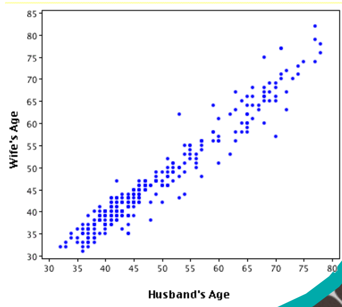
BELLWORK

1. Rank these correlations from weakest to strongest: -.29 , .62 , .99 , -.7 , -.33 , -1

2. Using the scatterplot describe:
   trend:
   strength:
   direction:



Sep 26-5:13 PM

WELCOME

ON YOUR DESK:
- hw ?'s
- bellwork
- notebook
- calculator
- colors

vocab

examples

Sep 26-5:13 PM

YESTERDAY?

- What is r?
- What makes it stronger?
- What does the sign of r tell you?
- So you type stuff into L1 and L2 and then press STAT -->CALC--> what??
- What does an r value (correlation coefficient) of -1.0 mean?
- If I tell you r is 2.4, you tell me...?

Sep 26-5:13 PM

Quick Review

Do you text while talking to a friend? 53 of 72 teens said yes, while only 14 of 87 adults said yes.

Create a contingency table! (You will need to know where to put the explanatory and response variables!)

Then find the marginal proportion of people who said they DO text while in conversation.

Find the conditional proportions by age to see if they text while talking or not.

Sep 26-5:13 PM

Quick Review

Do you text while talking to a friend? 53 of 72 teens said yes, while only 14 of 87 adults said yes.

Age/Texting→	Yes	No	Total
Teen	53	19	72
Adult	14	73	87
Total	67	92	159

DO text: 67/159 = 0.42

	yes	no
teen	.74	.26
adult	.16	.84

Sep 26-5:13 PM

Quick Review

How long did you study for our last test?

Study/Grade→	A or B	D or F	Total
1 or more hours	19	6	25
Less than 1 hour	5	21	26
Total	24	27	51

State the explanatory and response variables.

Find the conditional proportions by how many hours you studied to see what grade you got.

Sep 26-5:13 PM

Quick Review

How long did you study for our last test?

Study/Grade→	A or B	D or F	Total
1 or more hours	19	6	25
Less than 1 hour	5	21	26
Total	24	27	51

ex: # hours r: grade

	A/B	D/F
1 +	.76	.24
< 1	.19	.81

Sep 26-5:13 PM

3.3

How Can We Predict the Outcome of a Variable?

Sep 26-5:13 PM

Learning Objectives

- Definition of a regression line
- Use a regression equation for prediction
- Interpret the slope and y-intercept of a regression line
- Identify the least-squares regression line as the one that minimizes the sum of squared residuals
- Calculate the least-squares regression line

Sep 26-5:13 PM

Learning Objectives

- Compare roles of explanatory and response variables in correlation and regression
- Calculate r^2 and interpret

Sep 26-5:13 PM

Regression Analysis

The first step of a *regression analysis* is to identify the response and explanatory variables

We use *y* to denote the *response variable*

We use *x* to denote the *explanatory variable*

Sep 26-5:13 PM

Regression Line

A **regression line** is a straight line that describes how the response variable (y) changes as the explanatory variable (x) changes

A regression line predicts the value of the response variable (y) for a given level of the explanatory variable (x)

$$\hat{y} = a + bx$$

Sep 26-5:13 PM

Regression Line

- The y-intercept of the regression line is denoted by *a*

- The slope of the regression line is denoted by *b*

Sep 26-5:13 PM

We Do Example

How Can Anthropologists Predict Height Using Human Remains?

$\hat{y} = 61.4 + 2.4x$

$\hat{y}$ is the predicted height & x is the length of a femur (thigh bone) measured in cm

What if we found a femur with a length of 33 inches?

Sep 26-5:13 PM

Practice Example

Use the regression equation to predict the height of a person whose femur length was 50 centimeters

$$\hat{y} = 61.4 + 2.4x$$

$$\hat{y} = 61.4 + 2.4(50) = 181.4$$

Sep 26-5:13 PM

y-intercept

y-Intercept:

The predicted value for y when x = 0

Helps in plotting the line

May not have any interpretative value if no observations had x values near 0

Sep 26-5:13 PM

think about it...

Think about your shoe size against your height...how tall are you when you're born?

Sep 26-5:13 PM

slope

Slope: measures the change in the predicted variable (y) for a 1 unit increase in the explanatory variable in (x)

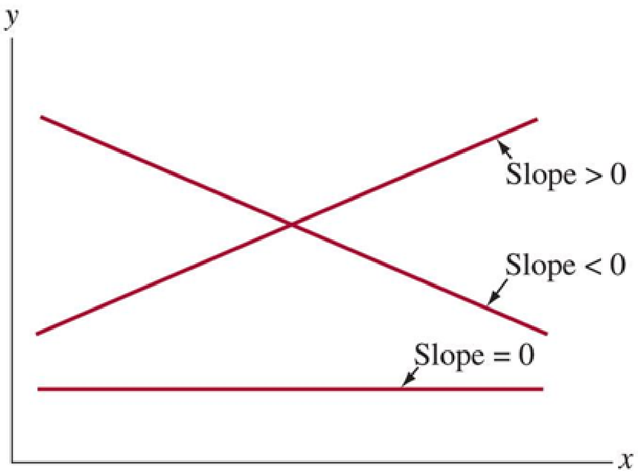
Example: A 1 cm increase in femur length results in a 2.4 cm increase in predicted height

Every 1 in x is the slope in y

$$\hat{y} = 61.4 + 2.4x$$

Sep 26-5:13 PM

slope values



Sep 26-5:13 PM

regression line

At a given value of x, the equation:

$$\hat{y} = a + bx$$

Predicts a single value of the response variable

But... we should not expect all subjects at that value of x to have the same value of y

Variability occurs in the y values!

Sep 26-5:13 PM

regression line

The regression line connects the estimated *means* of y at the various x values

In summary,

Describes the relationship between x and the *estimated means* of y at the various values of x

$$\hat{y} = a + bx$$

Sep 26-5:13 PM

residuals

Measures the size of the prediction errors, the vertical distance between the point and the regression line

Each observation has a residual

Calculation for each residual:

$$y - \hat{y}$$

A large residual indicates an unusual observation

Sep 26-5:13 PM

least squares regression

Residual sum of squares:

$$\sum (residuals)^2 = \sum (y - \hat{y})^2$$

The least squares regression line is the line that minimizes the vertical distance between the points and their predictions, i.e., it minimizes the residual sum of squares

Sep 26-5:13 PM

* That's it for today!

* If time remains, please get a book binder and catch up on homework

*Tomorrow, we will calculate the least squares regression line

Sep 26-6:12 PM

BELLWORK

1. What is the regression equation?

$\hat{y} = a + bx$

2. What does $\hat{y}$ represent?

prediction

3. What is the residual equation?

$y - \hat{y}$

Sep 26-5:13 PM

WELCOME

On your desk:

- notebook
- calculator

Sep 26-5:13 PM

regression formulas

- **Slope:** correlation

$$b = r\left(\frac{s_y}{s_x}\right)$$

slope ← st.dev.(y), st.dev.(x)

- **Y-Intercept:** mean(y), mean(x)

$$a = \bar{y} - b(\bar{x})$$

yint, slope

Regression line always passes through

$$(\bar{x}, \bar{y})$$

Sep 26-5:13 PM

We Do Example

Find slope, y-int, and write regression equation:

$\bar{x} = .275$

$\bar{y} = 4.979$

$s_x = 0.0091$

$s_y = 0.368$

$r = 0.653$

$$b = r\left(\frac{s_y}{s_x}\right) = \boxed{.653\left(\frac{.368}{.0091}\right)}$$

b=26.41

for every 1 change in x there is a 26.41 change in y

$$a = \bar{y} - b(\bar{x}) = \boxed{4.979 - 26.41(.275)}$$

a = -2.28

When x = 0, $\hat{y}$ is -2.28

$$\boxed{\hat{y} = -2.28 + 26.41x}$$

Sep 26-5:13 PM

Practice Example

Find slope, y-int, and write regression equation:

$\bar{x} = 2.88$

$\bar{y} = 85.25$

$s_x = 2.23$

$s_y = 13.65$

$r = .81$

$$b = .81\left(\frac{13.65}{2.23}\right) = 4.96$$

$$a = 85.25 - 4.96(2.88)$$
$$= 70.97$$

$$\boxed{\hat{y} = 70.97 + 4.96x}$$

x = # of hrs study
y = grade

Sep 26-5:13 PM

Calculator Commands

- Enter x data into L1
- Enter y data into L2
- STAT CALC menu
- Choose 8: LinReg(a+bx) (L_1, L_2)
- ~~1st number = x variable~~
- ~~2nd number = y variable~~
- Enter

Sep 26-5:13 PM

TRY IT!

TABLE 3.5: Team Batting Average and Team Scoring (Mean Number of Runs Per Game) for American League Teams in 2003[6]

Team	Batting Average	Team Scoring
Boston	.289	5.9
Toronto	.279	5.5
Minnesota	.277	4.9
Kansas City	.274	5.2
Seattle	.271	4.9
New York	.271	5.4
Anaheim	.268	4.5
Baltimore	.268	4.6
Texas	.266	5.1
Tampa Bay	.265	4.4
Chicago	.263	4.9
Oakland	.254	4.7
Cleveland	.254	4.3
Detroit	.240	3.6

$y = a + bx$

$\hat{y} = -6.25 + 41.58x$

Sep 26-5:13 PM

TRY IT!

Cereal	Sodium (mg)	Sugar (g)
Frosted Mini Wheats	0	7
Raisin Bran	210	12
All Bran	260	5
Apple Jacks	125	14
Capt Crunch	220	12
Cheerios	290	1
Cinnamon Toast Crunch	210	13
Crackling Oat Bran	140	10
Crispix	220	3
Frosted Flakes	200	11
Froot Loops	125	13
Grape Nuts	170	3
Honey Nut Cheerios	250	10
Life	150	6
Oatmeal Raisin Crisp	170	10
Honey Smacks	70	15
Special K	230	3
Wheaties	200	3
Corn Flakes	290	2
Honeycomb	180	11

Sep 26-5:13 PM

Correlation

Describes the strength of the linear association between 2 variables

Does not change when the units of measurement change

Does not depend upon which variable is the response and which is the explanatory

Sep 26-5:13 PM

Slope

Numerical value depends on the units used to measure the variables

Does not tell us whether the association is strong or weak

The two variables must be identified as response and explanatory variables

The regression equation can be used to predict values of the response variable for given values of the explanatory variable

Sep 26-5:13 PM

Proportion of the variation

$r^2 = .76$ or 76%

percentage

- r^2 measures the **proportion of the variation** in the y-values that is accounted for by the linear relationship of y with x
- A correlation of .9 means that 81% of the variation in the y-values can be explained by the explanatory variable, x

$$.9^2 = .81 = 81\%$$

Sep 26-5:13 PM

NEXT...

MIDWAY REVIEW

ANSWERS

Sep 26-5:13 PM

Sep 27-9:30 AM

Attachments

Mid-way Review Answer Key.doc