

Stat 101 – Lecture 11

Correlation Coefficient

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

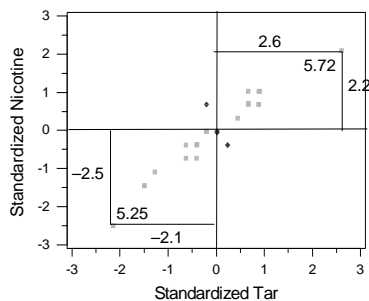
1

Standardized Values

- | | |
|-----------------------|------------------------|
| • Lucky Strike | • Now |
| -Tar = 24 | -Tar = 2 |
| -z _x = 2.6 | -z _x = -2.1 |
| -Nicotine = 1.5 | -Nicotine = 0.2 |
| -z _y = 2.1 | -z _y = -2.5 |

2

Nicotine Content vs. Tar Content



3

Stat 101 – Lecture 11

Correlation Coefficient

- Tar and nicotine

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

- $r = 0.956$

4

Correlation Coefficient

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

5

Correlation Coefficient

- Tar and nicotine

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

$$r = \frac{+29.889}{24(4.636)(0.281)} = +0.956$$

6

Stat 101 – Lecture 11

Correlation Conditions

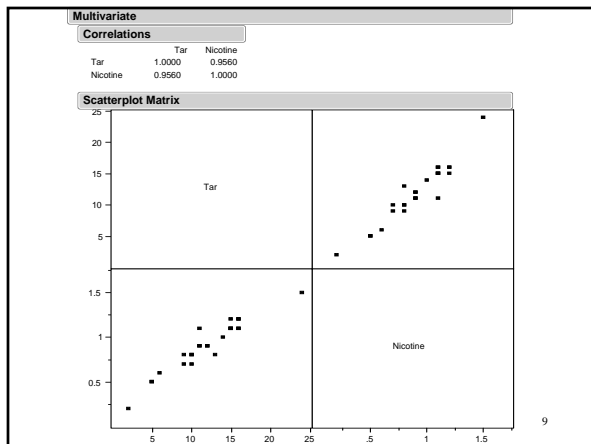
- Correlation applies only to quantitative variables.
- Correlation measures the strength of linear association.
- Outliers can distort the value of the correlation coefficient.

7

JMP

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

8



9

Stat 101 – 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 only linear.
 - Association – vague.
- 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
