ESMA 6205: Homework 2

Alejandro Ouslan

Table of contents

1 Introduction

Homework should be presented in a clear manner, not scratch notes paper will be allow. The home- work solution format will be .docx, .doc and .pdf. You can also add your code as part of the supple- mental material. You must include (if any) your code, but do not include a screenshot of the software input/output. Outputs may contain irrelevant information for the purpose of the question and can lead to a wrong interpretation. To be consider for full credit it will need to submitted before the due date.

2 Definition:

- **Sampling distribution of $\hat{\beta}$:**
- **Confidence interval:**
- **Predidiction interval:**

- **Hupothesis testing:**
- **Properties of the F-test:**

3 Problem 2

A member of a student team playing an interactive marketing game received the following computer output when studying the relation between advertising expenditures (X) and sales (Y) for one of the team's products - Estimated regression equation $\hat{y} = 350.7 - 0.18X$ - Two-sided p-value for estimated slope: 0.91

The student stated: "The message I get here is that the more we spend on advertising this product, the few units we sell!" Comment.

4 Problem 3

The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The dataset for this study is in call `CH01PR19.txt` you can download it from the authors website. Assume that the first-order regression model is appropriate.

- Obtain a 99 percent confidence interval for β_1. Interpret your confidence interval. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?
- Test, using the test statiistic t^*, whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshmn year (Y). Use a level of significance of $\alpha 0.01$ State the alternative, decision rule, and conclusion.
- What is the p-value of your test in part(b)? How does it support the conclusion reached in part(b)?
- Set up the ANOVA table.
- What is estimated by MS_{model} in you ANOVA table? by MS_{error}?
- Conduct an F test of whether or not $\beta_1 = 0$. State the alternative, decision rule, and conclusion.
- What is the absolute magnitude of the reduciton in the variation of Y when X is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?
- Obtain r and attach the appropriate sign.
- Which measure, R^2 or r, has the more clear-cut operation interpretation? Explain.

5 Problem 4

The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data `CH01PR20.txt` have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person.

- Obtain the estimated regression function.
- Plot the estimated regression function and the data. How well does the estiimated regression fucntion fit the data?
- Interpret $\hat{\beta}_0$ in your estimated regression function. Does $\hat{\beta}_0$ provide any relevant information here? Explain?
- Obtain a point estimate of the mean service time when $X = 5$ copiers are serviced.
- Estimate the change in the mean service time when the number of copiers serviced increased by one. Use a 90 percent confidence interval. interpret you confidence interval.
- Conduct a t test to determine whether or not there is a linear association between X and Y here; $\alpha = 0.1$. State the alternative, decision rule, and conclusion. What is the p-value of your test?
- Are your results in parts (e) and (f) consistent? Explain.
- The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Assume that $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. What is the p-value of the test?
- Does $\hat{\beta}_0$ give any relevant information here about the "start-up" time on calls-Le., about the time required before service work is begun on the copiers at a customer location?

6 Problem 5

For each of the following questions, explain whether a confidence interval for a mean response or a prediction interval for a new observation is appropriate

- What will be the humidity level in this greenhouse tomorrow when we set the temperature level at $30^{\circ}C$?
- How much do families whose disposable income is \$23,500 spend, on the average, for meals away from home
- How many kilowatt-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, given that the index of business activity for the area remains at its present level?

7 Problem 6

A criminologist studying the relationship between level of education-and crime rate in medium-sized U.S. counties collected the data `CH01PR28.txt` for a random sample of 84 counties; X is the percentage of individuals in the county having at least a high-school diploma, and Y is the crime rate (crimes reported per $100,000$ residents) last year. Assume that the normal bivariate model is appropriate.

- Set up the ANOVA table.
- Test whether or not there is a linear association between crime rate and percentage of high school graduates, by the F test. Is the p-value for the F test the same as that for the t test?
- By how much is the total variation in crime rate reduced when percentage of high school graduates is introduced into the analysis? Is this a relatively large or small reduction?
- Compute the Pearson product-moment correlation coefficient r.
- est whether crime rate and percentage of high school graduates are statistically independent in the population; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion.
- How do your estimates and conclusions in part (d) and (e) compare to those obtained in parts (a) and (b) respectively?
- Compute the Spearman rank correlation coefficient r_s.
- Test by means of the Spearman rank correlation coefficient whether an association exists between crime rate and percentage of high school graduates using test statistic and a level of significance 0.01. State the alternatives, decision rule, and conclusion.
- How do your estimates and conclusions in parts (g) and (h) compare to those obtained in parts

  (d) and (e), respectively?

- Computer the Kendall's tau correlation coefficient $\hat{\tau}$.
- Test by means of the Kendall's tau, correlation coefficient whether an association exists between crime rate and percentage of high school graduates using test statistic and a level of significance 0.01. State the alternatives, decision rule, and conclusion.
- ow do your estimates and conclusions in parts (j) and (k) compare to those obtained in parts (d), (e), (g), and (h) respectively?