Mathematical Statistics Exam, September 11th, 2018, Set A

Fill in the dotted spaces ["....."]. 1 question (•) = 1 point; maximum = 14 points. Only responses in the specified places will be checked, but you need to include your notes with calculations when you return your exam. Fill in your responses **after having verified them**; if illegible or larded with corrections and crossings-out, the answers will be treated as wrong. You can use a calculator, statistical tables and one a4 sheet of paper with helpful formulas. **Communication with others is not allowed**.

NAME: ..... student's number .....

Signature .....

1. A social networking service company wishes to determine whether the type (a static picture or a film) and the frequency (single per visit or changing with each reload) of advertising content shown to users affects the duration of time spent using the service. The collected data for a random sample of users are summarized in the table below.

Advertising settings	Static ad, once	Static ad, changes	Film ad, once	Film ad, changes
Average time spent (hours)	1.6	1	1.2	1
Variance of time spent (biased estimator)	0.16	0.25	0.16	0.25
Number of users	20	20	20	20

   We assume that the time spent using the service follows a normal distribution. We conduct an analysis of variance test for the four ways of advertising to verify the null hypothesis that they way advertising is effected does not impact the average time spent using the service.

   - The sum of squares between groups (SSB) is equal to ....., and the sum of squares within groups (SSW) is equal to ..... The value of the appropriate test statistic is equal to .....
   - The critical region for a 5% significance level is equal to ....., so for this significance level we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate). Based on this result we can say that THE TYPE /THE FREQUENCY /NEITHER THE TYPE NOR THE FREQUENCY /THE TYPE AND THE FREQUENCY COMBINED have impact on the duration of service usage (underline the appropriate).

2. The number of users of a social networking service X in 2017 amounted to 4299817, while the number of users o service Y amounted to 6220800. The average *yearly* growth rate for the number of users of X for the period 2013-2017 was equal to 44%, and was the same as the average *two-year* growth rate for Y for this period.

   - In 2013, the number of users of service X was equal to ..... and the number of users of service Y was equal to .....
   - Should the trends continue, the number of users of social networking service X will surpass the number of users of service Y in year .....

3. A researcher is interested in the time until a new social media user becomes bored with the service and switches to a new provider. We assume that this time (in months) is a random variable X with density $f(x) = \frac{1}{\alpha x^{1/\alpha+1}}$ for $x > 1$ and 0 otherwise, where $\alpha \in (0, \frac{1}{2})$ is an unknown parameter. We have: $EX = \frac{1}{1-\alpha}$, $EX^2 = \frac{1}{(1-2\alpha)}$, $E \ln X = \alpha$, $E(\ln X)^2 = 2\alpha^2$, $E\frac{1}{X} = \frac{1}{\alpha+1}$. We have a sample consisting of n observations $X_1, X_2, \ldots, X_n$.

   - The method of moments estimator for parameter α based on the first moment is equal to
     $\hat{\alpha}_{MM} =$ .....,
     and if the sample consists of the following observations: 3, 4, 7, 10, this estimator is equal to .....
   - The maximum likelihood estimator for parameter α is equal to
     $\hat{\alpha}_{MLE} =$ .....

4. We continue with the analysis introduced in the previous problem, for a sample consisting of one observation X. Let $\hat{\alpha}_1 = \ln X$ be an estimator of α.

   - The bias of the estimator $\hat{\alpha}_1$ is equal to ....., the variance of this estimator is equal to ....., so the MSE of this estimator is equal to .....
   - The Fisher information connected with the single observation is equal to ....., so the estimator $\hat{\alpha}_1$ IS EFFICIENT /IS NOT EFFICIENT /WE SHOULD NOT BE TALKING ABOUT EFFICIENCY IN THIS CASE (underline the appropriate).

5. A company wishes to determine whether the type of advertisement campaign influences the sales of its products. During sample 30 days when the first campaign was in effect, the average of daily sales amounted to $ 1000, with a variance (unbiased estimator) equal to 100^2. During sample 30 days when the second campaign was in effect, the average daily sales amounted to $ 1100, with a variance (unbiased estimator) equal to 120^2. We assume that the sales distribution is normal.

   - We begin by verifying whether the variances for the two periods are equal, for a significance level equal to 10%. The value of the appropriate test statistic is equal to ....., the critical region is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate).
   - We verify the null hypothesis that the campaign type did not influence average sales, for a significance level 1%. The value of the appropriate test statistic is equal to ....., the critical value for the test is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis /CAN NOT PERFORM THE TEST GIVEN THE RESULTS OF THE PREVIOUS POINT (•) (underline the appropriate).

6. A company wishes to determine whether a social media -based advertisement campaign influences the sales of its products. Before the campaign was launched, the average of daily sales amounted to $ 1000, with a variance of 100^2. For a sample of 30 days after the campaign was launched, the average daily sales amounted to $ 1100, with a variance (unbiased estimator) equal to 120^2. We assume that the sales distribution is normal.

- We begin by verifying whether the variance of sales remained constant, against the alternative that it increased when the campaign was launched, for a significance level equal to 10%. The value of the appropriate test statistic is equal to ....., the critical region is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate).
- We verify the null hypothesis that average sales were not affected by the campaign, against the alternative that they increased when the campaign was launched, for a significance level 1%. The value of the appropriate test statistic is equal to ....., the critical value for the test is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis /CAN NOT PERFORM THE TEST GIVEN THE RESULTS OF THE PREVIOUS POINT (•) (underline the appropriate).

7. An analytical company researched the propensity of users to click on links in pop-up advertisements. In a random sample of 1000 users, 150 clicked on an ad when it popped up.

   - Based on this sample, provide the realization of a 95% confidence interval for the fraction of users who click on advertisements: .....
   - Most of those who clicked on the ad did so accidentally, only some users clicked intentionally. A 95% confidence interval for the fraction of users who click on the ad and do so intentionally, with respect to a 95% confidence interval for the fraction of users who click, IS CENTERED AROUND THE SAME POINT /HAS THE CENTER MORE TO THE RIGHT /HAS THE CENTER MORE TO THE LEFT /IMPOSSIBLE TO SAY WITHOUT ADDITIONAL DATA (underline the appropriate), and it is WIDER /NARROWER /THE SAME WIDTH /IMPOSSIBLE TO SAY WITHOUT ADDITIONAL DATA (underline the appropriate).

Mathematical Statistics Exam, September 11th, 2018, Set B

Fill in the dotted spaces [“.....”]. 1 question (•) = 1 point; maximum = 14 points. Only responses in the specified places will be checked, but you need to include your notes with calculations when you return your exam. Fill in your responses **after having verified them**; if illegible or larded with corrections and crossings-out, the answers will be treated as wrong. You can use a calculator, statistical tables and one a4 sheet of paper with helpful formulas. **Communication with others is not allowed**.

NAME: ..... student's number .....

Signature .....

1. A company wishes to determine whether the type of advertisement campaign influences the sales of its products. During sample 10 days when the first campaign was in effect, the average of daily sales amounted to $ 2000, with a variance (unbiased estimator) equal to 110^2. During sample 10 days when the second campaign was in effect, the average daily sales amounted to $ 2100, with a variance (unbiased estimator) equal to 120^2. We assume that the sales distribution is normal.

    - We begin by verifying whether the variances for the two periods are equal, for a significance level equal to 5%. The value of the appropriate test statistic is equal to ....., the critical region is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate).
    - We verify the null hypothesis that the campaign type did not influence average sales, for a significance level 5%. The value of the appropriate test statistic is equal to ....., the critical value for the test is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis /CAN NOT PERFORM THE TEST GIVEN THE RESULTS OF THE PREVIOUS POINT (•) (underline the appropriate).

2. A company wishes to determine whether a social media -based advertisement campaign influences the sales of its products. Before the campaign was launched, the average of daily sales amounted to $ 2000, with a variance of 110^2. For a sample of 10 days after the campaign was launched, the average daily sales amounted to $ 2100, with a variance (unbiased estimator) equal to 120^2. We assume that the sales distribution is normal.

    - We begin by verifying whether the variance of sales remained constant, against the alternative that it increased when the campaign was launched, for a significance level equal to 5%. The value of the appropriate test statistic is equal to ....., the critical region is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate).
    - We verify the null hypothesis that average sales were not affected by the campaign, against the alternative that they increased when the campaign was launched, for a significance level 5%. The value of the appropriate test statistic is equal

to ....., the critical value for the test is equal to ....., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis /CAN NOT PERFORM THE TEST GIVEN THE RESULTS OF THE PREVIOUS POINT (•) (underline the appropriate).

3. A social networking service company wishes to determine whether the type (a static picture or a film) and the frequency (single per visit or changing with each reload) of advertising content shown to users affects the duration of time spent using the service. The collected data for a random sample of users are summarized in the table below.

Advertising settings	Static ad, once	Static ad, changes	Film ad, once	Film ad, changes
Average time spent (hours)	1.4	1	1.4	1
Variance of time spent (biased estimator)	0.25	0.36	0.25	0.36
Number of users	10	10	10	10

We assume that the time spent using the service follows a normal distribution. We conduct an analysis of variance test for the four ways of advertising to verify the null hypothesis that they way advertising is effected does not impact the average time spent using the service.

- The sum of squares between groups (SSB) is equal to ....., and the sum of squares within groups (SSW) is equal to ..... The value of the appropriate test statistic is equal to .....
- The critical region for a 10% significance level is equal to ....., so for this significance level we REJECT /DO NOT HAVE GROUNDS TO REJECT the null hypothesis (underline the appropriate). Based on this result we can say that THE TYPE /THE FREQUENCY /NEITHER THE TYPE NOR THE FREQUENCY /THE TYPE AND THE FREQUENCY COMBINED have impact on the duration of service usage (underline the appropriate).

4. A researcher is interested in the time until a new social media user becomes bored with the service and switches to a new provider. We assume that this time (in months) is a random variable X with density $f(x) = \frac{1}{\beta x^{1/\beta+1}}$ for $x > 1$ and 0 otherwise, where $\beta \in (0, \frac{1}{2})$ is an unknown parameter. We have: $EX = \frac{1}{1-\beta}$, $EX^2 = \frac{1}{(1-2\beta)}$, $E \ln X = \beta$, $E(\ln X)^2 = 2\beta^2$, $E\frac{1}{X} = \frac{1}{\beta+1}$. We have a sample consisting of n observations $X_1, X_2, \ldots, X_n$.

- The method of moments estimator for parameter β based on the second moment is equal to
  $\hat{\beta}_{MM} =$ .....,
  and if the sample consists of the following observations: 3, 4, 7, 10, this estimator is equal to .....
- The maximum likelihood estimator for parameter β is equal to
  $\hat{\beta}_{MLE} =$ .....

5. We continue with the analysis introduced in the previous problem, for a sample consisting of one observation X. Let $\hat{\beta}_1 = \ln X$ be an estimator of β.

- The bias of the estimator $\hat{\beta}_1$ is equal to ....., the variance of this estimator is equal to ....., so the MSE of this estimator is equal to .....

- The Fisher information connected with the single observation is equal to ....., so the estimator $\hat{\beta}_1$ IS EFFICIENT /IS NOT EFFICIENT /WE SHOULD NOT BE TALKING ABOUT EFFICIENCY IN THIS CASE (underline the appropriate).

6. An analytical company researched the propensity of users to click on links in pop-up advertisements. In a random sample of 1200 users, 120 clicked on an ad when it popped up.

   - Based on this sample, provide the realization of a 90% confidence interval for the fraction of users who click on advertisements: .....
   - Most of those who clicked on the ad did so accidentally, only some users clicked intentionally. A 95% confidence interval for the fraction of users who click on the ad and do so intentionally, with respect to a 90% confidence interval for the fraction of users who click, is CENTERED AROUND THE SAME POINT /HAS THE CENTER MORE TO THE RIGHT /HAS THE CENTER MORE TO THE LEFT /IMPOSSIBLE TO SAY WITHOUT ADDITIONAL DATA (underline the appropriate), and it is WIDER /NARROWER /THE SAME WIDTH /IMPOSSIBLE TO SAY WITHOUT ADDITIONAL DATA (underline the appropriate).

7. The number of users of a social networking service X in 2017 amounted to 2143589, while the number of users o service Y amounted to 2928200. The average *yearly* growth rate for the number of users of X for the period 2013-2017 was equal to 21%, and was the same as the average *two-year* growth rate for Y for this period.

   - In 2013, the number of users of service X was equal to ..... and the number of users of service Y was equal to .....
   - Should the trends continue, the number of users of social networking service X will surpass the number of users of service Y in year .....