Department of Mathematical Sciences
Examination paper for TMA4255

Academic contact during examination: Benjamin Dunn
Phone number for questions regarding exam content: **94147310**
Examination date: **27.05.2020**
Examination time: **09:00-13:30** (includes 30 minutes to upload answers)
Permitted examination support material: everything

Answer the questions on paper and submit a scan of your answers via the submission button.
You must show your work for each question.
You can answer the questions in English (preferred) or Norwegian.
There are a **total of 8 points possible**, with each question or question part worth the number of points stated in the parentheses.

1: To clarify the potential relationship between the flight speed (y-values) of the African swallow and the weight of its cargo (x-values), a knight gathered 10 data points. Assume a model of the form:

$$y_i = a + bx_i + \epsilon_i$$

for $i = 1, 2, ...10$, where each ϵ_i is independent and $\mathcal{N}(0, \sigma^2)$

The knight wished to make your life slightly easier and computed the following: $\sum_i x_i = 4.747$, $\sum_i x_i^2 = 3.440$, $\sum_i y_i = 32.924$, $\sum_i y_i^2 = 134.273$, $\sum_i x_i y_i = 10.108$, $SSE = 0.1898$ and $\sum_i (x_i - \bar{x})^2 = 1.187$. You should use a significance level of 0.05 for all of your work.

A (1 point): Perform an appropriate hypothesis test to check if there is evidence of a relationship between the flight speed and cargo weight.

We will assume a null hypothesis of $H_0 : b = 0$, $H_1 : b \neq 0$

$$\hat{b} = \frac{10 \times 10.108 - 4.747 \times 32.924}{10 \times 3.440 - 4.747^2} = -4.65$$
$$\hat{a} = \frac{32.924 - (-4.65) \times 4.747}{10} = 5.50$$
$$s = \sqrt{\frac{0.1898}{10 - 2}} = 0.154$$
$$SE(\hat{\beta}) = 0.154/\sqrt{1.187} = 0.141$$
$$t = -4.65/0.141 = -33.0$$

Since test statistic, t, is incredibly big (compare with value from table) we can safely reject the null hypothesis.

B (1 point): Someone claiming absolute knowledge of all things has stated that the actual mean flight speed of the African swallow carrying a load of 0.5 is 3. Perform an appropriate hypothesis test for this statement.

Let us make confidence intervals for our regression line at x_0:

$$\begin{aligned} CI &= \hat{a} + \hat{b} \times x_0 \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}} \\ &= 5.50 + (-4.65) \times 0.5 \pm 2.306 \times 0.154 \sqrt{\frac{1}{10} + \frac{(0.5 - 4.747/10)^2}{1.187}} \\ &= 3.175 \pm 0.113 \end{aligned}$$

Since 3 is outside the 95% confidence interval, we reject the H_0 with the note that that person was still pretty close so maybe not completely nuts.

C (1 point): It is well known that for European swallows the corresponding variance term, σ^2, for such an analysis is 0.02. Test the null hypothesis that $\sigma^2 = 0.02$ for African swallows.

We have that $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{n-2}$ so we can create confidence intervals for σ^2 as:

$$\begin{aligned} CI(\sigma^2) &= \left[\tfrac{(n-2) \times s^2}{17.535}, \tfrac{(n-2) \times s^2}{2.180}\right] \\ &= \left[\tfrac{0.1898}{17.535}, \tfrac{0.1898}{2.180}\right] \\ &= [0.011, 0.087] \end{aligned}$$

0.02 is in there so no reject guy.

D (1 point): After some discussion, the knights decided that the model was missing too many factors and came up with a new one with four factors: cargo weight (A), male or female (B), young or old (C) and cargo attachment method (D). They would like to do a proper 2^k factorial design but there just is not enough time. Two alternative designs have since been proposed with the first as shown below by the columns A, B, C and D. The second proposal has the same first three columns A, B, C but a different column, listed as D*, for the cargo attachment method. Which of the two proposals would you support and why?

A	B	C	D	D*
+	+	+	+	+
-	+	+	-	-
+	-	+	+	-
-	-	+	-	+
+	+	-	-	-
-	+	-	+	+
+	-	-	-	+
-	-	-	+	-

Note that D = AC and D* = ABC. If you check the alias structure for these guys, you will see all sorts of issues with using D = AC. The first obvious problem is that D is confounded with AC, while in the second proposal all of the main effects are aliased with three-factor interactions.

2 (1 point): A random variable X has a probability density function f given by $f(x) = \theta x^{\theta-1}$ for $0 < x < 1$ and $f(x) = 0$ for all other values of x, where $\theta > 0$ is a parameter. Find the maximum likelihood estimator of θ from n independent samples.

$$\begin{aligned}
\log L &= \log\left(\prod_i^n f(x_i)\right)\\
&= \sum_i^n \log(f(x_i))\\
&= \sum_i^n \log(\theta x^{\theta-1})\\
&= n\log(\theta) + \sum_i^n (\theta-1)\log(x_i)
\end{aligned}$$

$$\frac{\partial \log L}{\partial \theta} = n/\theta + \sum_i^n \log(x_i) = 0 \quad\longrightarrow\quad \theta = \frac{-n}{\sum_i^n \log(x_i)}$$

3: The people in the town near the castle Anthrax decided to compare the mean weight of the girls (G) with that of the local ducks (D). The weights are assumed to be independent samples from normal distributions (one for the duck and another for the girls) of unknown means and variances. The data is as follows:
3 girls, $N_G = 3$, with sample mean $\bar{x}_G = 50$ and sample variance $S_G^2 = 12$
4 ducks, $N_D = 4$, with sample mean $\bar{x}_D = 40$ and sample variance $S_D^2 = 14$

A (1 point): Assume, for just this part, that $\sigma_G^2 = \sigma_D^2$, find the 95% confidence interval for the difference between the mean weights of the ducks and girls ($\mu_D - \mu_G$).

$$\begin{aligned}
S_p^2 &= \frac{(3-1)\times 12 + (4-1)*14}{3+4-2} = 13.2\\
CI &= \bar{x}_D - \bar{x}_G \pm t_{\alpha/2} \times S_p\sqrt{1/N_D + 1/N_G}\\
&= 40 - 50 \pm 2.571\sqrt{13.2} \times \sqrt{1/4+1/3}\\
&= -10 \pm 7.13\\
&= [-17.13, -2.87]
\end{aligned}$$

B (1 point): Some people would prefer to further compare the mean weights of the ducks and girls with that of the local cats ($N_C = 4$, $\bar{x}_C = 20$ and $S_C^2 = 8$) but first they would like to test for equal variances, i.e. $H_0 : \sigma_D^2 = \sigma_C^2 = \sigma_G^2$, H_1 : the variances are not equal. Compute the appropriate test statistic to perform this hypothesis test for equal variances (note you do not need to compare this with the critical value or state if the H_0 should be rejected).

Bartlett's test!

$$\begin{aligned}
S_p^2 &= \frac{1}{N-3}\left[(N_D-1)S_D^2 + (N_G-1)S_G^2 + (N_C-1)S_C^2\right]\\
&= \frac{1}{3+4+4-3}\left[3\times 14 + 2\times 12 + 3\times 8\right] = 11.25\\
b &= \frac{\left[(S_D^2)^{N_D-1} \times (S_G^2)^{N_G-1} \times (S_C^2)^{N_C-1}\right]^{1/(N_D+N_G+N_C-3)}}{S_p^2}\\
&= \frac{\left[14^3 \times 12^2 \times 8^3\right]^{1/8}}{11.25} = 0.971
\end{aligned}$$

b is much bigger than the adjusted critical value, following the method in the book, so should end up rejected but that does not need to be stated.

4 (1 point): To better understand the migratory behavior of coconuts, the knights in the Kingdom of Mercia counted all of the coconuts that landed in the courtyard during the seasons X_1, X_2 and X_3 for five years. Note, in the Kingdom of Mercia there are conveniently only three seasons of equal length. Below are the counts of the coconuts during the three seasons over the corresponding years:

Year	X_1	X_2	X_3
516	15	30	8
517	9	4	10
518	16	11	6
519	2	3	17
520	4	7	9

Use a nonparametric test to test $H_0 : \mu_{X_1} = \mu_{X_2} = \mu_{X_3}$ with H_1 : the three means are not all equal.

Kruskal–Wallis test!

Year	X_1	X_2	X_3
516	15 (12)	30 (15)	8 (7)
517	9 (8.5)	4 (3.5)	10 (10)
518	16 (13)	11 (11)	6 (5)
519	2 (1)	3 (2)	17 (14)
520	4 (3.5)	7 (6)	9 (8.5)
R_i	38	37.5	44.5

If we follow the description in the book we would get:

$$\begin{aligned} h &= \tfrac{12}{N(N+1)} \sum_i^{seasons} R_i^2/n_i - 3(N+1) \\ &= \tfrac{12}{240}[38^2 + 37.5^2 + 44.5^2]/5 - 3 \times 16 \\ &= 48.305 - 48 \\ &= 0.305 \end{aligned}$$

h should be reasonably well approximated by a χ^2 distribution with 2 degrees of freedom. Since $h = 0.305$ is not larger than the estimated critical value of 5.991, we cannot reject the null hypothesis.

While the above method was taught in the course and is suggested in the book, there is also a slightly different form of the test statistic for the case of tied ranks (as we have here). Note: either answer is considered correct for this exam.

$$\begin{aligned} h &= \tfrac{1}{S^2}\left[\sum_i^{seasons} R_i^2/n_i - \tfrac{N(N+1)^2}{4}\right], \quad S^2 = \tfrac{1}{N-1}\left[\sum_i^{seasons}\sum_j^{years} R_{ij}^2 - \tfrac{N(N+1)^2}{4}\right] \\ h &= \tfrac{1}{S^2}[966.1 - 960], \qquad S^2 = \tfrac{1}{14}[12^2 + 8.5^2 + 13^2 \ldots + 8.5^2 - 960] \\ h &= \tfrac{6.1}{S^2}, \qquad S^2 = \tfrac{1}{14}[1239 - 960] \\ h &= 6.1/\left(\tfrac{279}{14}\right) \\ &= 0.306 \end{aligned}$$

As above, this is much smaller than 5.991 so we cannot reject the null hypothesis.