Stat 101L: Lecture 11

Correlation

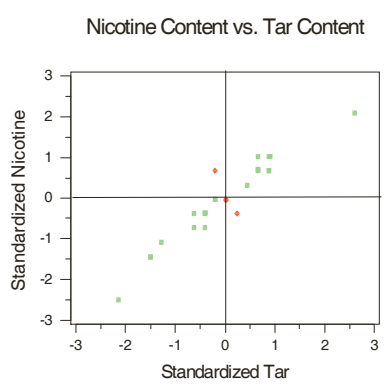
- Linear Association
  - How closely do the points on the scatter plot represent a straight line?
  - The correlation coefficient gives the direction of the linear association and quantifies the strength of the linear association between two quantitative variables.

1

Correlation

- Standardize y $$z_y = \frac{y - \bar{y}}{s_y}$$
- Standardize x $$z_x = \frac{x - \bar{x}}{s_x}$$

2



Stat 101L: Lecture 11

Correlation Coefficient

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

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

4

Correlation Conditions

- Correlation applies only to quantitative variables.
- Correlation measures the strength of linear association.
- Outliers can distort the value of the correlation coefficient.

5

Correlation Coefficient

- Tar and nicotine

$$r = \frac{\sum z_x z_y}{n-1} = \frac{22.9437}{24}$$

- $r = 0.956$

6

Stat 101L: Lecture 11

Correlation Coefficient

- There is a strong positive correlation, linear association, between the tar content and nicotine content of the various cigarette brands.

7

JMP

- Analyze – Multivariate methods – Multivariate
- Y, Columns
  - Tar
  - Nicotine

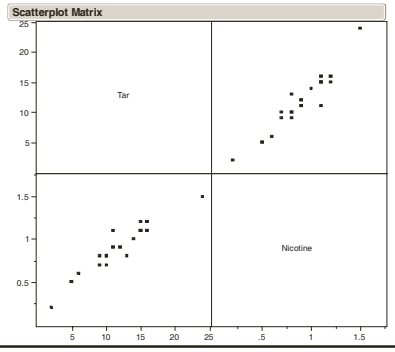
8

Multivariate

Correlations

	Tar	Nicotine
Tar	1.0000	0.9560
Nicotine	0.9560	1.0000

Scatterplot Matrix



9

Stat 101L: Lecture 11

Correlation Properties

- The sign of r indicates the direction of the association.
- The value of r is always between -1 and $+1$
- Correlation has no units.
- Correlation is not affected by changes of center or scale.

10

Correlation Cautions

- "Correlation" and "Association" are different.
  - Correlation – specific (linear).
  - Association – vague (trend).
- Don't correlate categorical variables.

11

Correlation Cautions

- Don't confuse correlation with causation.
  - There is a strong positive correlation between the number of crimes committed in communities and the number of 2nd graders in those communities.
- Beware of lurking variables.

12