**ANSWERS TO TEST NUMBER 6**

Question 1: (15 points)

A point estimate of $\mu_x$ - $\mu_y$ is

$$\bar{Y} - \bar{X} = 7.0 - 3.0 = 4.$$

Given the small samples, we are required to assume that the populations are normally distributed with the same variance—if this assumption does not hold, the test will be invalid. A point estimate of this common variance is

$$s_c^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_1^2}{n_1 + n_2 - 2} = \frac{(2)(1.0) + (4)(2.5)}{6} = \frac{12}{6} = 2.$$

The variance of $\bar{Y} - \bar{X}$ then becomes

$$s^2_{\bar{Y}-\bar{X}} = \frac{s_c^2}{n_1} + \frac{s_c^2}{n_2} = s_c^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right] = 2\left[\frac{1}{3} + \frac{1}{5}\right] = (2)(.2 + .3333) = 1.067.$$

Given the small samples, the test statistic

$$\frac{(\bar{Y} - \bar{X}) - (\mu_y - \mu_x)}{s_{\bar{Y}-\bar{X}}} = \frac{4}{\sqrt{1.067}} = \frac{4}{1.033} = 3.87$$

is distributed according to the $t$-distribution with $n_1 - n_2 - 2$ degrees of freedom. The probability of observing a value of $t$ larger than 3.87 (which is the same thing as the probability of observing a difference between $\bar{Y}$ and $\bar{X}$ as big as 4 when $\mu_x = \mu_y$) is (using XLISPSTAT)

```
> (- 1 (t-cdf 3.87 6))
0.004132939458758145
```

The $\alpha$-risk would thus have to be set at below $2 \times .004 = .008$ for the researcher to reject the null hypothesis.

It was noted at the beginning that one has to assume that the two populations have a common variance in order to do this test.

[Give yourself 4 points for knowing that the researcher had to assume that the two populations have the same variance and 3 points for knowing how to calculate that common variance. Then give yourself 4 points for understanding what the test statistic is and how to calculate it, 2 points for knowing that you need the area of the probability distribution of the sample mean to the right of the test statistic, 1 point for knowing how to calculate that probability and 1 point for knowing that you have to double it to get the P-value, which is the $\alpha$-risk below which the researcher would reject the null hypothesis.]

Question 2: (15 points)

Let $p_m$ and $p_f$ be the respective proportions of males and females that are left-handed. You want to test the null hypothesis $H_0 : p_m = p_f$ against the alternative hypothesis $H_1 : p_m \neq p_f$. If the null hypothesis is true, $p_m = p_f = p = .07$.

Let the sample proportions of males and females, respectively, that are left-handed be $\bar{p}_m$ and $\bar{p}_f$. Then, if the null hypotheses is true the variance of the difference between the sample proportions will be

$$\sigma^2\{\bar{p}_m - \bar{p}_f\} = \sigma^2\{\bar{p}_m\} + \sigma^2\{\bar{p}_f\} = \frac{p(1-p)}{n_m} + \frac{p(1-p)}{n_f}$$

where $n_m$ and $n_f$ are the numbers of males and females, respectively in the sample. If we sample the same number of males as females, $n$ of each, the variance of the difference between the sample proportions will become

$$\sigma^2\{\bar{p}_m - \bar{p}_f\} = p(1-p)\left[\frac{1}{n} + \frac{1}{n}\right] = 2\,\frac{p(1-p)}{n} = \frac{(2)(.07)(.93)}{n} = \frac{.1302}{n}.$$

and the standard deviation will be

$$\sigma\{\bar{p}_m - \bar{p}_f\} = \frac{.3608}{\sqrt{n}}$$

A 90% confidence interval will leave 5% of the probability weight on each tail of the sampling distribution—the distance to each confidence limit from $p$ will be 1.645 standard deviations. The researcher wants that distance to be .01. Hence

$$.01 = 1.645\,\sigma\{\bar{p}_m - \bar{p}_f\} = \frac{(1.645)(.3608)}{\sqrt{n}} = \frac{.59356925}{\sqrt{n}}$$

Hence

$$\sqrt{n} = \frac{.5983187}{.01} = 59.356925$$

with the result that the desired sample size will be approximately 3523.

[Give yourself 5 points for knowing how to calculate the variance of the difference in the sample proportions, 5 points for understanding that the distance to each confidence limit from $p$ will be 1.645 standard deviations and 5 points for knowing how to use these results to calculate the desired sample size.]

Question 3: (25 points)

a) The two populations are students in eco100 who are required to buy copies of the lecture notes and students in eco100 who are not so required. The populations consist of all possible potential students in the respective groups, not just the 86 and 60 students who represent samples drawn from those respective populations by selecting sections of the course offered in the current year. [Give yourself 5 points for understanding this.]

b) The difference between the mean responses of the first (required to buy notes) and second (not buying notes) is a point estimate of the difference between the means of the two populations.

$$E\{\bar{X}_1 - \bar{X}_2\} = \mu_1 - \mu_2$$

and

$$\bar{X}_1 - \bar{X}_2 = 8.48 - 7.80 = .68$$

The variance of the difference in the sample means will be

$$s^2_{\bar{X}_1 - \bar{X}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{.94}{86} + \frac{2.99}{60} = .01093 + .04983 = .0608$$

so the standard deviation of the difference between the means is the square root of .0608, which equals .2465. [Give yourself 2 points for understanding that the difference in the sample means will be a point estimate of the difference in population means and 4 points for knowing how to calculate the standard deviation of the difference in the sample means.]

The null hypothesis is

$$H_0 : \mu_1 = \mu_2$$

and the alternative hypothesis is

$$H_1 : \mu_1 \neq \mu_2$$

The $z$ statistic for the test is

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X} - \bar{Y}}} = \frac{.68}{.2465} = 2.7586.$$

Using XLISPSTAT, we can find the two critical values of $z$ for $\alpha = .01$ as follows

```
> (normal-quant .995)
2.5758293035489
> (normal-quant .005)
-2.5758293035489004
```

Note that $.005 = \alpha/2$. We can reject the null hypothesis of no difference of mean response at the 1% level. [Give yourself 2 points for knowing how to calculate the $z$ statistic and 2 points for understanding why the null hypothesis would have to be rejected.]

c) Using the sample statistics and critical values above, the 99% confidence interval will be

$$\begin{aligned}
U &= \bar{X}_1 - \bar{X}_2 + z_{(1-\alpha/2)} s_{\bar{Y}-\bar{X}} &= .68 + (2.576)(.2465) &= .68 + .634984 \\
&= 1.315 \\
L &= \bar{X}_1 - \bar{X}_2 - z_{(1-\alpha/2)} s_{\bar{Y}-\bar{X}} &= .68 - (2.576)(.2465) &= .68 - .634984 \\
&= 0.045
\end{aligned}$$

This says that there is a 99% chance that the interval [0.045, 1.315] will bracket $(\mu_1 - \mu_2)$. [Give yourself 6 points for being able to correctly calculate and interpret the confidence interval.]

d) If we were to calculate a 95% confidence interval $z_{(1-\alpha/2)}$ would equal

```
> (normal-quant .975)
1.9599639845400536
```

instead of 2.576, and so the confidence interval would be narrower. We have less confidence that the error in prediction will be small than large. [Give yourself 4 points for understanding why the interval would be narrower.]

Question 4: (25 points)

a) The sample mean must lie above 6.0 because, otherwise, we would have rejected the null hypothesis and it would have been unnecessary to calculate a $P$-value—were we to do so, that $P$-value would be greater than 0.5. [Give yourself 5 points for understanding this.]

b) The value of $z$ generated by the sample must have been that value to the left of which $1 - .0068 = .9932$ of the probability weight lies. This value will be

```
> (normal-quant  .9932)
2.4676584925406795
```

or approximately 2.47. [Give yourself 5 points for knowing how to calculate this.]

c) The estimated value of $z$ must have been equal to

$$z = \frac{\bar{X} - \mu_0}{s} = \frac{\bar{X} - 6.0}{2.16} = 2.47$$

which implies

$$(2.47)(2.16) = \bar{X} - 6.0$$

whence

$$(2.47)(2.16) + 6.0 = \bar{X} = 5.33 + 6.0 = 11.33$$

[Give yourself 3 points for understanding how to calculate $z$ and 7 points for undestanding how to calculate $\bar{X}$.]

d) $H_0$ will be rejected at any $\alpha$ less than .0068. It will thus be clearly rejected at the 1% level of significance. Of course, our choice of the significance level would depend on the costs of making the wrong decision. [Give yourself 5 points for understanding this.]

Question 5: (20 points)

a) We know that the variance of the sum of the two test grades will be

$$\sigma^2_{T_A} = \sigma^2_{T1} + \sigma^2_{T2} + 2\sigma_{T1,T2}$$

which implies

$$12^2 = 8^2 + 6^2 + 2\sigma_{T1,T2}$$
$$144 - 64 - 36 = 144 - 100 = 2\sigma_{T1,T2} = 44$$

so that $\sigma_{T1,T2} = 22$. Note that this covariance, and the variances as well, apply to a student in the course, not to the course average.

The correlation between the values obtained on the two tests is

$$r = \frac{\sigma_{T1,T2}}{\sigma_{T1}\sigma_{T2}} = \frac{22}{(8)(6)} = \frac{22}{48} = .45833$$

[Give yourself 5 points for understanding how to calculate the covariance and 2 points for understanding how to calculate the correlation.]

b) We want the mean and standard deviation of $T1_i - T2_i$. The mean will equal

$$\frac{\sum_{i=1}^{275} T1_i - T2_i}{275} = \frac{\sum_{i=1}^{275} T1_i}{275} - \frac{\sum_{i=1}^{275} T2_i}{275} = T1 - T2 = 32 - 36 = -4$$

and the variance will equal

$$\begin{aligned}
&\frac{\sum_{i=1}^{275}(T1_i + T2_i - 32 - 36)^2}{274} \\
&= \frac{\sum_{i=1}^{275}[(T1_i - 32)^2 + (T2_i - 36)^2 - 2(T1_i - 32)(T2_i - 36)}{274} \\
&= \sigma^2_{T1} + \sigma^2_{T2} - 2\sigma_{T1,T2} \quad = \quad 64 + 36 - (2)(22) = 100 - 44 = 56
\end{aligned}$$

This makes the standard deviation of the paired differences equal to $\sigma_D = 7.48$. [Give yourself 3 points for understanding how to calculate the mean paired difference and 7 points for understanding how to calculate the standard deviation of the paired difference.]

c) The $z$ statistic for the test of the null hypothesis that the students performed equally well on the two tests is

$$z = \frac{T1 - T2 - 0}{\sigma_D/\sqrt{n}} = \frac{-4}{7.48/\sqrt{275}} = \frac{-4}{7.48/16.6} = \frac{-4}{.4506} = 8.877.$$

We can see immediately that the null hypothesis will be rejected at any reasonable level of significance. The $P$-value is virtually zero. [Give yourself 3 points for understanding how to correctly calculate and interpret the $z$ statistic and reach the correct conclusion.]