Final Exam

Tuesday the 15th of December, 1998

This exam is open book. Write your answers on the exam sheets. If you need more space, continue your answer on the back of the page. Make sure you have all 10 pages!

The exam is 180 minutes long. There are six questions, worth a total of 100 points. They are not equally weighted, nor are they of equal difficulty. The number of points each question is worth is printed with the question. Read the questions carefully. If you are unsure of the interpretation, come ask. You must show your work to obtain full credit. If you use a result from class, state what result you are using. If you can’t complete a problem for any reason, explain what concepts are at issue, and how you would attack the problem. It is, in any case, a good idea to explain what your reasoning is in English. If I can’t tell that you understood what you were doing, I can’t give you credit, particularly if you get the wrong numerical answer. GOOD LUCK!

Question 1 (15 Points, 27 minutes)

Calculate the mean and variance for X and Y.

(a) X has the following distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

(b) Y has the following distribution:

\[ f(y) = \frac{x}{2} \quad 0 \leq x \leq 2 \]

Question 2 (20 Points, 36 minutes)

In a large population of married men (none of whom have ever been widowed), a fraction, p, have been divorced at least once. It is desired to estimate p but there is a concern that people may not be willing to admit they have been divorced. Therefore the following procedure is followed:

A box contains 100 envelopes. 100x of these contain the question:
“Have you ever been divorced?”
The other 100(1 − x) contain the question:
“Is this your first marriage?”
(Note that if a person would answer “yes” to the first question he would necessarily answer “no” to the second. We assume that everyone answers truthfully.) $N$ married men were selected at random from the population. Each of them in succession was asked to pick an envelope from the box, read its contents, and return it to the box. Each then answered, aloud, “yes” or “no” to the question he read. (Note that only the subject knows which question he read.) Let $Y$ denote the total number of men who answered “yes”. Assume that $x$ is known, $0 < x < 1$.

(a) Show that the distribution of $Y$ is $Bin(N, px + (1 - p)(1 - x))$.

(b) Find the method of moments estimator of $p$.

(c) Find the mean and variance of the estimator of part (b).

**Question 3 (10 Points, 18 minutes)**

Suppose that $X_1, X_2, X_3$ are independent $N(0, \sigma^2)$ random variables. State, giving reasons, the distributions of

(a) $\dfrac{(4X_1 - 3X_2)^2}{25\sigma^2}$

(b) $\dfrac{(X_1^2 + X_2^2)/2}{X_3^2}$

(c) $\dfrac{X_1}{\sqrt{(X_2^2 + X_3^2)/2}}$

**Question 4 (10 Points, 18 minutes)**

Suppose that $X_1, X_2, \ldots, X_n$ are random variables with common mean $\mu$. Consider the following additional conditions:

C1: $X_1, X_2, \ldots, X_n$ have common variance $\sigma^2$.
C2: $X_1, X_2, \ldots, X_n$ are independent.
C3: $X_1, X_2, \ldots, X_n$ are normally distributed.

State which (if any) of the three conditions are required for each of the following results to hold. (Note some may require more than one condition.)

(a) $E\bar{X} = \mu$

(b) $Var(\bar{X}) = \sigma^2/n$

(c) $\bar{X}$ is normally distributed.

(d) $T = \dfrac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$
Question 5 (15 Points, 27 minutes)

Suppose you are performing an analysis of CEO compensation across 8 different industries. For each of the 8 industries you randomly choose 10 companies and record the compensation for their CEO’s. (All numbers are in thousands of dollars.)

(a) Use this information to complete the partial One Way Anova table below. Is there significant evidence that there is a difference in compensation across the eight industries?

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups (Industries)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td></td>
<td></td>
<td>404.2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>46202.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Upon examining the average compensation for each industry you notice that the average for industry 2 is the largest ($440) and the average for industry 6 is the lowest ($402). Use Tukey’s method to calculate a 95% confidence interval for $\mu_2 - \mu_6$. Is this method justified, given that you used the data to decide on which confidence interval to calculate?

Question 6 (30 Points, 54 minutes)

Let $X_1, X_2, \ldots, X_n$ be an iid sample from a population with the following density;

$$f(x) = \frac{2x}{\theta^2} \exp(-x^2/\theta^2) \quad x > 0, \quad \theta > 0$$

(a) What is a sufficient statistic for $\theta$?

(b) What is the mle for $\theta$?

(c) Give the form of the Uniformly Most Powerful test for testing $H_0 : \theta = \theta_0$ vs $H_a : \theta > \theta_0$. Note you do not need to give the exact value of $c$.

(Hint: First construct the Neyman-Pearson Likelihood Ratio test for $H_0 : \theta = \theta_0$ vs $H_a : \theta = \theta_a$ where $\theta_a > \theta_0$.)

(d) Calculate $I(\theta)$.

(Hint: You may use the fact that $EX^2 = \theta^2$)

(e) Use the properties of mle’s and part d to show that

$$\hat{\theta} \pm 1.96 \frac{\hat{\theta}}{2\sqrt{n}}$$

is an approximate 95% confidence interval for $\theta$. (Note $\hat{\theta}$ is the mle for $\theta$.)