Math 140
Introductory Statistics

Professor B. Abrego

Lecture 23

Section 3.1

Significance Tests
Various Examples

- **1.** A job placement director claims that the average starting salary for nurses is $24,000. A sample of 10 nurses' salaries has a mean of $23,450 and a standard deviation of $400. Is there enough evidence to reject the director's claim at α=0.05.
- **2.** A statistician read that at least 77% of the population oppose replacing $1 bills with $1 coins. To see if this claim is valid, the statistician selected a sample of 80 people and found that 55 were opposed to replacing the $1 bills. At α=0.01, test the claim that at least 77% of the population are opposed to the change.
- **3.** A researcher reports that the average salary of an assistant professor is more than $42,000. A sample of 30 assistant professors has a mean salary of $43260 and a standard deviation of $5230. At α=0.05, test the claim that assistant professors earn more than $42,000 a year.

Significance Tests
Various Examples

- **4.** In a sample of 200 surgeons, 15% thought the government should control health care. In a sample of 200 general practitioners, 21% felt the same way. At α=0.01, is there a difference in the proportions?
- **5.** The average size of a farm in Indiana County, Pennsylvania, is 191 acres. The average size of a farm in Greene County, Pennsylvania, is 199 acres. Assume the data were obtained from two samples with standard deviations 38 and 12 acres, respectively, and sample sizes of 8 and 10, respectively. Can it be concluded at α=0.05 that the average size of the farms in the two counties is different? Assume the populations are normally distributed.
- **6.** A survey found that the average hotel room rate in New Orleans is $88.42 and the average room rate in Phoenix is $80.61. Assume that the data were obtained from two samples of 50 hotels each and that the standard deviations of these samples were $5.62 and $4.83 for New Orleans and Phoenix, respectively. At α=0.03, can it be concluded that there is a significant difference in the rates?

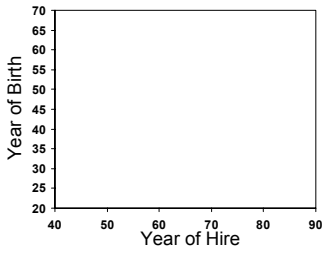
3.1 Scatterplots

- A scatterplot is a useful graphical representation of the relationship of two quantitative variables.
- Recall that:
  - Cases = objects/subjects of the statistical examination.
  - Variables (Characteristics) = information about the cases.

Scatterplots: How to make them.

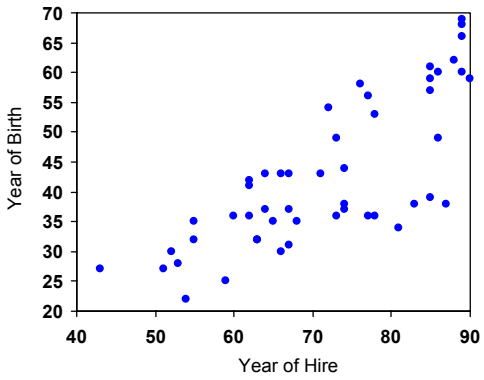
- On a scatterplot:
  - Each dot represents a case.
  - If the plot is Variable A versus or against Variable B then the y-coordinate equals Variable A and the x-coordinate equals Variable B.
- For example (page 102). If we want to do a scatterplot of Year of Birth, vs Year of Hire in the Westvaco case (data on page 5)

Year of Birth	Year of Hire



Scatterplots: How to make them.

- Here is the scatterplot with all the 50 employees. (data on page 5)



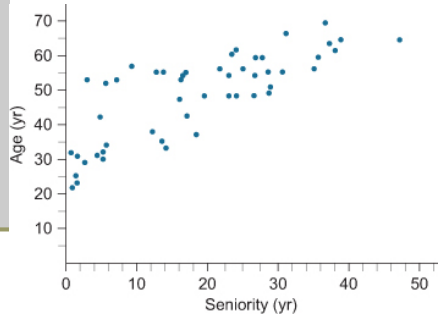
In your calculator, you can put the x-coordinates on L_1 and the y-coordinates on L_2.

Then you can see the scatterplot by selecting STAT PLOT -> ON and using L_1 for XList and L_2 for YList.

From data: Age vs. Seniority

Row	Job Title	Hourly (H) or Salaried (S)	Seniority (yr)	Round Terminated (6 = retained)	Age (yr)
1	Engineering Clerk	H	1.5	6	25
2	Engineering Tech II	H	12.4	6	38
3	Engineering Tech II	H	25.5	6	56
4	Secretary to Engin Manag	H	24.3	6	48
5	Engineering Tech II	H	16.3	1	53
6	Engineering Tech II	H	30.8	1	55
7	Engineering Tech II	H	27.9	1	59
8	Parts Crib Attendant	H	1.2	1	22
9	Engineering Tech II	H	13.8	2	55
10	Engineering Tech II	H	39.1	2	64
11	Technical Secretary	H	17.2	2	55
12	Engineering Tech II	H	28.8	3	55
13	Engineering Tech II	H	14.2	4	33
14	Engineering Tech II	H	13.7	4	35
15	Customer Serv Engineer	S	24.3	6	61
16	Customer Serv Engr Assoc	S	2.7	6	29
17	Design Engineer	S	23.3	6	48
18	Design Engineer	S	16.6	6	54
19	Design Engineer	S	12.9	6	55
20	Design Engineer	S	23.8	6	60

Discussion



- Why should the two variables plotted show a positive association?
- Why do all of the points lie above a diagonal line running from the lower left to the upper right?
- Is this sentence a reasonable interpretation of the display ? “As time passed, Westvaco tended to hire younger and younger people.”

Comparison: One variable vs. Two variables

One Variable	Two Variables
Shape (of Distribution)	Shape (of Scatterplot)
Center	**Trend**
Spread	**Strength** (correlation)

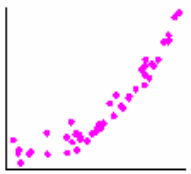
Describing the Pattern of a Scatterplot

- 1. Identify the **variables** and **cases.**
  - What exactly does each point represent?
  - Describe the the scale (units of measurement) and range of each variable.
- 2. Describe the overall **shape** of the relationship:
  - **linearity:** is the pattern linear (scattered about a line) or curved?
  - **clusters:** is there just one, or more than one?
  - **outliers:** are there any striking exceptions to the overall pattern?
- 3. Describe the **trend.**
  - If, as *x* gets larger, *y* tends to get larger, there is a **positive trend.** (The cloud of points tends to slope up as you go from left to right.)
  - If, as *x* gets larger, *y* tends to get smaller, there is a **negative trend.** (The cloud of points tends to slope down as you go from left to right.)

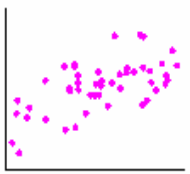
Describing the Pattern of a Scatterplot

- 4. Describe the **strength** of the relationship.
  - If the points cluster closely around an imaginary line, the association is **strong.** If the points are scattered farther from the line, the association is **weak.** The strength could vary on different places of the histogram or it could be constant.
- 5. Does the pattern **generalize** to other cases, or is the relationship an instance of "what you see is all there is"?
- 6. Are there plausible **explanations** for the pattern?
  - Is it reasonable to conclude that one variable causes the other?
  - Is there a third or **lurking variable** that might be causing both?

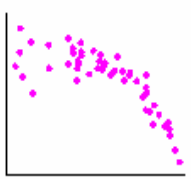
Examples



Curved and strong



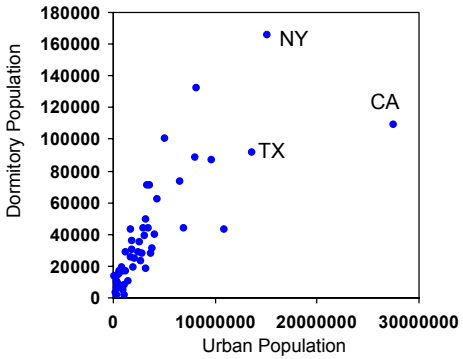
Linear and moderate



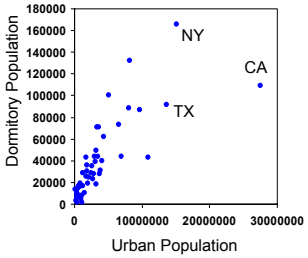
Curved with varying strength

Example Dormitory Populations

- The plot below, for the 50 United States, the number of people living in college dormitories versus the number of people living in cities, in thousands. Describe the pattern in the plot.

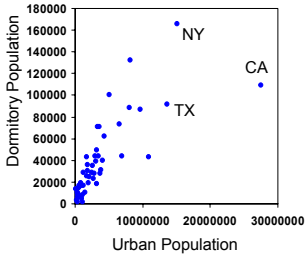


Example: *Dormitory Populations*



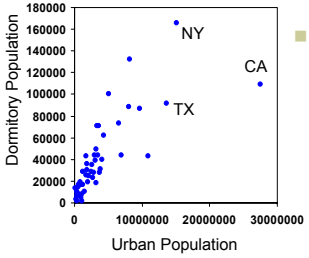
- 1. ***Variables and cases***. The scatterplot shows ***dormitory population*** versus ***urban population***, for the 50 U.S. states. Dormitory population ranges from near zero to a high of more than 165,000 in New York. The urban population ranges from near zero to about 15 million in New York and 28 million in California.

Example: *Dormitory Populations*



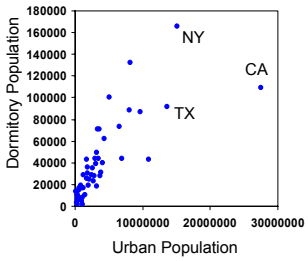
- 2. ***Shape***. While most states follow a linear trend, the three states with the highest urban population suggest curvature in the plot because for those states, the number of people in dormitories is proportionately lower than in the smaller states. California can be considered an outlier with respect to its urban population, which is much higher than in other states. It is also an outlier with respect to the overall pattern because it lies far below the generally linear trend.

Example: *Dormitory Populations*



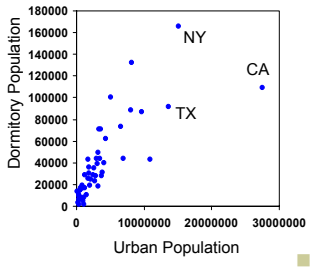
- 3. ***Trend***. The trend is positive—states with higher urban populations tend to have higher dormitory populations, and vice versa.
- 4. ***Strength***. The relationship varies in strength. As *x* increases, *y* becomes more variable—the cloud of points fans out.
  - A tight cluster in the lower left corresponds to a large number of states with small urban populations and small numbers of people living in dormitories.
  - Toward the upper right, the points are much more spread out. There are comparatively fewer states with large urban populations, and these states show much more variation in the number of people living in college dormitories.
  - The variation is roughly proportional to population size. For the states with smallest urban population, the points cluster rather closely to a line. For the states with the largest urban population, the states are more scattered from the line.
  - Overall, the strength of the relationship is **moderate**.

Example: *Dormitory Populations*



- 5. ***Generalization***. The 50 states aren't a sample from a larger population of cases, so the relationship here does not generalize to other cases. Because both variables tend to change rather slowly, however, we can expect the relationship in the scatterplot to be typical of other years.

Example: *Dormitory Populations*



- 6. ***Explanation***. It is tempting to attribute the positive relationship to the idea that cities attract colleges.
  - The main reason for the positive relationship, however, is not nearly so interesting: both variables are related to a **state's population**. The more people in a state, the more people live in dormitories and the more people live in cities.
- **Moral**: Interpreting association can be tricky, in part because the two variables you see in a plot will often be related to some lurking variable that you don't see.