Spring 2016 **Statistics 330** **Final Exam**

(100 points)

Name ______________________________

Instructions:

- Partial credit will be given only if you show your work.
- The questions are not necessarily ordered from easiest to hardest.
- You may only use your formula sheets (2 pages, both sides) and a calculator.
- You can find z- and t-tables attached to the back of the exam.
- You do not need to hand in your formula sheets along with your exam.

1. Let $c_1, \ldots, c_k$ denote the k components in a parallel system. Assume the k components operate independently, and denote $P(c_j \text{ works})$ as p_j. Derive the formula for the reliability of the system in terms of the probabilities p_j. Clearly state which axioms, definitions, or results you are using at each stage of the development.

   Note: Giving only the formula here will result in *no credit*; you need to show that you understand where the formula comes from and how it is derived from basic facts about probabilities.

   Answer:

2. The *Pareto distribution*, with parameters α and β, has pdf:

$$f(x) = \frac{\beta\alpha^{\beta}}{x^{\beta+1}} \text{ for } \alpha, \beta > 0$$

The image of a random variable with this distribution is (α, ∞).

(a) Verify that $f(x)$ is a pdf for any positive values of α and β.
Answer:

(b) Derive the mean and variance of this distribution when $\beta > 2$.
Answer:

3. Customers come to a bank teller's window according to a Poisson process at a rate of 1 customer every 15 minutes. Service times are independently exponentially distributed, with an average service time of 6 minutes. Customers that come while the teller is busy need to wait in line until the teller is free. There is only a single teller, and there is no limit to the length of the line.

   (a) Determine the average amount of time a customer spends waiting in line.

   Answer:

   (b) What fraction of the time is the teller busy with a customer?

   Answer:

   (c) What is the probability that an individual entering the bank needs to wait more than 10 minutes before being served?

   Answer:

4. Thomas the Bayesian is fitting the simple linear regression model to some data (x_i, y_i), for $i = 1, \ldots, n$. He has been told that the slope and the intercept of the regression line are 2 and -3, respectively, but he would like an estimate of σ^2 as well.

   Thomas is considering using the improper prior $\pi(\sigma^2) = \left(\frac{1}{\sqrt{\sigma^2}}\right)^p$ to represent his beliefs. Derive the posterior distribution of σ^2 using this prior; it should have the form of one of the common "named probability distributions" seen in class. Which values of p ensure that both parameter values are greater than 0?

   Answer:

5. Pigs are sent to a slaughterhouse according to a Poisson process with an unknown rate λ pigs per minute.

   (a) What distributions do the inter-arrival times $I_1, I_2, \ldots$ follow? Are they independent?

   Answer:

   (b) You arrive at time $t = 0$ and see pigs enter the facility at times $t = 2$, 3, 4.5, and 7 minutes. Calculate the inter-arrival times and find the MLE of λ.

   Answer:

6. CornCorp spends a lot of money creating new kinds of genetically modified corn. They have recently developed a new variety called "supercorn", and are now interested in testing how it compares to "normal corn".

   CornCorp scientists have grown 1000 plants for each of the two types of corn, measured their heights (in feet), and have summarized their results as follows:

	mean height	sample variance
normal corn	8	$\frac{1}{4}$
supercorn	11	$\frac{1}{2}$

   (a) We can imagine these 1000 supercorn plants as being a sample drawn from some infinite population of supercorn plants with mean height μ. Create a 98% confidence interval for μ.

   Answer:

   (b) Is there enough evidence in our data to conclude that their is indeed a difference between the mean supercorn height μ and the mean normal corn height η? Using a type 1 error rate of 0.05, determine this using either a hypothesis test or a confidence interval.

   Answer:

7. Suppose we have some data (x_i, y_i), for $i = 1, \ldots, n$ and we fit the simple linear regression model to these data using the least squares estimates for β_0 and β_1 and $\frac{1}{n-2}$SSE as our estimate for σ^2. That is, we are **not** putting a prior distribution on any of the parameters; we are estimating them using frequentist methods.

   (a) Describe the differences between $Var(\beta_0)$, $Var(\hat{\beta}_0)$, and $\hat{Var}(\hat{\beta}_0)$. Which one is used to create confidence intervals for β_0? Can the others be used instead?

   Answer:

   (b) Suppose we have a new observation with x-value x_{new}. Justify mathematically and intuitively why the prediction interval for Y_{new} will be wider than the confidence interval for $E[\hat{Y}_{new}] = \beta_0 + \beta_1 x_{new}$.

   Answer: