Mathematical Statistics Exam, June 11th, 2019, Set A

Fill in the dotted spaces [".........."]. 1 question (•) = 1 point; maximum = 16 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 simple calculator, statistical tables and one a4 sheet of paper with helpful formulas. **Communication with the rest of the world is not allowed**.

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

Signature ....

1. A researcher studied the amounts spent by car owners on car repairs. The results of a survey with 2000 individuals are summarized in the table below:

Amount spent (in dollars)	$[0, 1000)$	$[1000, 2000)$	$[2000, 3000)$	$[3000, 4000)$	$[4000, 5000)$
Number of owners of brand F cars	100	100	200	300	300
Number of owners of brand G cars	200	200	200	200	200

   - Based on the presented results, the average amount spent by owners of brand F cars may be approximated as ...., while the average amount spent by brand G car owners may be approximated as .... The variance of the amount spent by brand F car owners amounts to .... and is HIGHER /LOWER /THE SAME (underline the appropriate) as the variance for brand G car owners.
   - The median of the amounts spent by car owners of brand F cars may be approximated as .... Based on the skewness coefficient (using the median), which amounts to ...., we may say that the distribution of the amount spent by brand F car owners is LEFT SKEWED /RIGHT SKEWED /SYMMETRIC (underline the appropriate).

2. Based on the data above, we construct 95% confidence intervals.

   - The 95% confidence interval for the average amount spent by brand F car owners is equal to ....
   - The 95% confidence interval for the fraction of brand F car owners who spend more than 2000 dollars on repairs is equal to ...., and is WIDER /NARROWER /THE SAME LENGTH (underline the appropriate) as an analogous confidence interval for the fraction for brand G car owners.

3. Surveys were conducted in two countries to compare the behaviour of car owners. In country X, among 1024 individuals surveyed, 512 bought a car in the previous 3 years, while in country Y, among 1024 individuals surveyed, this number amounted to 550.

   - We conduct a test to verify whether the fractions of individuals who buy cars are equal for the two countries under study. The value of the appropriate test statistic is equal to ...., its p -value amounts to ...., so for a significance level $\alpha = 0.05$ we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

- Note that in country X we have: $512/1024 = \frac{1}{2}$. Is the test conducted in the previous point equivalent to verifying whether in country Y the fraction is equal to $\frac{1}{2}$? YES, it is the same test /YES, it is an equivalent test but for a different significance level /YES, it is an equivalent test but it is one-sided rather than two-sided /NO, this is not an equivalent test (underline the appropriate).

4. Let X be a single observation from a binomial distribution for $n = 6$ trials and an unknown probability of success $\theta \in (0, 1)$. Construct the most powerful test for the verification of the null hypothesis that $H_0 : \theta = \frac{1}{2}$ against the alternative that $H_1 : \theta = \frac{4}{5}$.

   - The critical region for the most powerful test for a significance level $\alpha = 0.109375$ is equal to ...., and for a significance level $\alpha = 0.1$ would be equal to ....
   - The power of the test for a significance level $\alpha = 0.109375$ amounts to .... The power of this test used to test against the alternative hypothesis that $\theta = \frac{9}{10}$ would be HIGHER /LOWER /THE SAME (underline the appropriate).

   Hint. The distribution of random variable X under the null and alternative hypotheses is:

$k =$	0	1	2	3	4	5	6
$P(X = k)$ for $\theta = \frac{1}{2}$	0.015625	0.09375	0.234375	0.3125	0.234375	0.09375	0.015625
$P(X = k)$ for $\theta = \frac{4}{5}$	0.000064	0.001536	0.01536	0.08192	0.24576	0.393216	0.262144

5. Let us assume that X – the number of times that an individual buys a new car througho-ut his lifetime – may be modeled by a Poisson distribution with unknown parameter $\theta > 0$, i.e. $P(X = x) = \frac{\theta^x}{x!}e^{-\theta}$ for $x = 0, 1, \ldots$. We have a sample of n observations: $X_1, X_2, \ldots, X_n$.

   - Researcher A constructs the method of moments estimator based on the mean $\hat{\theta}_M$. This estimator is equal to: $\hat{\theta}_M =$ .... and is the same as the maximum likelihood estimator for θ in this case YES /NO (underline the appropriate).
   - The estimator constructed in the previous point has the following characteristics:
     - The bias of this estimator is equal to ....
     - The asymptotic bias of this estimator is equal to ....
     - The asymptotic efficiency of this estimator is equal to ....

6. Continuing the analysis of the number of times an individual buys a new car as in the problem above, Researcher B constructs the maximum posterior probability estimator for θ, assuming a prior distribution for θ of the form $\pi(\theta) = 2e^{-2\theta}$ for $\theta > 0$.

   - This estimator is equal to: $\hat{\theta}_{MP} =$ ....
   - The estimator $\hat{\theta}_{MP}$ constructed in the previous point has the following characte-ristics:
     - The bias of this estimator is equal to ....
     - The variance of this estimator is equal to ....
     - The MSE of this estimator is equal to ....

7. Continuing the topic of the probability of buying new cars by individuals, Researcher C concentrated on the probability that an individual buys a new car during a calendar year. She conducted a survey with $n = 100$ respondents and asked them whether they bought a new car in 2017 and/or in 2018. The results of the survey are summarized in the table below:

Number of times respondent bought a new car in 2017 and 2018	0	1	2
number of observations in sample	$n_1 = 48$	$n_2 = 44$	$n_3 = 8$
theoretical probability of outcome	$(1-p)^2$	$2p(1-p)$	p^2

We conduct a chi-squared goodness of fit test for a significance level $\alpha = 0.05$ to determine whether it is possible that individuals have the same, unknown probability $p \in (0, 1)$ of buying a new car in 2017 and 2018 and that purchases in 2017 and 2018 are independent (resulting in a binomial distribution, as specified above).

- The Maximum Likelihood estimator for p has the formula $\hat{p}_{ML} =$ .... and for the observed values is equal to ....
- The test statistic of the chi-squared goodness of fit test has a value of .... The appropriate critical value of the chi-squared distribution with .... degrees of freedom is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

8. A sample of 16 car users were interrogated in order to determine the number of kilometers traveled daily by car. In the studied sample, the average amounted to 24, with a standard deviation (based on the unbiased estimator of the variance) equal to 9. We assume that distances traveled follow a normal distribution.

- We verify the null hypothesis that car users travel on average 25 km daily against the alternative that they travel less, for a significance level of 0.1. The value of the appropriate test statistic is equal to ...., the critical region of the appropriate test is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.
- We verify the null hypothesis that the variance of daily travels amounts to 100, against the alternative that it is smaller, for a significance level of 0.1. The value of the appropriate test statistic is equal to ...., the critical region of the appropriate test is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

Fill in the dotted spaces [“..........”]. 1 question (•) = 1 point; maximum = 16 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 simple calculator, statistical tables and one a4 sheet of paper with helpful formulas. **Communication with the rest of the world is not allowed**.

NAME: .... student’s number ....

Signature ....

1. A sample of 16 car users were interrogated in order to determine the number of kilometers traveled daily by car. In the studied sample, the average amounted to 20, with a standard deviation (based on the unbiased estimator of the variance) equal to 4. We assume that distances traveled follow a normal distribution.
   - We verify the null hypothesis that car users travel on average 15 km daily against the alternative that they travel more, for a significance level of 0.05. The value of the appropriate test statistic is equal to ...., the critical region of the appropriate test is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.
   - We verify the null hypothesis that the variance of daily travels amounts to 10, against the alternative that it is larger, for a significance level of 0.05. The value of the appropriate test statistic is equal to ...., the critical region of the appropriate test is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.
2. Surveys were conducted in two countries to compare the behaviour of car owners. In country X, among 900 individuals surveyed, 450 bought a car in the previous 3 years, while in country Y, among 900 individuals surveyed, this number amounted to 500.
   - We conduct a test to verify whether the fractions of individuals who buy cars are equal for the two countries under study. The value of the appropriate test statistic is equal to ...., its p -value amounts to ...., so for a significance level $\alpha = 0.1$ we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.
   - Note that in country X we have: $450/900 = \frac{1}{2}$. Is the test conducted in the previous point equivalent to verifying whether in country Y the fraction is equal to $\frac{1}{2}$? YES, it is the same test /YES, it is an equivalent test but for a different significance level /YES, it is an equivalent test but it is one-sided rather than two-sided /NO, this is not an equivalent test (underline the appropriate).
3. A researcher studied the amounts spent by car owners on car repairs. The results of a survey with 2000 individuals are summarized in the table below:

Amount spent (in dollars)	$[0, 1000)$	$[1000, 2000)$	$[2000, 3000)$	$[3000, 4000)$	$[4000, 5000)$
Number of owners of brand F cars	300	300	200	100	100
Number of owners of brand G cars	100	200	400	200	100

- Based on the presented results, the average amount spent by owners of brand F cars may be approximated as ...., while the average amount spent by brand G car owners may be approximated as .... The variance of the amount spent by brand F car owners amounts to .... and is HIGHER /LOWER /THE SAME (underline the appropriate) as the variance for brand G car owners.
- The median of the amounts spent by car owners of brand F cars may be approximated as .... Based on the skewness coefficient (using the median), which amounts to ...., we may say that the distribution of the amount spent by brand F car owners is LEFT SKEWED /RIGHT SKEWED /SYMMETRIC (underline the appropriate).

4. Based on the data above, we construct 90% confidence intervals.

   - The 90% confidence interval for the average amount spent by brand F car owners is equal to ....
   - The 90% confidence interval for the fraction of brand F car owners who spend more than 3000 dollars on repairs is equal to ...., and is WIDER /NARROWER /THE SAME LENGTH (underline the appropriate) as an analogous confidence interval for the fraction for brand G car owners.

5. Let us assume that X – the number of times that an individual buys a new car throughout his lifetime – may be modeled by a Poisson distribution with unknown parameter $\lambda > 0$, i.e. $P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}$ for $x = 0, 1, \ldots$. We have a sample of n observations: $X_1, X_2, \ldots, X_n$.

   - Researcher A constructs the method of moments estimator $\hat{\lambda}_M$ based on the mean. This estimator is equal to: $\hat{\lambda}_M =$ .... and is the same as the maximum likelihood estimator for λ in this case YES /NO (underline the appropriate).
   - The method of moments estimator constructed in the previous point has the following characteristics:
     - The bias of this estimator is equal to ....
     - The asymptotic bias of this estimator is equal to ....
     - The asymptotic efficiency of this estimator is equal to ....

6. Continuing the analysis of the number of times an individual buys a new car as in the problem above, Researcher B constructs the maximum posterior probability estimator for λ, assuming a prior distribution for λ of the form $\pi(\lambda) = 3e^{-3\lambda}$ for $\lambda > 0$.

   - This estimator is equal to: $\hat{\lambda}_{MP} =$ ....
   - The estimator $\hat{\lambda}_{MP}$ constructed in the previous point has the following characteristics:
     - The bias of this estimator is equal to ....
     - The variance of this estimator is equal to ....
     - The MSE of this estimator is equal to ....

7. Continuing the topic of the probability of buying new cars by individuals, Researcher C concentrated on the probability that an individual buys a new car during a calendar year. She conducted a survey with $n = 100$ respondents and asked them whether they

bought a new car in 2017 and/or in 2018. The results of the survey are summarized in the table below:

Number of times respondent bought a new car in 2017 and 2018	0	1	2
number of observations in sample	$n_1 = 66$	$n_2 = 28$	$n_3 = 6$
theoretical probability of outcome	$(1-p)^2$	$2p(1-p)$	p^2

We conduct a chi-squared goodness of fit test for a significance level $\alpha = 0.1$ to determine whether it is possible that individuals have the same, unknown probability $p \in (0,1)$ of buying a new car in 2017 and 2018 and that purchases in 2017 and 2018 are independent (resulting in a binomial distribution, as specified above).

- The Maximum Likelihood estimator for p has the formula $\hat{p}_{ML} =$ .... and for the observed values is equal to ....
- The test statistic of the chi-squared goodness of fit test has a value of .... The appropriate critical value of the chi-squared distribution with .... degrees of freedom is equal to ...., so we REJECT /have NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

8. Let X be a single observation from a binomial distribution for $n = 6$ trials and an unknown probability of success $\theta \in (0,1)$. Construct the most powerful test for the verification of the null hypothesis that $H_0 : \theta = \frac{1}{2}$ against the alternative that $H_1 : \theta = \frac{1}{5}$.

- The critical region for the most powerful test for a significance level $\alpha = 0.109375$ is equal to ...., and for a significance level $\alpha = 0.05$ would be equal to ....
- The power of the test for a significance level $\alpha = 0.109375$ amounts to .... The critical region of a test for the same null hypothesis and significance level but against the alternative hypothesis that $\theta = \frac{1}{10}$ would be WIDER /NARROWER /THE SAME (underline the appropriate).

Hint. The distribution of random variable X under the null and alternative hypotheses is:

$k =$	0	1	2	3	4	5	6
$P(X = k)$ for $\theta = \frac{1}{2}$	0.015625	0.09375	0.234375	0.3125	0.234375	0.09375	0.015625
$P(X = k)$ for $\theta = \frac{1}{5}$	0.262144	0.393216	0.24576	0.08192	0.01536	0.001536	0.000064