

1-Airline Arrival Study

(A random sample of 444 flights from LAX to JFK in one year.)

Scenario

You are employed as a research assistant to the Vice President of Sales at United Airlines and your supervisor, Alan Oppenheimer, asks you to study the flight characteristics from 444 trips between Los Angeles and New York City over the past year.

Descriptive Statistics

For each flight you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “ArrDelay” (i.e., the arrival delay in minutes), and one key categorical variable, “Time” (i.e., whether the flight departed in morning (a.m.) hours or in the afternoon/evening (p.m.) hours). You also decide to use a grouping categorical variable “Carrier” since each flight was made by American Airlines or United Airlines. This will enable you to make comparisons of both arrival delay minutes of the flight and whether it departed in the a.m. or p.m. based on the carrier – American Airlines or United Airlines.

Regression Analysis

You have also selected the numerical variable “DepDelay” (i.e., departure delay in minutes) to develop a simple linear regression model to predict “ArrDelay.”

2-AL & NL Baseball Batting Study

(Records of 146 Batters who Qualify for Annual Awards)

Scenario

You are employed as a research assistant to the Major League Baseball Players' Association and your supervisor, Greg Cant, asks you to assess the batting performance of 146 major league starting players so that the Association will be prepared to help these players negotiate new contracts with their baseball teams when their current contracts expire.

For each batter you collect data on various performance variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "W.A.R." (i.e., "Wins Above Replacement" by a substitute), and one key categorical variable, "BA Level" (i.e., whether the player's batting average is either at or above the median or below the median batting average of 0.271). You also decide to use a grouping categorical variable "League" since the player is either a member of a National League team or American League team. This will enable you to make comparisons of both the W.A.R. and the batting average level based on the league the player is in.

In addition, you have also selected the numerical variable "T.B." (i.e., "Total Bases" achieved by the batter through hits, walks, and getting to base after being hit by a pitch) to develop a simple linear regression model to predict the batter's "W.A.R."

3-AL & NL Starting Pitcher Study

(Records of 88 Starting Pitchers who Qualify for Annual Awards)

Scenario

You are employed as a research assistant to the Major League Baseball Players' Association and your supervisor, Greg Cant, asks you to assess the performance of 88 major league starting pitchers so that the Association will be prepared to help these pitchers negotiate new pitching contracts with their baseball teams when their current contracts expire.

For each pitcher you collect data on various performance variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "W.A.R." (i.e., "Wins Above Replacement" by a substitute), and one key categorical variable, "Throws" (i.e., whether the pitcher is left-handed or right-handed). You also decide to use a grouping categorical variable "League" since the pitcher is either a member of a National League team or American League team. This will enable you to make comparisons of both the W.A.R. and the pitching hand based on the league the player is in.

In addition, you have also selected the numerical variable "E.R.A." (i.e., "Earned Run Average" allowed by a pitcher) to develop a simple linear regression model to predict the pitcher's "W.A.R."

4-Albuquerque NM RE Study

(A random sample of 117 homes for resale in Albuquerque, NM.)

Scenario

You are employed as a research assistant to the Albuquerque Board of Realtors and your supervisor, Kimberly Hollister, asks you to study the resale prices of 117 single-family homes on the market in that city.

For each home you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Price" (i.e., the price of a home in thousands of dollars), and one key categorical variable, "Corner" (i.e., whether or not the house is situated on the corner of a street). You also decide to use a grouping categorical variable "Custom" since each house is either custom-designed or not. This will enable you to make comparisons of both the price of the home and whether it has a street corner location based on whether the home is custom-designed. In addition, you have also selected the numerical variable "Sqft" (i.e., "size of the house in square feet") to develop a simple linear regression model to predict the resale home "Price."

5- American Community Survey

(A random sample of 249 responses to the American Community Survey)

Scenario

You are employed as a research assistant at the United States Department of Commerce and your supervisor, Nicole Koppel, asks you to assess the work income (in thousands of dollars) based on a random sample of 249 adults extracted from the American Community Survey.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “WorkInc” (i.e., work Income in thousands of dollars), and one key categorical variable, “HowComm” (i.e., the individual’s most typical method of commuting to work – bus, car, subway/rail, taxi, other, works at home, or not applicable). You also decide to use a grouping categorical variable “Gender” since this will enable you to make comparisons of both the work income and how the individual typically commutes based on whether the individual is a male or a female. In addition, you have also selected the numerical variable “HrsWorked” (i.e., hours worked last week) to develop a simple linear regression model to predict “WorkInc.”

6-Automobile Study

(A random sample of 171 car models.)

Scenario

You are employed as a research assistant at the American Automobile Association (AAA) and your supervisor, Salvador Apari, asks you to study the features of recently manufactured car models. For this purpose you take a random sample of 171 car models.

For each automobile you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “MPG” (i.e., the achieved mileage per gallon), and one key categorical variable, “TypeAuto” (i.e., whether the auto is classified as a 4 door hatchback, 4 door SUV, coupe, minivan, sedan, or wagon). You also decide to use a grouping categorical variable “Origin” since each car model is manufactured in either Asia or Europe or in the US. This will enable you to make comparisons of both the “MPG” of the car model and the type of automobile based on its origin – manufactured in either Asia or Europe or manufactured in the US. In addition, you have also selected the numerical variable “Weight” (i.e., the weight of the automobile in pounds) to develop a simple linear regression model to predict the “MPG.”

7-Beer Study

(A random sample of 139 beers.)

Scenario

You are employed as a research assistant at the Alcohol and Tobacco Tax and Trade Bureau (U. S. Department of the Treasury) and your supervisor, R. Tyler Paterson, asks you to study the characteristics of beers sold throughout the United States. For this purpose you take a sample of 139 beers.

For each beer you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Pct Alcohol" (i.e., the alcoholic content in percentage), and one key categorical variable, "Light" (i.e., whether or not the beer product is considered to be light). You also decide to use a grouping categorical variable "Distribution Type" since each beer is distributed nationally or regionally. This will enable you to make comparisons of both the percent alcohol of the beer and whether or not the beer is considered to be light based on its distribution – national or regional. In addition, you have also selected the numerical variable "Calories" in the beer to develop a simple linear regression model to predict the "Pct Alcohol."

8-Birth Weight Study MA

(A random sample of 189 mothers were studied with respect to the birth weight of their child.)

Scenario

You are employed as a research assistant at the Harvard School of Public Health and your supervisor, Lorena Mejia, asks you to study various characteristics pertaining to the birthing process. For this purpose you take a random sample of 189 mothers with a newborn child.

For each mother-with-child you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “BWT” (i.e., the child’s birth weight in grams – under 2500 is considered low), and one key categorical variable, “MRACE” (i.e., whether or not the mother is white). You also decide to use a grouping categorical variable “SMOKE” (i.e., whether or not the mother smoked during her pregnancy). This will enable you to make comparisons of both the “BWT” of the child and whether or not the mother is white based on smoking status during pregnancy – smoked or did not smoke. In addition, you have also selected the numerical variable “WEIGHT” (i.e., the mother’s weight in pounds at her last menstrual period prior to pregnancy) to develop a simple linear regression model to predict the “BWT.”

9-Birth Weight Study NC

(A random sample of 1000 mothers were studied with respect to the birth weight of their child.)

Scenario

You are employed as a research assistant at the School of Public Health at the University of North Carolina and your supervisor, Lorena Mejia, asks you to study various characteristics pertaining to the birthing process. For this purpose you take a random sample of 1000 mothers with a newborn child.

For each mother-with-child you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "BirthWt" (i.e., the child's birth weight in pounds), and one key categorical variable, "MRace" (i.e., whether the mother is white or not white). You also decide to use a grouping categorical variable "MSmoke" (i.e., whether or not the mother smoked during her pregnancy). This will enable you to make comparisons of both the birth weight of the child and the mother's race based on the mother's smoking status during pregnancy – smoked or did not smoke. In addition, you have also selected the numerical variable "Weeks" (i.e., the number of weeks of gestation) to develop a simple linear regression model to predict the "BirthWt."

10-Bond Funds Study

(A random sample of 180 bond funds and their characteristics.)

Scenario

You are employed as a research assistant at the Lord-Meziani-Ozenbas Financial Consulting Company and your supervisor, Stanley Mamanen, asks you to study the characteristics of various bond funds. For this purpose you take a random sample of 180 bond funds.

For each bond fund you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “5-Year Return” (i.e., the percent five year return on investment), and one key categorical variable, “Risk” (i.e., whether the bond fund was considered above average, average, or below average risk). You also decide to use a grouping categorical variable “Category” since each bond fund is either classified as intermediate government or short term corporate. This will enable you to make comparisons of both the “5-Year Return” of the bond fund and its level of risk based on its “Category” – intermediate government or short term corporate. In addition, you have also selected the numerical variable “1-Year Return” (i.e., the percent one year return on investment) to develop a simple linear regression model to predict the “5-Year Return.”

11-Business Valuation Study

(A study of business valuations for 71 companies in the pharmaceutical industry.)

Scenario

You are employed as a research assistant at the Food and Drug Administration (FDA) and your supervisor, R. Jain, asks you to study the business valuations for 71 companies in the pharmaceutical industry.

For each company you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "ROE" (i.e., the percent return on equity), and one key categorical variable, "Debt/EBITDA Ratio Level" (i.e., a financial ratio classified as either "low" versus "moderate or high"). You also decide to use a grouping categorical variable "Type Co." since each pharmaceutical company studied is either classified as pharmaceutical preparation or biological products. This will enable you to make comparisons of both the "ROE" of the company and whether it has either a high or low "Debt/EBITDA Ratio Level" based on "Type Co." – pharmaceutical preparation or biological products. In addition, you have also selected the numerical variable "Price/Book Value Ratio" to develop a simple linear regression model to predict the "ROE."

12-Candidate Assessment Study

(An experimental study of competency rating and salary offered to 120 applicants for managing a science laboratory)

Scenario

You are employed as a research assistant at the American Association of University Professors (AAUP) and your supervisor, Kathy McKinney, asks you to evaluate an experimental study of competency rating and salary offered to 120 applicants for managing a science laboratory.

For each applicant you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Salary” (i.e., the hourly salary offered in dollars), and one key categorical variable, “School” (i.e., the type of education institution – private or public). You also decide to use a grouping categorical variable “Gender of Candidate” since this will enable you to make comparisons of both the salary offered and the type of educational institution based on whether the job applicant’s gender is male or female. In addition, you have also selected the numerical variable “Competency Rating” to develop a simple linear regression model to predict the hourly “Salary.”

13-Cereal Study

(A random sample of 76 cold cereal products.)

Scenario

You are employed as a research assistant at the Food and Drug Administration (FDA) and your supervisor, R. Jain, asks you to study the characteristics of cold cereals sold throughout the United States. For this purpose you take a sample of 76 cold cereal products.

For each cold cereal you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Calories” (i.e., the calories per serving), and one key categorical variable, “Cups” (i.e., whether the cold cereal product requires “one or more” cups per serving or “less than one” cup per serving). You also decide to use a grouping categorical variable “Manufacturer” since each cold cereal product is manufactured either by the well-known General Mills or Kelloggs or it is manufactured by some other company. This will enable you to make comparisons of both the calories per serving and whether or not the cold cereal product requires “one or more” versus “less than one” cup per serving based on its manufacturer – the well-known companies (General Mills and Kelloggs) or some other company. In addition, you have also selected the numerical variable “Weight” (in ounces of one serving) to develop a simple linear regression model to predict the “Calories” per serving.

14-College Football Study

(A financial data study based on a random sample of 105 college football programs.)

Scenario

You are employed as a research assistant at The New York Times and your supervisor, R. B. Drucker, asks you to study the financial records of college football programs. For this purpose you select a random sample of 105 football teams.

For each college football program you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Football Net Revenue” (i.e., the dollar amount generated by the program last year), and one key categorical variable, “Conference” (i.e., the conference the football program is affiliated with). You also decide to use a grouping categorical variable “Location” (i.e., whether the football program is located east or west of the Mississippi River). This will enable you to make comparisons of both the “Football Net Revenue” and “Conference” membership based on “Location” of the program – in the east or in the west. In addition, you have also selected the numerical variable “Total Pay” (i.e., the annual dollar amount paid to the head coach last year) to develop a simple linear regression model to predict the “Football Net Revenue.”

15- Community College Study

(Responses from a random sample of 562 students to a survey)

Scenario

You are employed as a research assistant to the Dean of Students, Dr. Betsy Rosini, at Berry Community College (BCC). She asks you to evaluate the responses to a survey about student characteristics and campus quality of life based on a random sample of 562 registered students this semester.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Working" (i.e., hours of work per week), and one key categorical variable, "Gender" (i.e., whether the student is female or male). You also decide to use a grouping categorical variable "Handed" (i.e., whether the student writes with the left hand or the right hand). This will enable you to make comparisons of both the working hours and gender of the student based on whether the individual is left or right handed. In addition, you have also selected the numerical variable "Credit Hrs" (i.e., number of credits enrolled in this semester) to develop a simple linear regression model to predict hours of work per week.

16-Commuting Study

(A random sample of 1000 commuters' characteristics.)

Scenario

You are employed as a research assistant in the United States Department of Transportation and your supervisor, S. Anestori, asks you to study commuter characteristics in two major cities, St. Louis and Atlanta. For this purpose you take a random sample of 1000 commuters – 500 in each city.

For each commuter you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Time” (i.e., the commuting time in minutes), and one key categorical variable, “Gender” (i.e., whether the commuter is a male or a female). You also decide to use a grouping categorical variable “City” since each trip was made either in Atlanta or in St. Louis. This will enable you to make comparisons of both commuting time and the gender of the commuter based on the city in which the commuter resides – Atlanta or St. Louis. In addition, you have also selected the numerical variable “Distance” (i.e., commuting distance in miles) to develop a simple linear regression model to predict commuting time in minutes.

17-Employee Survey

(A random sample of 400 employee characteristics.)

Scenario

You are employed as a research assistant to the Vice President of Human Resources at the ABDawn Corporation, a very large consumer products distributor with offices and plants throughout the United States. Your supervisor, Ram Misra, asks you to develop a profile of key characteristics of full-time employees at ABDawn. For this purpose you take a random sample of 400 full-time employees.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Salary" in thousands of dollars, and one key categorical variable, "Gender" (i.e., whether the full-time employee is male or female). You also decide to use a grouping categorical variable "BudgetDec" (i.e., whether or not the full-time employee is involved in budgetary decisions). This will enable you to make comparisons of both the salary and gender of the employee based on whether or not the employee is involved in budgetary decisions. In addition, you have also selected the numerical variable "WorkHrs" (i.e. number of hours the employee worked last week) to develop a simple linear regression model to predict "Salary."

18- Faculty Evaluations in TX

(A random sample of 182 faculty members' evaluations conducted by students at Texas universities.)

Scenario

You are employed as a research assistant at the American Association of University Professors (AAUP) and your supervisor, Kathy McKinney, asks you to assess faculty evaluations based on a random sample of 182 students attending Texas universities.

For each faculty member you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "fac_eval" (i.e., faculty evaluation by students using an instrument on a 1 = poor to 5 = excellent scale), and one key categorical variable "Rank" (i.e., whether the faculty member is tenured or not tenured). You also decide to use a grouping categorical variable "School" (i.e., whether the faculty member teaches at the University of Texas at Austin or some other Texas institution). This will enable you to make comparisons of both "fac_eval" and whether or not the faculty member is tenured based on whether or not the faculty member is teaching at the University of Texas at Austin. In addition, you have also selected the numerical variable "course_eval" (i.e., course evaluation by students using an instrument on a 1 = poor to 5 = excellent scale) to develop a simple linear regression model to predict the "fac_eval."

19-GSS Study

(A random sample concerning the socio-economic status of 904 adults in the U.S.A.)

Scenario

You are employed as a research assistant at the Roper Center for Public Opinion and your supervisor, Gail Taylor, asks you to develop a profile of the socio-economic status of adults in the U.S.A. For this purpose you take a random sample of 904 adults in the U.S.A.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "SEI" (i.e., a Socio-Economic index score of the individual – higher is considered better), and one key categorical variable, "GENDER" (i.e., whether the individual is male or female). You also decide to use a grouping categorical variable "RACE" (i.e., whether or not the individual is white). This will enable you to make comparisons of both the SEI and gender of the individual based on whether or not the person is white. In addition, you have also selected the numerical variable "PRESTIGE" (i.e. an Occupational Prestige Score) for the person in order to develop a simple linear regression model to predict "SEI."

20-Heart Stent Study

(A random sample of 100 coronary heart disease patients who had stents inserted.)

Scenario

You are employed as a research assistant at the Harvard School of Public Health and your supervisor, Lorena Mejia, asks you to study costs and outcomes of heart stent procedures. For this purpose you take a random sample of 100 coronary heart disease patients who had stents inserted.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “total_labor_cost” (i.e., the total labor cost in dollars for the heart stent procedure), and one key categorical variable, “gender” (i.e., whether the patient is male or female). You also decide to use a grouping categorical variable “outcome” (i.e., whether the heart stent procedure was deemed successful or whether there was no change or it failed). This will enable you to make comparisons of both the “total_labor_cost” and gender of the patient based on the “outcome” of the heart stent procedure. In addition, you have also selected the numerical variable “total_device_cost” (i.e., the cost for the heart stent device) in order to develop a simple linear regression model to predict “total_labor_cost.”

21-Human Resources Survey

(A random sample of 120 employee characteristics.)

Scenario

You are employed as a research assistant to the Vice President of Human Resources at the J. Myles-Miller Corporation, a very large manufacturer and distributor of women's apparel and accessories with offices and plants throughout the United States. Your supervisor, Mary Phillips, asks you to develop a profile of key characteristics of full-time employees at J.Myles-Miller. For this purpose you take a random sample of 120 full-time employees.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Salary" in thousands of dollars, and one key categorical variable, "Gender" (i.e., whether the full-time employee is male or female). You also decide to use a grouping categorical variable "Ethnicity" (i.e., whether or not the full-time employee is classified as minority or not minority). This will enable you to make comparisons of both the salary and gender of the employee based on whether or not the employee is classified as minority. In addition, you have also selected the numerical variable "Age" of the employee (in years) to develop a simple linear regression model to predict "Salary."

22-LI RE Study

(A random sample of 90 homes for sale in three Nassau County, NY towns.)

Scenario

You are employed as a research assistant to the Nassau County district manager of Weichert Realtors and your supervisor, Kimberly Hollister, asks you to study the resale prices of 90 single-family homes on the market in three towns in this county.

For each home you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Assessed Value” (i.e., the value of a home in thousands of dollars as assessed by Nassau County officials), and one key categorical variable, “Pool” (i.e., whether or not the house has an in-ground swimming pool). You also decide to use a grouping categorical variable “Town” since each house is either located in the neighboring towns of Glen Cove and Roslyn or it is located in Freeport. This will enable you to make comparisons of both the assessed value of the home and whether it has an in-ground swimming pool based on its location – in neighboring Glen Cove and Roslyn or in Freeport. In addition, you have also selected the numerical variable “House Size” (in square feet of living space) to develop a simple linear regression model to predict the “Assessed Value” of the home.

23-Loans and Debt Study

(A study of loans and debt from a random sample of 260 undergraduate college students.)

Scenario

You are employed as a research assistant at the Middle States Commission on Higher Education and your supervisor, R. B. Kobelova, asks you to study some key characteristics pertaining to student loans and debt. For this purpose you take a random sample of 260 undergraduate college students from across the United States.

For each student you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Loans” (i.e., the loans in dollars outstanding), and one key categorical variable, “Class Designation” – classified as either “freshmen or sophomore” versus “junior or senior.” You also decide to use a grouping categorical variable – the “Gender” of the student. This will enable you to make comparisons of both the “Loans” outstanding and student “Class Designation” based on “Gender.” In addition, you have also selected the numerical variable “CC Debt” (i.e., credit card in dollars outstanding) to develop a simple linear regression model to predict “Loans.”

24-Montclair & Millburn RE Study

(A random sample of 93 homes recently sold in two New Jersey towns.)

Scenario

You are employed as a research assistant to the district manager of Coldwell Banker Real Estate and your supervisor, Kimberly Hollister, asks you to study the prices and other characteristics of 93 single-family homes recently sold in two New Jersey towns – Montclair and Millburn.

For each home you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “PRICE(k)” (i.e., the sales price of a home in thousands of dollars), and one key categorical variable, “STYLE” (i.e., whether the house is Bi-level, Cape Cod, Colonial, Custom home, Ranch, Split-level, Tudor, or Victorian). You decide to use “TOWN” as a grouping categorical variable since each house is either located in Montclair or in Millburn. This will enable you to make comparisons of both the sales price of the home and its “STYLE” based on its location in either Montclair or Millburn. In addition, you have also selected the numerical variable “ASSESSVAL(k)” (i.e., the value of the home assessed by Essex County officials) to develop a simple linear regression model to predict the resale home “PRICE(k).”

25-Mutual Funds Study

(A random sample of 868 mutual funds and their characteristics.)

Scenario

You are employed as a research assistant at the Lord-Meziani-Ozenbas Financial Consulting Company and your supervisor, Stanley Mamanen, asks you to study the characteristics of various mutual funds. For this purpose you take a random sample of 868 mutual funds.

For each mutual fund you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “5-Year Return” (i.e., the percent five year return on investment), and one key categorical variable, “Fees” (i.e., whether or not managerial fees are charged by the mutual fund). You also decide to use a grouping categorical variable “Objective” since each mutual fund is either classified as having a growth objective or value objective. This will enable you to make comparisons of both the “5-Year Return” of the mutual fund and whether or not the mutual fund charges managerial fees based on its “Objective” – growth or value. In addition, you have also selected the numerical variable “3-Year Return” (i.e., the percent three year return on investment) to develop a simple linear regression model to predict the “5-Year Return.”

26-Nutrition Study

(A random sample of 314 patients.)

Scenario

You are employed as a research assistant at the Office of Nutrition and Fitness: State of New Jersey and your supervisor, H. Ravinder, asks you to evaluate nutrition-related characteristics of patients in an ongoing study. For this purpose you take a sample of 314 patients who are enrolled in the current study.

For each patient you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Calories” (i.e., the caloric intake or calories consumed per day), and one key categorical variable, “Gender” (i.e., whether the patient is a male or a female). You also decide to use a grouping categorical variable “Smoke” (i.e., whether or not the patient smokes cigarettes). This will enable you to make comparisons of both the daily caloric intake and whether or not the patient is male or female based on whether or not the patient smokes cigarettes. In addition, you have also selected the numerical variable “Fat” (i.e., the daily amount of fat consumed) to develop a simple linear regression model to predict the “Calories” consumed per day.

27-Public College Value Study

(A random sample of 100 public colleges and universities and their characteristics.)

Scenario

You are employed as a research assistant at the Middle States Commission on Higher Education and your supervisor, R. B. Kobelova, asks you to study some key characteristics of public colleges and universities. For this purpose you take a random sample of 100 public institutions of higher learning.

For each institution you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “4-Yr Grad. Rate” (i.e., the percentage of attending students who graduate in four years), and one key categorical variable, “DebtAbove20K” (i.e., whether or not the average student debt at graduations exceeds twenty thousand dollars). You also decide to use a grouping categorical variable “Region” since this will enable you to make comparisons of both “4-Yr Grad. Rate” and whether or not “DebtAbove20K” based on whether the institution is located in the Atlantic or eastern part of the USA or whether it is located elsewhere. In addition, you have also selected the numerical variable “Total Cost per Yr. (In State)” per student to develop a simple linear regression model to predict “4-Yr Grad. Rate.”

28-Retirement Funds Study

(A random sample of 316 retirement funds and their characteristics.)

Scenario

You are employed as a research assistant at the Lord-Meziani-Ozenbas Financial Consulting Company and your supervisor, Stanley Mamanen, asks you to study the characteristics of various retirement funds. For this purpose you take a random sample of 316 retirement funds.

For each retirement fund you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “3YrReturn%” (i.e., the percent three year return on investment) and one key categorical variable, “Risk” – whether the retirement fund has a risk level considered to be low, average, or high. You also decide to use a grouping categorical variable “Type” of retirement fund (i.e., growth objective or value objective). This will enable you to make comparisons of both the “3YrReturn%” of the retirement fund and the “Risk” level of the retirement fund based on its “Type” – either growth objective or value objective. In addition, you have also selected the numerical variable “1YrReturn%” (i.e., the percent one year return on investment) to develop a simple linear regression model to predict the “3YrReturn%.”

29-ROI College & University Study

(Return on Investment – ROI – data for a sample of 252 colleges and universities in the USA.)

Scenario

You are employed as a research assistant at the Middle States Commission on Higher Education and your supervisor, R. B. Kobelova, asks you to study some key financial characteristics at colleges and universities. For this purpose you take a random sample of 252 institutions of higher learning.

For each institution you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “30 YEAR NET ROI” (i.e., the 30-year net return on investment), and one key categorical variable, “CATEGORY” (i.e., the institution’s primary classification – Public, Private, Private Research, Liberal Arts, Business, Engineering, or Art-Music-Design). You also decide to use a grouping categorical variable “TYPE” since each institution is either private or public. This will enable you to make comparisons of both the “30 YEAR NET ROI” of the institution and its “CATEGORY” based on its “TYPE.” In addition, you have also selected the numerical variable “AVG AID AMOUNT” (i.e., the average dollar amount of aid per student) to develop a simple linear regression model to predict “30 YEAR NET ROI.”

30-Survey of Adult Workforce

(A random sample of 2537 adults in the workforce.)

Scenario

You are employed as a research assistant at the National Organization for Women (NOW) and your supervisor, Laurie K. Roda, asks you to study the responses to a national survey of 2537 adults in the workforce and make appropriate comparisons.

For each individual you collect data on various characteristics (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Donations (in K)” (i.e., the amount – in thousands of dollars – donated by the worker to charity last year), and one key categorical variable, “Job Satisfaction” (i.e., whether the worker claims to be very satisfied or not very satisfied with the job). You also decide to use as a grouping categorical variable “Gender” – whether the worker is male or female. This will enable you to make comparisons of both the “Donations (in K)” and “Job Satisfaction” of the worker based on whether or not the worker is male or female. In addition, you have also selected the numerical variable “Income (in K)” (i.e., the worker’s income last year in thousands of dollars) to develop a simple linear regression model to predict “Donations (in K).”

31-TV Set Study

(A sample of 174 TV models.)

Scenario

You are employed as a research assistant at Consumer Reports and your supervisor, Katrina Ashlinova, asks you to study the characteristics of TV models sold throughout the United States. For this purpose you take a sample of 174 recently manufactured TV models.

For each TV model you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Overall Score" (i.e., an overall evaluation provided to Consumer Reports by owners of the product), and one key categorical variable, "SizeGp" – whether the TV model's screen is small (below 40"), medium (40" to 49"), or large (50" or more). You also decide to use a grouping categorical variable "TypeTV" since each model is classified as either a plasma TV (P) or liquid TV (L). This will enable you to make comparisons of both "Overall Score" and "SizeGp" based on "TypeTV" – plasma or liquid. In addition, you have also selected the numerical variable "Quality Rating" (i.e., an overall evaluation provided by the Consumer Reports staff testing the product) to develop a simple linear regression model to predict the "Overall Score."

33-UT Speeding Violation Study (XL)

(A random sample of 118 individuals recently given a speeding ticket in Utah.)

Scenario

You are employed as a research assistant at the American Automobile Association (AAA) and your supervisor, Salvador Apari, asks you to study the characteristics of individuals who have received speeding tickets in Utah. For this purpose you take a random sample of 118 persons who were recently ticketed for speeding in Utah.

For each individual you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Age at Incident" (i.e., the age of the individual who received the speeding ticket), and one key categorical variable, "Type of Road" the driver was traveling on when ticketed – backroad, city, or freeway. You also decide to use a grouping categorical variable "Gender." This will enable you to make comparisons of both the "Age at Incident" and the "Type of Road" driven on when ticketed based on "Gender" -- whether the ticketed driver is male or female. In addition, you have also selected the numerical variable "Age at License" (i.e., the age of the ticketed driver when his/her driving license was obtained) to develop a simple linear regression model to predict "Age at Incident."

34-Wine Study

(A random sample of 100 wines and their characteristics.)

Scenario

You are employed as a research assistant at the Alcohol and Tobacco Tax and Trade Bureau (U. S. Department of the Treasury) and your supervisor, R. Tyler Paterson, asks you to study the characteristics of wines sold throughout the United States. For this purpose you take a sample of 100 wines.

For each wine you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, "Alcohol" (i.e., the alcoholic content in percentage), and one key categorical variable, "Quality Rating" since each wine is classified as "high" quality (if rated 6 or more on a 10-point scale) or "low" quality (if rated 5 or lower on a 10-point scale). You also decide to use a grouping categorical variable "Type" of wine (i.e., whether the wine is red or white). This will enable you to make comparisons of both the percent alcohol of the wine and its "Quality Rating" (i.e., high or low) based on the "Type" of wine (i.e., red or white). In addition, you have also selected the numerical variable "Density" in the wine to develop a simple linear regression model to predict "Alcohol."

35-Zagat NYC Restaurant Study

(A random sample of 66 NYC restaurants rated by Zagat with their characteristics.)

Scenario

You are employed as a research assistant at the New York State Department of Health and your supervisor, J. D. Chandler, asks you to study the Zagat restaurant ratings in the New York Metropolitan area. For this purpose you take a sample of 66 restaurants in New York City.

For each restaurant you collect data on several variables (see the Variable INFO tab) but for the reports you will be preparing you decide to focus on one key numerical variable, “Cost\$” (i.e., the average cost per meal), and one key categorical variable, “HighPopIndx” (i.e., whether or not the Zagat Popularity Index is 90 or higher). You also decide to use a grouping categorical variable “Type Food” since each restaurant is classified as specializing in Asian food or non-Asian food. This will enable you to make comparisons of both “Cost\$” and “HighPopIndx” based on “Type Food” – Asian or non-Asian. In addition, you have also selected the numerical variable “Décor&Service” (i.e., the combined Zagat ratings of Décor and Service where the total scores could range from 2 to 60) to develop a simple linear regression model to predict t “Cost\$.”

