Introduction to Econometrics (3rd Updated Edition)

by

James H. Stock and Mark W. Watson

Solutions to Odd-Numbered End-of-Chapter Exercises: Chapter 3

(This version August 17, 2014)

3.1. The central limit theorem suggests that when the sample size (n) is large, the distribution of the sample average ($\bar{Y}$) is approximately $N\left(\mu_Y, \sigma_{\bar{Y}}^2\right)$ with $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$. Given a population $\mu_Y = 100$, $\sigma_Y^2 = 43.0$, we have

(a) $n = 100$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{100} = 0.43$, and

$$\Pr(\bar{Y} < 101) = \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.43}} < \frac{101 - 100}{\sqrt{0.43}}\right) \approx \Phi(1.525) = 0.9364.$$

(b) $n = 64$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{64} = \frac{43}{64} = 0.6719$, and

$$\begin{aligned}\Pr(101 < \bar{Y} < 103) &= \Pr\left(\frac{101 - 100}{\sqrt{0.6719}} < \frac{\bar{Y} - 100}{\sqrt{0.6719}} < \frac{103 - 100}{\sqrt{0.6719}}\right) \\ &\approx \Phi(3.6599) - \Phi(1.2200) = 0.9999 - 0.8888 = 0.1111.\end{aligned}$$

(c) $n = 165$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{165} = 0.2606$, and

$$\begin{aligned}\Pr(\bar{Y} > 98) &= 1 - \Pr(\bar{Y} \le 98) = 1 - \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.2606}} \le \frac{98 - 100}{\sqrt{0.2606}}\right) \\ &\approx 1 - \Phi(-3.9178) = \Phi(3.9178) = 1.0000 \text{ (rounded to four decimal places).}\end{aligned}$$

3.3. Denote each voter's preference by Y. $Y = 1$ if the voter prefers the incumbent and $Y = 0$ if the voter prefers the challenger. Y is a Bernoulli random variable with probability $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. From the solution to Exercise 3.2, Y has mean p and variance $p(1 - p)$.

(a) $\hat{p} = \frac{215}{400} = 0.5375.$

(b) The estimated variance of $\hat{p}$ is $\widehat{\text{var}}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375 \times (1 - 0.5375)}{400} = 6.2148 \times 10^{-4}$. The standard error is $\text{SE}(\hat{p}) = (\text{var}(\hat{p}))^{\frac{1}{2}} = 0.0249.$

(c) The computed t-statistic is

$$t^{act} = \frac{\hat{p} - \mu_{p,0}}{\text{SE}(\hat{p})} = \frac{0.5375 - 0.5}{0.0249} = 1.506.$$

Because of the large sample size $(n = 400)$, we can use Equation (3.14) in the text to get the
p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5$:

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-1.506) = 2 \times 0.066 = 0.132.$$

(d) Using Equation (3.17) in the text, the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$ is

$$p\text{-value} = 1 - \Phi(t^{act}) = 1 - \Phi(1.506) = 1 - 0.934 = 0.066.$$

(e) Part (c) is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside ± (calculated t-statistic). Part (d) is a one-sided test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.

(f) For the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value 0.066 is larger than 0.05. Equivalently the calculated t-statistic 1.506 is less than the critical value 1.64 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

3.5.

(a) (i) The size is given by $\Pr(|\hat{p}-0.5|>.02)$, where the probability is computed assuming that $p=0.5$.

$$\begin{aligned}\Pr(|\hat{p}-0.5|>.02) &= 1-\Pr(-0.02\le \hat{p}-0.5\le .02)\\ &= 1-\Pr\left(\frac{-0.02}{\sqrt{.5\times.5/1055}}\le\frac{\hat{p}-0.5}{\sqrt{.5\times.5/1055}}\le\frac{0.02}{\sqrt{.5\times.5/1055}}\right)\\ &= 1-\Pr\left(-1.30\le\frac{\hat{p}-0.5}{\sqrt{.5\times.5/1055}}\le 1.30\right)\\ &= 0.19\end{aligned}$$

where the final equality using the central limit theorem approximation.

(ii) The power is given by $\Pr(|\hat{p}-0.5|>.02)$, where the probability is computed assuming that $p = 0.53$.

$$\begin{aligned}\Pr(|\hat{p}-0.5|>.02) &= 1-\Pr(-0.02\le \hat{p}-0.5\le .02)\\ &= 1-\Pr\left(\frac{-0.02}{\sqrt{.53\times.47/1055}}\le\frac{\hat{p}-0.5}{\sqrt{.53\times.47/1055}}\le\frac{0.02}{\sqrt{.53\times.47/1055}}\right)\\ &= 1-\Pr\left(\frac{-0.05}{\sqrt{.53\times.47/1055}}\le\frac{\hat{p}-0.53}{\sqrt{.53\times.47/1055}}\le\frac{-0.01}{\sqrt{.53\times.47/1055}}\right)\\ &= 1-\Pr\left(-3.25\le\frac{\hat{p}-0.53}{\sqrt{.53\times.47/1055}}\le -0.65\right)\\ &= 0.74\end{aligned}$$

where the final equality using the central limit theorem approximation.

(b) (i) $t=\frac{0.54-0.5}{\sqrt{0.54\times0.46/1055}}=2.61$, $\Pr(|t|>2.61)=.01$, so that the null is rejected at the 5% level.

(ii) $\Pr(t>2.61)=.004$, so that the null is rejected at the 5% level.

(iii) $0.54\pm1.96\sqrt{0.54\times0.46/1055}=0.54\pm0.03$, or 0.51 to 0.57.

(iv) $0.54\pm2.58\sqrt{0.54\times0.46/1055}=0.54\pm0.04$, or 0.50 to 0.58.

(v) $0.54\pm0.67\sqrt{0.54\times0.46/1055}=0.54\pm0.01$, or 0.53 to 0.55.

3.5 (continued)

(c) (i) The probability is 0.95 is any single survey, there are 20 independent surveys, so the probability if $0.95^{20} = 0.36$

(ii) 95% of the 20 confidence intervals or 19.

(d) The relevant equation is $1.96 \times \text{SE}(\hat{p}) < .01$ or $1.96 \times \sqrt{p(1-p)/n} < .01$. Thus n must be chosen so that $n > \frac{1.96^2 p(1-p)}{.01^2}$, so that the answer depends on the value of p. Note that the largest value that $p(1 - p)$ can take on is 0.25 (that is, $p = 0.5$ makes $p(1 - p)$ as large as possible). Thus if $n > \frac{1.96^2 \times 0.25}{.01^2} = 9604$, then the margin of error is less than 0.01 for all values of p.

3.7. The null hypothesis in that the survey is a random draw from a population with p = 0.11. The t-statistic is $t = \frac{\hat{p}-0.11}{SE(\hat{p})}$, where $SE(\hat{p}) = \hat{p}(1-\hat{p})/n.$ (An alternative formula for SE($\hat{p}$) is $0.11\times(1-0.11)/n$, which is valid under the null hypothesis that $p = 0.11$). The value of the t-statistic is -2.71, which has a p-value of that is less than 0.01. Thus the null hypothesis $p = 0.11$(the survey is unbiased) can be rejected at the 1% level.

3.9. Denote the life of a light bulb from the new process by Y. The mean of Y is μ and the standard deviation of Y is $\sigma_Y = 200$ hours. $\overline{Y}$ is the sample mean with a sample size $n = 100$. The standard deviation of the sampling distribution of $\overline{Y}$ is $\sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}} = \frac{200}{\sqrt{100}} = 20$ hours. The hypothesis test is $H_0 : \mu = 2000$ vs. $H_1 : \mu > 2000$. The manager will accept the alternative hypothesis if $\overline{Y} > 2100$ hours.

(a) The size of a test is the probability of erroneously rejecting a null hypothesis when it is valid.

The size of the manager's test is

$$\begin{aligned} \text{size} &= \Pr(\overline{Y} > 2100|\mu = 2000) = 1 - \Pr(\overline{Y} \le 2100|\mu = 2000) \\ &= 1 - \Pr\left(\frac{\overline{Y} - 2000}{20} \le \frac{2100 - 2000}{20} |\mu = 2000\right) \\ &= 1 - \Phi(5) = 1 - 0.999999713 = 2.87 \times 10^{-7}. \end{aligned}$$

$\Pr(\overline{Y} > 2100|\mu = 2000)$ means the probability that the sample mean is greater than 2100 hours when the new process has a mean of 2000 hours.

(b) The power of a test is the probability of correctly rejecting a null hypothesis when it is invalid. We calculate first the probability of the manager erroneously accepting the null hypothesis when it is invalid:

$$\begin{aligned} \beta &= \Pr(\overline{Y} \le 2100|\mu = 2150) = \Pr\left(\frac{\overline{Y} - 2150}{20} \le \frac{2100 - 2150}{20} |\mu = 2150\right) \\ &= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062. \end{aligned}$$

The power of the manager's testing is $1 - \beta = 1 - 0.0062 = 0.9938$.

(c) For a test with 5%, the rejection region for the null hypothesis contains those values of the t-statistic exceeding 1.645.

$$t^{act} = \frac{\overline{Y}^{act} - 2000}{20} > 1.645 \Rightarrow \overline{Y}^{act} > 2000 + 1.645 \times 20 = 2032.9.$$

The manager should believe the inventor's claim if the sample mean life of the new product is greater than 2032.9 hours if she wants the size of the test to be 5%.

3.11. Assume that n is an even number. Then $\tilde{Y}$ is constructed by applying a weight of $\frac{1}{2}$ to the $\frac{n}{2}$ "odd" observations and a weight of $\frac{3}{2}$ to the remaining $\frac{n}{2}$ observations.

$$E(\tilde{Y}) = \frac{1}{n}\left(\frac{1}{2}E(Y_1) + \frac{3}{2}E(Y_2) + \cdots \frac{1}{2}E(Y_{n-1}) + \frac{3}{2}E(Y_n)\right)$$

$$= \frac{1}{n}\left(\frac{1}{2}\cdot\frac{n}{2}\cdot\mu_Y + \frac{3}{2}\cdot\frac{n}{2}\cdot\mu_Y\right) = \mu_Y$$

$$\text{var}(\tilde{Y}) = \frac{1}{n^2}\left(\frac{1}{4}\text{var}(Y_1) + \frac{9}{4}\text{var}(Y_2) + \cdots \frac{1}{4}\text{var}(Y_{n-1}) + \frac{9}{4}\text{var}(Y_n)\right)$$

$$= \frac{1}{n^2}\left(\frac{1}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2 + \frac{9}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2\right) = 1.25\frac{\sigma_Y^2}{n}.$$

3.13 (a) Sample size $n = 420,$ sample average $\bar{Y} = 646.2$ sample standard deviation $s_Y = 19.5$. The standard error of $\bar{Y}$ is $\text{SE}(\bar{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{19.5}{\sqrt{420}} = 0.9515$. The 95% confidence interval for the mean test score in the population is

$$\mu = \bar{Y} \pm 1.96\text{SE}(\bar{Y}) = 646.2 \pm 1.96 \times 0.9515 = (644.34,\ 648.06).$$

(b) The data are: sample size for small classes $n_1 = 238,$ sample average $\bar{Y}_1 = 657.4,$ sample standard deviation $s_1 = 19.4;$ sample size for large classes $n_2 = 182,$ sample average $\bar{Y}_2 = 650.0,$ sample standard deviation $s_2 = 17.9.$ The standard error of $\bar{Y}_1 - \bar{Y}_2$ is $SE(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = 1.8281.$ The hypothesis tests for higher average scores in smaller classes is

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 > 0.$$

The *t*-statistic is

$$t^{act} = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{SE}(\bar{Y}_1 - \bar{Y}_2)} = \frac{657.4 - 650.0}{1.8281} = 4.0479.$$

The *p*-value for the one-sided test is:

$$p\text{-value} = 1 - \Phi(t^{act}) = 1 - \Phi(4.0479) = 1 - 0.999974147 = 2.5853 \times 10^{-5}.$$

With the small *p*-value, the null hypothesis can be rejected with a high degree of confidence. There is statistically significant evidence that the districts with smaller classes have higher average test scores.

3.15. From the textbook equation (2.46), we know that $E(\bar{Y}) = \mu_Y$ and from (2.47) we know that $\text{var}(\bar{Y}) = \frac{\sigma_Y^2}{n}$. In this problem, because Y_a and Y_b are Bernoulli random variables, $\hat{p}_a = \bar{Y}_a$, $\hat{p}_b = \bar{Y}_b$, $\sigma_{Ya}^2 = p_a(1-p_a)$ and $\sigma_{Yb}^2 = p_b(1-p_b)$. The answers to (a) follow from this. For part (b), note that $\text{var}(\hat{p}_a - \hat{p}_b) = \text{var}(\hat{p}_a) + \text{var}(\hat{p}_b) - 2\text{cov}(\hat{p}_a, \hat{p}_b)$. But, are independent (and thus have a $\text{cov}(\hat{p}_a, \hat{p}_b) = 0$ because $\hat{p}_a$ and $\hat{p}_b$ are indendepent (they depend on data chosen from independent samples). Thus $\text{var}(\hat{p}_a - \hat{p}_b) = \text{var}(\hat{p}_a) + \text{var}(\hat{p}_b)$. For part (c), this equation (3.21) from the text (replacing $\bar{Y}$ with $\hat{p}$ and using the result in (b) to compute the SE). For (d), apply the formula in (c) to obtain

95% CI is $(.859 - .374) \pm 1.96\sqrt{\frac{0.859(1-0.859)}{5801} + \frac{0.374(1-0.374)}{4249}}$ or 0.485 ± 0.017.

3.17. (a) The 95% confidence interval is $\overline{Y}_{m,2012} - \overline{Y}_{m,1992} \pm 1.96\,\mathrm{SE}(\overline{Y}_{m,2012} - \overline{Y}_{m,1992})$ where $\mathrm{SE}(\overline{Y}_{m,2012} - \overline{Y}_{m,1992}) = \sqrt{\frac{S^2_{m,2012}}{n_{m,2012}} + \frac{S^2_{m,1992}}{n_{m,1992}}} = \sqrt{\frac{12.09^2}{2004} + \frac{10.85^2}{1594}} = 0.38;$ the 95% confidence interval is (25.30 – 24.83) ± 0.75 or 0.47 ± 0.75.

(b) The 95% confidence interval is $\overline{Y}_{w,2012} - \overline{Y}_{w,1992} \pm 1.96\,\mathrm{SE}(\overline{Y}_{w,2012} - \overline{Y}_{w,1992})$ where $\mathrm{SE}(\overline{Y}_{w,2012} - \overline{Y}_{w,1992}) = \sqrt{\frac{S^2_{w,2012}}{n_{w,2008}} + \frac{S^2_{w,1992}}{n_{w,1992}}} = \sqrt{\frac{9.99^2}{1951} + \frac{8.39^2}{1368}} = 0.32;$ the 95% confidence interval is $(21.50 - 21.39) \pm 0.63$ or 0.11 ± 0.63.

(c) The 95% confidence interval is

$$(\overline{Y}_{m,2012} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2012} - \overline{Y}_{w,1992}) \pm 1.96\,\mathrm{SE}[(\overline{Y}_{m,2012} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2012} - \overline{Y}_{w,1992})],$$

where

$$\mathrm{SE}[(\overline{Y}_{m,2012} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2012} - \overline{Y}_{w,1992})]$$

$$= \sqrt{\frac{S^2_{m,2012}}{n_{m,2008}} + \frac{S^2_{m,1992}}{n_{m,1992}} + \frac{S^2_{w,2012}}{n_{w,2008}} + \frac{S^2_{w,2004}}{n_{w,2004}}}$$

$$= \sqrt{\frac{12.09^2}{2004} + \frac{10.85^2}{1594} + \frac{9.99^2}{1951} + \frac{8.39^2}{1368}} = 0.50.$$

The 95% confidence interval is

(25.30 – 24.83) − (21.50−21.39) ± 1.96 × 0.50 or 0.36 ± 0.98.

3.19. (a) No. $E(Y_i^2) = \sigma_Y^2 + \mu_Y^2$ and $E(Y_iY_j) = \mu_Y^2$ for $i \neq j$. Thus

$$E(\bar{Y}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(Y_i^2) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} E(Y_iY_j) = \mu_Y^2 + \frac{1}{n}\sigma_Y^2$$

(b) Yes. If $\bar{Y}$ gets arbitrarily close to μ_Y with probability approaching 1 as n gets large, then $\bar{Y}^2$ gets arbitrarily close to μ_Y^2 with probability approaching 1 as n gets large. (As it turns out, this is an example of the "continuous mapping theorem" discussed in Chapter 17.)

3.21. Set $n_m = n_w = n$, and use equation (3.19) write the squared SE of $\overline{Y}_m - \overline{Y}_w$ as

$$[SE(\overline{Y}_m - \overline{Y}_w)]^2 = \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2}{n} + \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n(n-1)}.$$

Similary, using equation (3.23)

$$[SE_{pooled}(\overline{Y}_m - \overline{Y}_w)]^2 = \frac{\frac{1}{2(n-1)}\left[\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2\right]}{2n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n(n-1)}.$$