750	0.5000	0.6250	0.7500	0.8750	1.0000

Consider the following two models:

$$\text{Model 1:} \quad Y = \beta_0 + \beta_1 x + \epsilon$$

and

$$\text{Model 2:} \quad Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$ and different measurements are assumed to be independent. The following calculations has already been carried out.

Model 1:

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	1.8036	0.0784
1	2.1242	0.1241

Analysis of variance

	Degrees of freedom	Square sum
REGR	1	2.9610
RES	6	0.0607
TOT	7	3.0217

Model 2:

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	2.0456	0.0813
1	0.9627	0.3315
2	1.0324	0.2876

Analysis of variance

	Degrees of freedom	Square sum
REGR	2	3.0047
RES	5	0.0170
TOT	7	3.0217

(a) Is the term in model 2 corresponding to x^2 meaningful? Carry out a test at the 5%-level. What is the interpretation of your result? (2p)

(b) Find a 99% confidence interval for β_1 using model 2. Does it differ from the corresponding confidence interval for β_1 using model 1? How do we interpret this? (1p)

5. Exodus runs into a problem of a certain species of fish that competes with the flamingos for food. The first idea for a solution was based on a chemical method called *Chemi-kill*, but due to fears of the effect on the flamingos this plan was scrapped. Fortunately, their close friend Susan stops by and claims that someone told her in a dream that introducing piranhas to the habitat would solve the problem.

   Gary contacts a Peruvian specialist Maria, who claims that there are two particularly ferocious types of red-bellied piranhas. Rick and Gary imports an equal amount of both types and devise an experiment where two identical tanks are filled with the different types of piranhas and 300 exemplars of the problem fish in each tank. What followed was a lesson in violence, where the starving piranhas went to attack. They stop the experiment after a day has passed and takes a count of the remaining problem fish. In the first tank there were 200 left and in the second 180. Let p_1 be the probability that a fish is eaten in the first tank and p_2 be the probability that a fish is eaten in the second tank.

   (a) Propose unbiased point estimators for p_1, p_2, and $p_2 - p_1$. Then find 95% confidence intervals for p_1, p_2 and $p_2 - p_1$. Can we reject that they're equally ferocious at this level? (2p)

   (b) After another call to Maria, she tells them that the second type might be a bit more aggressive. At the significance level 5%, can we reject the hypothesis that the types of piranhas are equally ferocious using the alternate hypothesis that the second type (corresponding to p_2) is more ferocious instead? (1p)

6. At the Exodus farm, a test is planned to verify certain conditions. To understand the test, Rick and Gary are going through the proof but got stuck at the following part. Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)^T$ be a probability vector, where $k \geq 2$ is an integer. Suppose that $\boldsymbol{Y} = (Y_1\ Y_2\ \cdots\ Y_k)^T \sim N(\boldsymbol{0}, \boldsymbol{C})$, where

$$\boldsymbol{C} = \begin{pmatrix} 1-p_1 & -\sqrt{p_1p_2} & -\sqrt{p_1p_3} & \cdots & -\sqrt{p_1p_k} \\ -\sqrt{p_2p_1} & 1-p_2 & -\sqrt{p_2p_3} & \cdots & -\sqrt{p_2p_k} \\ -\sqrt{p_3p_1} & -\sqrt{p_3p_2} & 1-p_3 & \cdots & -\sqrt{p_3p_k} \\ \vdots & \vdots & & \ddots & \vdots \\ -\sqrt{p_kp_1} & -\sqrt{p_kp_2} & -\sqrt{p_kp_3} & \cdots & 1-p_k \end{pmatrix}.$$

   It is stated in the proof that it now follows that $\boldsymbol{Y}^T\boldsymbol{Y} \sim \chi^2(k-1)$. Prove this. (2p)

Have a nice weekend!