

Problems and Exercises

For all hypothesis tests, assume that the level of significance is 0.05 unless otherwise stated.

1. Create an Excel workbook with worksheet templates (similar to the Excel workbook *Confidence Intervals*) for one-sample hypothesis tests for means and proportions. Apply your templates to the example problems in this chapter. (For subsequent problems, you should use the formulas in this chapter to perform the calculations, and use this template only to verify your results!)
2. A company is considering two different campaigns, A and B, for the promotion of their product. Two tests are conducted in two market areas with identical consumer characteristics, and in a random sample of 60 customers who saw campaign A, 18 tried the product. In a random sample of 100 customers who saw campaign B, 22 tried the product. What conclusion can management reach? (Assume that the population variance is not known.)
3. A management institute checked the past records of applicants and the mean score calculated was 350. The administration is interested to know whether the quality of new applicants has changed or not. From the recent scores of 100 applicants, the mean is 365 with a standard deviation of 38. Does this data provide statistical evidence that the quality of recent applicants has improved?
4. A retailer believes that its new advertising strategy will increase sales. Previously, the mean spending in 15 categories of consumer items in both the 18–34 and 35+ age groups was \$70.00.
 - a. Formulate a hypothesis test to determine if the mean spending in these categories has statistically increased.
 - b. After the new advertising campaign was launched, a marketing study found that the mean spending for 300 respondents in the 18–34 age group was \$75.86, with a standard deviation of \$50.90. Is there sufficient evidence to conclude that the advertising strategy significantly increased sales in this age group?
 - c. For 700 respondents in the 35+ age group, the mean and standard deviation were \$68.53 and \$45.29, respectively. Is there sufficient evidence to conclude that the advertising strategy significantly increased sales in this age group?
5. A financial advisor believes that the proportion of investors who are risk-averse (i.e., try to avoid risk in their investment decisions) is at least 0.7. A survey of 32 investors found that 20 of them were risk-averse. Formulate and test the appropriate hypotheses to determine whether his belief is valid.
6. Metropolitan Press hypothesizes that the average life of its largest Web press is 14,500 hours. They know that the standard deviation of press life is 2,100 hours. From a sample of 25 presses, the company find sample mean of 13,000 hours. At a 0.01 significance level, should the company conclude that the average life of the presses is less than the hypothesized 14,500 hours?
7. Ice Cream Manufacture is to produce a new ice cream flavor. The company's marketing research department surveyed 6,000 families and 335 of them showed interest in purchasing the new flavor. A similar study made two year ago showed that 5% of the families would purchase the flavor. What should the company conclude regarding the new flavor?
8. Call centers typically have high turnover. The director of human resources for a large bank has compiled data on about 70 former employees at one of the bank's call centers in the Excel file *Call Center Data*. In writing an article about call center working conditions, a reporter has claimed that the average tenure is no more than 2 years. Formulate and test a hypothesis using these data to determine if this claim can be disputed.
9. The manager of a store claims that 60% of the shoppers entering the store leave without making a purchase. Out of a sample of 50, it is found that 35 shoppers left without buying. Is the result consistent with the claim?
10. A sample of 400 athletes is found to have mean height of 171.38 cm. Can we call it a sample from a large population of mean height 171.17 and standard deviation of 3.30 cm?
11. The State of Ohio Department of Education has a mandated ninth-grade proficiency test that covers writing, reading, mathematics, citizenship (social studies), and science. The Excel file *Ohio Education Performance* provides data on success rates (defined as the percent of students passing) in school districts in the greater Cincinnati metropolitan area along with state averages. Test null hypotheses that the average scores in the Cincinnati area are equal to the state averages in each test and also for the composite score.
12. Formulate and test hypotheses to determine if statistical evidence suggests that the graduation rate for (1) top liberal arts colleges or (2) research universities in the sample *Colleges and Universities* exceeds 90%. Do the data support a conclusion that the graduation rates exceed 85%? Would your conclusions

XLMiner also provides **cross-validation**—a process of using two sets of sample data; one to build the model (called the training set), and the second to assess the model's performance (called the validation set). This will be explained in Chapter 10 when we study data mining in more depth, but is not necessary for standard regression analysis.

Key Terms

Autocorrelation	Multiple correlation coefficient
Best-subsets regression	Multiple linear regression
Coefficient of determination (R^2)	Overfitting
Cross-validation	Parsimony
Coefficient of multiple determination	Partial regression coefficient
Curvilinear regression model	Polynomial function
Dummy variables	Power function
Exponential function	R^2 (R-squared)
Homoscedasticity	Regression analysis
Interaction	Residuals
Least-squares regression	Significance of regression
Linear function	Simple linear regression
Logarithmic function	Standard error of the estimate, S_{YX}
Multicollinearity	Standard residuals

Problems and Exercises

- Each worksheet in the Excel file *LineFit Data* contains a set of data that describes a functional relationship between the dependent variable y and the independent variable x . Construct a line chart of each data set, and use the Excel *Trendline* tool to determine the best-fitting functions to model these data sets.
- A consumer products company has collected some data relating to the advertising expenditure and sales of one of its products:

Advertising cost	Sales
\$300	\$7000
\$350	\$9000
\$400	\$10000
\$450	\$10600

What type of model would best represent the data? Use the Excel *Trendline* tool to find the best among the options provided.

- Using the data in the Excel file *Demographics*, determine if a linear relationship exists between unemployment rates and cost of living indexes by constructing a scatter chart. Visually, do there appear to be any outliers? If so, delete them and then find the best-fitting linear regression line using the Excel *Trendline* tool. What would you conclude about the strength of any relationship? Would you use regression to make predictions of the unemployment rate based on the cost of living?
- Using the data in the Excel file *Weddings* construct scatter charts to determine whether any linear relationship appears to exist between (1) the wedding cost and attendance, (2) the wedding cost and the value rating, and (3) the couple's income and wedding cost only for the weddings paid for by the bride and groom. Then find the best-fitting linear regression lines using the Excel *Trendline* tool for each of these charts.
- Using the data in Excel file *Loans*, construct a scatter chart for monthly income versus loan amount and add a linear trendline. What is the regression model? If an individual has 7336 as monthly income, what would you predict the loan amount to be?
- Using the results of fitting the *Home Market Value* regression line in Example 8.4, compute the errors associated with each observation using formula (8.3) and construct a histogram.

7. Set up an Excel worksheet to apply formulas (8.5) and (8.6) to compute the values of b_0 and b_1 for the data in the Excel file *Home Market Value* and verify that you obtain the same values as in Examples 8.4 and 8.5.
8. The managing director of a consulting group has the following monthly data on total overhead costs and professional labor hours to bill to clients:⁴

Overhead Costs	Billable Hours
\$365,000	3,000
\$400,000	4,000
\$430,000	5,000
\$477,000	6,000
\$560,000	7,000
\$587,000	8,000

- a. Develop a trendline to identify the relationship between billable hours and overhead costs.
 - b. Interpret the coefficients of your regression model. Specifically, what does the fixed component of the model mean to the consulting firm?
 - c. If a special job requiring 1,000 billable hours that would contribute a margin of \$38,000 before overhead was available, would the job be attractive?
9. Using the Excel file *Weddings*, apply the Excel Regression tool using the wedding cost as the dependent variable and attendance as the independent variable.
 - a. Interpret all key regression results, hypothesis tests, and confidence intervals in the output.
 - b. Analyze the residuals to determine if the assumptions underlying the regression analysis are valid.
 - c. Use the standard residuals to determine if any possible outliers exist.
 - d. If a couple is planning a wedding for 175 guests, how much should they budget?
10. Using the Excel file *Weddings*, apply the Excel Regression tool using the wedding cost as the dependent variable and the couple's income as the independent variable, only for those weddings paid for by the bride and groom.
 - a. Interpret all key regression results, hypothesis tests, and confidence intervals in the output.
 - b. Analyze the residuals to determine if the assumptions underlying the regression analysis are valid.
 - c. Use the standard residuals to determine if any possible outliers exist.
 - d. If a couple makes \$70,000 together, how much would they probably budget for the wedding?
11. Using the data in Excel file *Crime*, apply the Excel regression tool using crime rate (CRIM) as the dependent variable and pupil-teacher ratio (PTRATIO) in the region as the independent variable.
 - a. Interpret all key regression results, hypothesis tests, and confidence intervals in the output.
 - b. Use the standard residuals to determine if any outliers exist.
12. Using the data in the Excel file *Student Grades*, apply the Excel Regression tool using the midterm grade as the independent variable and the final exam grade as the dependent variable.
 - a. Interpret all key regression results, hypothesis tests, and confidence intervals in the output.
 - b. Analyze the residuals to determine if the assumptions underlying the regression analysis are valid.
 - c. Use the standard residuals to determine if any possible outliers exist.
13. The Excel file *National Football League* provides various data on professional football for one season.
 - a. Construct a scatter diagram for Points/Game and Yards/Game in the Excel file. Does there appear to be a linear relationship?
 - b. Develop a regression model for predicting Points/Game as a function of Yards/Game. Explain the statistical significance of the model.
 - c. Draw conclusions about the validity of the regression analysis assumptions from the residual plot and standard residuals.
14. A deep-foundation engineering contractor has bid on a foundation system for a new building housing the world headquarters for a *Fortune* 500 company.

⁴Modified from Charles T. Horngren, George Foster, and Srikant M. Datar, *Cost Accounting: A Managerial Emphasis*, 9th ed. (Englewood Cliffs, NJ: Prentice Hall, 1997): 371.