MATH 10043 CHAPTER 10 DEFINITIONS and EXAMPLES

Section 10.2 *Correlation*

Definition: **Bivariate data** – the values of two different variables that are obtained from the same population. When we are working with two quantitative variables, we use **ordered pairs, (x, y)**, where we think x and y may be related in some way.

- x is the value of the first variable, called the independent variable
- y is the value of the second variable, called the dependent variable

Definition: A **scatter diagram** (or scatter plot) is a graph that plots the ordered (x, y) pairs as points on the x/y axes. A scatter plot has:

1. a title, identifying the source of the data.
2. a horizontal scale, identifying the x-variable.
3. a vertical scale, identifying the y-variable.

Ex. A) The following sample of college males was taken, with x representing heights in inches and y representing weights in pounds. Sketch a scatter plot of the data.

x	73	67	71	71	66	75	69	70	71	69	69
y	175	133	185	163	126	198	151	163	159	151	155

Definition: **Correlation analysis** measures the strength of a linear relationship between two variables.

Coefficient Of Linear Correlation --

-- also known as the **Correlation Coefficient, r**

r is the value of the strength of correlation.

r always has a value between –1 and 1.

Writing an Interpreting Statement for r – include four parts:

1. **strength of r.**
2. **sign of r.**
3. **what the variables represent (first x, then y).**
4. **population from which the sample was taken.**

Ex. B) Use the data from Example A above to find the correlation coefficient. Write a four-part statement to interpret the correlation coefficient.

•Using the Calculator to find the Correlation Coefficient:

****FIRST**: DO THIS STEP ONCE AND **ONLY** ONCE--
2ND CATALOG [this is the zero key]
scroll down to DIAGNOSTIC ON
ENTER ENTER

STAT ENTER (EDIT)
clear lists if necessary
enter x values into one list
enter y values into another list
STAT CALC option 4 (LinReg ax + b)
L#, L# ENTER

Section 10.2 Practice Problems

1. The owner of a health spa wished to study the relationship between the outdoor temperature (x) in degrees Fahrenheit at 11 a.m. and the number of customers (y) using the facilities at that time for randomly selected days during the summer. During the days sampled, the outdoor temperature ranged from 65 to 87 degrees Fahrenheit.
The sample yielded a correlation coefficient of r = 0.39. Interpret the correlation coefficient. [i.e., write a four-part statement.]

2. The ages (in years) and corresponding prices (in hundreds of dollars) of bottles of very fine wine selected at random were recorded during an auction. The bottles sold ranged from 18 to 60 years old. The correlation coefficient was found to be 0.87. Write an interpreting statement for the correlation coefficient given.

Section 10.3 *Regression*

Definition: **Regression analysis** gives an equation that provides values for y when given values of x. This equation can be called:
Regression equation Prediction equation Line-of-best-fit

Form of the regression equation: $\hat{y} = ax + b$

Definition: **Extrapolation** – making a prediction using an x value outside of the original range of x values.

•WHEN DO WE NOT MAKE PREDICTIONS?

1. When r is weak.

2. When we would be extrapolating.

Ex. C) Using data from Example A above: **(a)** Use linear regression to find the line of best fit. **(b)** If appropriate, predict the weight of a young-adult male whose height is 5 feet, 8 inches. **(c)** If appropriate, predict the weight of a young-adult male 4 feet, 11 inches tall. On parts b & c, if it is NOT appropriate to make the prediction, explain why.

Ex. D) The following data gives age in years, and price in hundreds of dollars of a sample of used Lexus GS300s. **(a)** Sketch a scatter plot of the data. **(b)** Find and interpret the correlation coefficient. **(c)** Use linear regression to find the line of best fit. **(d)** If appropriate, predict the price of a 6-year-old Lexus GS300. **(e)** If appropriate, predict the price of a 13-year-old Lexus GS300. On parts d and e, if it is not appropriate to make such a prediction, do not do so – write a sentence explaining why it is not appropriate.

x	5	9	11	10	7	5	4	2	2	3	4
y	210	150	57	70	167	250	295	352	375	240	267

Section 10.3 Practice Problems

1. The data below show the age in months, x, and height in inches, y, of Mrs. Owen's son. **(a)** Sketch a scatter diagram. **(b)** Find and interpret the correlation coefficient. **(c)** Write down the regression equation. (**d)** Is it appropriate to use the regression equation you found to predict his height when he is 27 months old? **(e)** Is it appropriate to use the regression equation you found to predict his height when he is 18 years old? On parts d & e, if it is NOT appropriate to make the prediction, explain why.

x	0	0.5	2.5	4	6	9	12	15	18	21	24
y	21	22	24	26	28	29	30	32	33	34.5	37

2. For Sherlock Holmes, the "science of deduction" was the key to solving every mystery. Distance between footprints (length of stride) told the master detective the height of a criminal he was pursuing. A random sample of 10 people provided the length of stride, x (in inches) and height, y (in inches) as given below. **(a)** Find and INTERPRET the correlation coefficient. **(b)** Find the regression equation. **(c)** Use the regression equation you found to predict the height of a person with a stride of 3 feet. **(d)** Use the regression equation to predict the height of a person with a stride of 31 inches. On parts c & d, if it is NOT appropriate to make the prediction, explain why.

Stride length -- x	22	24	26	30	23	29	24	32	22	32
Height -- y	60	61	68	75	64	70	67	77	62	73