Excel Descriptive Statistics

Excel Descriptive Statistics

A descriptive statistics is a measure of an individual variable.

- Mean: arithmetic average
- Median: middle point in distribution
- Mode: most common value (highest frequency of occurrence)
- Kurtosis: Are the data peaked higher or lower than normal?
- Skewness: Is the peak shifted left or right?
- Standard deviation: measure of spread
- Range: highest value minus lowest value

Business Analytics: Excel Descriptive Statistics

The End

Excel Correlations

Excel Correlations

Correlation measures the direction of movement of two variables.

- If two variables move in the same direction, the correlation is positive.
- If two variables move in opposite directions, the correlation is negative.
- If change in one variable is not related to change in another variable, the correlation is zero.
- Correlations range from −1 to +1.
- Correlations close to zero are weak correlations; correlations far from zero are strong correlations.
- A correlation does not measure how much each variable changes, just the direction.

Business Analytics: Excel Correlations

The End

Regression Overview

Regression

Regression measures the direction and magnitude of change of a dependent (Y) variable when an independent (X) variable changes.

- First, check the significance of F (also called the p value of the F-statistic).
  - If it is 0.05 or less, then the equation is significant, and you can interpret it.
  - If it is more than 0.05, then the equation is not significant, and you can say nothing.
- Second, interpret the R^2.
  - The R^2 measures the percent of change in Y explained by the change in X.
- Third, check the p value of the coefficient.
  - If it is 0.05 or less, then the coefficient is significant, and you can interpret it.
  - If it is more than 0.05, then the coefficient is not significant, and you can say nothing.

Linear Regression Assumptions

Linear regression is only valid if the following assumptions are true:

1. The relationship between X and Y is linear.
   - This assumption is violated when you have nonlinear data; this is called nonlinearity.

2. The Xs are fixed numbers, not random variables (nonstochastic) and not related to each other (i.e., independent: $Corr(X_i, X_j) = 0$).
   - This assumption is violated when two or more X variables are correlated; this is called multicollinearity.

Linear Regression Assumptions (cont.)

3. The error terms (also called residuals):

   a. Have zero mean and constant variance: $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$
      - This assumption is violated when the spread of residuals increases or decreases when X changes; this is called heteroscedasticity. The residuals often appear as a wedge or sideways V shape.

   b. Are independent: $Cov(\varepsilon_i, \varepsilon_j) = 0$
      - This assumption is violated when the residuals are correlated as X changes; this is called serial correlation. The residuals often appear as a U shape or upside-down U shape.

   c. Are normally distributed: $\sim N(0, \sigma^2)$
      - This assumption is violated when one or more data points are far from the regression line; this is called an outlier.

Exponential Regression

- Exponential regression is used when there is compounding over time.
- Data that are often compounded include:
  - Sales data: sales might increase 10% per year, compounded over time
  - Savings accounts: most banks offer compounded interest

Power Regression

- Power regression is often used when there is a learning curve.

- A learning curve means that the time to accomplish a task decreases over time.

- A learning curve applied to manufacturing yields economies of scale (i.e., the more a factory produces, the less time it takes to produce each item).

Time Series With Moving Average

- Sales data often are affected by seasonality.
  - Seasonality occurs when one time of year affects the trend of data.
    - House sales tend to increase in the summer.
    - Retail sales tend to increase during the holidays.
- A moving average smooths the data to eliminate seasonality.

Business Analytics: Regression Overview

The End

Overview of Forecasting

Forecasting

- Forecasting projects trends into the future.
- Excel forecasting preserves the trend and seasonality.
- The forecasting technique allows you to accommodate missing data and duplicate data.

The End

Excel Multivariate Regression

Multivariate Regression

- Multivariate regression has two or more X variables.
- Multivariate regression measures the direction and magnitude of change of a dependent (Y) variable when **each** independent (X) variable changes.
  - First, check the significance of F (also called the p value of the F-statistic).
    - If it is 0.05 or less, then the equation is significant, and you can interpret it.
    - If it is more than 0.05, then the equation is not significant, and you can say nothing.
  - Second, interpret the R^2.
    - The R^2 measures the percent of change in Y explained by the change in X.
  - Third, check the p value of the coefficient.
    - If it is 0.05 or less, then the coefficient is significant, and you can interpret it.
    - If it is more than 0.05, then the coefficient is not significant, and you can say nothing.

Business Analytics: Excel Multivariate Regression

The End

Overview of Prediction Models, Sensitivity Analysis, and Conditional Formatting

Prediction Models, Sensitivity Analysis, and Conditional Formatting

- A prediction model allows you to estimate the outcome of a model (dependent variable) when the X-variables (independent variables) take on specific values.

- A sensitivity analysis estimates predictions for a range of values.

- Conditional formatting can highlight low to high values using colors, icons, and bar graphs.

Business Analytics: Overview of Prediction Models,
Sensitivity Analysis, and Conditional Formatting

The End

Overview of Categorical Variables

Categorical Variables

- Categorical variables are variables that have labels instead of numbers.
  - For example, a university rank can be assistant professor, associate professor, or professor.
- Categorical variables can be included in a regression by converting them to dummy variables.
  - A dummy variable is either 0 or 1.
- There are always n-1 dummies for n categories.
  - If there are five vegetables, then there are four dummy variables for vegetables.
  - For the dummy variable Carrot, Carrot = 1 if it is a carrot and Carrot = 0 if it's another vegetable.
- There will be one category missing in the regression; its value is part of the intercept, called the base case.
- Other categories have an intercept, which is the base case plus the coefficient for that category's dummy.

The End