PRACTICE final exam, Math 170S
Instructor: Elizaveta Rebrova

Printed name: ______________________________

Signed name: ______________________________

Student ID number: ______________________________

Instructions:

- Read problems very carefully. If you have any questions please ask.
- **The correct final answer alone is not sufficient for full credit - you should explain your solutions,** saving enough time to attempt all the problems.
- You final answers does not need to be completely simplified (unless otherwise stated), e.g. a sum of several decimal fractions would be completely fine. Ask if in doubt.
- Books, calculators, phones are not allowed, please put them away. If you need more paper please raise your hand.

Question	Points	Score
1	10	
2	10	
3	8	
4	12	
5	10	
6	10	
7	10	
8	10	
9	10	
10	10	
Total:	100	

1. ***For this problem only***: *give a specific, but short (one sentence) answer. If several things can be done with the help of a specific test, it is enough to clearly identify one of them.* ***Explain all the notations you use.***

   (a) (2 points) What hypothesis can be tested with the help of Wilcoxon test?

   (b) (2 points) What hypothesis can be tested with the help of ANOVA test(s)?

   (c) (2 points) In the model for simple linear regression we assume that data points (x_i, Y_i) are close to a line, namely, $Y_i = \alpha_1 + \beta x_i + \varepsilon_i$. What assumption do we make about the noise ε_i?

   (d) (2 points) If we plot sample quantiles of some data versus quantiles of standard normal distribution (q-q plot) and the resulting points are close to a straight line, what conclusion can we make?

   (e) (2 points) Suppose we are using some test for a hypothesis, and we would like to improve both Type I and Type II errors with the same test. What can we do?

2. Consider the distributions $N(\mu_X, 400)$ and $N(\mu_Y, 225)$. Let $\theta = \mu_X - \mu_Y$. Let $\bar{x}$ and $\bar{y}$ denote the observed means of two independent random samples, each of size n, from the respective distributions. Say we reject $H_0 : \theta = 0$ and accept $H_1 : \theta > 0$ if $\bar{x} - \bar{y} \geq c$. Let $K(\theta)$ be the power function of the test.

   (a) (1 point) By definition, power function

   $$K(t) = \mathbb{P}\,(\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad)\,.$$

   (b) (2 points) The distribution of

   $$\bar{x} - \bar{y} \sim$$

   Give the parameters of the distribution too.

   (c) (7 points) Find n and c, so that $K(0) = 0.05$ and $K(10) = 0.90$, approximately.

3. Let $X_1, X_2, ..., X_{50}$ be a random sample from Poisson distribution with parameter $\lambda > 0$.

   (a) (4 points) We will reject H_0: $\lambda = 2$ and accept H_1: $\lambda > 2$ if the observed sum $\sum_{i=1}^{50} X_i$ is bigger than 110. Write a numerical expression that computes the exact Type I error probability.

   (b) (4 points) Give a normal approximation for the probability obtained in part (a). In this part you should get a numerical answer.

   Hint: The Poison distribution with parameter λ has pmf $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. Also, $\mathbb{E}(X) = \mathrm{Var}(X) = \lambda$. Also, if $X_i \sim Poi(\lambda)$ and independent, then $\sum_{i=1}^{m} X_i \sim Poi(m\lambda)$

4. Let $X_1, \ldots, X_{10}$ be a random sample from $N(0, \sigma^2)$.

   (a) (4 points) Show that most powerful critical region for testing $H_0 : \sigma^2 = 9$ against $H_1 : \sigma^2 = 2$ can be defined using statistic $\sum_{i=1}^{10} x_i^2$.

   (b) (4 points) Find the best critical region of the size $\alpha = 0.05$.

   (c) (4 points) For the test defined by this critical region find probability of Type II error.

   Hint: Recall that $\sum_{i=1}^{n} X_i^2/\sigma^2 \sim \chi^2(n)$.

5. (10 points) On some magical island, there were elves, dwarfs, orcs and goblins. A researcher arrives to the island and decides to interview all four spices about their history and traditions. She invites 10 representatives of each folk for an interview. However, she soon finds out that at least some of the island inhabitants are quite mean and aggressive and don't interview well (whereas the rest are quite nice and friendly). So, the researcher wonders: are there more and less aggressive spices on the island? She records the data she got so far with in the following table:

	nice	not nice
elves:	3	7
dwarfs:	5	5
orcs:	4	6
goblins:	8	2

How can she test on the significance level 10% a hypothesis that being nice is a spices independent feature?

Specify what test you will use. Show all the steps. Make a conclusion (whether the researcher should accept or reject her hypothesis about independence).

6. Density function of X is defined as

$$f_X(t) = \begin{cases} \frac{4}{\theta^2}t, & \text{when } 0 < t < \theta/2, \\ -\frac{4}{\theta^2}t + \frac{4}{\theta}, & \text{when } \theta/2 < t < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Unknown parameter $\theta \in (0, 2]$.

(a) (5 points) Find estimator of θ by the method of moments.

(b) (3 points) Is it biased or not?

(c) (2 points) For the following observations, give a point estimate for θ:

$$0.3 \quad 0.2 \quad 0.5 \quad 0.3 \quad 0.2 \quad 0.2$$

7. (10 points) Suppose that the observation X has exponential density function $f_X(t) = \theta e^{-\theta t}$ for $t \geq 0$, where θ is an unknown parameter. You know that θ is either 2 or 4. Prior the experiment, you have no idea which one is more likely (so, we assume that two options are equally likely). Then you observe in 3 experiments that $X = 7, 2, 1$. Find the posterior distribution of Θ and MAP estimator of Θ. Give numerical values as an answer.

   Hint 1: Posterior distribution (as well as prior) is discrete and can take only two distinct values. So, posterior distribution = new probabilities of each possible value of θ.

   Hint 2: $e^{20} \gg 8$.

8. A sample size of $n = 100$ is taken from the production of light switches. A quality control engineer discovered that 10 switches does not meet the quality standard.

   (a) (5 points) Suppose that another 500 switches are to be produces using the same technology. What are minimal and maximal amounts of defective switches should we expect with 95% confidence?

   Hint: Use confidence interval for the defect rate of the production (=proportion of bad switches) to answer this question.

   (b) (5 points) Estimate the confidence (=probability) that among the next 500 switches will be at most 60 defective switches.

9. You test a random number generator which supposed to give 0, 1 or 2 equally likely. You observe 80 zeros, 105 ones, and 115 twos, using the same generator many times in a row.
   (a) (5 points) Using the Chi-Square test, do you accept the hypothesis that your generator is good (gives 0, 1 and 2 equally likely indeed) at the 10% significance level? What about 1% significance level?
   (b) (3 points) Approximate p-value for this test.
   (c) (2 points) What is the largest significance level on which you would accept the above hypothesis (that the generator is good)?

10. Let $X_{(1)} < X_{(2)} < \ldots < X_{(5)}$ be the order statistics of five independent observations $X_1, \ldots, X_5$ taken from normal $N(0,4)$ distribution. Compute the probabilities

    (a) (5 points) $\mathbb{P}(X_{(1)} < 1)$

    (b) (5 points) $\mathbb{P}(X_{(4)} < 5)$

    You should give a **numerical** answer in this problem. But the answer does not have to be completely simplified.