12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation

Section 12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation

Learning Objective:

In this section, you will:

- Understand and interpret linear correlation
- Apply hypothesis testing about linear correlation

Scatter plots are particularly helpful graphs when we want to see if there is a linear relationship among data points. They indicate both the direction of the relationship between the x variables and the y variables, and the strength of the relationship. We calculate the strength of the relationship between an independent variable and a dependent variable using linear regression.

A **correlation** exists between two variables when the values of one variable are associated with the values of the other variable.

The **linear correlation coefficient, r,** measures the strength of the linear correlation between the independent variable x and the dependent variable y.

What the VALUE of *r* tells us:

- The value of r is always between –1 and +1: $–1 \leq r \leq 1$.
- The size of the correlation r indicates the strength of the linear relationship between x and

y. Values of r close to –1 or to +1 indicate a stronger linear relationship between x and y.

- If $r = 0$ there is absolutely no linear relationship between x and y **(no linear correlation)**.
- If $r = 1$, there is perfect positive correlation. If $r = –1$, there is perfect negative correlation. In both these cases, all of the original data points lie on a straight line. Of course, in the real world, this will not generally happen.

What the SIGN of *r* tells us

- A positive value of r means that when x increases, y tends to increase and when x decreases, y tends to decrease **(positive correlation)**.
- A negative value of r means that when x increases, y tends to decrease and when x decreases, y tends to increase **(negative correlation)**.

Example 1a: In Europe and Asia, m-commerce is popular. M-commerce users have special mobile phones that work like electronic wallets as well as provide phone and Internet services. Users can do everything from paying for parking to buying a TV set or soda from a machine to banking to checking sports scores on the Internet. For the years 2000 through 2004, was there a relationship between the year and the number of m-commerce users? Construct a scatter plot. Let x = the year and let y = the number of m-commerce users, in millions.

x (year)	y (# of users in millions)
2000	0.5
2002	20.0
2003	33.0
2004	47.0

Finding the linear correlation coefficient:

We can calculate the linear correlation using the graphing calculator, **F:LinRegTTest**. First enter the data into two lists (L1 and L2), then under STAT-TESTS, find F:LinRegTTest. Enter your lists, and choose the “not equal” option. Scroll down to find r.

We can draw a conclusion about the relationship between the entire populations by using a hypothesis testing.

- Method 1: Using the p-value
- Method 2: Using a table of critical values.

Example 1b: Use the data from example 1a, and use 0.01 significance level to test the claim that there is a linear correlation between the year and the number of m-commerce users 1) Null and Alternative Hypothesis

2. Calculator Work

3. Test Statistic, P-Value and Linear correlation coefficient r

4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

Example 2: A random sample of 11 statistics students produced the following data, where *x* is the third exam score out of 80, and *y* is the final exam score out of 200. Use a 0.05 significance level to test the claim that there is a linear correlation between the third exam score and the final exam score.

x (third exam score)	*y* (final exam score)
65	175
67	133
71	185
71	163
66	126
75	198
67	153
70	163
71	159
69	151
69	159

1. Null and Alternative Hypothesis

2. Calculator Work

3. Test Statistic, P-Value and Linear correlation coefficient r

4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

For more information and examples see online textbook OpenStax Introductory Statistics pages 682-685, 691-696.