Lecture 4
Segment 2

Calculating correlations

Correlations 2

- Important topics
  - Pearson product-moment correlation coefficient (r)
  - Covariance

Correlations 2

- Correlation coefficient (*r*)
  - Pearson product-moment correlation coefficient
- *r* = the degree to which X and Y vary together, relative to the degree to which X and Y vary independently
- *r* = (covariance of X & Y) / (variance of X & Y)

Correlations 2

- Here comes the math!
- Formulae for *r*
  - Raw score formula
  - z-score formula

Correlations 2

- Let’s quickly review calculations from Lecture 2 on descriptive statistics

Linsanity!

Jeremy Lin (10 games)

Points per game	(X-M)	$(X-M)^2$
28	5.3	28.09
26	3.3	10.89
10	-12.7	161.29
27	4.3	18.49
20	-2.7	7.29
38	15.3	234.09
23	0.3	0.09
28	5.3	28.09
25	2.3	5.29
2	-20.7	428.49
M = 227/10 = 22.7	M = 0/10 = 0	M = 922.1/10 = 92.21

Results

- M = mean = 22.7
- SD = standard deviation = 9.6
- SD^2 = variance = 92.21

Notation

- M = mean
- SD = standard deviation
- SD^2 = variance (also known as MS)
    - MS stands for Mean Squares
    - SS stands for Sum of Squares

Just one new concept ☺

- SP = Sum of cross Products

Just one new concept ☺

- Review: To calculate SS
  - For each subject in the sample, calculate their deviation score
    - $(X - M_x)$
  - Square the deviation scores
    - $(X - M_x)^2$
  - Sum the squared deviation scores
    - $SS_x = \Sigma[(X - M_x)^2] = \Sigma[(X - M_x) \text{ x } (X - M_x)]$

Just one new concept ☺

- To calculate SP
  - For each subject in the sample, calculate their deviation scores on both X and Y
    - $(X - M_x)$
    - $(Y – M_y)$
  - Then, for each subject, multiply the deviation score on X by the deviation score on Y
    - $(X – M_x) \times (Y – M_y)$
  - Then sum the “cross products”
    - $SP = \Sigma[(X – M_x) \times (Y – M_y)]$

Formulae to calculate r

Raw score formula:

$$r = SP_{xy} / \sqrt{SS_x SS_y}$$

$$SS_x = \Sigma(X - M_x)^2 = \Sigma[(X - M_x)(X - M_x)]$$

$$SS_y = \Sigma(Y - M_y)^2 = \Sigma[(Y - M_y)(Y - M_y)]$$

$$SP_{xy} = \Sigma[(X - M_x)(Y - M_y)]$$

Formulae to calculate *r*

Raw score formula continued:

$$r = SP_{xy} / \sqrt{SS_x SS_y}$$

$$r = \Sigma[(X - M_x)(Y - M_y)] / \sqrt{\Sigma(X - M_x)^2 \, \Sigma(Y - M_y)^2}$$

Formulae to calculate r

z-score formula:

$r = \Sigma(z_x z_y) / N$

$z_x = (X - M_x) / SD_x$
$z_y = (Y - M_y) / SD_y$

$SD_x = \sqrt{\Sigma(X - M_x)^2 / N}$
$SD_y = \sqrt{\Sigma(Y - M_y)^2 / N}$

Formulae to calculate *r*

z-score formula continued:

$z_x = (X - M_x) / SD_x$
$z_y = (Y - M_y) / SD_y$

$SD_x = \sqrt{\Sigma(X - M_x)^2 / N}$
$SD_y = \sqrt{\Sigma(Y - M_y)^2 / N}$

Formulae to calculate *r*

Proof of equivalence:

$z_x = (X - M_x) / \sqrt{\Sigma(X - M_x)^2 / N}$

$z_y = (Y - M_y) / \sqrt{\Sigma(Y - M_y)^2 / N}$

$$r = \Sigma \{ [(X - M_x) / \sqrt{\Sigma(X - M_x)^2 / N}] \\ [(Y - M_y) / \sqrt{\Sigma(Y - M_y)^2 / N}] \} / N$$

Formulae to calculate *r*

Proof of equivalence continued:

$$r = \Sigma \{ [(X - M_x) / \sqrt{\Sigma(X - M_x)^2 / N}]$$
$$[(Y - M_y) / \sqrt{\Sigma(Y - M_y)^2 / N}] \} / N$$

$$r = \Sigma [(X - M_x)(Y - M_y)] /$$
$$\sqrt{\Sigma(X - M_x)^2 \, \Sigma(Y - M_y)^2}$$

$$r = SP_{xy} / \sqrt{SS_x SS_y}$$

Variance and covariance

- Variance = MS = SS / N
- Covariance = COV = SP / N

- Correlation is standardized COV
  - Standardized so the value is in the range -1 to 1

Note on the denominators

- Correlation for descriptive purposes
  - Divide by N
- Correlation for inferential purposes
  - Divide by N-1

Correlations 2: Review

- Important topics
  - Pearson product-moment correlation coefficient (*r*)
  - Covariance

Image in slide 6 is from Nathaniel S. Butler, NBAE/Getty Images.