Math 140
Introductory Statistics

Professor Bernardo Ábrego
Lecture 25
Section 3.3

Pizza again

- **Pizza again.** The table below shows the calorie and fat content of 5 ounces of various kinds of pizza.

Pizza	Fat (g)	Calories
Pizza Hut's Hand Tossed	9	230
Domino's Deep Dish	19.5	385
Pizza Hut's Pan	14	280
Domino's Hand Tossed	12	305
Little Caesar's Original Round	8	230
Little Caesar's Deep Dish	14.2	350
Pizza Hut's Stuffed Crust	15	370

Pizza again

- **Pizza again.** The table below shows the calorie and fat content of 5 ounces of various kinds of pizza.

Pizza	Fat (g)	Calories
Pizza Hut's Hand Tossed	9	230
Domino's Deep Dish	19.5	385
Pizza Hut's Pan	14	280
Domino's Hand Tossed	12	305
Little Caesar's Original Round	8	230
Little Caesar's Deep Dish	14.2	350
Pizza Hut's Stuffed Crust	15	370

- Draw a scatterplot of calories vs. fat for the seven pizzas described.
- What is the response variable? What is the predictor, or explanatory variable?
- Calculate the regression line.
- What is the meaning of the slope of this line?

3.3 Correlation:
The strength of a linear trend

- The correlation is a useful summary statistic that measures trend and strength.
- Formally, it is defined as:

$$\text{Correlation} = r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$

$$= \frac{1}{n-1}\sum z_x \cdot z_y$$

- It is the average of the products of the z-scores of x and y coordinates.

Properties of Correlation

- $-1 \leq r \leq 1$.
- r positive means: positive trend
- r negative means: negative trend
- r close to zero means:
    - correlation (strength) is weak
- r close to 1 or -1 means:
    - correlation (strength) is strong.

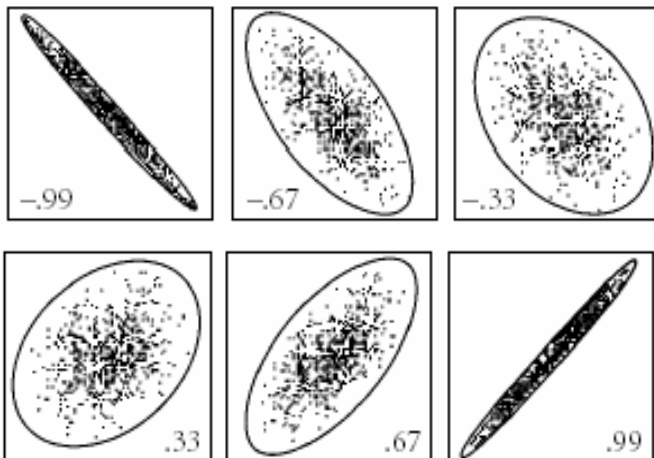# Correlation Graphically

- It is useful to draw an ellipse around the points in a scatter plot. Then
    - $r > 0$ means that the ellipse slants upward.
    - $r < 0$ means that the ellipse slants downward.
    - $|r|$ close to 0 means that the ellipse is fat.
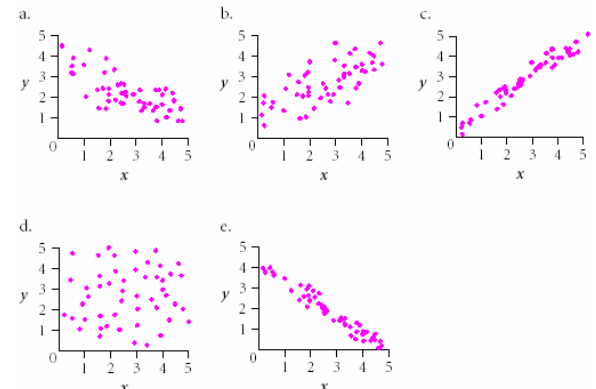   - $|r|$ close to 1 means that the ellipse is skinny.

Example

Practice Problem.

- P10. Match each of the five scatterplots with its correlation, choosing from –.95, –.5, 0, .5, .95

More Pizza

- This table from gives the amount of fat and calories in various pizzas. Use your calculator to find the correlation, r.

Pizza	Calories	Fat (g)	Cost ($)
Pizza Hut's Hand Tossed	305	9	1.51
Domino's Deep Dish	382	16	1.53
Pizza Hut's Pan	338	14	1.51
Domino's Hand Tossed	327	9	1.90
Little Caesars' Pan! Pan!	309	10	1.23
Little Caesars' Pizza! Pizza!	313	11	1.28
Pizza Hut's Stuffed	349	13	1.23

Using the Calculator

- First, input the x-values in list L_1 and the y-values in list L_2 (or any two other lists, just remember which)
- Then on the STAT/CALC menu select LinReg(ax+b), and type L_1, L_2 after it.
- Then observe that b=b_0 and a=b_1.
- If r does not show make sure you have Diagnostic On, you can use Catalog to find it.

Relationship between r and b_1

- The correlation r, and the slope of the regression line b_1 are related according to the following:

$$b_1 = r\frac{s_x}{s_y}$$

- Where s_x and s_y are the standard deviations of the x and y coordinates of the points

Example

- In 2000, the mean verbal score for all SAT I test takers was 505 with a standard deviation of 111. For math, the mean was 514 with a standard deviation of 113. The correlation between the two scores was not given but is known to be quite high. If you can estimate this correlation as, say, .7, you can find the equation of the regression line and use it to estimate the math score from a student's verbal score.
- [Source: *2000 College Bound Seniors: A Profile of SAT Program Test Takers,* The College Board, 2000, p. 7.]