

Week 6 Notes

Describing Scatterplots

- Look for overall pattern and unusual features.
- The overall pattern of a scatterplot can be described by the form, direction, and strength of the relationship.

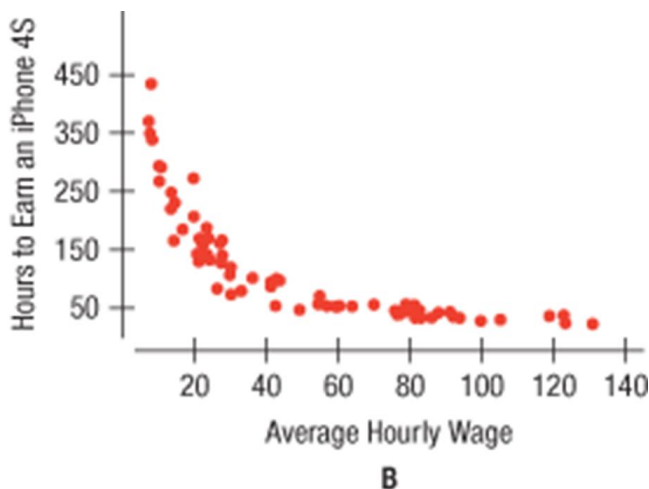
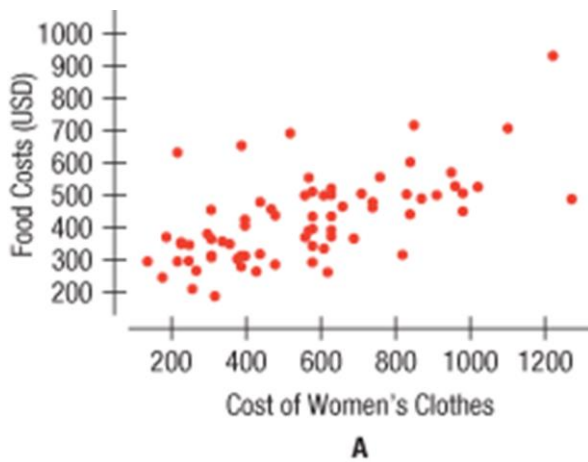
Form: straight (linear relationship), curved (curvilinear relationship), exotic, no pattern?

Direction: positive, negative, neither?

Strength: perfect, strong, moderate, weak, no relationship?

- An important kind of unusual feature is an outlier, an individual value that falls outside the overall pattern.

e.g. Describe the patterns shown by each of these plots.



In Plot A, the association between Food Costs and Cost of Women's Clothes is positive and straight with a few high outliers, and moderately strong. In Plot B, the association between Hours to Earn an iPhone4S and Average Hourly Wage is negative and strong, but the form is not straight.

Linear Regression

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

$$b_0 = \bar{y} - b_1 \bar{x}$$

The equation of the least squares regression line of y on x is

$$\hat{y} = b_0 + b_1 x$$

with slope $b_1 = r \frac{s_y}{s_x}$ and intercept $b_0 = \bar{y} - b_1 \bar{x}$.

The coefficient of determination r^2 is the square of the correlation coefficient (r). It represents the fraction of the variation in the values of y that is explained by the least squares regression of y on x .

Exercise 2: The manager wants to predict Sales from Number of Sales People Working.

Number of Sales People Working	Sales (in \$1000)
2	10
3	11
7	13
9	14
10	18
10	20
12	20
15	22
16	22
20	26
$\bar{x} = 10.4$	$\bar{y} = 17.6$
$SD(x) = 5.64$	$SD(y) = 5.34$
$r = 0.965$	

- Find the slope estimate, b_1 .
- What does it mean, in this context?
- Find the intercept, b_0 .
- What does it mean, in this context? Is it meaningful?
- Write down the equation that predicts Sales from Number of Sales People Working.
- If 18 people are working, what sales do you predict?
- What proportion of the variation in sales is explained by this model?