MATH 476 Statistics

Fred J. Hickernell **Spring 2006**
Test **Thursday, March 23**

Instructions:

i. *This exam consists of FOUR questions. Answer all of them.*

ii. *The data and situations portrayed in this test are fictitious, but realistic.*

iii. *This exam is closed book, but you may use 2 double-sided letter-size sheets of notes, the statistical tables from the text, and JMP.*

iv. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*

1. (22 marks)
Let $X_1, \ldots, X_n$ be i.i.d. $\sim$ Exponential(μ) random variables. Consider the sample mean, $\bar{X} = (X_1 + \cdots + X_n)/n$, as an estimator of μ.

   a) Is $\bar{X}$ unbiased?
   b) What is the mean square error of $\bar{X}$?
   c) What is the maximum likelihood estimator of μ? Is it the sample mean?

Answer:

$$\begin{aligned} E(\bar{X}) &= [E(X_1) + \cdots + E(X_n)]/n = \mu \\ \mathrm{var}(\bar{X}) &= [\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)]/n^2 = \mu^2/n, \\ \mathrm{MSE}(\bar{X}) &= E[(\bar{X} - \mu)^2] = \mathrm{var}(\bar{X}) = \mu^2/n \end{aligned}$$

So, $\bar{X}$ is unbiased and its mean square error is μ^2/n. The log-likelihood function for μ is

$$\log(L(x_1, \ldots, x_n|\mu)) = \log\left(\frac{e^{-x_1/\mu} \cdots e^{-x_n/\mu}}{\mu \cdots \mu}\right) = \log\left(\frac{e^{-n\bar{x}/\mu}}{\mu^n}\right) = \frac{-n\bar{x}}{\mu} - n\log(\mu),$$

$$0 = \frac{d\log(L(x_1, \ldots, x_n|\mu))}{d\mu} = \frac{n\bar{x}}{\mu^2} - \frac{n}{\mu} = \frac{n(\bar{x} - \mu)}{\mu^2}.$$

So, the maximum likelihood estimate of μ is the sample mean, $\bar{x}$.

2. (22 marks)
Consider again the situation in the previous problem.

   a) If $\bar{x}$ is observed to be 60 seconds for $n = 50$ observations, what is an approximate 95% confidence interval for μ using the central limit theorem?
   b) If one requires a 95% confidence interval with a half-width of 5 seconds, how large must the sample size be?

Answer: Since $n = 50$ is rather large, the approximate 95% confidence interval is $\bar{x} \pm 1.96\bar{x}/\sqrt{n} = 60 \pm 1.96 \times 60/\sqrt{50} \approx 60 \pm 17$. For a 95% confidence interval with a half-width of 5, we need $1.96 \times 60/\sqrt{n} \le 5$ or $n \ge (1.96 \times 60/5)^2 \approx 553$.

3. (28 marks)
A class of 15 students are tested at the beginning and end of the semester on their knowledge of Illinois state geography. The students' scores are

Beginning	78	52	43	66	91	32	55	77	20	31	88	65	58	70	62
End	85	50	60	84	98	40	50	85	50	30	95	78	65	72	60

a) What is the evidence for some difference in district-wide student scores after having spent one semester in school, assuming this class to be a typical sample? State the null and alternative hypotheses, the p-value, and your conclusion.

b) Suppose that the identification on the test exam papers is lost, and so the scores in the two rows of a particular column do not correspond to the same student, even though they do still represent class performance at the beginning and end of the semester. How does this change your analysis and conclusions?

Answer: a. Input the two sets of scores in two columns (variables), labeled "before" and "after" in JMP. Then compute the difference of these columns in a third column labeled "improvement". Use the `Analyze` *command and the* `Test Mean` *option. Alternatively, you may use the* `Matched Pairs` *command from the JMP Starter Window. Let μ_d denote the mean do the difference or improvements. The null hypothesis is $H_0 : \mu_d = 0$, the alternative is $H_a : \mu_d \neq 0$. The p-value for this alternative from JMP is 0.0063, so we conclude that there is a difference.*

b. Now these two sets of scores are treated as independent observations from different samples. Input the two sets of scores in one column, labeled "scores" and in the next column give labels "before" and "after". Use the `Two Sample t-Test` *command in the JMP Starter Window. Let μ_B and μ_A denote the means of classes of students before and after the semester. The null hypothesis is $H_0 : \mu_B = \mu_A$, the alternative is $H_a : \mu_B \neq \mu_A$. The p-value for this alternative from JMP is 0.3236. So, we cannot conclude any difference. One may interpret these results as that individual students will improve on average, but the change in the mean of the whole class is insignificant because of small sample size.*

4. (28 marks)
The Chicago city council, faced with the important and contentious issue of banning smoking in public places, decides to poll 1000 residents at random to determine θ, the proportion favoring the ban. If at least 2/3 of the students support the ban, then the city council will move forward.

a) The city council decides to reject the null hypothesis, $H_0 : \theta \leq 2/3$, i.e., that no more than two-thirds of all Chicago residents support the proposed smoking ban, if at least 700 of those residents polled support the ban. What is the significance level of this hypothesis test?

b) What are the probabilities of rejecting the null hypothesis $H_0 : \theta \leq \theta_0$, for $\theta_0 = 0.6, 0.7, 0.8$.

c) Sketch the power function for this test.

Answer: a. Input into JMP one column labeled "Opinion" with the entries "For" and "Against" and in a second column labeled "Frequency" the numbers 700 and 300. Then choose the option `Test Probabilities` *and in the row "For" enter 0.66667. The p-value is 0.0133, which is the significance level of the test.*

b. Repeat `Test Probabilities` *with the entries* $0.6, 0.7$ *and* 0.8 *in the row "For". These are the probabilities of rejecting the null hypothesis for different values of* θ_0*, which is also the values of the power of the test. They are,* $< 0.0001, 0.5156,$ *and* 1.0000*, respectively.*

c. $\mathrm{power}(\theta)$ *starts at zero when* $\theta = 0$*, is nearly zero to* $\theta = 0.6$*, and then increases to nearly one by* $\theta = 0.8$ *and becomes one at* $\theta = 1$*.*