C.16 Two independent food scientists are researching the shelf-life (Y) of "Bill's Big Red" spaghetti sauce. The first collects a random sample of 25 jars and finds their average shelf life to be $\overline{Y} = 48$ months. The second researcher collects a random sample of 100 jars and finds their average shelf life to be $\overline{Y}_2 = 40$ months.

a. Find the ratio of the standard error of $\overline{Y}_1$ relative to the standard error of $\overline{Y}_2$.

b. A combined estimate can be obtained by finding the weighted average $\tilde{Y} = c\overline{Y}_1 + (1-c)\overline{Y}_2$. Is there any value of c that makes this estimator of μ unbiased?

c. What value of c yields the combined estimate with the smallest standard error? Explain the intuition behind your solution, and why weighting the two means equally, with $c = 0.5$, is not the best choice.

C.17 Suppose school children are subjected to a standardized math test each spring. In the population of comparable children, the test score Y is normally distributed with mean 500 and standard deviation 100, $Y \sim N(\mu = 500, \sigma^2 = 100^2)$. It is claimed that reducing class sample size will increase test scores.

a. How can we tell if reducing class size actually does increase test scores? Would you be convinced if a sample of $N = 25$ students from the smaller classes had an average test score of 510? Calculate the probability of obtaining a sample mean of $\overline{Y} = 510$, or more, even if smaller classes actually have no effect on test performance.

b. Show that a class average of 533 will be reached by chance only 5% of the time, if the smaller class sizes have no effect. Is the following statement correct or incorrect? "We can conclude that smaller classes raise average test scores if a class of 25 students has an average test score of 533 or better, with this result being due to sampling error with probability 5%."

c. Suppose that smaller classes actually do improve the average mean population test score to 550. What is the probability of observing a class of 25 with an average score of 533 or better? If our objective is to determine whether smaller classes increase test scores, is it better for this number to be larger or smaller?

d. If smaller classes increase average test score to 550, what is the probability of having a small class average of less than or equal to 533?

e. Draw a figure showing two normal distributions, one with mean 500 and standard deviation 100, and the other with mean 550 and standard deviation 100. On the figure locate the value 533. In part (b) we showed that if the change in class size has no effect on test scores, we would still obtain a class average of 533 or more by chance 5% of the time; we would incorrectly conclude that the smaller classes helped test scores, which is a Type I error. In part (d) we derived the probability that we would obtain a class average test score of less than 533, making us unable to conclude that smaller classes help, even though smaller classes did help. This is a Type II error. If we push the threshold to the right, say 540, what happens to Type I and Type II errors? If we push the threshold to the left, say 530, what happens to the probability of Type I and Type II errors?

C.11.2 Computer Exercises

C.18 Does being in a small class help primary school students learning, and performance on achievement tests? Use the sample data file *star5_small* to explore this question.

a. Consider students in regular-sized classes, with $REGULAR = 1$. Construct a histogram of *MATHSCORE*. Carry out the Jarque–Bera test for normality at the 5% level of significance. What do you conclude about the normality of the data?

b. Calculate the sample mean, standard deviation and standard error of the mean for *MATHSCORE* in regular-sized classes. Use the t-statistic in equation (C.16) to test the null hypothesis that the population mean (the population of students who are enrolled in regular-sized classes) μ_R is 490 versus the alternative that it isn't. Use the 5% level of significance. What is your conclusion?

c. Given the result of the normality test in (a), do you think the test in part (b) is justifiable? Explain your reasoning.

d. Construct a 95% interval estimate for the mean μ_R.

e. Repeat the test in (b) for the population of students in small classes, $SMALL = 1$. Denote the population mean for these students as μ_S. Use the 5% level of significance. What is your conclusion?

f. Let μ_R and μ_S denote the population mean test scores on the math achievement test, *MATHSCORE*. Using the appropriate test, outlined in Section C.7.2, test the null hypothesis $H_0{:}\mu_S - \mu_R \le 0$ against the alternative $H_1{:}\mu_S - \mu_R > 0$. Use the 1% level of significance. Does it appear that being in a small class increases the expected math test score, or not?

C.23 Does additional education have as large a payoff for females as males? Use the data file *cps5* to explore this question. If your software does not permit using this larger sample use *cps5_small*.

a. Calculate the sample mean wage of females who have 12 years of education. Calculate the sample mean wage of females with 16 years of education. What did you discover?

b. Calculate a 95% interval estimate for the population mean wage of females with 12 years of education. Repeat the calculation for the wages of females with 16 years of education. Do the intervals overlap?

c. Calculate the sample mean wage of males who have 12 years of education. Calculate the sample mean wage of males with 16 years of education. What did you discover? How does the difference in wages for males compare to the difference of wages for females in part (a)?

d. Calculate a 95% interval estimate for the population mean wage of males with 12 years of education. Repeat the calculation for the wages of males with 16 years of education. Does the interval for males with 12 years of education overlap with the comparable interval for females? Does the interval for males with 16 years of education overlap with the comparable interval for females?

e. Denote the population means of interest by μ_{F16}, μ_{F12}, μ_{M16}, μ_{M12} where F and M denote female and male, and 12 and 16 denote years of education. Estimate the parameter $\theta = (\mu_{F16} - \mu_{F12}) - (\mu_{M16} - \mu_{M12})$ by replacing population means by sample means.

f. Calculate a 95% interval estimate of θ. Based on the interval estimate, what can you say about the benefits of the addition of four years of education for males versus females? Use the 97.5 percentile from the standard normal, 1.96, when calculating the interval estimate.

C.24 How much does the variation in wages change when individuals receive more education? Is the variation different for males and females? Use the data file *cps5* to explore this question. If your software does not permit using this larger sample use *cps5_small*.

a. Calculate the sample variance of wages of females who have 12 years of education. Calculate the sample variance of wages of females who have 18 years of education. What did you discover?

b. Carry out a two-tail test, using a 5% level of significance, of the hypothesis that the variance of wage is the same for females with 12 years of education and females with 18 years of education.

c. Calculate the sample variance of wages of males who have 12 years of education. Calculate the sample variance of wages of males who have 18 years of education. What did you discover?

d. Carry out a two-tail test, using a 5% level of significance, of the hypothesis that the variance of wage is the same for males with 12 years of education and males with 18 years of education.

e. Carry out a two-tail test of the null hypothesis that the mean wage for males with 18 years of education is the same as the mean wage of females with 18 years of education. Use the 1% level of significance.

C.25 What happens to the household budget share of necessity items, like food, when total household expenditures increase? Use data file *malwai_small* for this exercise.

a. Obtain the summary statistics, including the median and 90th percentile, of total household expenditures.

b. Construct a 95% interval estimate for the proportion of income spent on food by households with total expenditures less than or equal to the median.

c. Construct a 95% interval estimate for the proportion of income spent on food by households with total expenditures more than or equal to the 90th percentile.

d. Summarize your findings from parts (b) and (c).

e. Test the null hypothesis that the population mean proportion of income spent on food by households is 0.4. Use a two-tail test and the 5% level of significance. Carry out the test separately using the complete sample, and using the samples of households with total expenditures less than or equal to the median, and again for households whose total expenditures are in the top 10%.

C.26 At the famous Fulton Fish Market in New York City sales of Whiting (a type of fish) vary from day to day. Over a period of several months, daily quantities sold (in pounds) were observed. These data are in the data file *fultonfish*.

a. Using the data for Monday sales, test the null hypothesis that the mean quantity sold is greater than or equal to 10,000 pounds a day, against the alternative that the mean quantity sold is less than 10,000 pounds. Use the $\alpha = 0.05$ level of significance. Be sure to (i) state the null and alternative hypotheses, (ii) give the test statistic and its distribution, (iii) indicate the rejection region, including