SOLUTIONS FINAL STAT101
FALL 2004

1. A. Plainly we take $H_0 : \mu \leq 60$ and $H_a : \mu > 60$. Now

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{62.6 - 60}{7.4/\sqrt{334}} = 6.42.$$

Since n is large it matters not if we compare with z_1^* or t_1^*. In any case $t = 6.42$ is off the scale and $t >> z_1^* = 2.33$ so we reject H_0.

B. (i) Here we want the probability of getting a type II error. The acceptance region is given by $\bar{x} \leq \mu_0 + z_1^* \frac{\sigma}{\sqrt{n}} = 60 + 2.33 \frac{7.4}{\sqrt{334}} = 60.9434$. And so we want $P(\bar{x} \leq 60.9434 | \mu = 61) = P(Z \leq \frac{60.9434-61}{7.4/\sqrt{334}}) = 0.4443$.

(ii) Using the formula we gave in class

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2} = \frac{(2.33 + 1.645)^2 (7.4)^2}{61.60} = 865.24,$$

so 866 will do.

C. Here we take $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$. Comparing

$$t = \frac{3.25 - 0}{7.15/\sqrt{20}} = 2.033$$

with $t^* = 2.093$ (taking 19 degrees of freedom) we see that we must retain H_0.

2. A. (i) Of course we take a normal with the same mean and standard deviation: $N(2.77, 2.108^2)$.

(ii) As the saying goes, all roads lead to Roma: it is plain from the histogram or the normal quantile plot that the data is right leaning. Comparing the mean and the median supports this. It *may* be the case that a normal approximation is sufficiently accurate in the middle hunk of data.

B. (i) This is just $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ or $2.77 \pm 1.96 \frac{2.108}{\sqrt{400}} = 2.77 \pm 0.2068$. Thus the confidence interval runs from 2.563 to 2.977, which we could just read off the JMP output to a sufficiently high degree of accuracy!

(ii) Using the formula we discussed in class with $E = 0.2$ we have

$$n = \frac{{z^*}^2 \sigma^2}{E^2} = \frac{1.96^2 2.11^2}{0.2^2} = 427.58,$$

so that 428 will suffice.

C. Again with wanton abandon we apply the formula we learn't in class. We have $\hat{p} = \frac{66}{400} = 0.165$, and so the confidence interval interval is estimated by

$$\hat{p} \pm z^* \frac{\sqrt{\hat{p}(1-\hat{p}}}{\sqrt{n}} = 0.16 \pm 1.96 \frac{\sqrt{(0.165)(0.835)}}{\sqrt{400}} = 0.165 \pm 0.036.$$

Alternatively we could use $\tilde{p} = \frac{66+2}{400+4} = 0.1683$ and obtain 0.1683 ± 0.0365.

D. (i) Here we look at the contingency table and extract conditional probabilities. For example, the probability of not getting a job given that one is a finance major is 9.52%. The others are around 20%; specifically, management is the highest at 21.74%, marketing at 19.44% and 'other' at 18.9970.

(ii) Literally the P-value in this context says that if the fraction of students without jobs were uniform across the majors the probability that we would observe the discrepancy in the data in one more extreme (i.e. fractions unequal) happens about 3.6% of the time. If, for example, $\alpha = 0.05$ we would reject the null hypothesis in favour of a relationship between major and students not getting a job offer.

3. A. We have $\mu = 0(0.15) + 1(0.15) + \cdots + 10(0.1) = 2.57$ which one can check against the JMP output.

B. Without thinking we write down $H_0 : p = 0.15$ and $H_a : p \neq 0.15$. We compute

$$z = \frac{0.165 - 0.15}{\sqrt{0.15(0.85)/400}} = 0.84$$

and since $|z| < z^* = 1.96$ we regretfully retain the null hypothesis.

C. (i) Of course we have an instance of our old friend the central limit theorem, and just as a week is a long time in politics, a thousand is a big number in probability; anyhow a cursory perusal of the normal quantile plot confirms once again our faith in this marvellous result.

(ii) This is what one might term a 'derived' hypothesis test. A moment's reflection tells us that we take $H_0 : p = 0.95$ and $H_a : p \neq 0.95$. Here we have $\hat{p} = 961/1000 = 0.961$ and

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.961 - 0.95}{\sqrt{0.95(0.05)/1000}} = 1.59,$$

and since $|z| = 1.59 < z^* = 1.96$ we retain H_0.

4 A. One could begin by inspecting the equations for each fit, and note that the gradients in particular are vastly different. One could also compare specific instances of the two predictors: take, for example, $X = 5$. Then $\hat{Y}_1 = 29.89 + 9.4378(5) = 74$ and $\hat{Y}_2 = 56.73 + 2.069(5) = 67.1$. We could be more quantitative and compare RMSE: in the first case we have 17.85 and the second 6.71. What's happening here is the associating 0 values for 0 jobs inflates the variance.

B. Plainly the line which fits all 400 data points does not provide a deft fit. The RMSE tells us this, but also the residual plots have a much more appealing distribution for the 334 data points.

5. A. We approximate the probability using our friend the CLT: the mean profit is $10,000(60) = \$60,000$ and the variance is $10,000(1,000)^2$ which gives a standard deviation of 100,000. Thus $P(\text{Total} > 500,000) \approx P(Z > \frac{500,000-600,000}{100,000} = -1) = 0.8413$, which is appealing.

   B. A little contemplation tells us that the middle is the best position for the biased coin: since we want two consecutive heads the middle one must be a head in any relevant outcome and so we choose the middle to be the biased coin. Alternatively we may compute the probabilities for each contender, viz. HHH, HHT and THH. Plainly it matters not if the biased coin is in first or third place, and we compute a probability of 0.4375 for the event consisting of either of these outcomes. In the case that the biased coin is in middle place we compute a probability of 0.5625.

   C. Using our now familiar recipe we take $H_0 : p \leq 0.1$ and $H_a : p > 0.1$, and with $\hat{p} = 11/63 = 0.1587$ we have

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.1587 - 0.1}{\sqrt{0.1(0.9)/63}} = 1.55$$

   and since $z < z_1^* = 1.645$ we retain H_0, and so conclude that getting a question correct does not increase the chance of being in the "Top 10% club".

6. To answer these you just need to have a feel for the transformations we considered in class.
   a. lower.
   b. lower.
   c. higher.
   d. and don't know.