Stat 257.01 Fall 2004
Assignment #3 (Selected) Solutions

(3.30) Very briefly:

(a) As noted in exercise 3.28, there were 9965 teenagers aged 12–18 years old and living in households who were randomly selected for this survey. Those teenagers were classified as NS (non-smoker), EX (experimenter with smoking), FS (former smoker), or CS (current smoker). Each table lists the percentage of respondents for that question WITHIN EACH OF THESE FOUR CLASSIFICATIONS. Note that we are not told the number of respondents in each classification. These are all conditional proportions, given smoking status. (Observe that each column sums to 100%.)

(b) The data indicate that 12% of non-smokers sampled think that smoking helps reduce stress, while 46.5% of current smokers sampled believe that smoking helps reduce stress.

(c) From the second table, it appears that regardless of their smoking status, teenagers believe that almost all doctors are strongly against smoking.

(4.1) There are 10 possible samples of size $n = 2$ which are as follows:

$$\{0,1\}; \{0,2\}; \{0,3\}; \{0,4\}; \{1,2\}; \{1,3\}; \{1,4\}; \{2,3\}; \{2,4\}; \{3,4\}.$$

The population variance σ^2 is

$$\sigma^2 = E((y-\mu)^2) = E(y^2) - (\mu)^2 = \sum_y y^2 p(y) - \left[\sum_y y p(y)\right]^2$$

where $\mu = E(y)$. Thus,

$$\sigma^2 = \frac{1}{5}(0^2 + 1^2 + 2^2 + 3^2 + 4^2) - \left[\frac{1}{5}(0+1+2+3+4)\right]^2 = 6 - 4 = 2.$$

As in equation (4.1) on page 83

$$V(\overline{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2}{2}\left(\frac{5-2}{5-1}\right) = \frac{3}{4} = 0.75.$$

It is also possible to find $V(\overline{y})$ directly by computing the 10 sample means and finding the variance. In this case, the 10 sample means are as shown in the table below, and we find that

$$E(\overline{y}) = \frac{1}{10}(0.5 + 1 + 1.5 + 2 + 1.5 + 2 + 2.5 + 2.5 + 3 + 3.5) = 2$$

and

$$E(\overline{y}^2) = \frac{1}{10}(0.5^2 + 1^2 + 1.5^2 + 2^2 + 1.5^2 + 2^2 + 2.5^2 + 2.5^2 + 3^2 + 3.5^2) = \frac{19}{4} = 4.75.$$

Hence,

$$V(\overline{y}) = E(\overline{y}^2) - [E(\overline{y})]^2 = 4.75 - 2^2 = 0.75.$$

(4.2) For any data sample, recall that $s^2 = \frac{1}{n-1}\sum(y_i - \overline{y})^2$ where $\overline{y} = \frac{1}{n}\sum y_i$. Hence we have

Sample	$\overline{y}$	s^2
$\{0,1\}$	$1/2 = 0.5$	$1/2 = 0.5$
$\{0,2\}$	$2/2 = 1$	2
$\{0,3\}$	$3/2 = 1.5$	$9/2 = 4.5$
$\{0,4\}$	$4/2 = 2$	8
$\{1,2\}$	$3/2 = 1.5$	$1/2 = 0.5$
$\{1,3\}$	$4/2 = 2$	2
$\{1,4\}$	$5/2 = 2.5$	$9/2 = 4.5$
$\{2,3\}$	$5/2 = 2.5$	$1/2 = 0.5$
$\{2,4\}$	$6/2 = 3$	2
$\{3,4\}$	$7/2 = 3.5$	$1/2 = 0.5$

Thus we compute directly that $E(s^2)$ is

$$E(s^2) = \frac{1}{10}(0.5 + 2 + 4.5 + 8 + 0.5 + 2 + 4.5 + 0.5 + 2 + 0.5) = \frac{25}{10} = 2.5.$$

Notice that this agrees with the formula given in the problem, namely

$$E(s^2) = \frac{N}{N-1}\sigma^2 = \frac{5}{5-1} \cdot 2 = \frac{10}{4} = 2.5.$$

(4.7) In this problem, we are required to estimate the population mean μ. This requires the formulæ on page 85. As always, the sample average $\overline{y}$ is used as our (point) estimate of μ. Thus, $\hat{\mu} = 12.5$. We can bound the error of estimation using $2\sqrt{\hat{V}(\overline{y})}$ so that

$$B = 2\sqrt{\hat{V}(\overline{y})} = 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)} = 2\sqrt{\frac{1252}{100}\left(\frac{10000-100}{10000}\right)} \approx 7.04.$$

In other words, 12.5 ± 7.04 is an approximate 95% confidence interval for μ.

(4.8) In order to estimate τ, the total gallons of water used, we employ the formulæ on pages 91–92. As a (point) estimate of τ we use $\hat{\tau} = N\overline{y} = 10\,000 \times 12.5 = 125\,000$. We can bound the error of estimation using $2\sqrt{\hat{V}(\hat{\tau})}$ so that

$$B = 2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{N^2\left(\frac{s^2}{n}\right)\left(\frac{N-n}{N}\right)} = 2\sqrt{10\,000^2\left(\frac{1252}{100}\right)\left(\frac{10\,000-100}{10\,000}\right)} \approx 70\,412.5.$$

In other words, $125\,000 \pm 70\,412.5$ is an approximate 95% confidence interval for τ.