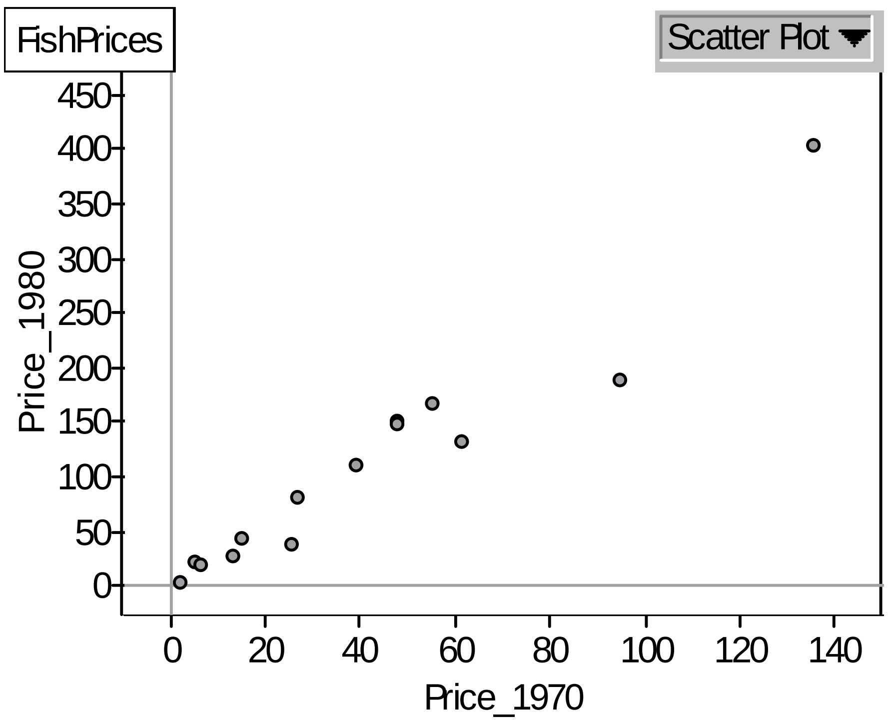
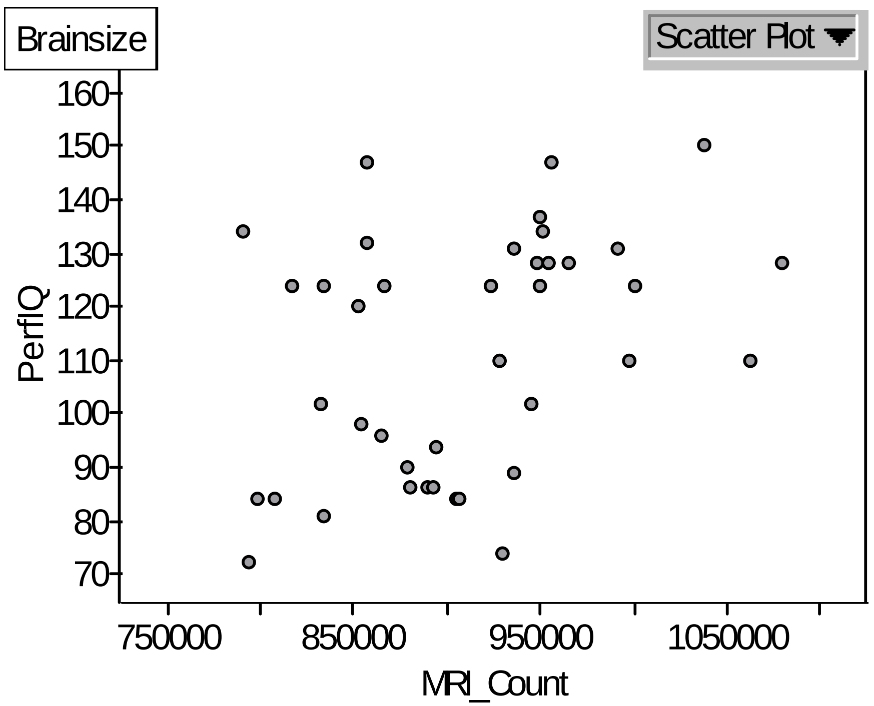
Correlation: The Strength of a Linear Trend

Linear Trend

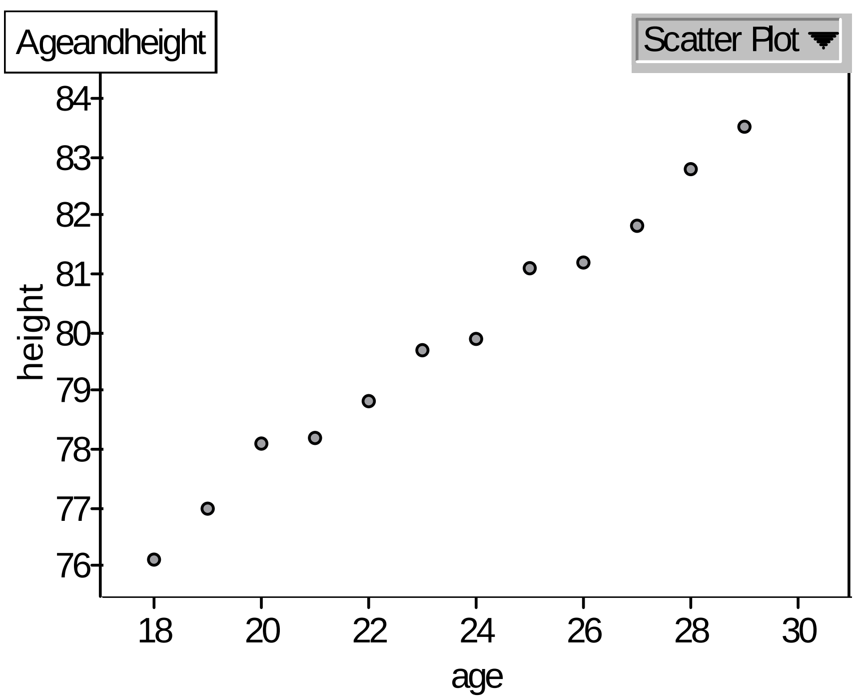
- Strong linear trend
  - Points tightly packed around regression line
- Weak linear trend
  - Points scattered around regression line

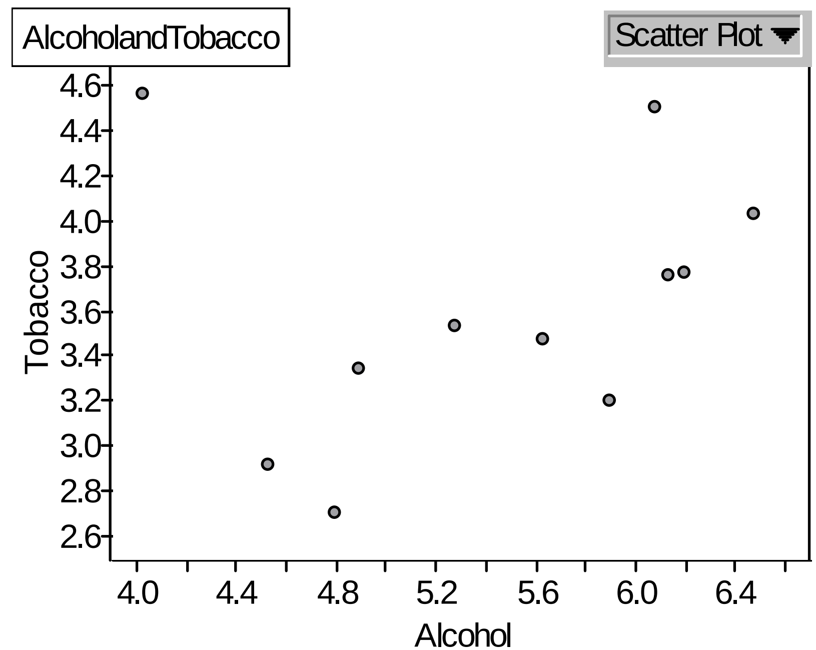
Linear Trend





Linear Trend





Linear Correlation Coefficient

- Also known as the *Pearson product moment correlation coefficient*

Linear Correlation Coefficient

- Properties
  - The correlation coefficient is always between –1 and 1, inclusive, that is, $-1 \leq r \leq 1$
  - r is a measure of the strength of linear relation

Linear Correlation Coefficient

- Properties
  - *The correlation coefficient is always between –1 and 1, inclusive, that is, $-1 \le r \le 1$*
  - $r = 1$ is the perfect positive linear relation between two variables

Linear Correlation Coefficient

- Properties
  - *The correlation coefficient is always between –1 and 1, inclusive, that is, $-1 \le r \le 1$*
  - r = -1 is the perfect negative linear relation between two variables

Trend/Association

- Positive Trend
  - Two variables are positively associated if the values of the output increase as the values of the input increase
- Negative Trend
  - Two variables are negatively associated if the values of the output decrease as the values of the input increase

Linear Correlation Coefficient

- Properties
  - The closer r is to 1 the stronger the evidence of positive association between two variables
  - The closer r is to –1 the stronger the evidence of negative association between two variables

Linear Correlation Coefficient

- Properties
  - If r is close to 0 then there is evidence of no linear relation between the two variables
  - ***CAUTION:*** r close to zero provides evidence that there is no *linear* relation but it does not imply that there is no relation between the variables

Linear Correlation Coefficient

- Also known as the *Pearson product moment correlation coefficient*

- *Guessing Correlations*

http://www.stat.uiuc.edu/courses/stat100//java/GCApplet/GCAppletFrame.html

Linear Correlation Coefficient

- Hand calculations or table calculations -

$$r=\frac{n\left(\sum xy\right)-\left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right)-\left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right)-\left(\sum y\right)^2}}$$

Linear Correlation Coefficient

- Properties
  - The linear correlation coefficient is a unit-less measure of association
  - The units of measure of x and y play no role in the interpretation of r

Linear Correlation Coefficient

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

$$r = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}$$

Linear Correlation Coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

$$r = \frac{\sum \left(\frac{x-\overline{x}}{S_x}\right)\left(\frac{y-\overline{y}}{S_y}\right)}{n-1}$$

Linear Correlation Coefficient

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{S_x}\right)\left(\frac{y-\bar{y}}{S_y}\right)$$

$$r = \frac{1}{n-1}\sum\left\{\left(\frac{x-\bar{x}}{S_x}\right)\left(\frac{y-\bar{y}}{S_y}\right)\right\}$$

Recall for

Least Squares Regression Line

- For the least squares regression line y = mx + b,
  - Slope:
  $$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$
  - Y-coordinate of y-intercept:
  $$b = \bar{y} - m\bar{x}$$

Least Squares Regression Line

- For the least squares regression line $y = mx + b$,
  - Slope: $$m = r\frac{S_y}{S_x}$$
  - Y-coordinate of y-intercept: $$b = \bar{y} - m\bar{x}$$

Error and Correlation

- Sum of the Squared Errors (SSE)
  - Using residuals for the least squares line

$$SSE = \sum \left(y - \hat{y} \right)^2$$

y – data value; $\hat{y}$ - predicted value

Error and Correlation

- Sum of the Squared Errors (SSE)
  - Using residuals for the least squares line

$$SSE = \sum \left(y - \hat{y} \right)^2$$

(x,y) –corresponding pair of data values

$\hat{y} = mx + b$ -predicted value

Error and Correlation

- Total Sum of the Squared Error (SST)
  - Using deviations from the mean

$$SST = \sum \left(y - \bar{y} \right)^2$$

y - data value; $\bar{y}$ - mean

Error and Correlation

- If $\frac{SSE}{SST}$ is near 0 then the variables have a strong linear relationship

- If $\frac{SSE}{SST}$ is near 1 then the variables have a weak linear relationship

Error and Correlation

- How is $\frac{SSE}{SST}$ related to the correlation coefficient?

Error and Correlation

- How is $\frac{SSE}{SST}$ related to the correlation coefficient?
- $r^2 = 1 - \frac{SSE}{SST}$

or, if you prefer,

$$r^2 = \frac{SST - SSE}{SST}$$