Stat 257.01 Fall 2004
Assignment #4 Solutions

(4.10) An estimate of the proportion of students on campus in favour of converting from the semester system to the quarter system is simply the ratio of those sampled who are in favour to the total number sampled; hence,

$$\hat{p} = \frac{25}{40} = \frac{5}{8} = 0.625.$$

A bound on the error of estimation is given by equation (4.17) on page 97. Thus,

$$2\sqrt{\hat{V}(\hat{p})} = 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{N-n}{N}\right)} = 2\sqrt{\frac{(0.625)(0.375)}{40-1}\left(\frac{2000-40}{2000}\right)} \approx 0.1535.$$

In other words, 0.625 ± 0.1535 is an approximate 95% confidence interval for p.

(4.14) In this case, we find $\hat{p} = 20/50 = 0.4$ is the sample proportion of vouchers filed incorrectly. Thus, an estimate for the total number of vouchers filed incorrectly is

$$N\hat{p} = 250 \times 0.4 = 100.$$

A bound on the error of estimation is $2\sqrt{\hat{V}(N\hat{p})} = 2\sqrt{N^2\hat{V}(\hat{p})}$ where $\hat{V}(\hat{p})$ is as in equation (4.17) on page 97. Thus,

$$2\sqrt{\hat{V}(N\hat{p})} = 2\sqrt{N^2\left(\frac{\hat{p}\hat{q}}{n-1}\right)\left(\frac{N-n}{N}\right)} = 2\sqrt{250^2\left(\frac{0.4\cdot 0.6}{50-1}\right)\left(\frac{250-50}{250}\right)} \approx 31.3.$$

In other words, 100 ± 31.3 is an approximate 95% confidence interval for Np, the total number of vouchers filed incorrectly.

(4.15) As always, we use the sample mean $\overline{y}$ as an estimator of the population mean μ. Thus,

$$\hat{\mu} = \overline{y} = 2.1.$$

A bound on the error of estimation is given by equation (4.4) on page 85. Thus,

$$2\sqrt{\hat{V}(\overline{y})} = 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)} = 2\sqrt{\frac{(0.4)^2}{20}\left(\frac{200-20}{200}\right)} \approx 0.17.$$

In other words, we can tell the psychologist that 2.1 ± 0.17 is an approximate 95% confidence interval for μ, the average reaction time to stimulus among 200 patients in a hospital specializing in nervous disorders.

(4.16) In order to estimate μ with a bound of 1 second on the error of estimation, we should take the sample size at least as large as

$$n = \frac{N\sigma^2}{(N-1)B^2/4 + \sigma^2} = \frac{200\cdot 1}{(200-1)\cdot 1/4 + 1} \approx 3.94.$$

Thus, we need to have a sample that contains at least 4 individuals.

(4.25) Be sure to review your Stat 151 notes on the interpretation of a confidence interval. Remember, a 95% confidence interval for μ like 2.1 ± 0.17 in **(4.15)** DOES NOT SAY that the probability that μ is in $(1.93, 2.27)$ is 0.95. Rather, the interpretation is that "this confidence interval was produced by a method that yields correct results 95% of the time." We simply *hope* that the interval we computed is not one of the unlucky 5%. However, we generally have no way of knowing! Essentially this is what the Gallup explanation is saying, although to me, it is a little too wordy. Also note that the $\pm 4\%$ is referring to the bound on the error of estimation.

(4.26) It is VERY important to be careful when reading or hearing statements like this; it is easy to lie with statistics. Consider first a population of size $N = 100\,000$. Let p be the true proportion of red beans. We KNOW that $p = \hat{p} = 0.3$. Hence, a bound on the error of estimation is given by

$$2\sqrt{\hat{V}(\hat{p})} = 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{N-n}{N}\right)} = 2\sqrt{\frac{0.3 \cdot 0.7}{1000-1}\left(\frac{100\,000-1000}{100\,000}\right)} \approx 0.03.$$

Thus, a 95% confidence interval for p is 0.3 ± 0.03 or $(0.27, 0.33)$. In other words, if we draw a sample of size $n = 1000$, then we conclude that a 95% confidence interval for the number of red balls is $(270, 330)$. Consider next a population of size $N = 80$ million. In this case the fpc is extremely close to 1, so that a bound on the error of estimation is given by

$$2\sqrt{\hat{V}(\hat{p})} \approx 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}} = 2\sqrt{\frac{0.3 \cdot 0.7}{1000-1}} \approx 0.3.$$

Again we conclude that a 95% confidence interval for p is $(0.27, 0.33)$, so that a 95% confidence interval for the number of red balls is $(270, 330)$. Whence, the Nielsen claim that the "basic statistical law wouldn't change" is, in fact, accurate.

(4.27) In order to decide whether or not the nickname *Mr. October* was justified for Reggie Jackson, we can simply compute approximate 95% confidence intervals for his cumulative batting average in each of the regular season, League Championship Series, and World Series. In the following formulæ, let N be the number of at bats, let n be the number of hits, and let

$$B = 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{N-n}{N}\right)}.$$

Regular season: We have $N = 9864$, $n = 2584$, $\hat{p} = 2584/9864 \approx 0.262$, $B = 0.015$ so that an approximate 95% confidence interval is

$$(0.247, 0.277).$$

League Championship Series: We have $N = 163$, $n = 37$, $\hat{p} = 37/163 \approx 0.227$, $B = 0.123$ so that an approximate 95% confidence interval is

$$(0.154, 0.350).$$

World Series: We have $N = 98$, $n = 35$, $\hat{p} = 35/98 \approx 0.357$, $B = 0.132$ so that an approximate 95% confidence interval is

$$(0.225, 0.489).$$

Since the three confidence intervals are NOT DISJOINT (that is, they all share a common set) the data do not provide sufficient evidence to suggest that Reggie Jackson's World Series performance was unusual. Thus, his nickname was unjustified. (Perhaps he should have been called *Mr. Consistency*!)