
Final Project

Regression, ANOVA, and Categorical Data Analysis

Purpose

This project brings together model building, prediction, diagnostics, ANOVA, and chi-square analysis. Use the project workbook posted in WebCampus (Canvas). It contains the worksheets **Sales**, **Supermarket**, **Restaurant**, **Savings**, and **Magazine Preference**.

Part 1: Product Sales

The **Sales** worksheet contains monthly sales, advertising expenditure, number of competing products, and warranty period.

1. Identify the response and predictors. Explain the role of each variable.
2. Fit a multiple linear regression model for sales. Write the fitted equation and interpret each slope.
3. Evaluate overall fit, individual predictors, assumptions, multicollinearity, and influential observations.
4. Predict sales when advertising expenditure is \$400, the number of competitors is 1, and the warranty period is 6 months. Report a point prediction and prediction interval. State whether this predictor combination is within the scope of the data.

Part 2: Supermarket Profit

The **Supermarket** worksheet contains profit, food sales, nonfood sales, and store size. Treat profit as the response.

1. Fit and interpret a multiple linear regression model using all three predictors.
2. Check the model assumptions and identify any influential observations.
3. Predict profit when food sales are 300, nonfood sales are 80, and store size is 30. Report a prediction interval.
4. A supermarket with those predictor values reports profit of 24. Use the model and prediction interval to assess whether the value is unusual.
5. Calculate variance inflation factors. If multicollinearity is a concern, propose and defend a better model. Compare it with the original.

Part 3: Restaurant Promotions

The **Restaurant** worksheet contains sales under no promotion, a Sunday newspaper flyer, a radio advertisement, and discount coupons.

1. Identify the experimental or observational design from the workbook structure. Select an ANOVA method that matches it and justify the choice.
2. State the null and alternative hypotheses. Test at $\alpha = 0.05$.
3. If the overall result is significant, conduct appropriate follow-up comparisons and identify which promotions differ from no promotion.
4. Check assumptions and write a recommendation that considers both statistical evidence and the observed size of the sales differences.

Part 4: Savings During Wartime and Peacetime

The **Savings** worksheet contains personal savings and personal income for 1935 through 1949, with years identified as wartime or peacetime.

1. Create a scatterplot of savings against income. Use different symbols or colors for wartime and peacetime.
2. Create an indicator variable coded 0 for peacetime and 1 for wartime. Fit a regression model with income and the indicator as predictors.
3. Add the two fitted lines to the scatterplot and interpret the wartime coefficient.
4. Fit a second model with an income-by-wartime interaction. Test whether the slopes differ, then select and defend the more appropriate model.
5. Check assumptions and discuss the limits of drawing conclusions from 1935 through 1949 data.

Part 5: Magazine Preference and Age

The **Magazine Preference** worksheet contains magazine preference and age group for 510 respondents.

1. Create and display the contingency table with row or column percentages that help answer the question.
2. Test whether magazine preference is independent of age group at $\alpha = 0.05$.
3. Check the expected-count condition. If it fails, explain and use a defensible alternative.
4. Identify the cells that contribute most to the chi-square statistic and explain the pattern in plain language.

Final Management Summary

End the report with a one-page summary of the most useful findings from the five analyses. Write for a manager who understands the business context but does not need to see software commands.

Submission Checklist

Submit through WebCampus (Canvas) by the deadline posted there. Submit both of the following files:

- a rendered PDF report; and
- the complete `.Rmd` source file used to produce the report.

The PDF must include your name and assignment title, answer every question in the order presented, show readable and labeled tables and figures, and explain results in complete sentences. The `.Rmd` file must contain all code needed to reproduce the report. Both files must run from the same data supplied for the assignment. Do not submit unexplained software output. You are responsible for checking all results and explaining every choice you made.

A Note About AI

AI can be used as a resource, but it cannot do the assignment for you. I do not want to read AI slop. Do not paste pages of generated writing, code, or interpretations into your report and expect me to sort through it. In the real world, you cannot send AI garbage to a boss, client, or customer. You should not submit it in this class either.

If you use AI, use it to help you learn. Check what it gives you. Make sure the code works. Internalize the material, interpret it, and then explain it in your own words and voice. You should understand every part of your submission and be able to defend the choices you made. Include a short note at the end of the report that names the tool and tells me how you used it. If the work reads like unreviewed AI output, it will receive a 0%, and you will have the opportunity to redo it under the policy in the syllabus.

How Your Work Will Be Evaluated

Credit is based on the statistical reasoning, correct use of R, appropriate checks of assumptions, quality of visual and numerical evidence, and accuracy of the interpretation. A correct number without an explanation is not a complete answer.