MATH 476 Statistics

Fred J. Hickernell **Spring 2006**
Test **Thursday, April 21**

Instructions:

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

ii. *The data and situations portrayed in this test are all realistic. Some are fictitious.*

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 and consider the sample mean, $\bar{X} = (X_1 + \cdots + X_n)/n$. Recall that $E(X_i) = \mu$, $\text{var}(X_i) = \mu^2$, $\bar{X}$ is an unbiased estimator of μ, and $\bar{x}$ is the maximum likelihood estimator of μ.

   a) Is $\bar{X}^2$ an unbiased estimator of $\text{var}(X_i)$?
   b) Is $\bar{x}^2$ the maximum likelihood estimator of $\text{var}(X_i)$?

Answer:

$$\begin{aligned}
E(\bar{X}^2) &= E\left\{[(X_1 + \cdots + X_n)/n]^2\right\} \\
&= \frac{1}{n^2} E[X_1^2 + \cdots + X_n^2 + 2X_1X_2 + \cdots + 2X_{n-1}X_n] \\
&= \frac{1}{n^2}\left[nE(X_1^2) + n(n-1)E(X_1X_2)\right] \\
&= \frac{1}{n^2}\left[n(\mu^2 + \mu^2) + n(n-1)\mu^2\right], \quad \textit{since } E(X_1^2) = [E(X_1)]^2 + \text{var}(X_1) \\
&= \frac{\mu^2(n+1)}{n}.
\end{aligned}$$

So, $\bar{X}^2$ is a biased (but asymptotically unbiased) estimate of $\text{var}(X_i)$.

Since $\bar{x}$ is the maximum likelihood estimator of μ, $\bar{x}^2$ is the maximum likelihood estimator of $\mu^2 = \text{var}(X_i)$. We can also show it directly by maximizing the likelihood function. The log-likelihood function for μ^2 is

$$\log(L(x_1, \ldots, x_n|\mu^2)) = \log\left(\frac{e^{-x_1/\mu} \cdots e^{-x_n/\mu}}{\mu \cdots \mu}\right) = \log\left(\frac{e^{-n\bar{x}/\mu}}{\mu^n}\right) = \frac{-n\bar{x}}{\mu} - n\log(\mu).$$

The value of μ that maximizes this is $\bar{x}$, so the value of μ^2 that maximizes it is $\bar{x}^2$, which is not the sample variance.

2. (25 marks)
Two hundred children are asked to name their favorite Chicago sports team. The answers are:

Bears	Bulls	Cubs	White Sox
40	52	45	63

a) Is there evidence to refute the assertion that these four teams are equally popular?

b) If the sample is increased to 2000 students with the same proportions observed for each team, how does your conclusion change, if at all? Why?

Answer: a. Enter the four team names in one column and the frequencies in another. Choose `Analyze` $\rightarrow$ `Distribution` *and then* `Test Probabilities`*. Enter equal probabilities of* 0.25*. The Pearson chi-squared test gives a p-value of* 0.1136*, which is too large to refute the assertion that these four teams are equally popular.*

b. If we increase the observed frequencies by a factor of ten, then the chi-square statistic increases by a factor of ten, and the Pearson chi-squared p-value is < 0.0001*. We can refute the assertion that these four teams are equally popular.*

3. (25 marks)
You may suspect that as you become accustomed with university life, your grades get better. The grades of students in several applied mathematics classes were observed and categorized by the year of study of the students who received them:

	A	B	C
Frosh	12	30	18
Sophomores	40	65	19
Juniors	35	40	10
Seniors	26	22	5

a) Is there evidence of a relationship between year of study and grades received?

b) Besides the hypothesis that becoming more accustomed to university life improves your grades, what other reasons could there be for a relationship between year of study and grades received?

Answer: a. Enter each of the three grades each four times in one column. In the second column enter each year of study three times, so each pair of grade and year occurs once. In the third column enter the frequency. Choose `Fit Y by X`*, and then choose the first two columns as Y and X, and the third as frequency. The p-value for the Pearson chi-square test is* 0.0048*, so there is strong evidence of a relationship between year of study and grades received.*

b. This relationship could be due to a number of things: weaker students drop out, a junior taking a class with a freshman is likely to do better because he is more mature, professors in the upper level classes give higher grades, etc.

4. (28 marks)
Hot dogs may be tasty, but they are also unhealthy. The table below shows the calorie per hot dog for different brands of hot dogs categorized by kind.

Calories in Hot Dogs

Beef	Meat	Poultry
186	173	129
181	191	132
176	182	102
149	190	106
184	172	94
190	147	102
158	146	87
139	139	99
175	175	107
148	136	113
152	179	135
111	153	142
141	107	86
153	195	143
190	135	152
157	140	146
131	138	144
149		
135		
132		

a) Is there any significant difference in the mean calories over different brands for different kinds of hot dogs?

b) If you go to a barbecue, without asking the host or hostess to read the individual label, what kind of hot dog should you choose if you want to consume fewer calories?

Answer: a. In JMP enter all the calorie counts in one column and the kind of hot dog in another column. Choose `Fit Y by X`*, and then choose calories as Y and hot dog type as X. Next choose* `Means/ANOVA`*. The p-value is* < 0.0001*, indicating a significant difference in the mean calorie counts of the three kinds of hot dogs.*

b. Poultry hot dogs have the least number of calories on average.