PDE PROVIDER

Professional Development for the Expert

Course Title

Regression Analysis

Instructor

Alan Fata, DBA

Credit

1 PDU

Questions

10

Professional Development for the Expert, Inc.
(571) 619-5550, info@pdeprovider.com

I. Introduction

Regression analysis is one of the most widely used statistical techniques in business, economics, finance, medicine, engineering, psychology, and the social sciences. It enables researchers to examine the relationship between one dependent (outcome) variable and one or more independent (predictor) variables. Unlike correlation, which only measures the strength and direction of an association, regression goes a step further by estimating how changes in predictor variables influence the outcome variable and by allowing future values to be predicted.

For example, a bank may wish to determine whether customer satisfaction, service quality, and online banking usability influence customer loyalty. A university may investigate whether study hours, attendance, and prior GPA predict students' final grades. In both cases, regression provides a mathematical model that quantifies these relationships while testing whether they are statistically significant.

Regression is therefore both a **descriptive** and **predictive** statistical tool. Researchers use it not only to explain existing relationships but also to forecast future outcomes and support evidence-based decision-making.

Learning Objectives
After completing this lesson, students should be able to:

- Explain the purpose of regression analysis.
- Distinguish between simple and multiple regression.
- Interpret regression equations.
- Understand regression coefficients.
- Interpret R^2 and Adjusted R^2.
- Explain hypothesis testing in regression.
- Interpret ANOVA tables.
- Evaluate regression assumptions.
- Interpret SPSS regression output.
- Draw meaningful research conclusions.

II. What is Regression Analysis?

Regression analysis is a statistical method used to determine how one or more independent variables affect a dependent variable. The dependent variable (Y) is the outcome that researchers wish to explain or predict. The independent variables (X) are the factors believed to influence that outcome.

For example:

- **Dependent variable:** Employee performance
- **Independent variables:** Training, motivation, leadership, work experience

Regression attempts to answer questions such as:

- Does training improve performance?
- How much does motivation contribute?
- Which predictor has the strongest effect?
- Can performance be predicted from these variables?

Unlike simple comparison tests, regression allows researchers to examine several predictors simultaneously while controlling for the influence of the others.

III. Types of Regression

a. Simple Linear Regression

Simple regression examines the relationship between **one independent variable** and **one dependent variable**.

Example: A researcher wants to determine whether study hours predict exam scores.

- Independent variable: Study hours
- Dependent variable: Exam score

The regression equation is: $Y = a + bX$

where:

- **Y** = predicted value
- **a** = intercept
- **b** = regression coefficient (slope)
- **X** = independent variable

A positive slope indicates that as X increases, Y also increases, while a negative slope indicates an inverse relationship.

---

b. Multiple Regression

Multiple regression examines the effect of **two or more independent variables** on a dependent variable.

Example: A university investigates whether GPA depends on:

- Study hours
- Attendance
- Class participation
- Previous GPA

The regression equation becomes

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$$

This model estimates the unique contribution of each predictor while holding all other variables constant.

IV. Understanding the Regression Equation

Suppose SPSS produces the following equation: $\hat{Y} = 20 + 3X$

This equation means:

- The intercept is 20.
- Every additional unit of X increases Y by 3 units.
- If X increases from 5 to 6, Y is expected to increase by 3.

For example, if a student studies for 10 hours,
Predicted score
= 20 + (3 × 10)
= 50

The regression equation therefore provides predictions as well as information about the strength of relationships.

---

a. Visualizing the Line of Best Fit

$$\hat{y} = b_0 + b_1x$$

The fit is $\hat{y} = 8.49 - 0.54x$; $R^2 = 0.72$.

The least-squares regression line is chosen to minimize the squared vertical distances (residuals) between the observed data points and the fitted line. A steeper positive slope indicates a stronger positive relationship, while a negative slope indicates that the outcome decreases as the predictor increases.

---

b. Regression Coefficients

Regression coefficients indicate the amount of change in the dependent variable associated with a one-unit increase in an independent variable.

There are two commonly reported coefficients.

- **Unstandardized Coefficient (B):** The unstandardized coefficient represents the actual change in the dependent variable.
  - Example: B = 2.8
  - Interpretation: Each additional hour of training increases employee performance by 2.8 points.

- **Standardized Coefficient (Beta):** Beta coefficients are standardized and therefore allow comparison between predictors measured on different scales.

  - Example:

Variable	Beta
Training	0.60
Motivation	0.35
Experience	0.12

Training has the greatest influence because it has the largest standardized Beta coefficient.

---

c. Coefficient of Determination (R^2)

R^2 measures how much variation in the dependent variable is explained by the regression model. The formula is

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}}$$

R^2 ranges between 0 and 1.

Interpretation examples:

R^2	Interpretation
0.20	Model explains 20% of the variation
0.45	Model explains 45% of the variation
0.70	Model explains 70% of the variation
0.90	Excellent explanatory power

Example: $R^2 = 0.58$
Interpretation: The independent variables explain 58% of the variation in employee performance. The remaining 42% is explained by other factors not included in the model.

Adjusted R^2: Adjusted R^2 is an improved version of R^2 for multiple regression.
Unlike R^2, Adjusted R^2 penalizes the inclusion of unnecessary variables that do not improve the model.

Researchers generally report Adjusted R^2 because it provides a more realistic estimate of model performance.

V. Hypothesis Testing in Regression

Regression tests whether the independent variables significantly predict the dependent variable.

The hypotheses are:

- **Null hypothesis (H_0):** The regression coefficient equals zero. There is **no significant relationship**.

- **Alternative hypothesis (H_1):** The regression coefficient is not zero. There **is a significant relationship**.
    - Decision rule:
        - $p < 0.05 \rightarrow$ Reject H_0
        - $p \geq 0.05 \rightarrow$ Fail to reject H_0

ANOVA in Regression

Regression uses Analysis of Variance (ANOVA) to determine whether the regression model as a whole is statistically significant.

The ANOVA table compares:

- Variation explained by the regression model
- Variation due to random error

The F-statistic is calculated as

$$F = \frac{\text{Regression Mean Square}}{\text{Residual Mean Square}}$$

A large F-value combined with $p < 0.05$ indicates that the regression model significantly predicts the dependent variable.

VI. Assumptions of Regression

For regression results to be reliable, several assumptions should be satisfied:

1. The relationship between variables should be linear.
2. Observations should be independent.
3. Residuals should have constant variance (homoscedasticity).
4. Residuals should be approximately normally distributed.
5. Independent variables should not exhibit severe multicollinearity.

6. Extreme outliers should not unduly influence the model.

Violating these assumptions may produce biased or unreliable estimates.

VII. Interpreting SPSS Regression Output

A typical SPSS regression output includes several key tables.

- **Model Summary** reports the strength of the model through R, R^2, Adjusted R^2, and the standard error of the estimate.

- **ANOVA Table** indicates whether the overall regression model is statistically significant using the F-test and its associated p-value.

- **Coefficients Table** provides the regression coefficients (B), standardized coefficients (Beta), t-statistics, significance levels (p-values), and confidence intervals for each predictor. Variables with p-values below the chosen significance level (commonly 0.05) are considered statistically significant contributors to the model.

VIII. Conclusion

- **Advantages of Regression Analysis:** Regression analysis offers several important benefits. It enables researchers to predict future outcomes, quantify relationships between variables, compare the relative importance of predictors, and control for multiple variables simultaneously. Because of these strengths, regression is one of the most frequently used statistical techniques in academic research and professional practice.

- **Limitations of Regression Analysis:** Despite its usefulness, regression has limitations. It identifies statistical associations but does not by itself establish causation. The results depend on the quality of the data and the appropriateness of the selected variables. Violations of regression assumptions, omitted variables, multicollinearity, or influential outliers can reduce the validity and accuracy of the model.

- **Summary:** Regression analysis is a powerful statistical technique used to explain and predict the relationship between variables. Simple regression analyzes one predictor, whereas multiple regression examines the combined effects of several predictors. Researchers evaluate regression models using coefficients, R^2, Adjusted R^2, ANOVA, t-tests, and p-values while ensuring that the underlying assumptions are satisfied. Mastering regression enables students to conduct rigorous empirical research and interpret statistical findings with confidence across disciplines such as business, economics, finance, management, healthcare, and the social sciences.