Bivariate Analysis

Project ENABLE

May 31, 2019

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Bivariate Analysis

- Bivariate analysis is a statistical methods to describe relationship between two variables (e.g., cholesterol and blood pressure).
- Bivariate analysis helps to find other variables which move in concert with the variable we want to understand.

Bivariate Analysis

- Bivariate analysis is a statistical methods to describe relationship between two variables (e.g., cholesterol and blood pressure).
- Bivariate analysis helps to find other variables which move in concert with the variable we want to understand.

Correlation Coefficient for Numeric Data

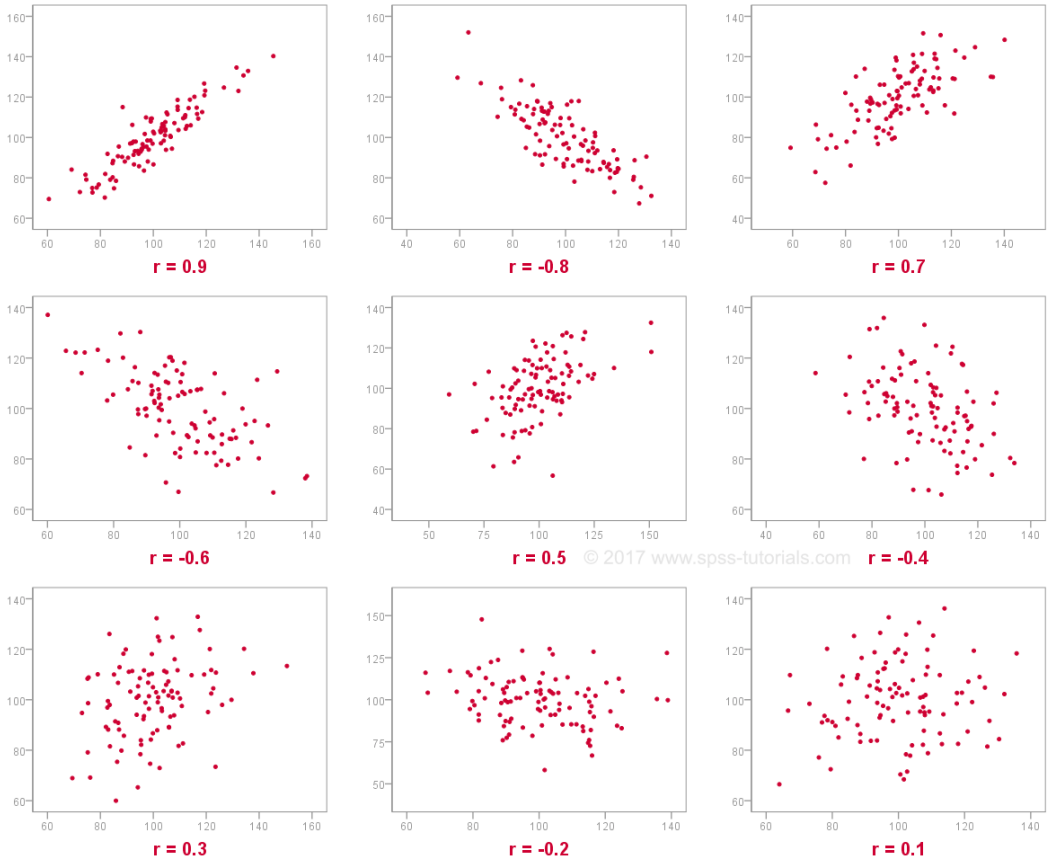
- Correlation coefficient ranges from -1 to 1.
- There are two properties: magnitude and sign.
- Equation (Pearson's product moment coefficient) :

$$\gamma_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{n\sum_{i=1}^{n}x_i y_i - \sum_{i=1}^{n}x_i \sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i{}^2 - (\sum_{i=1}^{n}x_i)^2}\sqrt{n\sum_{i=1}^{n}y_i{}^2 - (\sum_{i=1}^{n}y_i)^2}}$$

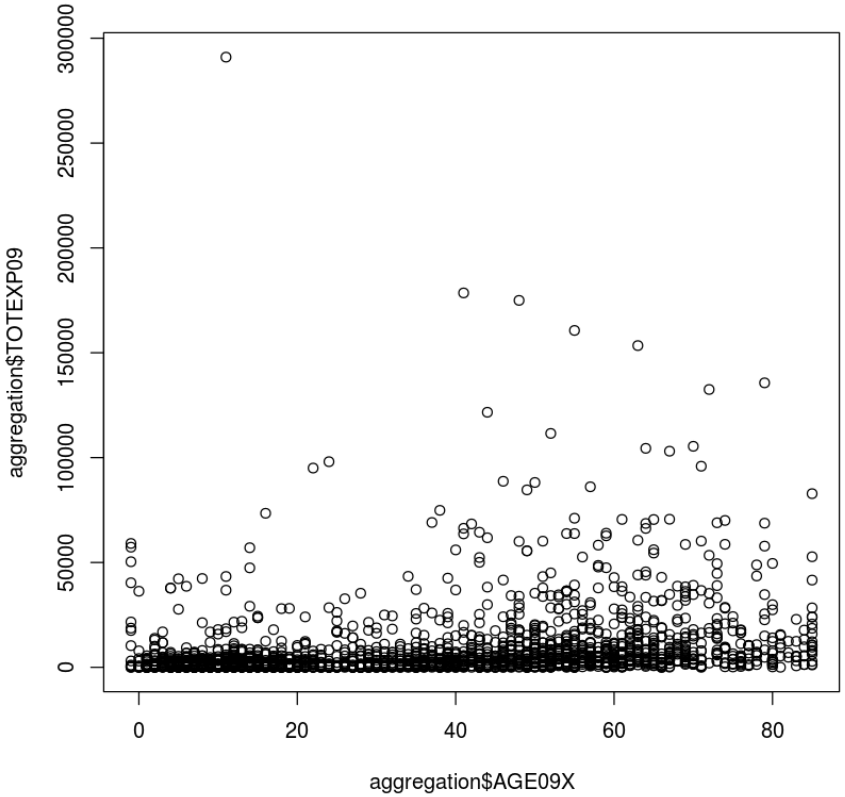
where:

  - n is sample size
  - x_i, y_i are the individual sample points with index i
  - $\bar{x}$ represents the sample mean

Pearson Correlation Visualized as Scatterplot



A Scatterplot of Total Expenditure and Age



χ^2 Correlation Test for Nominal Data

- Chi-squared (χ^2) examines whether the number of observations (counts) within categories depart from expected counts for the corresponding categories.
- Equation (Pearson's χ^2 correlation):

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Where:

- o_{ij} is observed frequency
- e_{ij} is expected frequency
- $e_{ij} = \frac{r_i \times c_j}{n}$, where r_i is respective row total, c_j is respective column total, and n is total number of observations

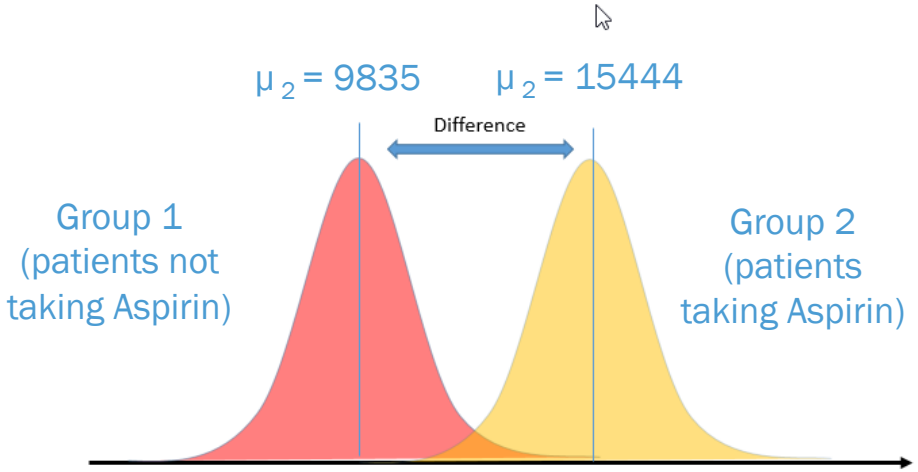
χ^2 Correlation Computation

	Northeast	Midwest	South	West	Total
Taking Aspirin	56 (62.78)	64 (72.79)	144 (115.32)	74 (87.12)	338
Not Taking Aspirin	220 (213.22)	256 (247.21)	363 (391.68)	309 (295.89)	1148
Total	276	320	507	383	1486

- $$\chi^2 = \frac{(56-62.78)^2}{62.78} + \frac{(220-213.22)^2}{213.22} + \frac{(64-72.79)^2}{72.79} + \frac{(256-247.21)^2}{247.21} + \frac{(144-115.32)^2}{115.32} + \frac{(363-391.68)^2}{391.68} + \frac{(74-87.12)^2}{87.12} + \frac{(309-295.89)^2}{295.89} = 14.11$$

T-test

- *T*-test is used to compare means of two groups and to determine how likely the difference between two means occur by chance.



T-test

- *T*-test is used to compare means of two groups and to determine how likely the difference between two means occur by chance.
  - Dependent (paired) *t*-test: compares means of a single sample of individuals (same subjects) at two different times (control vs. treatment) or on two different measures.
  - Independent *t*-test: compares means of two samples that are independent
    - Equal variance: when the number of subjects in two samples is the same or the variance of two samples is similar.
    - Unequal variance: when the number of subjects in two samples is different and the variance of two samples is different.

ANOVA

- The Analysis of Variance (ANOVA) is an extension of the *t*-test.
- ANOVA is used to compare means of more than two groups and to determine how likely the difference of means among them occur by chance.

Any Questions?

Prescriptive Analysis

Next Class

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL