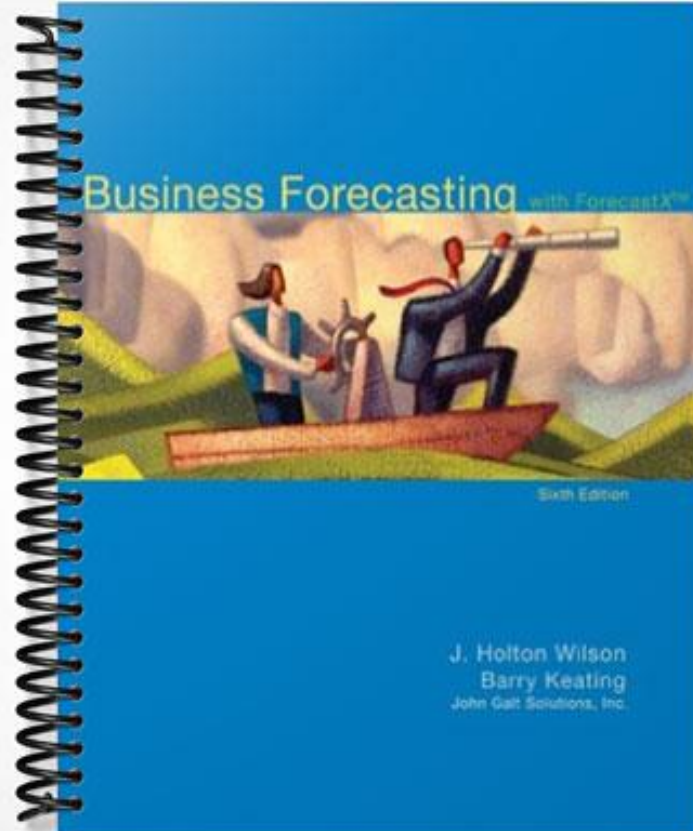


TEST BANK



Business Forecasting with ForecastX™

Sixth Edition

J. Holton Wilson
Barry Keating
John Galt Solutions, Inc.

Chapter 2

MULTIPLE CHOICE TEST BANK

Note: The correct answer is denoted by **.

1. Why are forecasting textbooks full of applied statistics?
 - A) Statistics is the study of uncertainty.
 - B) Real-world business decisions involve risk and uncertainty.
 - C) Forecasting attempts to reduce the uncertainty for uncertain events.
 - D) Forecasting ultimately deals with probability.
 - E) All the above. **

2. Which of the following is not part of the recommended nine-step forecast process?
 - A) What role do forecasts play in the business decision process?
 - B) What exactly is to be forecast?
 - C) How urgent is the forecast?
 - D) Is there enough data?
 - E) All the above are parts of the process. **

3. Of the following model selection criteria, which is often the most important in determining the appropriate forecast method?
 - A) Technical background of the forecast user.
 - B) Patterns the data have exhibited in the past. **
 - C) How much money is in the forecast budget?
 - D) What is the forecast horizon?
 - E) When is the forecast needed?

4. In a time series plot of a typical The GAP store which of the following data patterns would be most prominent?
 - A) Trend.
 - B) Seasonal. **
 - C) Cyclical.
 - D) Random.
 - E) All the above.

5. Which of the following is incorrect?
 - A) The forecaster should be able to defend why a particular model or procedure has been chosen.
 - B) Forecast errors should be discussed in an objective manner to maximize management's confidence in the forecast process.

- C) Forecast errors should not be discussed since most people know that forecasting is an inexact science. **
 - D) You should tailor your presentation to the sophistication of the audience to maximize credibility in the forecast process.
 - E) None of the above.
6. In the model-testing phase of the nine-step process, which of the following refers to that portion of a sample that is best to use to evaluate model-forecast accuracy?
- A) Fit.
 - B) Forecast horizon.
 - C) Holdout period. **
 - D) Accuracy.
 - E) None of the above.
7. The text presents a guide to selecting an appropriate forecasting method based on
- A) data patterns.
 - B) quantity of historical data available.
 - C) forecast horizon.
 - D) quantitative background of the forecast user.
 - E) All the above. **
8. Which time-series component is said to fluctuate around the long-term trend and is fairly irregular in appearance?
- A) Trend.
 - B) Cyclical. **
 - C) Seasonal.
 - D) Irregular.
 - E) None of the above.
9. Forecasting January sales based on the previous month's level of sales is likely to lead to error if the data are _____.
- A) Stationary.
 - B) Non-cyclical.
 - C) Seasonal. **
 - D) Irregular.
 - E) None of the above.
10. The difference between seasonal and cyclical components is:
- A) Duration.
 - B) Source.
 - C) Predictability.

- D) Frequency.
E) All the above. **
11. For which data frequency is seasonality not a problem?
- A) Daily.
B) Weekly.
C) Monthly.
D) Quarterly.
E) Annual. **
12. One can realistically not expect to find a model that fits any data set perfectly, due to the _____ component of a time series.
- A) Trend.
B) Seasonal.
C) Cyclical.
D) Irregular. **
E) None of the above.
13. When a time series contains no trend, it is said to be
- A) nonstationary.
B) seasonal.
C) nonseasonal.
D) stationary. **
E) filtered.
14. Stationarity refers to
- A) the size of the RMSE of a forecasting model.
B) the size of variances of the model's estimates.
C) a method of forecast optimization.
D) lack of trend in a given time series. **
E) None of the above.
15. Which of the following is not a measure of central tendency in a population?
- A) Mean.
B) Mode.
C) Median.
D) Range. **
16. Which of the following is not a descriptive statistic?
- A) Median.

- B) Mean.
 - C) Range.
 - D) Variance.
 - E) None of the above. **
17. Which of the following is not an important part of classical statistics?
- A) Summary measures of probability distributions called descriptive statistics.
 - B) Probability distribution functions, which characterize all outcomes of a variable.
 - C) The use of sampling distributions, which describe the uncertainty in making inference about the population on the basis of a sample.
 - D) The concept of expected value.
 - E) All of the above are important. **
18. The standard normal probability table
- A) is equivalent to a t distribution if the sample size is less than 30.
 - B) shows a normal distribution with standard deviation equal to zero.
 - C) is used to make inference for all normally distributed random variables. **
 - D) All the above.
 - E) None of the above.
19. The median and mode may be more accurate than the sample mean in forecasting the populations mean when
- A) The sample size is small.
 - B) The sample size is large.
 - C) The sample has one large outlier. **
 - D) The population is assumed to be normally distributed.
 - E) All the above.
20. The arithmetic average of the occurrence of some random variable is also called the _____.
- A) Range.
 - B) Mean. **
 - C) Variance.
 - D) Standard deviation.
 - E) None of the above.
21. In finance, an investor who ignores risk is termed "risk neutral." What descriptive statistic is our risk neutral investor ignoring when she generates stock portfolios?
- A) Median.
 - B) Mean.
 - C) Mode.

- D) Standard deviation. **
E) None of the above.
22. In calculating the sample variance we subtract one from the sample size. This is because
- A) the population mean is unknown.
B) of using the sample mean to estimate the population mean.
C) the sum of deviations about the sample mean is zero.
D) the sample mean is employed.
E) All the above. **
23. Which statistic is correctly interpreted as the "average" spread of data about the mean?
- A) Mode.
B) Range.
C) Variance.
D) Standard deviation. **
E) Mean.
24. Which measure of dispersion in a data set is the most intuitive and represents an average?
- A) Range.
B) Mode.
C) Standard deviation. **
D) Variance.
E) Mean.
25. Which of the following is not an attribute of a normal probability distribution?
- A) It is symmetrical about the mean.
B) Most observations cluster around the mean.
C) Most observations cluster around zero. **
D) The distribution is completely determined by the mean and variance.
E) All the above are correct.
26. Which of the following is not a foundation of classical statistics?
- A) Summary measures of probability distribution called descriptive statistics.
B) Probability distribution function which characterize all possible outcomes of a random variable.
C) The knowledge of thousands and thousands of normal probability tables required for statistical inference of normally distributed random variables. **
D) The concept of expected value, which is the average value of a random variable taken over a large number of samples.

27. A company claims that the rubber belts, which it manufactures, have a mean service life of at least 800 hours. A random sample of 36 belts from a very large shipment of the company's belts shows a mean life of 760 hours and a standard deviation of 90 hours. Which of the following is the most appropriate on the basis of the sample results?

- A) The sample results do not warrant rejection of the company's claim if the risk of a Type I error is specified at .05.
- B) The sample results do warrant rejection of the company's claim if the risk of Type I error is specified at .05. **
- C) Since the sample mean falls below the company's claim, the sample results indicate that the company claim is incorrect.
- D) The sample results are indeterminate since the magnitude of the sample standard deviation is greater than the difference between the company's claimed figure and the sample mean.

28. Based upon ten years of monthly data, the monthly rate of return of the DOW Jones 30 composite stock portfolio was normally distributed with mean .0084 and variance .0014. What is the probability, that in any given month, we observe a rate of return on the DOW above 10 percent?

- A) Less than one percent. **
- B) Two percent.
- C) Three percent.
- D) Not enough information is provided to answer the question.

29. Suppose you observe the entire population of a random variable and you wish to test some hypothesis about the mean. To perform your hypothesis test you

- A) apply a sampling distribution to the problem.
- B) obtain sample estimates of population parameters.
- C) simply find the population mean and compare it to the hypothesized value. **
- D) Apply the t distribution.
- E) There is no answer to this question.

30. If two large random samples are drawn from two populations, each having a mean of \$100, the relevant sampling distribution of their difference has a mean of:

- A) \$200.
- B) The sum of the two sample means.
- C) 0. **
- D) The difference between the two sample means.

31. The sampling distribution of the sample mean, when sampling from a normal population with mean μ and variance σ^2 is:

- A) normally distributed with mean μ and variance σ^2 .

- B) normally distributed with mean μ and variance s^2 .
- C) normally distributed with mean μ and variance σ^2/n . **
- D) normally distributed with mean 0 and variance 1.
32. Type I error
- A) is said to arise when we reject a true null hypothesis.
- B) has a probability value equal to the significance level of any statistical test.
- C) is a measure of the uncertainty associated with rejecting any null hypothesis on the basis of sample data.
- D) both a) and b) are correct.
- E) A, B, and C are all correct. **
33. The sampling distribution of the sample mean is
- A) normally distributed with mean μ and variance σ^2 .
- B) normally distributed with mean μ and variance s^2 .
- C) distributed as a t distribution with variance 1.
- D) normally distributed with mean 0 and variance 1.
- E) None of the above. **
34. Sampling distributions
- A) are the distributions of all possible values of a sample statistic based upon repeated sampling.
- B) are used to make inference when the population of a variable is unobservable.
- C) exhibit important properties for the ranking of alternative estimators such as unbiasedness and efficiency.
- D) All the above. **
35. The null hypothesis is that there is no significant linear relationship between prices and sales. Type I error is:
- A) To conclude that there is a significant linear relationship when there is.
- B) To conclude that there is not a significant linear relationship when there is.
- C) To conclude that there is a significant linear relationship when there is not. **
- D) To conclude that the correlation coefficient is equal to zero.
36. An unbiased model
- A) is one that does not consistently over-estimate or under-estimate the true value of a parameter. **
- B) is one that consistently produces estimates with the smallest RMSE.
- C) is one, which contains no independent variable; it depends solely on time-series pattern recognition.
- D) is one made up by a team of forecasters.

37. Suppose that you mistakenly move the decimal point to the right one digit in data from a normal population with a mean of zero. What happens to the standard deviation?

- A) Data with mistake has standard error ten times the original. **
- B) Data with mistake has same standard error as the original.
- C) Data with the mistake has twice the standard error of the original.
- D) Data with the mistake has one hundred times the standard error of the original.
- E) None of the above.

38. Which test statistic is appropriate for statistical inference about the population mean when both the population mean and variance are unknown?

- A) $(\bar{X} - \mu)/\sigma$.
- B) $(\bar{X} - \mu)/s$.
- C) $(\bar{X} - \mu)/s$.
- D) None of the above. **

39. Which statement is incorrect?

- A) Confidence intervals depend on sample size.
- B) The sample mean is the best estimator if sampling from a normal population.
- C) The sample mean is an unbiased estimator.
- D) Confidence intervals provide no more information than point estimates. **
- E) The sample variance is an unbiased estimator.

40. Which of the following statements is incorrect about the family of normal distributions?

- A) $P[\mu \pm 1\sigma] = .68$.
- B) $P[\mu \pm 2\sigma] = .95$.
- C) $P[\mu \pm 3\sigma] = .99$.
- D) None of the above. **

41. The standard normal transformation

- A) involves subtracting the mean so as to center the transformed distribution on zero.
- B) involves division by the standard deviation so as to standardize the transformed variance to one.
- C) is used to make inference about all normal distributions based upon the standard normal distribution.
- D) is the number of standard deviations by which a random variable differs from its mean.
- E) All the above. **

42. A machine fills ketchup bottles. One of the requirements is that the mean content of the bottles should be 10 ounces. Management wishes to set up a decision rule to decide whether or

not this is true based on a random sample of bottles. The risk of type I error is specified at .05. A sample of 100 bottles will be taken; it is believed that the standard deviation of fills is .3 ounces. If it is decided that $Z = 2$, the decision rule boundary values are:

- A) 9.60 and 10.40.
 - B) 9.10 and 9.90.
 - C) 9.94 and 10.06. **
 - D) 9.40 and 10.60.
 - E) None of the above.
43. Last year's midterm results showed a mean of 51 points and a variance of 46. An approximate confidence interval is closest to:
- A) 44.2 to 57.8.
 - B) 37.4 to 64.6. **
 - C) 5 to 97.
 - D) None of the above.
44. A difference between the population standard deviation of the random variable X , and the standard deviation of the sampling distribution of the sample mean is
- A) one is based upon the other.
 - B) dependence on sample size.
 - C) the possibility of sampling error.
 - D) application to the t distribution.
 - E) All the above. **
45. Which probability distribution is appropriate for testing hypotheses concerning an unknown population mean when the sample variance is used to estimate the population variance?
- A) The normal distribution with mean μ and variance σ^2 .
 - B) The normal distribution with mean 0 and variance 1.
 - C) The standard normal distribution.
 - D) The t distribution with $n-1$ degrees of freedom. **
 - E) None of the above.

Note: The next two questions are a pair.

46. A random sample of bolts is taken from inventory and their length is measured. The average length in the sample is 5.3 inches, with a standard deviation of .2 inches. The sample size was 50. The point estimate for the mean length of all bolts in inventory is:
- A) 5.3 inches. **
 - B) .2 inches.
 - C) 4.908 to 5.692 inches.
 - D) 5.3 inches plus or minus .2

47. In the previous question, a 95% confidence interval for the unknown population mean is closest to:
- A) 5.3 inches.
 - B) 4.9 to 5.7 inches.
 - C) 5.3 inches plus or minus .056. **
 - D) 4.784 to 5.816 inches.
 - E) None of the above.
48. Which of the following statements about the probability of Type I and Type II error is not correct?
- A) Type I error cannot occur if the null hypothesis is false.
 - B) Type II error cannot occur if the null hypothesis is true.
 - C) If the null hypothesis is true, the results of the test will either be a correct conclusion or a Type I error.
 - D) It is not possible to specify both the probabilities of Type I and II errors since only one of them can occur. **
49. A sample of 100 selected at random from a process with a mean of 500.52 and a standard deviation of 4.0. Estimate the probability that a sample of 100 would have a mean equal to or greater than 500.52 if the true population mean is really 500.0.
- A) about .4938.
 - B) about .9032.
 - C) about .0968. **
 - D) about .4032.
50. A random sample of employee files is drawn revealing an average of 2.8 overtime hours worked per week with a standard deviation of .7; the sample size is 500. The resulting 90% confidence interval is:
- A) 2.1 to 3.5.
 - B) 2.6 to 3.5.
 - C) 2.75 to 2.85. **
 - D) 2.6 to 3.0.
 - E) None of the above.
51. A hypothesis test requires a two-tailed critical region if the
- A) alternative hypothesis is one of the true mean being above some number.
 - B) alternative hypothesis is one of the true mean being below some number.
 - C) alternative hypothesis is one of the true mean not being equal to some number. **
 - D) null hypothesis involves an inequality.

52. In statistical hypothesis testing, the approach is to see whether you find sufficient evidence to reject the null hypothesis. This implies
- A) the null is framed such that its rejection confirms a conjecture.
 - B) we can set the probability of rejecting a true null.
 - C) we can set the probability of Type I error to our satisfaction.
 - D) we seek to reject the null to confirm a belief.
 - E) All the above. **
53. A medical researcher has just calculated a correlation coefficient of zero for two particular random variables. Which of the following statements is most accurate?
- A) There is no significant linear difference between the two variables.
 - B) There is no significant relationship between the two variables.
 - C) There is no significant linear relationship between the two variables. **
 - D) There is a significant linear relationship between the two variables.
54. The correlation coefficient is an extremely important descriptive statistic because
- A) It provides a unit-free measure of how two random variables move together.
 - B) It provides a measure of the linear association between a pair of random variables.
 - C) It provides the forecaster with a diagnostic tool of when regression analysis is appropriate for the business-forecasting problem.
 - D) All the above. **
55. A large sample of X-Y data values are analyzed and reveal a correlation coefficient of -.88. Which statement is correct?
- A) If r had been +.88, the correlation would have been much stronger.
 - B) The correlation is weak because r is less than -1.
 - C) A fairly strong negative linear relationship exists. **
 - D) A weak negative relationship exists.
56. Which of the following is not true regarding the Central Limit Theorem?
- A) For a sufficient large sample, the sampling distribution is approximately normal.
 - B) The sampling distribution converges to the normal distribution as the sample size increases.
 - C) Regardless of the population distribution from which the sample is drawn, if the sample size is sufficiently large, the normal curve can be used.
 - D) Regardless of the sample size used, the normal curve can be used as the sampling distribution. **
57. Suppose two random variables X and Y are related as follows: $Y = 1/X^2$. The population Pearson correlation coefficient should be:

- A) +1.
 - B) 0. **
 - C) -1.
 - D) .5.
 - E) None of the above.
58. Which functions are not appropriate for use of the Pearson correlation coefficient to estimate the correlation between a pair of random variables?
- A) Cubic polynomials.
 - B) Quadratic polynomials.
 - C) Higher-order polynomials.
 - D) Functions involving a variable raised to the one-half power.
 - E) Reciprocal functions.
 - F) All the above. **
59. The sampling distribution of the sample Pearson correlation coefficient
- A) has a mean of zero.
 - B) has a mean of ρ . **
 - C) is biased.
 - D) has a mean of r .
 - E) None of the above.
60. If we were to know the true population correlation, confidence intervals for the population correlation can be constructed using the _____ distribution.
- A) t distribution.
 - B) standard normal distribution. **
 - C) chi-square distribution.
 - D) F distribution.
 - E) All the above.
61. If the scatterplot of two variables has a circular pattern, this suggests the two variables have a population correlation coefficient of
- A) -1.
 - B) -.5.
 - C) 0. **
 - D) +.5.
 - E) +1.
62. Which of the following is not used to calculate the sample Pearson correlation coefficient for the variables X and Y?
- A) Sample mean of X.

- B) Sample mean of Y.
 - C) Sample covariance of X and Y.
 - D) Sample standard deviation of X.
 - E) All the above are used to calculate correlation coefficients. **
63. Which of the following is not a benefit of a scatter diagram?
- A) The nature of the X-Y relationship (linear or nonlinear) may be revealed.
 - B) The strength of the relationship may be revealed.
 - C) The sign of the correlation coefficient will be revealed.
 - D) Displaying the population size. **
64. In order to conduct a correlation analysis, the collected data must be:
- A) Related to the real world.
 - B) Numerical. **
 - C) Consist of categories.
 - D) Highly Correlated.
 - E) All the above.
65. Which of the following is not a reason for testing the population correlation coefficient is zero?
- A) To see if r and ρ are equal. **
 - B) To make inference from sample to population.
 - C) To bring sample size into the analysis.
 - D) To determine if a significant X-Y relationship exists.
 - E) All the above are correct.
66. Suppose the sample Pearson correlation coefficient (r) is estimated to be .75 with a sample size of 35. The correct calculated value of the test statistic for a null of zero correlation is:
- A) 6.5. **
 - B) 8.6.
 - C) 44.1.
 - D) 2.5
 - E) None of the above.
67. When testing the null hypothesis that the population correlation between a pair of variables is zero
- A) the normal sampling distribution is used.
 - B) the chi-square distribution is used.
 - C) the standard normal distribution is used.
 - D) The t distribution is used for small samples. **

68. For a collection of 15 X-Y data values, the sample correlation coefficient was estimated at $-.63$ from a sample of size 15. The calculated t value for a null of zero correlation is:
- A) 2.92.
 - B) 1.92.
 - C) -2.92 . **
 - D) -1.92 .
 - E) None of the above.
69. When the correlation coefficient is negative, it means:
- A) there is a weak relationship.
 - B) when X goes down, Y does too.
 - C) X will not be a good predictor of Y.
 - D) when X goes down, Y tends to go up. **
 - E) None of the above.
70. When forecasting with time-series data, it is highly recommended to test for the presence of a trend in the data. Testing for trend at the 10% level of significance
- A) can be accomplished by use of a standard 95% correlogram.
 - B) requires use of the standard normal probability distribution.
 - C) can be accomplished by comparing the estimated autocorrelation coefficient with the number 2 divided by the square root of sample size.
 - D) requires use of the t distribution. **
71. Quarterly time-series data with a trend can be applied to models that assume stationary data by
- A) Averaging the data over time.
 - B) Taking the first difference of the original series. **
 - C) Taking the fourth difference of the original series.
 - D) Using a moving average.
72. Which of the following is not consistent with the presence of a trend in a time series?
- A) The autocorrelation function declines quickly to zero as the lag increases. **
 - B) The autocorrelation function of the first-differences declines quickly to zero as the lag increases.
 - C) The autocorrelation function declines slowly towards zero as the lag increases.
 - D) The autocorrelation function of the first-differences quickly declines to zero.
73. Serial correlation refers to the correlation between a variable and:
- A) itself.
 - B) another very similar variable.

- C) itself when lagged one or more periods. **
 D) another variable when the analysis is done on a computer.
 E) None of the above.
74. Which of the following is appropriate for testing the null hypothesis of zero autocorrelation at lag k at the approximate 95 percent level?
- A) Reject null if $|r_k| > 1/n$.
 B) Reject null if $|r_k| > 2/n$.
 C) Reject null if $|r_k| < 1/n$.
 D) Reject null if $|r_k| < 2/n$.
 E) None of the above. **
75. A time series whose 24-quarter lag correlogram shows no tendency to diminish towards zero can be said to
- A) have a trend term.
 B) be nonstationary.
 C) have a long memory.
 D) be serially correlated.
 E) All the above. **
76. Which of the following null hypotheses is not consistent with a test for data seasonality?
- A) $H_0: \rho_4 = \rho_8 = 0$ for quarterly data.
 B) $H_0: \rho_{12} = \rho_{24} = 0$ for monthly data.
 C) $H_0: \rho_1 = \rho_2 = 0$ for annual data. **
 D) $H_0: \rho_7 = 0$ for weekly data.
 E) All the above.

ESSAY/PROBLEM EXAM QUESTIONS

1. Volatility in exchange rates leads speculators to bet on market trends in foreign exchange markets. To examine this investment opportunity, data were obtained from 1980M1 through 1990M12 on the Japanese-yen US-dollar exchange rate, defined as number of yen per one US dollar.

The data were converted to monthly percentage changes by first-differencing the logarithms of the level. Treating our data as a population, we then calculated the monthly mean and variance for the percentage change in exchange rate. We found the percentage change in (yen/\$) exchange rate is distributed normally with mean $-.004386$ and variance $.00093585$.

What is the probability that, in any given month, we observe a 5% or more increase in the exchange rate? Formally, find: $P[\%(yen/\$) \geq .05]$.

ANSWER: The appropriate Z value for $\%(\text{yen}/\$) = .05$ is:

$$Z = [.05 - (-.004386)]/.03059 = 1.78.$$

Using Table 2-4, the probability can be found as follows:

$$P[Z \geq 1.778] = P[Z \geq 0] - P[0 \leq Z < 1.778] = .5 - .4625 = .0375.$$

Accordingly, there is about a 4% chance of a 5% or more change in the exchange rate in a given month.

2. Prove that, when sampling from a normally distributed population, the sample mean is an unbiased estimator of the population mean.

ANSWER: Suppose we draw a random sample from a random variable X having a normal distribution with mean μ and variance σ^2 . If the population is large we can assume that each drawing X_i is identically and independently normally distributed with mean μ and variance σ^2 .

To show that the sample mean is an unbiased estimator, we need to find the expected value of the sample mean:

$$E(\bar{X}) = E\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \left(\sum_{i=1}^n E(X_i)\right) = \frac{1}{n} n\mu = \mu.$$

Accordingly, the sample mean is an unbiased estimator of the population mean when sampling from a large normal population. This implies that, when using the sample mean to make inference about the population mean, we are correct on average.

3. Prove that the sum of deviations about the sample mean is zero.

ANSWER: The question asks you to show that for sample size n, that the sum of deviations about the sample mean is zero:

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

Using the definition of the sample mean, this can easily be shown:

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - \sum_{i=1}^n \bar{X} = n \bar{X} - n \bar{X} = 0.$$

4. A random sample of twelve automobiles showed the following figures for miles achieved on a gallon of gas. Assume the population distribution is normal. From the data:

$$n = 12, \sum_{i=1}^{12} X_i = 232.9, \sum_{i=1}^{12} X_i^2 = 4514.1, \text{ and } \sum_{i=1}^{12} (X_i - \bar{X})^2 = 13.2867.$$

(a) Find the expected miles per gallon of gas.

ANSWER: Sample mean = $232.9/12 = 19.4$ miles per gallon.

(b) Find the sample variance.

ANSWER: $13.2867/11 = 1.208$ is the estimated sample variance.

Note that we divide sum of squared deviations by 11, not 12. This is because we are using the sample mean to calculate deviations.

(c) Find an approximate 95% confidence interval for the population mean.

ANSWER: The approximate 95% confidence interval is two standard deviations from the mean using the sampling distribution of the sample mean:

$$P\left[\bar{X} - 2\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + 2\left(\frac{s}{\sqrt{n}}\right)\right] = .95.$$

After substituting in the appropriate estimates:

$$P\left[19.4 - 2\left(\frac{1.099}{\sqrt{12}}\right) \leq \mu \leq 19.4 + 2\left(\frac{1.099}{\sqrt{12}}\right)\right] = .95,$$

we can solve for the confidence interval:

$$P[18.765 < \mu < 20.03] = .95.$$

5. In 1985, the government bond yield in the United States was 10.62 percent. A random sample of government bond yields in nine foreign countries was:

11.04, 6.34, 10.94, 13.00, 7.34, 13.09, 4.78, 10.62, 6.87.

The mean foreign bond yield was 9.34 with variance 9.31. Assume that government bond yields are normally distributed.

At the 5 percent level of significance, test whether the government bond yields in the rest of the world during 1985 were lower than in the United States.

ANSWER: Following the four steps of hypothesis testing we have:

Step #1: Hypotheses: We formulate null so that its rejection verifies our assertion.

$$H_0: \mu_F \geq 10.62 \quad H_A: \mu_F < 10.62$$

Step #2: Test Statistic: Inference about an unknown population mean when the variance is unknown requires use of the t distribution:

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = t_{n-1}.$$

Step #3: Critical Region: Since we have a one-tailed alternative the critical region lies in the negative tail of the sampling distribution. In addition, since our sample size is 9 and the level of significance is 5%, we have $P(t_8 \leq t_{8,.05}) = .05$, hence $t_{8,.05} = -1.86$.

Step #4: Decision Rule:

$$\text{Reject } H_0 \text{ if } \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right| > 1.860.$$

Inserting the values of sample mean and standard error we have:

$$\frac{9.34 - 10.62}{\frac{\sqrt{9.31}}{\sqrt{9}}} = -1.26.$$

Accordingly, we cannot reject the null and conclude that world rates of interest were **not** significantly lower than in the United States.

6. WABC radio is the home of a popular morning talk show called "Imus in the Morning," featuring talk show host Don Imus. WABC was concerned that many listeners found Don Imus too offensive. Accordingly, WABC conducted a market research survey in which a sample of listeners were asked their opinion of the show on a scale of 1 to 10, with 10 being the most favorable ranking. The mean response for a sample of 400 people was 7.25 with a sample standard deviation of 2.51. Using this same survey technique, the Howard Stern Show had a mean rating of 7.85.

Does Don Imus have statistically the *same* approval rating as Howard Stern? Test this assertion using a 5 percent level of significance.

ANSWER: Formulate the null as if Imus's was equal to Stern's:

$$H_0: \mu = 7.85, \text{ and } H_A: \mu \neq 7.85$$

Since both the population mean and variance were unknown, the appropriate critical region, applying the t distribution is with 399 degrees of freedom is:

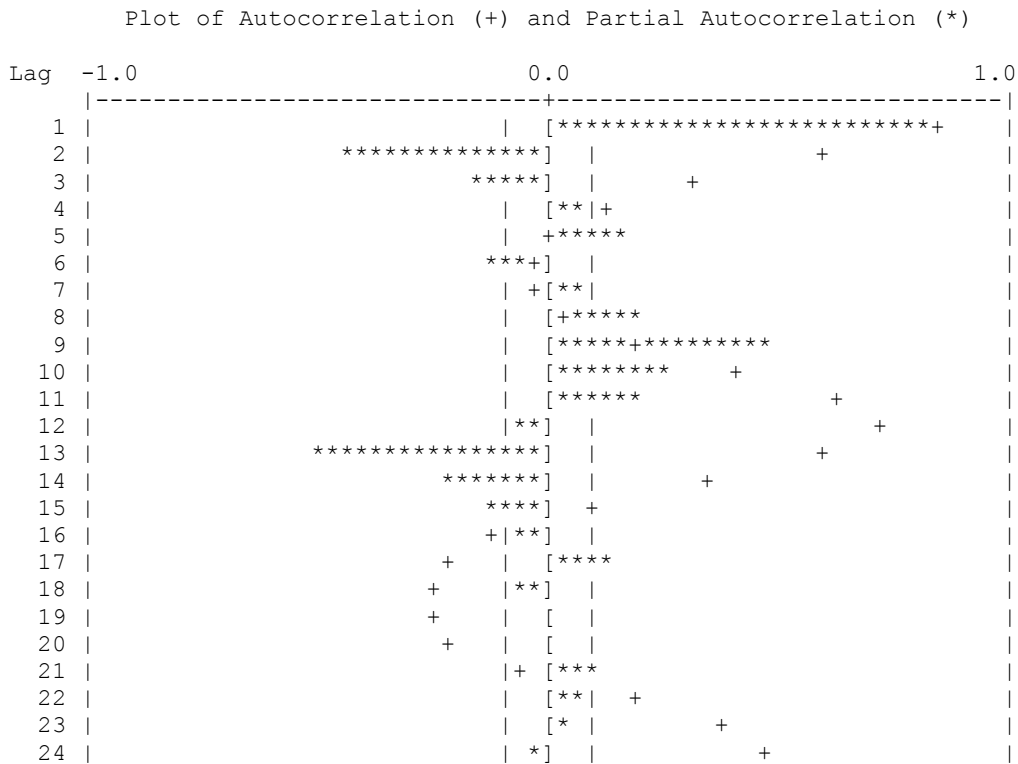
$$\text{Reject } H_0 \text{ if } \left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right| > 1.965.$$

The calculated value of the t-statistic is:

$$\frac{7.25 - 7.85}{\frac{2.51}{\sqrt{400}}} = -4.78.$$

Since this lies in the critical region, we reject the null hypothesis and conclude that Imus's approval rating is statistically *different* than Limbaugh's. (Note that we would get the same result using a one-tailed alternative).

7. Private housing starts are considered leading indicators of future economic activity. Using monthly data on private housing starts over the period 1959M1-1997M4, the estimated correlograms for a 24-month lag structure are reported below.



Using the approximate 95% rule, do private housing starts data exhibit seasonal variation? Explain your inferences clearly and note what hypotheses you are testing in your answer.

ANSWER: Seasonality refers to variation in the level of a series that occurs at the same time each year. Accordingly, seasons should be correlated over time.

To test for seasonality we can use estimated autocorrelation coefficients under the following null hypothesis for monthly data:

$$H_0: \rho_k = 0 \text{ for } k \text{ of } 12 \text{ and } 24.$$

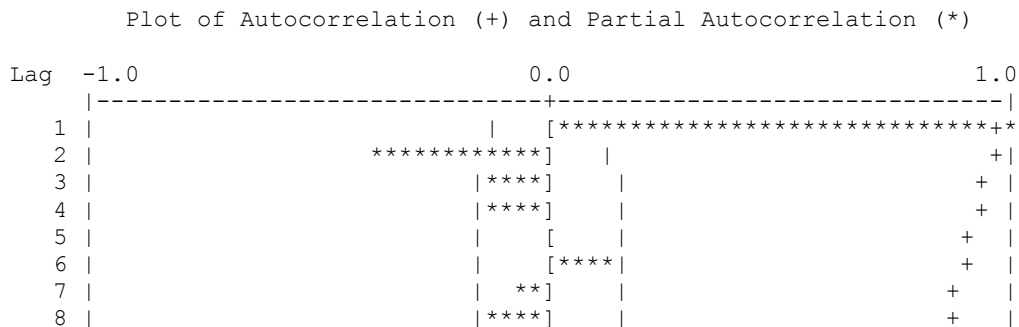
Our data series on private housing starts has 460 observations. Accordingly, using the approximate 95% rule, the absolute critical value of r_k is:

$$|r_k| = \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{460}} = .0932.$$

Since the estimated values of r_{12} and r_{24} are above the critical value of .0932, we reject the null of zero autocorrelation at lags of 12- and 24-months. Accordingly, there is sufficient evidence to conclude the data is seasonal. Indeed, as shown in the correlogram, the seasonal patterns are quite significant.

8. Gross Domestic Product (GDP) is a measure of the current dollar value of all goods and services produced in the United States. It is essentially a base measure of national income and is at the center of most debates about the well being of the U.S. economy. Accordingly, the forecasting of GDP is done by many agencies -- Office of Management and Budget, Federal Reserve Board, Department of Labor, Department of Commerce -- to name a few.

A necessary first-step in preliminary model selection is to examine whether the data is stationary. To examine this statistically, quarterly seasonally adjusted GDP data were used to obtain the following correlogram were estimated using a two-year lag.



a) Is GDP stationary? Explain using the approximate 95% rule.

ANSWER: Stationarity refers to the absence of any long-term trend in the level of a time series. Stationarity can be tested by examination of the autocorrelation coefficients over time. Ability to reject the null:

$$H_0: \rho_k = 0 \text{ for } k \text{ within } 1\text{-}2 \text{ years,}$$

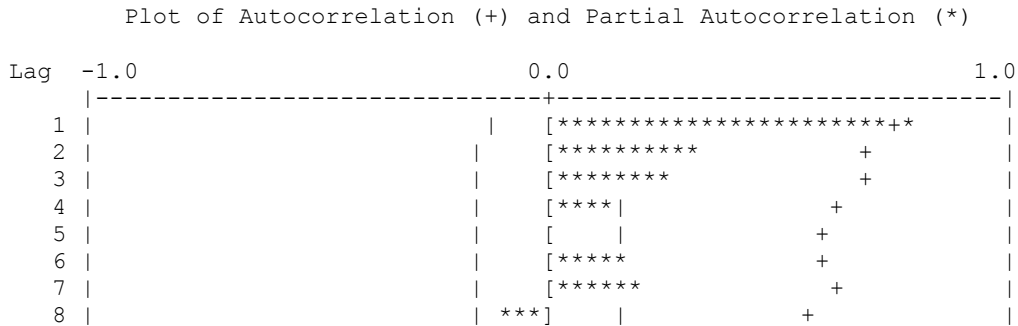
implies that the time series is nonstationary.

Our data series on GDP has 205 observations. Accordingly, using the approximate 95%, the absolute critical value of r_k is:

$$|r_k| = \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{205}} = .139686.$$

Since all of the estimated coefficients exceed this critical value, we can reject each individual null of zero autocorrelation, and conclude that the data are nonstationary.

b) As a further examination of the behavior of GDP we next first-differenced the data to remove any trend. The resulting series is DGDP and a 2-year autocorrelation correlogram is reported below.



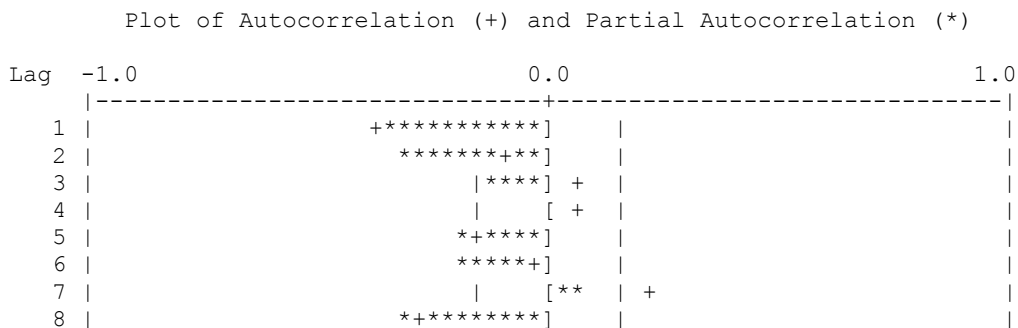
Are the first differences of GDP (DGDP) stationary? Explain.

ANSWER: The first-differences of GDP (DGDP) are not stationary as shown by the ability to reject the null:

$$H_0: \rho_k = 0 \text{ for } k = 1, 2, \dots, 8.$$

Accordingly, the first-differenced series still has trend. This is consistent with a nonlinear trend in the levels and may require further differencing.

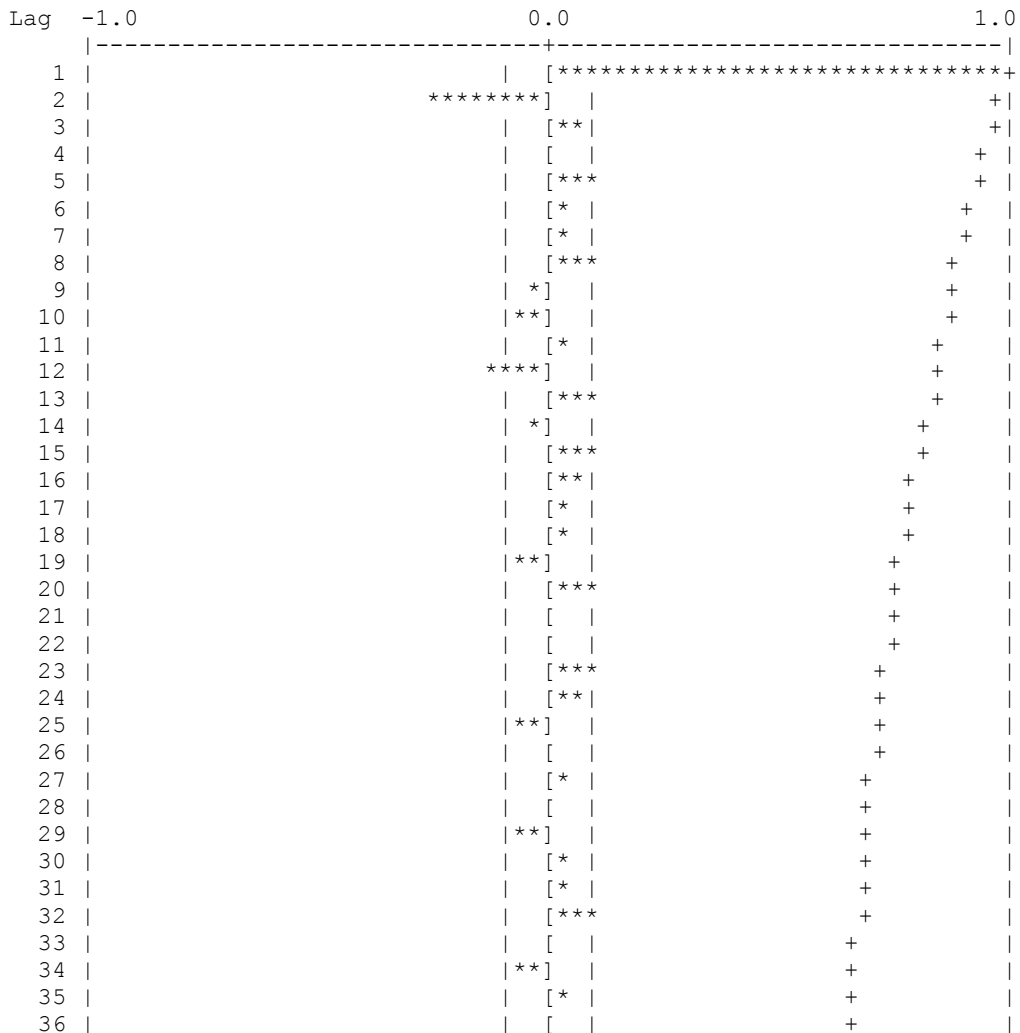
Specifically, a correlogram and autocorrelation function for differences of the first-differenced series is reported below.



As shown in the correlogram, this series is now stationary. Students should note that, while most economic data is first-difference stationary, a few series will require second differencing to remove any trend in the data.

9. Stationarity refers to whether or not time-series data has an upward or downward trend over time. To examine this issue with regard to stock prices, monthly data were obtained for the period 1947M1 through 1997M4 on the Standard and Poor's 500 composite. Using a 36-month lag, the following correlogram was estimated.

Plot of Autocorrelation (+) and Partial Autocorrelation (*)



Using the correlogram, are stock prices stationary? Explain.

ANSWER: Since we can reject all of the autocorrelations being zero for lags up to three years, stock prices are clearly nonstationary. Specifically, we can reject the following null with level of significance 5% using the correlogram:

$$H_0: \rho_k = 0 \text{ for } k = 1 \text{ to } 36.$$

This is why financial analysts think in terms of rates of return, which allows a comparison of stock price changes over time and across stocks.

10. Explain why a 95% confidence interval for a population parameter is wider than a 90% confidence interval for that parameter based on the same information.

ANSWER: This is a general result. Based on the same information, the greater the probability content the wider will be the confidence interval for any population parameter. This is to be

expected; the surer we want to be that a computed interval will contain the parameter, the wider the interval that will be required.

11. Daimler-Chrysler Motor Company is testing a new engine for miles per gallon (MPG). Based upon testing under normal conditions for 100,000 miles, the following sample mileages were obtained: 30.7, 31.8, 30.2, 32.0, and 31.3.

a) Find a point estimate for MPG.

ANSWER: The point estimate is simply the sample mean of 31.2.

b) What is the sample variance of MPG?

ANSWER: Summing the squared deviations about the sample mean and dividing by sample size minus one gives us .565 as our estimate of the standard error of MPG.

c) Suppose Chrysler wanted to advertise the new engine as obtaining at least 30 miles-per-gallon under normal driving conditions. Can they proceed with the advertising campaign at the 95% level of confidence?

ANSWER: Perhaps the easiest way to examine this is using a 95% confidence interval. The formula for an exact 95% confidence interval is:

$$P\left[\bar{X} - t_{n-1,\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{n-1,\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right] = 1 - \alpha.$$

Given our extremely small sample size of 5, the appropriate t-value for $\alpha = .05$ is 2.776. Plugging in the values of the sample mean and standard deviation, our confidence interval becomes:

$$P\left[31.2 - 2.776\left(\frac{.7517}{\sqrt{5}}\right) \leq \mu \leq 31.2 + 2.776\left(\frac{.7517}{\sqrt{5}}\right)\right] = .95.$$

Accordingly, our 95% confidence interval is 30.3 to 32.1 MPG. Since the lower bound for this interval is above 30 MPG, we can be 95% confident that the true mean MPG exceeds 30 MPG. Go ahead with the advertising campaign.