10.2 & 10.3 CORRELATION AND REGRESSION

●● *A relationship between two variables is called a correlation.*●●

Paired Data - x and y

Sketch Scatter plots

... shoe size vs. height ... hrs exercise/week vs. weight ... shoe size vs. salary

Scatter plots pg. 519

– Positive correlation

– Negative correlation

– No correlation

– Nonlinear correlation

USING THE CALCULATOR TO DRAW A SCATTER PLOT

***** construct a scatter plot:** (data from chapter problem)

Year	1960	1973	1986	1995	2002	2003
Cost of Pizza (x)	.15	.35	1.00	1.25	1.75	2.00
Subway Fare (y)	.15	.35	1.00	1.35	1.50	2.00

Enter Data

Enter data x and y into lists, say L_1 and L_2.

Scatter Plot

Establish the picture desired:

2nd - StatPlot -TYPE: is first one, and Xlist is L_1 and Ylist is L_2

Clear any functions defined in Y=

View the picture:

ZOOM-9: ZoomStat to see picture

PEARSON (OR LINEAR) CORRELATION COEFFICIENT

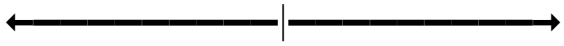
Measures the strength of the linear relationship.

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$

We $COULD$ actually compute r - but we'll use the calculator

– **Round to 3 decimal places**

– r-values range from -1 to 1



Draw scatter plots above the number line

– If $|r| >$ value in Table A-6, $\alpha = 0.05$ (page 760), linear correlation is significant

– If $|r| <$ value in Table A-6, $\alpha = 0.05$ (page 760), linear correlation is not significant

***** Use table A-6 (α = 0.05) and determine whether there is a significant linear correlation between x and y.**

a) n = 22, r = – .087 b) n = 40, r = 0.294 c) n = 25, r = – .401

●● *When there is a significant linear correlation between two variables, the equation describing the relationship, $y = ax + b$, is called the regression equation, and its graph is the regression line* ●●

COMPUTING THE CORRELATION COEFFICIENT WITH THE CALCULATOR

Turn Diagnostics On to show correlation coefficient, r.
This process only needs to be done ONCE, before you do your first linear regression.
2nd 0 (for CATALOG) Type D - and arrow down to DiagnosticsOn- press ENTER

***** use the data shown below (from the first page on this section)**

Year	1960	1973	1986	1995	2002	2003
Cost of Pizza (x)	.15	.35	1.00	1.25	1.75	2.00
Subway Fare (y)	.15	.35	1.00	1.35	1.50	2.00

(1) **Enter data** in L_1 & L_2

(2) **Draw a scatter plot** of the data (CLEAR any functions defined on Y=)
by looking at the scatter plot, do you think the linear correlation is significant?

(3) **Compute r and interpret (Use Table A-6, pg. 760, for $\alpha = 0.05$)**
STAT →CALC 4:LinReg(ax + b) L_1,L_2 ENTER

If you have turned Diagnostics on earlier, r and r^2 values will appear after a and b

(4) **Write equation of the line if appropriate**

(5) **Graph Equation of Line**:
to paste the equation of line into $Y_1 =$ do the following:
Y =
CLEAR Y_1
VARS 5:Statistics → → EQ 1:RegEQ

(6) Predict y for Additional Values of x Find subway fare when Cost of Pizza is $1.45

1. If the r value is significant, then put the particular x into the equation and find y.

   With the calculator, if you have the equation of the line in Y=, then
   GRAPH, 2nd CALC (on top row) 1:VALUE ENTER and enter the x value.

2. If the r value is NOT significant, then $\overline{y}$ is the best prediction that can be made.

   Find the mean of L_2 with the calculator

***** #18 pg. 533. Casino Size and Revenue**

Size (in thousands of square feet)	160	227	140	144	161	147	141
Revenue (in millions of dollars)	189	157	140	127	123	106	101

(1) **Enter data** in L_1 (size) & L_2 (revenue)

(2) **Draw a scatter plot** of the data (CLEAR any functions defined on Y=)
by looking at the scatter plot, do you think the linear correlation is significant?

(3) **Compute r and interpret (Use Table A-6, pg. 760, for $\alpha = 0.05$)**
STAT ➞CALC 4:LinReg(ax + b) L_1,L_2 ENTER

If you have turned Diagnostics on earlier, r and r^2 values will appear after a and b

(4) **Write equation of the line if appropriate**
Record $y = ax + b$
Specific values of a and b, are given in your calculator as you compute r.

(5) Predict y for Additional Values of x Find revenue when the casino is 182,000 sq. ft.

1. If the r value is significant, then put the particular x into the equation and find y.

   With the calculator, if you have the equation of the line in Y=, then
   GRAPH, 2nd CALC (on top row) 1:VALUE ENTER and enter the x value.

2. If the r value is NOT significant, then $\overline{y}$ is the best prediction that can be made.

   Find the mean of L_2 with the calculator

***** The table lists weights (in hundreds of pounds) and highway fuel usage rates (in mi/gal) for a sample of domestic new cars. Based on the result, can you expect to pay more for gas if you buy a heavier car? Use Table A-6 for (α=.05)**

x Weight (hundreds of pounds)	29	35	28	44	25	34	30	33	28	24
y Fuel (miles per gallon)	31	27	29	25	31	29	28	28	28	33

Use steps 1 – 6 from the previous page and answer the following:

i) Record and interpret r:

ii) Write equation if appropriate:

iii) Find the fuel usage if the car weight is 4,200 pounds.

***** #16 pg. 532 Heights of Presidents and Runners-Up (in inches)**

Winner	69.5	73	73	74	74.5	74.5	71	71
Rinner-Up	72	69.5	70	68	74	74	73	76

Use steps 1 – 6 from the previous page and answer the following:

i) Record and interpret r:

ii) Write equation if appropriate:

iii) What is the best predicted height of a runner-up if the winners height is 72 inches?

COMMENTS:

Common Errors Involving Correlation

- **Avoid concluding that a significant linear correlation between two variables implies causality**
- Data based on rates or average is a source of potential error
- No linear relationship does not imply no relationship at all. There are possibilities of nonlinear relationships

Round off rules

- **r to 3 decimal places**
- **a and b to 3 significant digits**

Predictions

- If there is a significant linear correlation use equation
- Otherwise the best predicted y value is $\overline{y}$

Guidelines for using Regression Equation - pg. 539 & 540

- If there is no significant linear correlation, don't use the regression equation to make predictions
- When using the regression equation for predictions, stay within the scope of the available sample data
- A regression equation based on old data is not necessarily valid now
- Don't make predictions about a population that is different from the population from which the sample data was drawn