Mathematical Statistics Exam, June 2015, set B

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 ....

1. Mortgage values from applications in Bank ABC were analyzed. Previous analyses have shown that the mortgage value may be described by a normal distribution with a standard deviation of 100 thousand dollars. The mean mortgage value in a sample of 36 consumers was equal to 440 thousand dollars.

   - The realization of a 95% confidence interval for the mean mortgage value is

     ....

   - Data were analyzed further, and it appeared that indeed, in the studied sample the standard deviation (calculated on the base of the unbiased estimator of the variance) was equal to 100 thousand dollars. One of the bank employees proposed to use the sample standard deviation to calculate the confidence interval. In this case, the realization of a 95% confidence interval for the mean mortgage value is

     ....

     and this interval is LONGER /THE SAME LENGTH /SHORTER (underline the appropriate) than the confidence interval from the previous point.

2. Let $X_1, \ldots, X_n$ be a random sample from a distribution with density

   $$f_\beta(x) = 3e^{-3(x-\beta)}, \text{ for } x > \beta,$$

   and 0 otherwise, for an unknown parameter $\beta > 0$.

   - The Maximum Likelihood Estimator of β is given by:

     $\hat{\beta} =$ ....

   - Let $\hat{\beta}_2 = \bar{X} + \frac{1}{3}$ be an estimator of β. The bias of this estimator is equal to

     ....

3. Monthly output of a coal mine was analyzed. Data in the form of a series of chained simple indices for the year 2014 is presented below:

Month	1	2	3	4	5	6	7	8	9	10	11	12
Index	1.1	1	0.9	0.9	1	0.9	0.9	1	0.9	0.9	1	1.1

- The lowest output in 2014 was observed in month .... and provided that in December 2013 the output was 400 thousand tons, this lowest value was equal to ....
- Chain indices describing output changes for the I, II, III and IV quarters of the year are equal to

  I : ...., II : ...., III : ...., IV : ....,

  and the average rate of change for these quarters was ....

4. Let $X_1, \ldots, X_n$ be independent random observations from a distribution with density

$$f_\theta(x) = e^{\theta - x} e^{-e^{\theta - x}}$$

for an unknown $\theta > 0$.

- The Method of Quantiles Estimator for θ, based on the median of the distribution, is equal to

  ....

- Fisher information connected with a sample size on n is equal to

  ....

Hint. $\int \left(e^{\theta-x}\right)^k e^{-e^{\theta-x}} dx = e^{-e^{\theta-x}} \left(e^{\theta-x}+1\right)^{k-1} + c$, *for* $k = 1, 2$.

5. The duration of functioning of online shops (in months) in a given trade was analyzed, with respect to the shop's marketing strategy (prices significantly lower than in stationary shops, prices similar to stationary shops). Aggregate results, divided into four classes of duration, are presented below:

Time in months	(0,12]	(12,24]	(24,36]	(36,48]	Total
Number of shops with lower prices	40	60	80	60	240
Number of shops with equal prices	60	40	40	20	160
Total	100	100	120	80	400

We verify the hypothesis that the probability of being in a given class of duration does not depend on the marketing strategy, with the use of a chi-squared test.

- The distribution of the test statistic has .... degrees of freedom, and the value of the test statistic for the sample is

  ....

- For significance level $\alpha = 0.05$, the critical region of the test is equal to ...., so we REJECT /DO NOT HAVE GROUNDS TO REJECT H_0 (underline the appropriate).

6. On the base of data from the previous problem, the hypothesis that the fractions of shops which functioned less than two years are the same for both groups is tested against the alternative that for shops having prices similar to stationary shops this fraction is higher.

   - The test statistic for the verification of the above hypothesis is equal to ...., so for a significance level $\alpha = 0.05$ we REJECT /DO NOT HAVE GROUNDS TO REJECT H_0 (underline the appropriate).
   - The *p-value* of the test statistic form the previous point is equal to .... and is HIGHER /THE SAME /LOWER (underline the appropriate), than if the alternative were that the fractions are different.

7. Again based on the data presented in Problem 5, the hypothesis that the mean duration of a shop in the given trade is equal to 24 months is tested against the alternative that it is shorter. The standard deviation calculated on the base of the data equals 9 months (based on the unbiased estimator of the variance).

   - The mean duration time is equal to ...., and the critical region for the verification of the hypothesis for significance level $\alpha = 0.01$ is equal to

     ....

   - The test statistic is equal to ...., so for significance level $\alpha = 0.01$ we REJECT /DO NOT HAVE GROUNDS TO REJECT H_0 (underline the appropriate).

8. The time spent on a bus stop waiting for the bus (in minutes) has a uniform distribution over the interval $(0, \theta)$, where $\theta > 0$ is an unknown parameter. Two statisticians use the bus; they have observed the following waiting times (denoted. $X_1, \ldots, X_5$):

   $$4,\ 3,\ 6,\ 1,\ 9.$$

   - Let us assume that the first statistician wants to verify the null hypothesis that $\theta = 8$ against the alternative that $\theta = 10$ with a test such that the critical region is $\{X_{5:5} > c\}$ for a constant c. For a test with significance level $\alpha = 0.1$, the constant c should be equal to:

     ....

     and the decision based on the observations is to REJECT /NO GROUNDS TO REJECT H_0 (underline the appropriate).

   - Let us assume that the second statistician, based on previous experience, supposes that the distribution of θ is given by a density

     $$\pi(\theta) = \frac{4}{6}\left(\frac{6}{\theta}\right)^5 \quad \text{for } \theta > 6,$$

     and zero otherwise. The Bayesian Most Probable estimate of θ, based on this *a priori* distribution and the observed sample, is equal to

     ....