 Sardar Patel Institute of Technology,Mumbai

Department of Electronics and Telecommunication Engineering

T.E. Sem-V (2018-2019)

ETL54-Statistical Computational Laboratory

**Lab-5: Analysis of Covariance (Part-I)**

**Name:Yash Patil Roll No.2016120036**

**Objective:**To carry out analysis of variance and compare with regression model

**Outcomes:**

To carry out analysis of (co)variance

To compare ANOVA with linear regression model

To use ANCOVA for significance of models

To evaluate the linear models

To extract information from models

**System Requirements:** Ubuntu OS with R and RStudio installed

**Introduction to Analysis of Variances:**

To build linear models we use regression analysis which describe the effect of variation in predictor variables on the response variable. Sometimes, if we have a categorical variable with values like Yes/No or Male/Female etc. The simple regression analysis gives multiple results for each value of the categorical variable. In such scenario, we can study the effect of the categorical variable by using it along with the predictor variable and comparing the regression lines for each level of the categorical variable. Such an analysis is termed as Analysis of Covariance also called as ANCOVA. [1] [2][3]

**Procedure:**

**Part-I: Motor Trend Car Road Tests [mtcars dataset]**

Read the online tutorial on analysis of variance.

[1] <https://www.tutorialspoint.com/r/r_analysis_of_covariance.htm>

Perform the steps given and write interpretation in your own words.

> input <- mtcars[,c("am","mpg","hp")]

> print(head(input))

am mpg hp

Mazda RX4 1 21.0 110

Mazda RX4 Wag 1 21.0 110

Datsun 710 1 22.8 93

Hornet 4 Drive 0 21.4 110

Hornet Sportabout 0 18.7 175

Valiant 0 18.1 105

> input=mtcars

> result=aov(mpg~hp\*am,data=input)

> print(summary(result))

Df Sum Sq Mean Sq F value Pr(>F)

hp 1 678.4 678.4 77.391 1.50e-09 \*\*\*

am 1 202.2 202.2 23.072 4.75e-05 \*\*\*

hp:am 1 0.0 0.0 0.001 0.981

Residuals 28 245.4 8.8

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> result=aov(mpg~hp+am,data=input)

> print(summary(result))

Df Sum Sq Mean Sq F value Pr(>F)

hp 1 678.4 678.4 80.15 7.63e-10 \*\*\*

am 1 202.2 202.2 23.89 3.46e-05 \*\*\*

Residuals 29 245.4 8.5

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> input=mtcars

> result1=aov(mpg!hp\*am,data=input)

Error: unexpected '!' in "result1=aov(mpg!"

> result1=aov(mpg~hp\*am,data=input)

> result2=aov(mpg~hp+am,data=input)

> print(anova(result1,result2))

Analysis of Variance Table

Model 1: mpg ~ hp \* am

Model 2: mpg ~ hp + am

Res.Df RSS Df Sum of Sq F Pr(>F)

1 28 245.43

2 29 245.44 -1 -0.0052515 6e-04 0.9806

**Part-II: Agriculture Experiment: Effectiveness of Insect Sprays**

[2] R for Dummies-Second Edition by Andrie de Vries and Joris Meys

Read the Chapter-15 of R for Dummies- Pages 313 to 325

Perform the steps given in the book and complete the part-II

**Analyzing variances**

An analysis of variance (ANOVA) is a very common technique used to com-

pare the means between different groups. To illustrate this, take a look at the

**Code:**

dataset InsectSpray:

> str(InsectSprays)

'data.frame': 72 obs. of 2 variables:

$ count: num 10 7 20 14 14 12 10 23 17 20 . . .

$ spray: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 . . .

This dataset contains the results of an agricultural experiment. Six insecti-

cides were tested on 12 fields each, and the researchers counted the number

of pesky bugs that remained on each field. Now the farmers need to know if

the insecticides make any difference, and if so, which one they should use.

You answer this question by using the aov() function to perform an ANOVA.

**Building the model**

For this simple example, building the model is a piece of cake. You essentially

want to model the means for the variable count as a function of the variable

spray. You translate that to R like this:

**Code:**

**> AOVModel <- aov(count ~ spray, data = InsectSprays)**

You pass two arguments to the aov() function in this line of code:

✓ The formula count ~ spray, which reads as “count as a function of

spray”

✓ The argument data, where you specify the data frame in which the

variables in the formula can be found.

Every modeling function returns a model object with a lot of information

about the fitted model. Always put this model object in a variable. This

way you don’t have to refit the model when you need to perform extra

calculations.

**Looking at the object**

As with every object, you can look at a model object just by typing its name

in the console. If you do that for the object Model that you created in the

preceding section, you see the following

**Output:**

> AOVModel

Call:

aov(formula = count ~ spray, data = InsectSprays)

Terms:

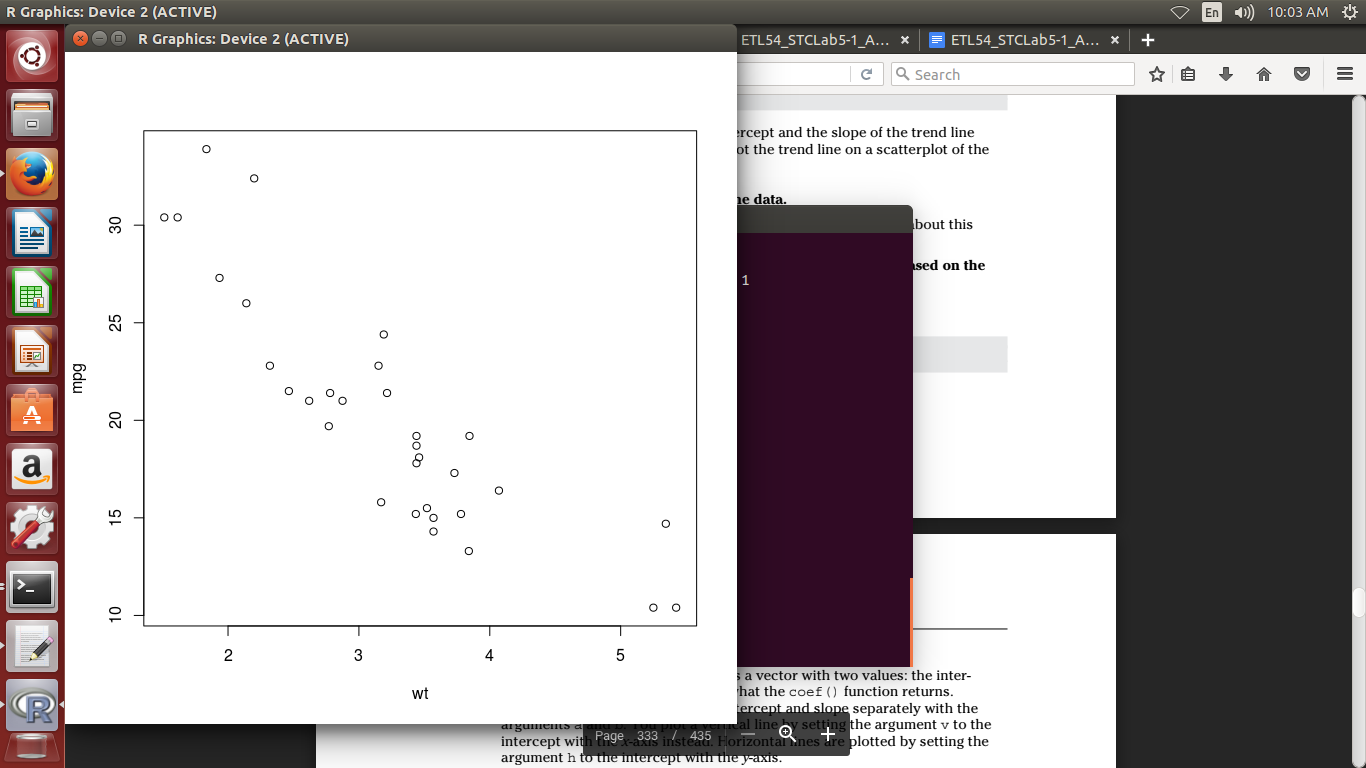
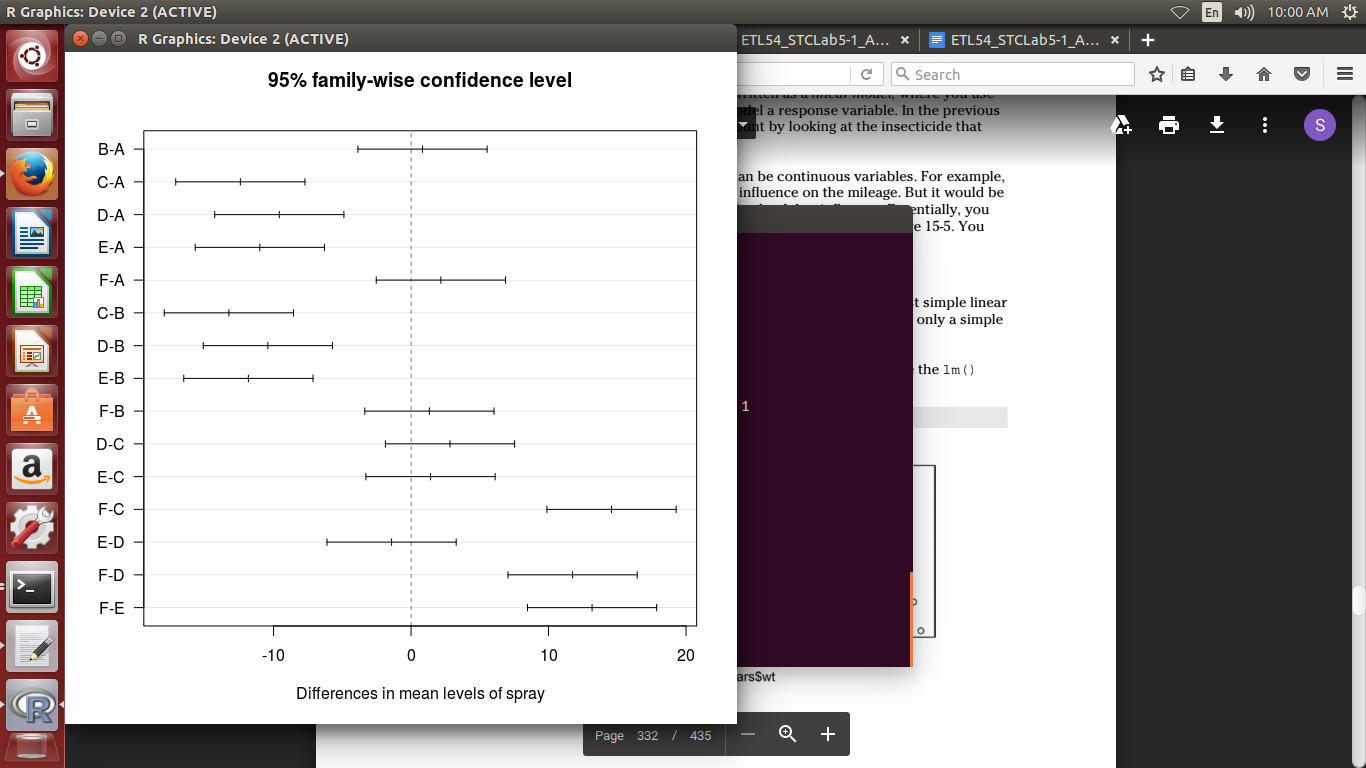
spray Residuals

Sum of Squares 2668.833 1015.167

Deg. of Freedom 5 66

Residual standard error: 3.921902

Estimated effects may be unbalanced



This doesn’t tell you that much, apart from the command (or the call) you

used to build the model and some basic information on the fitting result.

In the output, you also read that the estimated effects may be unbalanced.

This isn’t a warning— it’s a message that’s

built in by the author of the aov() function. This one can pop up in two

situations:

✓ You don’t have the same number of cases in every group.

✓ You didn’t set orthogonal contrasts.

In this case, it’s the second reason. You can continue with this model as we

do now (that’s also how those models are fitted in SPSS and SAS by default),

or you can read the nearby sidebar, “Setting the contrasts,” and use contrasts

as the statistical experts who wrote R think you should.

## **Describe the following with respect to Analysis of Variance and linear models**

What is analysis of variance?

Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among group means in a sample. ANOVA was developed by statistician and evolutionary biologist Ronald Fisher. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether the population means of several groups are equal, and therefore generalizes the *t*-test to more than two groups. ANOVA is useful for comparing (testing) three or more group means for statistical significance. It is conceptually similar to multiple two-sample t-tests, but is more conservative, resulting in fewer type I errors, and is therefore suited to a wide range of practical problems.

How ANCOVA used test the a linear Models?

**F-test:** The F-test of significance is used to test each main and interaction effect, for the case of a single interval dependent and multiple (>2) groups formed by a categorical independent. F is between-groups variance divided by within-groups variance. If the computed p-value is small, then significant relationships exist.

Adjusted means are usually part of ANCOVA output and are examined if the F-test demonstrates significant relationships exist. Comparison of the original and adjusted group means can provide insight into the role of the covariates. For k groups formed by categories of the categorical independents and measured on the dependent variable, the adjustment shows how these k means were altered to control for the covariates. Typically, this adjustment is one of linear regression of the type: Yadj.mean = Ymean – b\*(Xith.mean-Xmean), where Y is the interval dependent, X is the covariate, i is one of the k groups, and b is the regression coefficient. There is no constant when Y is standardized. For multiple covariates, of course, there are additional similar X terms in the equation.

**t-test**: A test of significance of the difference in the means of a single interval dependent, for the case of two groups formed by a categorical independent.

How to extract information from models?

1. Create variables for difference in date, time and addresses
2. Create new ratios and proportions.
3. Apply standard transformations
4. Include effect of influencer
5. Check variables for seasonality and create the model for right period

**Conclusion:**

1. Created linear regression models.
2. Derived the significance of the model using **cov(Analysis of Covariance)**
3. Compared two models using **anova**.

References:

[1] <https://www.tutorialspoint.com/r/r_analysis_of_covariance.htm>

[2] R for Dummies-Second Edition by Andrie de Vries and Joris Meys

[3] http://onlinestatbook.com/2/analysis\_of\_variance/intro.html