Spring 2020 **Statistics 330** **Final Exam**
(75 points)

Please write your first and last name here:

Name _____

Instructions:

- Partial credit will be given only if you show your work.
- Reason out your answers. In many cases, a line or two of justification is enough.
- The questions are roughly in the order in which the material is presented in class, so they are not necessarily ordered easiest to hardest.
- If you get stuck on one, it may be a good idea to move on and come back to that question at the end.
- You may use your prepared notes, a calculator, and tables only.

1. **Descriptive Statistics**

   A quantitative variable has the following statistics:

   $$Min = 0 \quad Q_1 = 20 \quad Med = 25 \quad Q_3 = 34 \quad Max = 70$$

   Circle all values below that would be considered an outlier. (2 points)

   a. 0
   b. 8
   c. 18
   d. 48
   e. 58

2. Consider the data set: 24 32 33 37 49 35 31

   (a) Calculate the mean, median and IQR (5 points)

   (b) Check for outliers using the 1.5(IQR) rule, and indicate which data points are outliers. (4 points)

   (c) Remove any outliers and recalculate the mean, median, IQR. If there are no outliers, then you can say "answer same as (a)" (5 points)

3. **Bias/Unbiased Estimators and Mean Square Error (MSE)**

Let $X_1, X_2, \ldots, X_n$ be iid pois(λ) random variables. Recall $\mathbb{E}(X) = \lambda$ and $var(X) = \lambda$. Consider three estimators of λ:

$$\hat{\lambda}_1 = \frac{X_1+2X_2}{3}, \qquad \hat{\lambda}_2 = \overline{X}_n, \qquad \hat{\lambda}_3 = 5$$

(a) Calculate the expected value of each estimator. Which estimators are biased and which are unbiased? (4 points)

(b) Calculate the variance of each estimator. (4 points)

(c) Give the Mean Square Error of each estimator. Recall: $MSE(\hat{\theta})$ can be written as $\{\mathbb{E}(\hat{\theta}) - \theta\}^2 + Var(\hat{\theta})$. (4 points)

4. **MOM and MLE Estimation**

Let $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$ where:

$$f_X(x) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} & 0 \leq x < \infty \\ 0, & \text{otherwise} \end{cases}$$

Suppose we observe four values: $x_1 = 2.15, x_2 = 2.68, x_3 = 2.17, x_4 = 2.28$

(a) It turns out that $\mathbb{E}(X) = \sqrt{\frac{\theta\pi}{2}}$ where $\pi \approx 3.14....$ Give the Method of Moments estimator for θ and report the numerical estimate based on the four data values. (4 points)

(b) The likelihood function is: $L(\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta} e^{-\frac{x_i^2}{2\theta}} = \prod_{i=1}^{n} x_i \cdot \theta^{-n} \cdot e^{-\frac{\sum x_i^2}{2\theta}}$

i. Give the log-likelihood function $\ell(\theta) = log(L(\theta))$. (3 points)

ii. Take the derivative of $\ell(\theta)$ with respect to θ, set it equal to zero, solve for θ, and report the Maximum Likelihood estimator of θ. (Assume 2^{nd} derivative test passes) (5 points)

iii. Based of the four data values above, report the Maximum Likelihood estimate of θ. (2 points)

5. **Hypothesis Testing**

We are given random samples from two separate populations. Let μ_1 and μ_2 be the population means of Population 1 and Population 2 respectively. We summarize several statistics of the two samples as follows.

	Population 1	Population 2
sample size	80	10
sample mean	52	49
sample variance	64	130

(a) Consider a hypothesis test to test whether population mean 1 is different than the value 50.

i. Give the null and alternative hypotheses. (4 points)

ii. Calculate the test statistic. (3 points)

iii. Calculate the $p-$value. Give your decision and conclusion. (3 points)

(b) Consider a hypothesis test to test whether populations mean 1 is larger than population mean 2.

i. State the null and alternative hypotheses. (4 points)

ii. Calculate the test statistic. (3 points)

iii. Calculate the $p-$value. Give your decision and conclusion. (3 points)

6. **Confidence Intervals**

A poll was conducted by Gallup from March 28-April 2 to understand Americans' attitude regarding medical visits. The poll was conducted using a random sample of 7,276 U.S adults.

(a) In poll, 42% said they would be "very concerned" about being exposed to the novel coronavirus (COVID-19) at a doctor's office or hospital. **Calculate and interpret** a 98% confidence interval for the true proportion of all U.S. adults that are very concerned about being exposed to COVID-19 at a doctor's office or hospital. (5 points)

(b) 49% out of 3638 women, and 35% out of 3638 men sampled said they were "very concerned" about being exposed to COVID-19 at a doctor's office or hospital. **Calculate and interpret** a 95% confidence interval for the difference in proportion of those "very concerned" between women and men. (5 points)

(c) Is it likely that the proportions of men and women that are "very concerned" about being exposed to COVID-19 from a doctor's office or hospital are equal? Explain. (Use your answer in part (b) or conduct a hypothesis test to make your argument). (3 points)