UNIVERSITY OF HOUSTON
MIDTERM EXAMINATION

Term: Fall    Year: 2009

<table>
<thead>
<tr>
<th>Student: First Name</th>
<th>Max</th>
<th>Last Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>UH Student ID Number</td>
<td>---------------</td>
<td>-----------</td>
<td>-------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course Abbreviation</th>
<th>Math 4397/6397 and Number</th>
<th>Date of Exam</th>
<th>October 20, 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Title</td>
<td>Biostatistics</td>
<td>Time Period</td>
<td>Start time: 2:30pm</td>
</tr>
<tr>
<td>Section(s)</td>
<td>001</td>
<td>End time:</td>
<td>3:50pm</td>
</tr>
<tr>
<td>Sections of Combined Course(s)</td>
<td></td>
<td>Duration of Exam</td>
<td>1 hour 20 mins</td>
</tr>
<tr>
<td>Section Numbers of Combined Course(s)</td>
<td></td>
<td>Number of Exam Pages</td>
<td>8 pages</td>
</tr>
<tr>
<td>Instructor(s)</td>
<td>B. G. Bodmann</td>
<td>(including this cover sheet)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exam Type</td>
<td>Closed book</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional Materials Allowed</td>
<td>Approved sheet</td>
</tr>
</tbody>
</table>

Notes:

- Please use only the basic mathematical functions on your calculator.
- Show your work on all questions. Simple “yes” or “no” answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- The last page contains popular formulas, including a table of selected quantiles for the $t$-distribution.
- Good luck!
1. Short answers (9 points each).

a. A well known soccer player historically scores a goal in 90% of his games. What is the probability that the player scores in two consecutive games? Explain your answer briefly stating any assumptions made.

Assuming the performances in each game are independent,

\[ P(\text{score in 2}) = P(\text{score in 1})^2 = 0.9^2 = 81\%. \]

b. Suppose that \( f_1, f_2, f_3 \) are all densities and let \( \pi_1, \pi_2, \pi_3 \) be positive numbers so that \( \pi_1 + \pi_2 + \pi_3 = 1 \). Show that \( g(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x) \) is a valid density.

Since \( \pi_1, \pi_2, \pi_3 \geq 0 \), and \( f_1(x), f_2(x), f_3(x) \geq 0 \) we have \( g(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x) \geq 0 \).

Also,
\[
\int_{-\infty}^{\infty} g(x) \, dx = \pi_1 \int_{-\infty}^{\infty} f_1(x) \, dx + \pi_2 \int_{-\infty}^{\infty} f_2(x) \, dx + \pi_3 \int_{-\infty}^{\infty} f_3(x) \, dx
= \pi_1 + \pi_2 + \pi_3 = 1.
\]

c. A random variable \( X \) which is exponentially distributed with rate parameter \( \lambda = 3 \) has the cumulative distribution function

\[ F(x) = \begin{cases} 
1 - e^{-3x}, & x \geq 0 \\
0, & \text{else} 
\end{cases} \]

Compute the median of \( X \).

We want \( x_{0.5} \). i.e., \( F(x_{0.5}) = 1 - e^{-3x_{0.5}} = \frac{1}{2} \)

\[ e^{-3x_{0.5}} = \frac{1}{2} \]

\[ -3x_{0.5} = -\ln \frac{1}{2} \]

\[ x_{0.5} = \frac{1}{3} \ln 2 \approx 0.23 \]
2. You and your friend are playing a game where you roll a fair die. If it comes up a 1, 2, 3 or 4 he gives you a dollar. If it comes up a 5 or 6, you have to give him a dollar.

   a. (4 points) What are your expected winnings, $X_1$, after the first round?

   \[ P(X_1 = 1) = \frac{4}{6} = \frac{2}{3}, \quad P(X_1 = -1) = \frac{2}{6} = \frac{1}{3} \]

   \[ \implies E[X_1] = \frac{2}{3}(1) + \frac{1}{3}(-1) = \frac{1}{3} \]

   b. (8 points) How many rounds do you have to play the game until you have a total expected winnings greater than or equal to $5$? (Show some work.)

   \[ T = X_1 + X_2 + \ldots + X_n \]

   \[ E[T] = E[X_1 + X_2 + \ldots + X_n] \]

   \[ = n E[X_1] = \frac{n}{3} \geq 5 \]

   \[ \implies n \geq 15, \text{ so 15 rounds.} \]

3. (9 points) Suppose that the US intelligence quotients (IQs) are normally distributed with mean 100 and standard deviation 16. What IQ score represents the 5th percentile? (Explain your calculation.)

   Normal quantiles

   \[ z_{0.05} = -z_{0.95} = -1.645 \]

   \[ \implies IQ_{0.05} = \mu - 1.645 \times 16 \]

   \[ = 100 - 1.645(16) \]

   \[ = 73.7 \]
4. (12 points) Suppose that a sample of 100 subjects are drawn from the population of another
country and that 60 of the sampled subjects have an IQ below 116. Give a 95% confidence
interval estimate of the true probability $p$ of drawing a subject from this population with an
IQ below 116. State appropriate assumptions and use approximations.

We use Slutsky’s theorem, approx. $\sigma \approx \delta$,
and Wald’s estimate $S \approx \sqrt{\hat{p}(1-\hat{p})}$
with $\hat{p} = \frac{60}{100} = 0.6$.

Consequently, with $\sqrt{\text{normal}}$ distributed
IQ, $\text{CI} \approx S$ for $p$ is

$$
\hat{p} \pm z_{0.975} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
$$

$$
= 0.6 \pm 1.96 \sqrt{\frac{(0.6)(0.4)}{100}}
$$

which gives

$$
[0.50, 0.70]
$$
5. (9 points) You are in desperate need to simulate standard normal random variables yet do not have a computer available. You do, however, have a standard six sided die. Knowing that the mean of a single die roll is 3.5 and the standard deviation is 1.71, describe how you could use the die to approximately simulate the outcomes of a standard normal random variable.

Roll die a large number of times, say $n$. Take avg. of outcomes from single rolls. Expected value of avg $\bar{X}_n$ is $E[\bar{X}_n] = 3.5$.

and std. error is $\sigma / \sqrt{n} = 1.71 / \sqrt{n}$.

By CLT, for $n$ large

$$\frac{\bar{X}_n - 3.5}{1.71 / \sqrt{n}}$$

is approximately normally distributed.
The next questions involve the following setting: Suppose that 18 obese subjects were randomly assigned, 9 each, to a new diet pill or to a placebo. Subjects' body mass indices (BMIs) were measured at the beginning of the study and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the initial value (followup - beginning value) was $-3.0 \text{ kg/m}^2$ for the treated group and $1.0 \text{ kg/m}^2$ for the placebo group. The corresponding sample standard deviations of the differences was $1.5 \text{ kg/m}^2$ for the treatment group and $1.8 \text{ kg/m}^2$ for the placebo group.

6. (9 points) Calculate and interpret a 95% confidence interval for the change in BMI for the treated group; assume normality of the random BMI values in the group.

\[ n = 9 \Rightarrow t - \text{dist} \quad Cl \]

\[-3.0 \pm t_{8,0.975} \frac{5}{\sqrt{9}} \]

\[= -3.0 \pm 2.306 \frac{1.5}{3} \Rightarrow [-4.15, -1.85] \]

BMI appears to decrease, Cl excludes mean zero.

7. (12 points) Does the change in BMI over the four week period appear to differ between the treated and placebo groups? Create the relevant 95% confidence interval and interpret. Assume normality of the BMI values in each group and a common variance.

Pooled variance, $X$ treated, $Y$ placebo

\[ S_p^2 = \frac{S_{x}^2 + S_{y}^2}{16} = \frac{1}{2} \left( 1.5^2 + 1.8^2 \right) \]

\[= 2.745 \]

\[ Cl \quad (@ 95\%) \]

\[ \bar{x} - \bar{y} \pm t_{16,0.975} S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \]

\[= -4.0 \pm 2.12 \sqrt{\frac{1}{9} + \frac{1}{9}} \]

\[i.e. \quad \left[ -5.66, -2.34 \right] \]

Since 0 is not in Cl, the change appears to differ.