PSYCHOLOGY 210: INTRODUCTION TO STATISTICS

Correlation & Regression

May 9, 2011

1

AGENDA

- Announcements
- Chapter 15: Correlation & Regression
- Next time: Quiz 6; Review for Final

2

ANNOUNCEMENTS

- Exam 3: Let's Review It!
- Quiz 6 Thursday May 12th (Correlation & Regression; Chapter 15)
- Lab 10 will be due at the BEGINNING of LECTURE on Thurs. May 19th
- Final Thur. May 19th 10:15-12:15pm (extra credit will be available!)

3

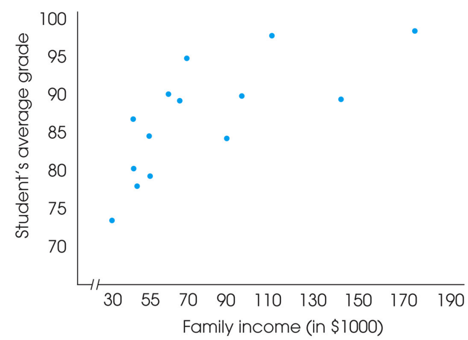
CORRELATION INTRO

- Measures and describes a relationship between two variables.
  - Requires 2 scores for each individual
- Characteristics of relationships
  - Direction
    - Positive: Variables tend to change in the same direction
    - Negative: Variables tend to go in opposite directions
  - Form (linear is most common)
  - Strength
    - Perfect correlation: 1.00 (+ or -)

4

CORRELATION INTRO

Person	Family Income (in $1000)	Student's Average Grade
A	31	72
B	38	86
C	42	81
D	44	78
E	49	85
F	56	80
G	58	91
H	65	89
I	70	94
J	90	83
K	92	90
L	106	97
M	135	89
N	174	95



5

PEARSON CORRELATION

- Measures the degree and the direction of the linear relationship between two variables
- Perfect linear relationship
  - Every change in X has a corresponding change in Y
  - Correlation will be –1.00 or +1.00

$$r = \frac{covariability\ of\ X\ and\ Y}{variablity\ of\ X\ and\ Y\ separately}$$

6

PEARSON CORRELATION

- Sum of Products of Deviations:
  - Similar to SS (sum of squared deviations)
  - Measures the amount of covariability between two variables

$$SP = \sum (X - M_X)(Y - M_Y)$$

Definitional Formula

$$SP = \sum XY - \frac{\sum X \sum Y}{n}$$

Computational Formula

7

PEARSON CORRELATION

- Ratio comparing the covariability of X and Y (numerator) with the variability of X and Y separately (denominator)

$$r = \frac{SP}{\sqrt{SS_X SS_Y}}$$

8

PEARSON CORRELATION

- Example:

X	Y	X - M_x	Y - M_y	$(X – M_x)^2$	$(Y - M_y)^2$	Product
0	2	0 - 6 = -6	2 – 4 = -2	$(-6)^2$	$(-2)^2$	12
10	6	10 – 6 = 4	6 – 4= 2	$(4)^2$	$(2)^2$	8
4	2	4 – 6 = -2	2 – 4 = -2	$(-2)^2$	$(-2)^2$	4
8	4	8 – 6 = 2	4 – 4 = 0	$(2)^2$	$(0)^2$	0
8	6	8 – 6 = 2	6 – 4 = 2	$(2)^2$	$(2)^2$	4

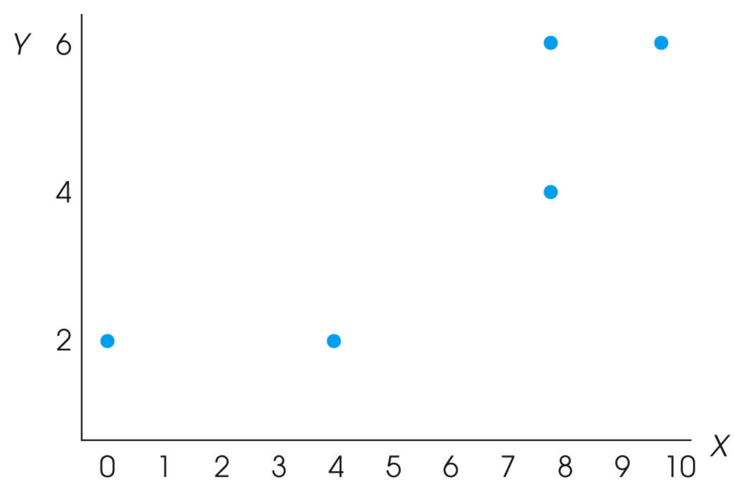
M_x:6 SS_x: 64 SS_y: 16 SP: 28

M_y: 4

$$r = \frac{SP}{\sqrt{SS_X SS_Y}} \qquad r = \frac{28}{\sqrt{(64)\ (16)}} = 0.875$$

9

PEARSON CORRELATION



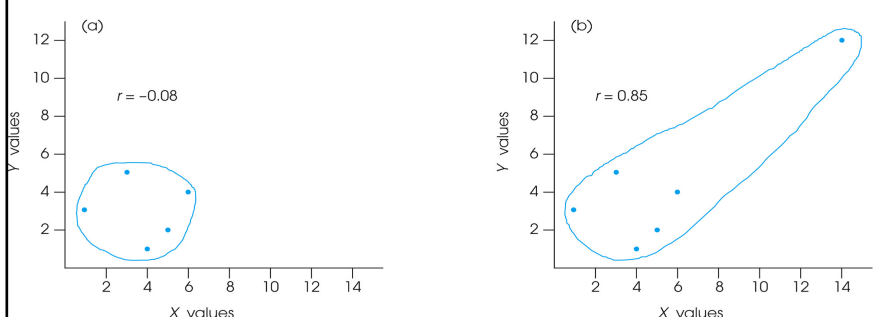
10

PEARSON CORRELATION

- When & why correlations are used:
  - Prediction
  - Validity
  - Reliability
  - Theory verification
- Interpreting:
  - Correlation does not demonstrate causation; only describes relationship b/t 2 variables
  - Value of correlation is affected by the range of scores in the data
  - Extreme points – outliers – have an impact
  - Correlation cannot be interpreted as a proportion.
    - To show the shared variability, need to square the correlation

11

PEARSON CORRELATION



Original Data

Subject	X	Y
A	1	3
B	3	5
C	6	4
D	4	1
E	5	2

Data with Outlier Included

Subject	X	Y
A	1	3
B	3	5
C	6	4
D	4	1
E	5	2
F	14	12

12

COEFFICIENT OF DETERMINATION

- Coefficient of determination measures the proportion of variability in one variable that can be determined from the relationship with the other variable.

$$\text{Coefficient of Determination} = r^2$$

13

HYPOTHESIS TESTING WITH THE PEARSON CORRELATION

- Pearson correlation is usually computed for sample data, but used to test hypotheses about the relationship in the population.
- Population correlation shown by Greek letter rho (ρ)
- Nondirectional: H_0: $\rho = 0$ and H_1: $\rho \neq 0$
- Directional: H_0: $\rho \leq 0$ and H_1: $\rho > 0$

14

HYPOTHESIS TESTING WITH THE PEARSON CORRELATION

- Sample correlation used to test population ρ
- Degrees of freedom (df) = $n – 2$
- Hypothesis test can be computed using either t or F.
- Critical Values have been computed
  - See Table B.6
  - A sample correlation beyond ± Critical Value is very unlikely
  - A sample correlation beyond ± Critical Value leads to rejecting the null hypothesis.

15

CORRELATION: IN LITERATURE

- E.g., " A correlation for the data revealed that the amount of education and annual income were significantly related, r = +.65, n = 30, p < .01, two-tails."
- Includes:
  - Sample size
  - Probability level
  - Type of test used (one or two-tailed
  - Value of r

16

PARTIAL CORRELATION

- A partial correlation measures the relationship between two variables while controlling the influence of a third variable by holding it constant
  - 3 variables: X, Y, Z, can compute 3 individual Pearson correlations:
    - (1) r_{xy} measures correlation between X & Y
    - (2) r_{xz} measures correlation between X & Z
    - (3) r_{yz} measures correlation between Y & Z

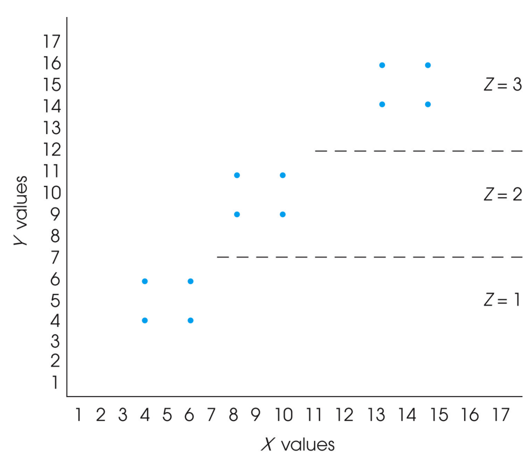
$$r_{xy \cdot z} = \frac{r_{xy} - (r_{xy} r_{yz})}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

17

PARTIAL CORRELATION

- X: Churches; Y: Crimes; Z: Population
  - r_{xy}: 0.943
  - r_{xz}: 0.971
  - r_{yz}: 0.971

X	Y	Z
4	4	1
4	6	1
6	4	1
6	6	1
9	9	2
9	11	2
11	9	2
11	11	2
14	14	3
14	16	3
16	14	3
16	16	3



18

PARTIAL CORRELATION

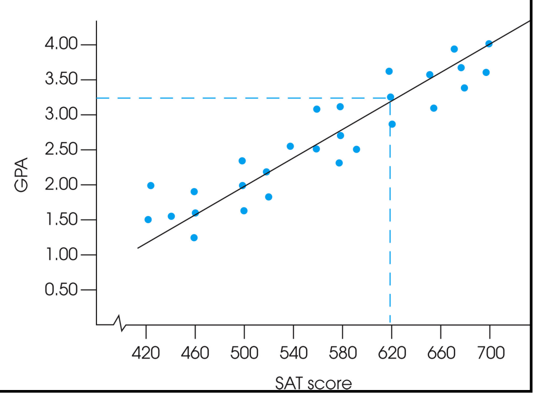
- X: Churches; Y: Crimes; Z: Population
  - rxy: 0.943
  - rxz: 0.971
  - ryz: 0.971

$$r_{xy\cdot z} = \frac{r_{xy} - (r_{xy} r_{yz})}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}} = \frac{0.943 \; - (0.971)(0.971)}{\sqrt{(1-0.971^2)(1-0.971^2)}} = 0$$

19

REGRESSION

- The Pearson correlation measures a linear relationship between two variables
- The line through the data
  - Makes the relationship easier to see
  - Shows the central tendency of the relationship
  - Can be used for prediction



20

REGRESSION

- Linear Equations
- General equation for a line
  - Equation: $Y = bX + a$
  - X and Y are variables
  - a and b are fixed constant
    - b : Slope. Determines how much the Y variable changes when X is increased by 1 point
    - a : Y-intercept. The point where the line intercepts the Y-axis.

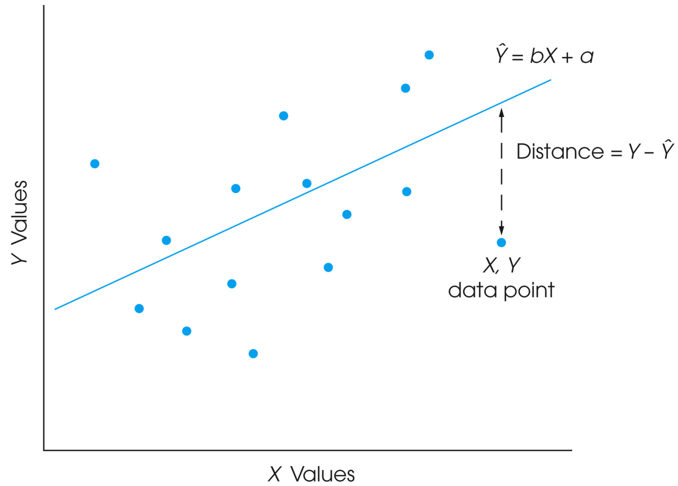
21

REGRESSION

- **Regression** is the method for determining the best-fitting line through a set of data
  - The line is called the **regression line**
- $\hat{Y}$ is the value of Y predicted by the regression equation for each value of X
- $(Y - \hat{Y})$ is the distance of each data point from the regression line: the **error** of prediction
  - Can be positive or negative
- **Regression** minimizes **total squared error:** $\sum(Y - \hat{Y})^2$
  - Total Squared Error: Overall squared error of prediction between the line & the data
  - Best-fit line has the smallest total squared error

22

REGRESSION



23

REGRESSION

- Regression line equation: $\hat{Y} = bX + a$
- The slope of the line, b, can be calculated

$$b = \frac{SP}{SS_X} \quad or \quad r\frac{s_Y}{s_X}$$

- The line goes through (M_X, M_Y) so

$$a = M_Y - bM_X$$

24

REGRESSION

- Regression equation makes prediction
- Precision of the estimate is measured by the ***standard error of estimate***
  - Standard Error of Estimate: A measure of the standard distance between a regression line and the actual data points

$$\sqrt{\frac{SS_{residual}}{df}} = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-2}}$$

- Unpredicted variability in Y scores:
  $SS_{residual} = (1 - r^2)\, SS_Y$
- Predicted variability in Y scores:
  $SS_{regression} = r^2\, SS_Y$

REGRESSION

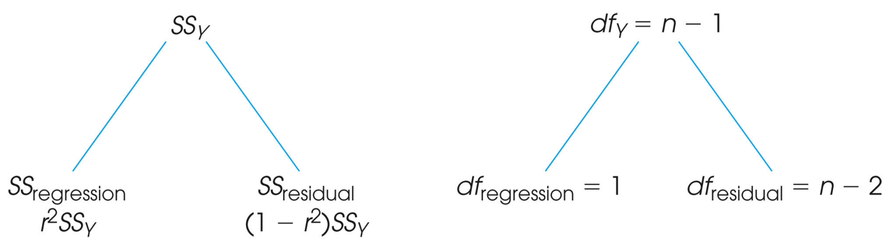
- Analysis of Regression
  - Similar to Analysis of Variance
  - Uses an *F*-ratio of two Mean Square values
  - Each *MS* is a *SS* divided by its *df*

$$MS_{regression} = \frac{SS_{regression}}{df_{regression}}$$

$$MS_{residual} = \frac{SS_{residual}}{df_{residual}}$$

$$F = \frac{MS_{regression}}{MS_{residual}}$$

REGRESSION



27

NEXT TIME…

- Quiz 6: Chapter 15
- Extra Info. on Correlation & Regression NOT found in book
  - Examples of Correlation & Regression
- Review for Final Exam
- Homework Chapter 15:
  - #s: ALL Even Problems
    - You'll want as much practice as possible!
    - Recommend: Odd numbers too (answers are in back of book)

28

RICCA'S REALLY BAD STATS JOKE

- **Knock! Knock! Who's there?**
  - **Willie and Boris.**
  - **Willie Boris who?**
  - **Willie Boris with his stat lecture today?**

- **What did one regression coefficient say to the other regression coefficient?**
  - I'm partial to you!

29