Chapter 11: Analyzing Association Between Quantitative Variables: Regression Analysis

Section 11.3: How Can We Make Inferences About the Association?

Learning Objectives

1. Descriptive and Inferential Parts of Regression
2. Assumptions for Regression Analysis
3. Testing Independence between Quantitative Variables
4. A Confidence Interval for β

Learning Objective 1: Descriptive and Inferential Parts of Regression

- **The sample regression equation, r, and r^2 are *descriptive parts* of a regression analysis**
- **The *inferential parts* of regression use the tools of confidence intervals and significance tests to provide inference about the regression equation, the correlation and r-squared in the population of interest**

Learning Objective 2: Assumptions for Regression Analysis

- **Basic assumption for using regression line for description:**
  - **The population means of y at different values of x have a straight-line relationship with x, that is:**

$$\mu_y = \alpha + \beta x$$

  - **This assumption states that a straight-line regression model is valid**
  - **This can be verified with a scatterplot.**

Learning Objective 2:
Assumptions for Regression Analysis

- **Extra assumptions for using regression to make statistical inference:**
  - **The data were gathered using randomization**
  - **The population values of y at each value of x follow a normal distribution, with the same standard deviation at each x value**

5

Learning Objective 2:
Assumptions for Regression Analysis

- **Models, such as the regression model, merely *approximate* the true relationship between the variables**
- **A relationship will not be *exactly* linear, with *exactly* normal distributions for y at each x and with *exactly* the same standard deviation of y values at each x value**

6

Learning Objective 3:
Testing Independence between Quantitative Variables

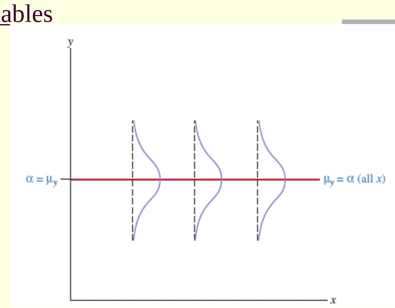
- **Suppose that the slope β of the regression line equals 0**

Then…

  - **The mean of y is identical at each x value**
  - **The two variables, x and y, are *statistically independent:***
    - **The outcome for y does not depend on the value of x**
    - **It does not help us to know the value of x if we want to predict the value of y**

7

Learning Objective 3:
Testing Independence between Quantitative Variables



8

Learning Objective 3: Testing Independence between Quantitative Variables

- **Steps of Two-Sided Significance Test about a Population Slope β:**

1. Assumptions:

- **The population satisfies regression line:**

$$\mu_y = \alpha + \beta x$$

- **Randomization**
- **The population values of y at each value of x follow a normal distribution, with the same standard deviation at each x value**

9

Learning Objective 3: Testing Independence between Quantitative Variables

- **Steps of Two-Sided Significance Test about a Population Slope β:**

2. Hypotheses:

H_0: β = 0, H_a: β ≠ 0

3. Test statistic: $t = \frac{b - 0}{se}$

- **Software supplies sample slope b and its se**

$$se = \sqrt{\frac{\sum \text{residuals}^2}{(n-2)\sum (x-\bar{x})^2}},$$

see instructor for program

10

Learning Objective 3: Testing Independence between Quantitative Variables

- **Steps of Two-Sided Significance Test about a Population Slope β:**

4. P-value: Two-tail probability of t test statistic value more extreme than observed:

Use t distribution with *df* = n-2

5. Conclusions: Interpret P-value in context

- **If decision needed, reject H_0 if P-value ≤ significance level**

11

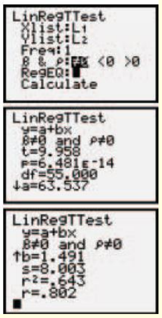
Learning Objective 3: Example: Is Strength Associated with 60-Pound Bench Press?

TABLE 12.4: MINITAB Printout for Regression Analysis of y = Maximum Bench Press (BP) and x = Number of 60-Pound Bench Presses (BP_60)

The regression equation is BP = 63.5 + 1.49 BP_60

Predictor	Coef	SE Coef	T	P
Constant	63.537	1.956	32.48	0.000
BP_60	1.4911	0.150	9.96	0.000

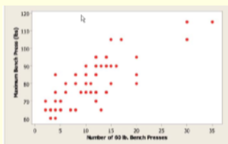
R-Sq = 64.3%



12

Learning Objective 3: Example: Is Strength Associated with 60-Pound Bench Press?

- **Conduct a two-sided significance test of the null hypothesis of independence**
- **Assumptions:**
  - **A scatterplot of the data revealed a linear trend so the straight-line regression model seems appropriate**
  - **The scatter of points have a similar spread at different x values**
  - **The sample was a convenience sample, not a random sample, so this is a concern**



Learning Objective 3: Example: Is Strength Associated with 60-Pound Bench Press?

- **Hypotheses: H_0: $\beta = 0$, H_a: $\beta \neq 0$**
- **Test statistic:**

$$t = \frac{b-0}{se} = \frac{(1.49-0)}{0.150} = 9.96$$

- **P-value: 0.000**
- **Conclusion: An association exists between the number of 60-pound bench presses and maximum bench press**

Learning Objective 4: A Confidence Interval for β

- **A small P-value in the significance test of H_0: $\beta = 0$ suggests that the population regression line has a nonzero slope**
- **To learn how far the slope β falls from 0, we construct a confidence interval:**

$$b \pm t_{.025}(se) \text{ with } \; df = n-2$$

Learning Objective 4: Example: Estimating the Slope for Predicting Maximum Bench Press

- **Construct a 95% confidence interval for β**

$$1.49 \pm 2.00(0.150) \text{ which is :}$$

$$1.49 \pm 0.30 \text{ or } (1.2, 1.8)$$

- **Based on a 95% CI, we can conclude, on average, the maximum bench press increases by between 1.2 and 1.8 pounds for each additional 60-pound bench press that an athlete can do**

Learning Objective 4:
Example: Estimating the Slope for Predicting Maximum Bench Press

- **Let's estimate the effect of a 10-unit increase in x:**
  - **Since the 95% CI for β is (1.2, 1.8), the 95% CI for 10β is (12, 18)**
  - **On the average, we infer that the maximum bench press increases by at least 12 pounds and at most 18 pounds, for an increase of 10 in the number of 60-pound bench presses**