Solution for Homework 11

Extra Problem Sheet:

9 (a) X = amount of cereal in the small bowl
Y= amount of cereal in the large bowl
T = X + Y (total amount)
E (T) = 4.0 and SD (T) = 0.5.
P (T > 4.5) = P (Z > 1) = 0.1587
(b) D = Y – X (difference)
E (D) = 1 and SD (D) = 0.5
P (D < 0) = P (Z < -2) = 0.0228

10 (Note: The R.V. D in this problem is different from the one in the previous problem.)
E (D) = 1 and SD (D) = 0.2408
P (D > 0.5) ≈ P (Z > -2.076) = 0.9812

11 X = CU student's IQ
Y = Nameless U student's IQ
D = X – Y (difference)
E (D) = 10 and SD (D) ≈ 15.62
P (D > 5) ≈ P (Z > - 0.32) = 0.6255

12 (Note: The R.V. D in this problem is different from the one in the previous problem.)
(a) E (D) = 10 and SD (D) ≈ 8.246
(b) P (D > 5) ≈ P (Z > -0.6063) ≈ 0.7291

18.4 (a) The distribution is approximately normal with mean 0.14 and standard deviation 0.0155.
(b) 20% or more Harley owners is unlikely: P ($\hat{p}$ > 0.20) ≈ P (Z > 3.87) < 0.0002. There is a fairly good chance of finding at least 15% Harley owners: P ($\hat{p}$ > 0.15) ≈ P (Z > 0.64) < 0.2611.

18.6 The population is too small.

18.8 There were only 5 or 6 "successes" in the sample (because 5/2673 and 6/2673 both round to 0.2%).

18.11 (a) The methods can be used here because we assume we have a large SRS from a much larger population.
(b) $\hat{p}$ = 692/1048 ≈ 0.6603, and the 95% confidence interval is 0.632 to 0.689.
(c) The margin of error for a 95% confidence interval ("19 cases out of 20") was (slightly less than) 3%.

18.28 (a) We find $\hat{p}$ ≈ 0.5397, so the 95% confidence interval is 0.5397 ± 0.0306 = 0.5090 to 0.5703, and $\tilde{p}$ = 0.5396, so the plus four interval is 0.5396 ± 0.0305 = 0.5090 to 0.5701. By either the standard or the plus four method, the margin of error is roughly 3%.
(b) We were not given sample sizes for each gender. (However, by solving the system x + y = 1019 and 0.65x +0.43y = 550: approximately 508 men and 511 women.)
(c) The margin of error for women alone would be greater than 0.03 because the sample size is smaller.

18.31 We find $\hat{p}$ = 0.4202, so the 99% confidence interval is 0.4202 ± 0.0301 ≈ 0.3901 to 0.4503, and the plus four interval is 0.4203 ± 0.0301 ≈ 0.3903 to 0.4504.

18.32 For testing H_0: p = 0.5 vs. Ha: p < 0.5, we have z ≈ -6.75. This gives P < 0.0002 – very strong evidence that less than half the population attended church or synagogue in the preceding week. Additionally the intervals from the previous exercise do not include 0.50 or more.

18.33 n = (2.576/0.01)^2*(0.5)*(0.5) ≈ 16589.4 – use n = 16590. The use of p^{*} = 0.5 is reasonable because our confidence interval shows that the actual p is in the range 0.3 to 0.7, so that this conservative approach will not greatly inflate the sample size.

19.11 We find $\hat{p}_1 \approx 0.7105$ and $\hat{p}_2 \approx 0.5700$, SE ≈ 0.0395, and so the 95% confidence interval is $\hat{p}_1 - \hat{p}_2$ ± 1.96SE = 0.1405 ± 0.0774 ≈ 0.0631 to 0.2179. Using the plus four method: $\tilde{p}_1 \approx 0.7083$ and $\tilde{p}_2 \approx 0.5698$, SE ≈ 0.0394, so the interval is $\tilde{p}_1 - \tilde{p}_2$ ± 1.96SE = 0.1386 ± 0.0772 ≈ 0.0614 to 0.2158.

19.20 (a) To test H_0: $p_1 = p_2$ vs. Ha: $p_1 \neq p_2$, we find $\hat{p}_1$ = 52/65 = 0.8 and $\hat{p}_2$ = 30/55 ≈ 0.5455, and $\hat{p}$ = (52+30) / (65+55) ≈ 0.6833. Then SE ≈ 0.08523, so z = ($\hat{p}_1 - \hat{p}_2$) / SE ≈ 2.99. This gives P = 0.0028 -- strong evidence that there is a difference (specifically, that urban/suburban students are more likely to succeed).

(b) For a confidence interval, SE ≈ 0.08348, so the 90% confidence interval for $p_1 - p_2$ is ($\hat{p}_1 - \hat{p}_2$) ± 1.645*SE = 0.2545 ± 0.1373 ≈ 0.1172 to 0.3919. Using the plus four method, the interval is 0.1113 to 0.3930.

19.23 (a) $\hat{p}_1 \approx 0.2137$, $\hat{p}_2 \approx 0.4149$ and SE ≈ 0.014995, the 99% confidence interval is ($\hat{p}_1 - \hat{p}_2$) ± 2.576*SE ≈ 0.1626 to 0.2398. Using the plus four method, the interval is 0.1623 to 0.2395.

(b) Because the 99% confidence interval for the difference does not include 0, the P-value against the two-sided alternative will be smaller than 0.01.

(c) The count of other types of nonresponse from January to April was 491 – 333 = 158, so $\hat{p}_1 \approx$ 0.1014; in July and August, the count was 1174 – 861 = 313, so $\hat{p}_2 \approx 0.1508$. For testing H_0: $p_1 = p_2$ vs. Ha: $p_1 \neq p_2$, we have $\hat{p}$ = (158 + 313) / (1558 + 2075) ≈ 0.1296 and SE ≈ 0.01126, so z = ($\hat{p}_1 - \hat{p}_2$) / SE ≈ -4.39. This gives P < 0.0004 -- very strong evidence that other nonresponse rates also differ between the seasons.