This is a take home midterm. You can use your notes, books or any other source you need. You are supposed to work on your own text without external help. I'll be available to answer question in person or via email.

To solve the Exam problems, I have not collaborated with anyone nor sought external help and the material presented is the result of my own work.

Signature: ____________________

Name: ____________________

Question:	1	2	Total
Points:	50	50	100
Score:			

Question:	1	2	Total
Bonus Points:	0	20	20
Score:			

Question 1 . . . . *50 point*

Let x_i, $i = 1, \ldots, 48$, be the result of 48 independent measurements of the amount of pollen grains in the air in Atlanta in the spring of 2020. You know that

$$\sum_{i=1}^{48} x_i = 492.02 \qquad \sum_{i=1}^{48} x_i^2 = 5420.4$$

(a) (10 points) Assume that the results come from a normal distribution with both expected value μ and variance σ^2 unknown. Find the Confidence Interval for μ at a confidence level of 99%.

Solution: We have

$$\bar{x} = \frac{492.02}{48} = 10.25 \qquad s^2 = \frac{1}{47}\left(5420.4 - \frac{492.02^2}{48}\right) = 8.02$$

so that, using that $t_{47,0.005} = 2.68$ we get

$$10.25 - \frac{2.68\sqrt{8.02}}{\sqrt{48}} \leq \mu \leq 10.25 + \frac{2.68\sqrt{8.02}}{\sqrt{48}}$$

or

$$9.15 \leq \mu \leq 11.35$$

(b) (10 points) You decide to hold the following test for the true population mean μ

- H_0: $\mu \geq \mu_0$ where $\mu_0 = 11$
- H_a: $\mu < \mu_0$

at a significance level of 0.01. Give the 0.01 significance level rejection region. Will you reject H_0 at this significance level?

Solution: If

$$T = \frac{\sqrt{N}(\bar{X} - \mu_0)}{\sqrt{\frac{S^2}{N-1}}}$$

then the test is $T < -t_{47,0.01} = -2.40$. With our data we get

$$t = -1.83$$

so that we do not reject H_0.

(c) (10 points) Given the results in part (a), compute the p-value of the test.

Solution: The p-value of the test is $\Pr(T < -1.83) = 0.037$ where T is a t r.v. with 47 d.o.f. Thus

$$P = 0.037\,.$$

(d) (10 points) Compute the probability $\beta(10, 8)$ of a Type 2 error when the true expected value is $\mu = 10$ and the true value of $\sigma^2 = 8$. Both Matlab and R can compute the c.d.f. and inverse c.d.f. of a non central t distribution (see page 579 in the textbook).

Solution: Calling $T_N(x|\psi) = \Pr(T \leq x)$ when T is a noncentral t r.v. with N d.o.f. and noncentrality ψ we have, with $\psi = -\sqrt{48/8} = -2.45$,

$$\beta(10, 8) = 1 - T_{47}(-2.40| - 2.45) = 0.48$$

(e) (10 points) How big should your sample be in such a way that $\beta(10,8) < 0.2$. Give an estimate or lower bound.

Solution: We need N such that

$$\beta(10,8) = 1 - T_{N-1}\left(T_{N-1}^{-1}(0.01 \,|\, 0)\,\Big|\, - \sqrt{\frac{N}{8}}\right) < 0.2\,.$$

The smallest such N is $N = 84$.

Question 2 . . . . *50 point*

A random sample of size N is selected from a population with distribution

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{(x-A)}{\theta}} & x \geq A \\ \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\theta > 0$ and A are parameter.

(a) (10 points) Use Maximum Likelihood to find estimators $\hat{\theta}$ for θ and $\hat{A}$ for A.

Solution: Let X_i be the r.v. forming your sample. The likelihood function is

$$l(A, \theta) = \theta^{-N} \exp\left(-\sum_{i=1}^{N} \frac{(x_i - A)}{\theta}\right)$$

if $x_i > A$ foe every i and 0 otherwise. Thus the maximum over A is reached when A is equal to the smallest of the x_i that is

$$\hat{A} = \min_i X_i := \underline{X}$$

fixed A we can differentiate over θ and get

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} X_i - \hat{A} = \overline{X} - \underline{X}$$

(b) (10 points) Assuming that θ is known, find the p.d.f. of $\hat{A}$.

Solution: Since teh X_i are independent we have

$$\Pr(\hat{A} > a) = \prod_i \Pr(\hat{X}_i > a) = \begin{cases} \exp\left(-\frac{n(a-A)}{\theta}\right) & a > A \\ 0 & \text{otherwise} \end{cases}$$

(c) (10 points) Still assuming that θ is known, use the properties of

$$R = \frac{\hat{A} - A}{\theta/N}$$

to create a test δ_α for the hypotheses

- H_0: $A \leq 1$
- H_a: $A > 1$

with a given level α.

Solution: Clearly a value of $\hat{A}$ much bigger than $A_0 = 1$ would be a strong evindence against H_0. This means a value of

$$R = \frac{\hat{A} - A_0}{\theta/N}$$

greater than 0. Thus our test is $\delta : R > c$. Observe that $R - N(A - A_0)/\theta$ is a exponential r.v. with mean 1. Thus

$$\sup_{A<A_0} \Pr(R > c \,|\, A) = \sup_{A<A_0} e^{-(c-N(A-A_0)/\theta)} = e^{-c}$$

so that

$$\delta_\alpha : \; R > -\log(\alpha)\,.$$

(d) (10 points) Find the power function $\pi(A, \delta_\alpha)$ and the p-value (as a function of the sample result) for the test in part (c).

Solution: The power function is

$$\Pr(R > c \mid A) = \begin{cases} \alpha e^{N(A-A_0)/\theta} & A < A_0 + \theta \log(\alpha)/N \\ 1 & \text{otherwise} \end{cases}$$

If r is the result of the sample for the R statistics we get

$$P = \sup_{A<A_0} \Pr(R > r \mid A) = e^{-r}$$

(e) (10 points (bonus)) Assume now that θ is unknown, find the joint p.d.f. of $\hat{\theta}$ and $\hat{A}$. (**Hint:** if X_1 and X_2 are two independent exponential r.v. with the same expected value, find the j.p.d.f. of $Y_1 = \min(X_1, X_2)$ and $Y_2 = X_1 + X_2$.)

Solution:

(f) (10 points (bonus)) Still assuming that θ is unknown, use the result of part (e) to find a test δ for the hypotheses in part (d) at a given level α. Discuss the power function and p-value of this test.