Mathematical Statistics Exam, September 2016, set A

Fill in the dotted spaces [".........."]. 1 question (•) = 1 point; maximum = 16 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. Cyclical market surveys are conducted in two seaside resorts. Each year, sales in 15 randomly chosen establishments in each location are analyzed. Aggregate results for selected sales articles are presented in the table below.

		Fish and chips		A bottle of mineral water	
		Average price (zł.)	Average daily sales (number of items)	Average price (zł.)	Average daily sales (number of items)
Dębki	2014	21	70	3	100
	2015	22	70	4	150
	2016	23	80	5	100
Sopot	2014	25	70	4	100
	2015	27	80	5	120
	2016	29	90	6	120

- The average annual growth rate of mineral water bottle prices in Sopot during the 2014-2016 period is equal to .... Assuming that the growth rate will prevail next year (in 2017), the average price will amount to ....
- Based on the PRICE /QUANTITY index according to the LASPEYRES /PAASCHE formula (underline the appropriate) we can say that the aggregate sales value for the two studied articles in Sopot in 2016 rose .... times with respect to 2014 due to changes in the sales quantities, assuming 2014 prices.

2. Deepening the analysis of data presented in Problem 1 above, price variances were calculated. It appears that both in Dębki and in Sopot, in all the studied years, the variance (unbiased estimator) amounted to 9. We assume that prices are distributed normally.

- We verify the null hypothesis H that in 2014, fish and chips were equally expensive in Sopot as in Dębki, against the alternative that Sopot was more expensive. The value of the test statistic for the appropriate test is equal to

..............................,

the critical region of the test for a significance level 0.01 has the form

....................,

so we REJECT /DO NOT HAVE GROUNDS TO REJECT H (underline the appropriate).

- If the researchers did not calculate the variance, but rather used the results of previous research (we know that fish and chips prices have invariably standard deviations equal to 3 zł, regardless of location), the value of the test statistic for the appropriate test would be equal to

  ....,

  and the critical region of the test for a significance level 0.01 would have the form

  ....

3. Continuing the analysis of data presented in Problem 1, we conduct an analysis of variance test to verify the hypothesis that the average price level of mineral water in Sopot during the years 2014 - 2016 did not change. Standard deviations (calculated on the base of unbiased estimators) of prices, for all the three studied years, amounted to 3 zł.

   - The sum of squares within groups is equal to ....,

     while the sum of squares between groups is equal to ....,

     so the value of the test statistic is equal to ....,
     and in effect for a significance level of 0.05 we REJECT /DO NOT HAVE GROUNDS TO REJECT the hypothesis of equality of means (underline the appropriate).
   - Which statements are true? Underline the appropriate:
     - In our case the analysis of variance test can't be conducted without the assumption that the variables have normal distributions. TRUE /FALSE
     - In our case the analysis of variance test can't be conducted without the assumption that the establishments do not function for more than one season. TRUE /FALSE
     - In our case the analysis of variance test can't be conducted without the assumption that the number of establishments functioning each year in Sopot is the same. TRUE /FALSE

4. Analyzing the sample of 15 fish and chips prices in Dębki in 2016, the researchers assumed that these can be described with a uniform distribution over the interval $(23 - \theta, 23 + \theta)$ for an unknown parameter $\theta > 0$.

   - The method of quantiles estimator for θ, based on the first quartile, has the form

     $$\hat{\theta}_Q = \ldots\ldots\ldots\ldots$$

     and assuming that in our sample this quartile is equal to 20.5, the estimated value is

     ....

- The method of moments estimator for θ, based on the variance, has the form

$$\hat{\theta}_M = \dots\dots\dots\dots\dots\dots$$

and assuming that the sample variance is equal to 9, the estimated value is

$$\dots\dots\dots\dots\dots\dots$$

5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with density $f(x) = \frac{1}{\mu}e^{-\frac{x-1}{\mu}}$ for $x > 1$ and 0 otherwise, where $\mu > 0$ is an unknown parameter.

   - Let m be an estimator of μ given by the formula $m = \bar{X}$.

     Calculate the bias of estimator m: ....
   - Fisher Information related to a single observation from this distribution is equal to
     $I_1(\mu) =$ ....

   Hint. $\int_c^\infty xe^{-\lambda x}dx = \frac{(e^{-c\lambda})(c\lambda+1)}{\lambda^2}$

6. Let us assume that the price levels of mineral water in Sopot in 2016 may be described by a distribution with density

$$g(x) = \begin{cases} \frac{1}{\theta^2}\left(x - (6-\theta)\right) & x \in (6-\theta, 6], \\ \frac{1}{\theta^2}(6+\theta-x) & x \in (6, 6+\theta), \\ 0 & \text{otherwise,} \end{cases}$$

   for an unknown parameter $\theta > 0$. Based on a single observation we verify the hypothesis that $\theta = 3$ against the alternative that $\theta = 2$ using a test with a critical region of the form

$$|X - 6| > c$$

   for a constant c selected appropriately to obtain a significance level 0.1.

   - The value of c is equal to ....
   - The power of the test for $\theta = 2$ is equal to ...., so the proposed test is VERY GOOD /SO SO /BAD (underline the appropriate).

7. 100 randomly chosen individuals in Sopot were interrogated on a given day with respect to the number of fish and chips portions they have eaten that day. The answers are summarized in the table below:

Number of portions	0	1	2
Number of respondents	$k_0 = 25$	$k_1 = 50$	$k_2 = 25$
Assumed probability	p^2	$2p(1-p)$	$(1-p)^2$

   - Assuming that the number of portions eaten by an individual follows a binomial distribution with parameters 2 and $1-p$, the maximum likelihood estimator for p was calculated. The likelihood function, using the notation form the table above, has the form

$$L(p) = \binom{k_0 + k_1 + k_2}{k_0} \binom{k_2 + k_1}{k_1} p^{2k_0} (2p(1-p))^{k_1} (1-p)^{2k_2},$$

so the value of the ML estimator of p in the studied sample is equal

to ....

- The hypothesis that the distribution of the number of fish and chips portions eaten is binomial was tested using a chi-squared test for a significance level 0.05. The value of the test statistic is equal to

  ...., the critical value of the test with .... degrees of freedom is equal to ...., so we REJECT /DO NOT HAVE GROUNDS TO REJECT the null (underline the appropriate).

8. Using data from the previous problem, we analyze the fraction of respondents who ate two portions of fish and chips.

   - The hypothesis that this fraction is equal to $\frac{3}{10}$ was tested against the alternative that it is not equal to $\frac{3}{10}$. The value of the appropriate test

     statistic is equal to ...., the corresponding $p - value$ is equal to ...., so for a significance level of 0.1 we REJECT /DO NOT HAVE GROUNDS TO REJECT the null (underline the appropriate).

   - A 95-percent confidence interval for the fraction was built. The realization is equal to:

     ....