Stat 160 Fall 2008
Solutions to Assignment #9

22.40 (a) The basic design is as follows. The 16 pieces of identical fabric were randomly assigned to one of two groups with 8 pieces of fabric in each group. Each piece of fabric in the first group was dyed by Method B, whereas each piece of fabric in the second group was dyed by Method C. The colours from the two methods were then compared.

22.40 (b) Let μ_B denote the true mean lightness score for Method B, and let μ_C denote the corresponding true mean for Method C. We want to compare the difference between the two means by testing

$$H_0 : \mu_B = \mu_C \quad \text{against} \quad H_a : \mu_B \neq \mu_C.$$

Stemplots for each method are shown below.

Method B

40.	88	98			
41.	27	28	30	39	50
41.	66				

Method C

42.	20	28	30	50
42.	43	45	65	
43.	13			

Even though the stemplots show that the distributions are somewhat irregular, as is common for small samples, the use of the t procedure appears justified. (There are no outliers and the distributions are roughly symmetric.) The summary statistics are given as follows.

	sample size	sample mean	sample standard deviation
Method B	8	41.2825	0.2550
Method C	8	42.4925	0.2939

Even though Method B gives a somewhat lower lightness score (i.e., a somewhat darker colour), the difference between the methods is small. The t test statistic is given by

$$t = \frac{(\overline{x}_B - \overline{x}_C)}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_C^2}{n_C}}} = -8.79.$$

With $df = 7$, we see from Table C that the P-value is incredibly small. In fact, P-value$<$ 0.001. Thus, we conclude that there is overwhelming evidence that Method B gives darker colour. (However, we need to be careful since the difference in mean lightness scores may be too small to be important in practice.)

21.22 (a) Let p_1 denote the true proportion of Illinois high school freshman who use anabolic steroids, and let p_2 denote the corresponding proportion for seniors. In order to draw conclusions about these populations, the samples should be randomly chosen from a variety of Illinois schools.

21.22 (b) Using the plus four method, a 95% confidence interval for p_1 is given by

$$\tilde{p}_1 \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+4}} \quad \text{where} \quad \tilde{p}_1 = \frac{X_1+2}{n_1+4}.$$

Since

$$\tilde{p}_1 = \frac{34+2}{1679+4} = 0.0214$$

the required 95% confidence interval is

$$0.0214 \pm 1.96\sqrt{\frac{(0.0214)(1-0.0214)}{1679+4}} = 0.0214 \pm 0.00691 = [0.0145, 0.0283].$$

21.22 (c) In order to test $H_0 : p_1 - p_2 = 0$ against $H_a : p_1 - p_2 \neq 0$ we compute the z test statistic given by

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where

$$\hat{p}_1 = \frac{X_1}{n_1}, \quad \hat{p}_2 = \frac{X_2}{n_2}, \quad \text{and} \quad \hat{p} = \frac{X_1+X_2}{n_1+n_2}.$$

We compute

$$\hat{p}_1 = \frac{34}{1679} = 0.02025, \quad \hat{p}_2 = \frac{24}{1366} = 0.01757, \quad \text{and} \quad \hat{p} = \frac{34+24}{1679+1366} = 0.01905$$

so that

$$z = \frac{(0.02025 - 0.01757)}{\sqrt{(0.01905)(1-0.01905)\left(\frac{1}{1679}+\frac{1}{1366}\right)}} = 0.54.$$

From Table A, we see that this corresponds to a P-value of 0.5892. This indicates that there is no evidence of a difference in steroid usage rates between Illinois high school freshmen and seniors.

3. (a) A 96% confidence interval for $p_1 - p_2$, the true difference in proportions, using the plus-four method is given by

$$(\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$$

where

$$\tilde{p}_1 = \frac{X_1+1}{n_1+2} \quad \text{and} \quad \tilde{p}_2 = \frac{X_2+1}{n_2+2}.$$

If we let population 1 consist of teenagers who use instant messaging online, and we let population 2 consist of adults who use instant messaging online, then

$$\tilde{p}_1 = \frac{736+1}{981+2} = 0.7497 \quad \text{and} \quad \tilde{p}_2 = \frac{511+1}{1217+2} = 0.4208$$

so that the required 96% confidence interval is

$$(0.7497 - 0.4208) \pm 2.054\sqrt{\frac{(0.7497)(1-0.7497)}{981+2} + \frac{(0.4208)(1-0.4208)}{1217+2}} = [0.2883, 0.3695].$$

3. (b) In order to test $H_0 : p_1 - p_2 = 0$ against $H_a : p_1 - p_2 > 0$ we compute the z test statistic given by

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$\hat{p}_1 = \frac{X_1}{n_1}, \quad \hat{p}_2 = \frac{X_2}{n_2}, \quad \text{and} \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}.$$

Since

$$\hat{p}_1 = \frac{736}{981} = 0.7503, \quad \hat{p}_2 = \frac{511}{1217} = 0.4199, \quad \text{and} \quad \hat{p} = \frac{736 + 511}{981 + 1217} = 0.5673,$$

we find

$$z = \frac{0.7503 - 0.4199}{\sqrt{(0.5673)(1 - 0.5673)\left(\frac{1}{981} + \frac{1}{1217}\right)}} = 15.54$$

Since this z statistic corresponds to a P-value of essentially 0, we see that there is overwhelming evidence to suggest that the proportion of teenagers who use IM online is greater than the proportion of adults who use IM online.

4. (a) The sample size n needed to estimate a proportion at the 95% confidence level within a margin of error m satisfies

$$m = 1.96\sqrt{\frac{p^*(1-p^*)}{n}} \quad \text{and so} \quad n = \left(\frac{1.96}{m}\right)^2 p^*(1-p^*).$$

With $m = 0.01$ and $p^* = 0.5$ we find

$$n = \left(\frac{1.96}{0.01}\right)^2 (0.5)(0.5) = 9604.$$

Since there is a cost of \$5 per observation, such a sample size would cost the biologist \$48 020 which is well above his \$1500 limit.

4. (b) If we now choose $m = 0.01$ and $p^* = 0.75$ we find

$$n = \left(\frac{1.96}{0.01}\right)^2 (0.75)(0.25) = 1801.$$

Since there is a cost of \$5 per observation, such a sample size would cost the biologist \$9005 which is also well above his \$1500 limit.

4. (c) If the biologist will be satisfied with a margin of error $m = 0.05$, then with $p^* = 0.5$, we find

$$n = \left(\frac{1.96}{0.05}\right)^2 (0.5)(0.5) = 385.$$

In this case, since there is a cost of \$5 per observation, such a sample size would cost the biologist \$1925 which is slightly above his \$1500 limit.

Finally, using $m = 0.05$ and $p^* = 0.75$, we find

$$n = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 289$$

which, at a cost of \$5 per observation, would cost the biologist \$1445. This is affordable given the biologist's budget!