STAT 401 **MIDTERM 2** **Dec. 7 2006**
J.Millson

1. The pre-med advisor at State University is trying to decide whether he should encourage juniors to sign up for a private review course as a means for preparing for the MCAT's. Among the 15 students who had taken the exam most recently, five had enrolled in the review course and ten had not. The average MCAT scores for the first group were 10,8,9,8 and 11; the average scores for the other were 8,7,7,9,10,8,7,11,8 and 8. Assume that the distributions of MCAT scores are normal.

(a) Test whether the variances of the scores in each group are equal using the two-sample F-test on your calculator. Make your decision on the basis of whether or not the resulting P-value for the F-test is large or small.

(b) Do the *appropriate* one-sided two-sample t-test to test whether taking the review course significantly *improves* scores on the MCAT (use the level $\alpha = .1$).

(10 points)

2. For many years sodium nitrite has been used as a curing agent for bacon and until recently it was thought to be perfectly harmless. But now it appears that during frying, sodium nitrite induces the formation of nitrosopyrrolidine (NPy), a substance suspected of being a carcinogen. In one study focusing on the problem, measurements were made of the amount of NPy recovered after the frying of three slices of four commercially available brands of bacon.

A	20	40	18
B	75	25	21
C	15	30	21
D	25	30	31

Determine whether we accept or reject H_0: there is no difference between the average amounts of NPy between the four brands at level $\alpha = .05$.

(10 points)

3. This whole page and the beginning of the next page is material you will need to do the problem on the next page. It is also on pages 393-394 of your text. The resulting test is called the two proportion z-test. It is available on your calculator.

Suppose $\hat{p}_1$ is the sample proportion associated to a sample of size m from a Bernoulli population with $p = p_1$ and $\hat{p}_2$ is the sample proportion associated to a sample of size n from a Bernoulli population with $p = p_2$. Let $\hat{p}$ be the weighted average sample proportion

$$\hat{p} = \frac{m}{m+n}\hat{p}_1 + \frac{n}{m+n}\hat{p}_2$$

and

$$\hat{q} = 1 - \hat{p}.$$

To test

$$H_0 : p_1 - p_2 = 0$$
$$against$$
$$H_a : p_1 - p_2 \neq 0.$$

use the following decision rule. Put

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})}}.$$

The decision rule is then: reject H_0 if either $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$.

This is the test you must perform to do part (a) of the problem on the next page.

To do part (b) of the problem on the next page you will need the following formula from the text pg. 394 for $\beta(p_1, p_2)$. First let $\overline{p}$ resp. $\overline{q}$ be the weighted averages of p_1 and p_2 resp. q_1 and q_2 defined by

$$\overline{p} = \frac{m}{m+n}p_1 + \frac{n}{m+n}p_2$$

and

$$\overline{q} = \frac{m}{m+n}q_1 + \frac{n}{m+n}q_2.$$

Hence $\overline{p}+\overline{q}=1$.

We also define σ by

$$\sigma=\sqrt{\frac{p_1q_1}{m}+\frac{p_2q_2}{n}}.$$

Then we have the *beta formula*

$$\begin{aligned}\beta(p_1,p_2)=&\Phi(\frac{z_{\alpha/2}\sqrt{\overline{pq}(\frac{1}{m}+\frac{1}{n})}-(p_1-p_2)}{\sigma})\\ &-\Phi(\frac{-z_{\alpha/2}\sqrt{\overline{pq}(\frac{1}{m}+\frac{1}{n})}-(p_1-p_2)}{\sigma})\end{aligned}$$

Now here is the problem

A sample of 300 urban adult residents of particular state revealed 63 who favored increasing the speed limit from 55 to 65 miles per hour whereas a sample of 180 rural residents yielded 75 who favored the increase.

(a) Apply the two proportion z test (above) at level $\alpha=.05$ to decide whether the sentiment for increasing the speed limit is different between the two groups.

(b) If the true proportions favoring the the increase are $p_1=.2$ (urban) and $p_2=.4$ (rural) what is the probability that H_0 will be rejected using the above level .05 test with 100 urban residents and 100 rural residents (in other words what is $\beta(.2,.4)$)?

(10 points)

4. The point of the following problem is to find a confidence interval and a test for β_0, the *y- intercept* of the regression line. The theory will be very similar to the one we developed in class for β_1, the *slope* of the regression line.

Let $Y\sim N(\beta_0+\beta_1x,\sigma^2)$. Suppose that $(x_1,Y_1),(x_2,Y_2),\cdots,(x_n,Y_n)$ is a random sample from the space of Y (i.e. we are in the framework of simple linear regression). Define a random variable $S^2_{\widehat{\beta_0}}$ (the S^2 that belongs to the estimator $\widehat{\beta_0}$) by the formula

$$S^2_{\widehat{\beta_0}}=(\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n(x_i-\overline{x})^2})\quad S^2.$$

Then define $S_{\widehat{\beta_0}}$ to be the square root of $S^2_{\widehat{\beta_0}}$. Assume the following theorem from probability theory.

Theorem 1. *The random variable* $\frac{\widehat{\beta_0}-\beta_0}{S_{\widehat{\beta_0}}}$ *has* t- *distribution with* $n-2$ *degrees of freedom.*

(a) Write down a two-sided $100(1-\alpha)\%$ confidence interval for β_0 and prove that your interval has the correct confidence level.
(b) Write down a decision rule for testing

$$H_0 : \beta_0 = (\beta_0)_0$$
$$against$$
$$H_a : \beta_0 \neq (\beta_0)_0.$$

and prove that your test has significance level α.

(Hint: look at the formulas we proved in class for the slope β_1 and try to carry them over to the y-intercept β_0.)

(20 points)