

PSYC 60 – Statistics
Spring 2012
Notes #9

Correlation

Book: Chapter 6 - Describing Relationships
Sections 6.1-6.6

Relationships / Scatterplot

Direction of relationship

Positive - as one variable increases,
the other variable _____

Negative - as one variable increases,
the other variable _____

Strength of relationship

a change in one variable is _____ accompanied by a predictable change in the other

Strong

Weak

None

Shape of relationship

Linear

Curvilinear

Example: Factors affecting PSYC60 quiz grades

The quiz has a mean $M = 80$ pts. and standard deviation $SD = 17$ pts.

Assume that each quiz grade is a combination of these factors:

25 pts for guessing on test

+ 5 pts per hour class/studying

+ 20 pts if high school stats

+ 2 pts per year in college

students range from 0 to "unlimited"

only 1% of students took high school stats

all students have 0-4 years in college

Correlation variable, r

Pearson correlation - a value that describes the linear relationship between pairs of quantitative variables

Example values:

Symbol:

Sign:

Magnitude:

r Formula (Theoretical or z Score Formula)

r =

Example: Calculate the r value for the following quiz grades based on hours studied

	<u>Hours</u>	<u>Grade</u>	<u>zHours</u>	<u>zGrade</u>	<u>zHours * zGrade</u>
A:	10	90	?	?	?
B:	2	75	-1.15	-0.93	1.07
C:	9	81	0.46	-0.13	-0.06
M =	7	82			
s =	4.36	7.55			

r Formula (Computational)

In the book, but you will not be tested on this formula

Is the r value statistically significant?

If there is no relationship between two variables, in theory r = _____

However, r will not be this exactly due to _____

Values are more likely to be significant the closer they are to _____ (the further from _____)

How many data points (n) do you need to calculate r?

n = 1

n = 2

n = 3

For correlation, degrees of freedom (df) = _____

We can calculate a t value that indicates how likely an r value is by chance

t =

The typical rules apply for hypothesis testing, finding critical t, and making a decision.

Example: A scientist calculates an r value for the relationship between drug dosage and cell lifespan. For 10 cell cultures, the r value is 0.2 Is the correlation significant? If she had performed the experiment with 120 cell cultures and found $r = 0.2$, is that significant? Is the relationship stronger? What caused the difference?

r - Issues

r^2 = percent of variance in one variable that is explained/predicted by the other variable

Not effected by units of measurement or size of units

Range restriction

Correlation not causation

Outliers

Regression

Book: Chapter 7 - Regression

Background: Formula for a line

Given degrees Celsius ($^{\circ}\text{C}$), what are the degrees Fahrenheit ($^{\circ}\text{F}$)?

Slope

Known: For a change of 9°F , there is a change of 5°C

Like an "exchange rate"

Intercept

Known: Freezing is 32°F and 0°C

0°C times the slope = 0°F , but should be 32°F

There is an "adjustment/offset"

Formula for a line is _____

We will use the symbols $y = b * x + a$ to match the book

(We will not use $y = m * x + b$, though you may have previously learned this)

Factors affecting PSYC60 quiz grades

Continuing with the example from the Correlation section above

Goal: Given the number of hours worked, predict someone's grade

We need a _____, which means we need "best" values for ____ and ____

Setting the slope wrong

Setting the intercept wrong

Ways that our formula could be "best":

Through as many
points as possible

Minimize
total error

Minimize
squared error

Least squares regression line: Solving

Example: Determine the formula for the least squares regression line for predicting quiz grades based on hours studied

	<u>Hours</u>	<u>Grade</u>	
A:	10	90	SS Hours = 38
B:	2	75	SS Grades = 114
C:	9	81	

Formula for the line:

b =

a =

Least squares regression line: Quantifying error

Once a least square regression line is determine, this formula defines the typical amount that the _____ values deviates from the _____ values

Like past variables that quantify how much _____ values deviate from _____ values, this variable is another _____

Name:

Formula:

Example: Determine the standard error of the estimate for the least squares regression line for predicting quiz grades based on hours studied

	<u>Hours</u>	<u>Grade</u>	<u>Predicted (Y')</u>	Y - Y'	(Y - Y') ²
A:	10	90			
B:	2	75			
C:	9	81			

Where have we been / where are we going next?

1 & 2-sample z & t tests	Correlation/ Regression	Chi-squared
-----------------------------	----------------------------	-------------

1st variable:

2nd variable:

Graphs:

Goal: