

Below are descriptions of the data sets available for the model building project for Stat 401. You can choose any one of the data sets to analyze for your project. Data sets are available on the Web, go to the course homepage.

1. Percentage body fat as it relates to body measurements.

Percentage body fat is determined from a persons density. The density is obtained from the displacement of water in a large tub. This is a time consuming and inconvenient activity. It would be more convenient if the percentage body fat could be predicted from simple body measurements, ones that can be obtained with a scale and a tape measure. Data are obtained on 80 men. A list of the variables appears below and the data are saved as **bodyfat.JMP** on the course homepage.

- **BodyFat:** Percentage Body Fat
- **Age:** Age in years
- **Weight:** Weight in pounds
- **Height:** Height in inches
- **Adiposity:** $\text{Weight}/(\text{Height}*\text{Height})$
- **Neck:** Neck circumference (cm)
- **Chest:** Chest circumference (cm)
- **Abdomen:** Abdomen circumference (cm) at the umbilicus
- **Hip:** Hip circumference (cm)
- **Thigh:** Thigh circumference (cm)
- **Knee:** Knee circumference (cm)
- **Ankle:** Ankle circumference (cm)
- **Biceps:** Extended biceps circumference (cm)
- **Forearm:** Forearm circumference (cm)
- **Wrist:** Wrist circumference (cm)

2. Baseball salaries as they relate to performance.

Our population of interest are Major League baseball players who played at least one game in both the 1991 and 1992 seasons, excluding pitchers. A random sample of 80 players is taken from the population of interest. The 1992 salary is the response variable. The explanatory variables relate to various performance measures. A list of the variables appears below and the data are saved as **baseball.JMP** on the course homepage.

- **Salary:** 1992 Salary in thousands of dollars
- **BA:** Batting average
- **OBP:** On base percentage
- **Runs:** Number of runs scored
- **Hits:** Number of hits
- **Doubles:** Number of doubles
- **Triples:** Number of triples
- **HRs:** Number of home runs
- **RBI:** Number of runs batted in
- **Walks:** Number of walks
- **SOs:** Number of strike outs
- **SBs:** Number of stolen bases
- **Errors:** Number of errors
- **FAElig:** Indicator of Free Agent Eligibility (Yes=1, No=0)
- **FA91/2:** Indicator of Free Agent 1991/92 (Yes=1, No=0)
- **ArbElig:** Indicator of Arbitration Eligibility (Yes=1, No=0)
- **Arb91/2:** Indicator of Arbitration 1991/92 (Yes=1, No=0)
- **Name:** Player's name

3. Home Sales Prices.

The city tax assessor is interested in predicting residential home sale prices in a mid-western city. A random sample of 80 homes that sold in 2002 was obtained. The sale price is the response variable. The explanatory variables involve characteristics of the house and surrounding property. A list of the variables appears below and the data are saved as **homesales.JMP** on the course homepage.

- **SalesPrice**: Price of home when sold in 2002 (\$)
- **SqFt**: Finished living area (sq. ft.)
- **Bedrooms**: Number of bedrooms
- **Baths**: Number of bathrooms
- **AC**: Air Conditioning: 1 Yes, 0 No
- **Garage**: Number of cars garage will hold
- **Pool**: Swimming Pool: 1 Yes, 0 No
- **YearBuilt**: Year home was constructed
- **Quality**: Index of quality: 1 High, 2 Medium, 3 Low
- **LotSize**: Lot Size (sq. ft.)
- **Highway**: Next to highway: 1 Yes, 0 No

4. College tuition, room & board and fees.

These data are selected from the 1995 U.S. News & World Report's Guide to America's Best Colleges. Eighty colleges are selected at random from a subset of the approximately 1300 colleges in the guide. We are interested in predicting the cost of college (the total of tuition, room & board and fees) per year from the other information available. A list of the variables appears below and the data are saved as **College80.JMP** on the course homepage

- **ID**: Federal ID number
- **Name**: College name
- **State**: State
- **Pub/Pri**: Public=1, Private=2
- **ACT**: Average ACT score
- **ACTQ1**: First quartile ACT
- **ACTQ3**: Third quartile ACT
- **AppsRec**: Number of applications received.
- **AppsAcc**: Number of applications accepted.
- **AppsEnrol**: Number of new students enrolled.
- **PctTop10**: Percentage of new students in top 10% of H.S. class.
- **PctTop25**: Percentage of new students in top 25% of H.S. class.
- **Enroll**: Number of undergraduates enrolled.
- **Tuition**: In state tuition in dollars.
- **RoomBoard**: Room and Board in dollars.
- **Fees**: Fees in dollars.
- **PctFacPhD**: Percentage of faculty with Ph.D.s
- **Stud/Fac**: Student to faculty ratio.
- **GradRate**: Graduation rate.