Mathematical Statistics Exam, June 20th, 2018, Set A

Fill in the dotted spaces [".........."]. 1 question (•) = 1 point; maximum = 14 points. Only responses in the specified places will be checked, without justifications or calculations. 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. The results of a survey conducted with unemployed individuals in two small towns are presented in the table below:

Length of unemployment (in months)	$[0,6)$	$[6,12)$	$[12,18)$	$[18,24)$	$[24,30)$
Number of respondents in city A	15	20	30	20	15
Number of respondents in city B	25	20	10	20	25

   - In city A, the average duration of unemployment is equal to ...., the variance of the length of unemployment is equal to ...., and the median value of unemployment duration is equal to ....
   - In city B, the average duration of unemployment is HIGHER /LOWER /THE SAME as in A (underline the appropriate), the variance of the duration of unemployment is HIGHER /LOWER /THE SAME as in A (underline the appropriate), and the median value of unemployment duration is HIGHER /LOWER /THE SAME as in A (underline the appropriate).

2. We want to verify whether the duration of unemployment (in years) in a certain town can be described by a $\Gamma(1,2)$ distribution, against the alternative that the distribution is $\Gamma(3,2)$, on the base of a single observation. We construct the most powerful test for these two hypotheses, for a significance level $\alpha = 0.2$.

   - The critical region of the most powerful test in this case is equal to ....
   - If the observed duration of unemployment was equal to 1 month, the decision would be to REJECT /NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis, and if the observed value was equal to 2 months, the decision would be to REJECT /NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

   Hint. The density of a $\Gamma(k,\lambda)$ distribution is equal to $f(x) = \frac{\lambda^k}{(k-1)!}x^{k-1}e^{-\lambda x}$ for $x > 0$, for integer values of k and $\lambda > 0$.

3. We continue the topic of unemployment duration in small towns; we assume that in general it follows a $\Gamma(k,\lambda)$ distribution, where k is a fixed, known integer, and $\lambda > 0$ is an unknown parameter. Our sample consists of observations $X_1, X_2, \ldots, X_n$.

- Researcher A constructs the Maximum Likelihood estimator for λ. This estimator has the form $\hat{\lambda}_{ML} =$ ...., and if $k = 2$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 3$, the estimate is equal to $\hat{\lambda}_{ML} =$ ....
- Researcher B constructs the Bayesian Most Probable Estimator, based on a prior distribution with density $f(\lambda) = 2e^{-2\lambda}$ for $\lambda > 0$. This estimator has the form $\hat{\lambda}_{BMP} =$ ...., and if $k = 2$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 3$, the estimate is equal to $\hat{\lambda}_{BMP} =$ ....

4. A HR specialist is especially interested in two traits of workers: whether they are diligent (or lazy) and whether they are creative (or not). We assume that the two traits are independent, and that the probabilities that a worker is diligent and lazy are each equal to $\frac{1}{2}$. We assume that the probability that a worker is creative is equal to p, where $p \in (0, 1)$ is unknown. A group of workers were asked to take a test; the results are summarized in the table below:

Type of worker	Diligent and creative	Diligent and not creative	Lazy and creative	Lazy and not creative
Number of workers	$N_1 = 10$	$N_2 = 30$	$N_3 = 20$	$N_4 = 40$
Assumed probabilities	$\frac{1}{2}p$	$\frac{1}{2}(1-p)$	$\frac{1}{2}p$	$\frac{1}{2}(1-p)$

   - Find the formula for the MLE of p (depending on N_1, N_2, N_3, N_4):

     $\hat{p} =$ .... and the value of the estimator for the above sample: ....
   - Using the chi-squared test (and the estimate from the previous point), verify whether the assumed probability scheme is valid for a significance level 0.05. The value of the appropriate test statistic is equal to ...., under the null hypothesis the statistic has a distribution with .... degrees of freedom and a critical value equal to ...., so we REJECT /HAVE NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

5. We analyze the wages of workers, depending on their traits. In a random sample of workers, the average wage for a group of 200 workers identified as lazy amounted to \$ 2400, with a standard deviation equal to \$ 500. Meanwhile, for a group of 200 workers identified as diligent, the average amounted to \$ 2600, with a standard deviation also equal to \$ 500. Standard deviations were calculated on the base of the unbiased estimator of the variance.

   - We want to verify whether the average wages in the two groups are the same. The value of the appropriate test statistic is equal to ...., the critical region of the test for a significance level $\alpha = 0.01$ has the form ...., so we REJECT /HAVE NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis for this significance level.
   - If the same averages and variances were found for a sample consisting of 8 lazy workers and 8 diligent workers, the result of a testing procedure analogous to the previous point WOULD BE THE SAME /WOULD BE DIFFERENT /WE CAN'T SAY – IT WOULD DEPEND ON THE ASSUMED DISTRIBUTION OF WAGES (underline the appropriate).

6. We analyze the fraction of employees who arrive late in the office. On Monday, out of 240 individuals employed in a company, 180 were late.

- Provide the realization of a 90% confidence interval for the fraction of employees who arrive late, based on this sample: ....
- We verify the null hypothesis that the fraction of workers who arrive late is equal to 75%, against the alternative that it is lower. The value of the appropriate test statistic is equal to ...., the p-value of this result is equal to ...., so for a significance level of $\alpha = 0.05$ we REJECT /HAVE NO GROUNDS TO REJECT the null (underline the appropriate).

7. A researcher is interested in the duration of employment of managers in public companies. We assume that this duration (in months) may be described by a uniform distribution over the interval $(1, \theta)$, where $\theta > 1$ is an unknown parameter. We construct $\hat{\theta}_{MM}$, the method of moments estimator for θ, based on the first moment of the distribution for a sample of size n.

   - This estimator has the form $\hat{\theta}_{MM} =$ .... If in a random sample of 10 individuals, the average duration was equal to 12 months, with a variance equal to 900, the estimator of the parameter θ would be equal to ....
   - The bias for $\hat{\theta}_{MM}$ is equal to ...., the variance is equal to ....,

     so the Mean Square Error for the formula is equal to ....

Mathematical Statistics Exam, June 20th, 2018, Set B

Fill in the dotted spaces ["..........."]. 1 question (•) = 1 point; maximum = 14 points. Only responses in the specified places will be checked, without justifications or calculations. 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. We analyze the fraction of employees who arrive late in the office. On Monday, out of 360 individuals employed in a company, 240 were late.
   - Provide the realization of a 95% confidence interval for the fraction of employees who arrive late, based on this sample: ....
   - We verify the null hypothesis that the fraction of workers who arrive late is equal to 50%, against the alternative that it is higher. The value of the appropriate test statistic is equal to ...., the p-value of this result is equal to ...., so for a significance level of $\alpha = 0.01$ we REJECT /HAVE NO GROUNDS TO REJECT the null (underline the appropriate).
2. The results of a survey conducted with unemployed individuals in two small towns are presented in the table below:

Length of unemployment (in months)	$[0, 6)$	$[6, 12)$	$[12, 18)$	$[18, 24)$	$[24, 30)$
Number of respondents in city A	25	20	10	20	25
Number of respondents in city B	15	20	30	20	15

   - In city A, the average duration of unemployment is equal to ...., the variance of the length of unemployment is equal to ...., and the median value of unemployment duration is equal to ....
   - In city B, the average duration of unemployment is HIGHER /LOWER /THE SAME as in A (underline the appropriate), the variance of the duration of unemployment is HIGHER /LOWER /THE SAME as in A (underline the appropriate), and the median value of unemployment duration is HIGHER /LOWER /THE SAME as in A (underline the appropriate).
3. We analyze the wages of workers, depending on their traits. In a random sample of workers, the average wage for a group of 200 workers identified as lazy amounted to \$ 2300, with a standard deviation equal to \$ 1000. Meanwhile, for a group of 200 workers identified as diligent, the average amounted to \$ 2500, with a standard deviation also equal to \$ 1000. Standard deviations were calculated on the base of the unbiased estimator of the variance.

- We want to verify whether the average wages in the two groups are the same. The value of the appropriate test statistic is equal to ...., the critical region of the test for a significance level $\alpha = 0.05$ has the form ...., so we REJECT /HAVE NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis for this significance level.
- If the same averages and variances were found for a sample consisting of 8 lazy workers and 8 diligent workers, the result of a testing procedure analogous to the previous point WOULD BE THE SAME /WOULD BE DIFFERENT /WE CAN'T SAY – IT WOULD DEPEND ON THE ASSUMED DISTRIBUTION OF WAGES (underline the appropriate).

4. A HR specialist is especially interested in two traits of workers: whether they are diligent (or lazy) and whether they are creative (or not). We assume that the two traits are independent, and that the probabilities that a worker is diligent and lazy are each equal to $\frac{1}{2}$. We assume that the probability that a worker is not creative is equal to p, where $p \in (0, 1)$ is unknown. A group of workers were asked to take a test; the results are summarized in the table below:

Type of worker	Diligent and creative	Diligent and not creative	Lazy and creative	Lazy and not creative
Number of workers	$N_1 = 20$	$N_2 = 40$	$N_3 = 10$	$N_4 = 30$
Assumed probabilities	$\frac{1}{2}(1-p)$	$\frac{1}{2}p$	$\frac{1}{2}(1-p)$	$\frac{1}{2}p$

   - Find the formula for the MLE of p (depending on N_1, N_2, N_3, N_4):

     $\hat{p}$ = ... and the value of the estimator for the above sample: ....
   - Using the chi-squared test (and the estimate from the previous point), verify whether the assumed probability scheme is valid for a significance level 0.01. The value of the appropriate test statistic is equal to ...., under the null hypothesis the statistic has a distribution with .... degrees of freedom and a critical value equal to ...., so we REJECT /HAVE NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

5. A researcher is interested in the duration of employment of managers in public companies. We assume that this duration (in months) may be described by a uniform distribution over the interval $(2, \theta)$, where $\theta > 2$ is an unknown parameter. We construct $\hat{\theta}_{MM}$, the method of moments estimator for θ, based on the first moment of the distribution for a sample of size n.

   - This estimator has the form $\hat{\theta}_{MM}$ = ... If in a random sample of 12 individuals, the average duration was equal to 10 months, with a variance equal to 900, the estimator of the parameter θ would be equal to ....
   - The bias for $\hat{\theta}_{MM}$ is equal to ...., the variance is equal to ....,

     so the Mean Square Error for the formula is equal to ....

6. We want to verify whether the duration of unemployment (in years) in a certain town can be described by a $\Gamma(1, 2)$ distribution, against the alternative that the distribution is $\Gamma(5, 2)$, on the base of a single observation. We construct the most powerful test for these two hypotheses, for a significance level $\alpha = 0.3$.

- The critical region of the most powerful test in this case is equal to ....
- If the observed duration of unemployment was equal to 2 months, the decision would be to REJECT /NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis, and if the observed value was equal to 4 months, the decision would be to REJECT /NO GROUNDS TO REJECT (underline the appropriate) the null hypothesis.

Hint. The density of a $\Gamma(k, \lambda)$ *distribution is equal to* $f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x}$ *for* $x > 0$, *for integer values of* k *and* $\lambda > 0$.

7. We continue the topic of unemployment duration in small towns; we assume that in general it follows a $\Gamma(k, \lambda)$ distribution, where k is a fixed, known integer, and $\lambda > 0$ is an unknown parameter. Our sample consists of observations $X_1, X_2, \ldots, X_n$.

   - Researcher A constructs the Maximum Likelihood estimator for λ. This estimator has the form $\hat{\lambda}_{ML} =$ ...., and if $k = 3$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 4$, the estimate is equal to $\hat{\lambda}_{ML} =$ ....
   - Researcher B constructs the Bayesian Most Probable Estimator, based on a prior distribution with density $f(\lambda) = 3e^{-3\lambda}$ for $\lambda > 0$. This estimator has the form $\hat{\lambda}_{BMP} =$ ...., and if $k = 3$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 4$, the estimate is equal to $\hat{\lambda}_{BMP} =$ ....