***ANOVA***

***(Analysis of Variance)***

***Gauri Mahadik***

***PGP-DSBA May 2021***

Table of Contents:

Contents:

Problem 1……………………………………………………………………………………………… 1

Sample dataset ………………………………………………………………………………………...1 Information about the data………………………………………………………………………………1

Que1.1: State the null and the alternate hypothesis for conducting one-way ANOVA for both Education and Occupation individually……………………………………………………………….. 2 Hypothesis for Education ……………………………………………………………………………...2 Hypothesis for Occupation………………………………………………………………………………2

Que1.2: Perform one-way ANOVA for Education with respect to the variable ‘Salary’. State whether the null hypothesis is accepted or rejected based on the ANOVA results. Table 1 ………………………………………………………………………………………………....2 Fig. 1 ……………………………………………………………………………………………………2

Que1.3: Perform one-way ANOVA for variable Occupation with respect to the variable ‘Salary’. State whether the null hypothesis is accepted or rejected based on the ANOVA result…………3 Table 2 …………………………………………………………………………………………………3 Fig 2 …………………………………………………………………………………………………3

Que1.4: If the null hypothesis is rejected in either (1.2) or in (1.3), find out which class means are significantly different. Interpret the result. ……………………………………………………….4 Fig.3 ………………………………………………………………………………………………….4

Que1.5: What is the interaction between the two treatments? Analyze the effects of one variable on the other (Education and Occupation) with the help of an interaction plot. …………………..5 Table 3 …………………………………………………………………………………..5 Fig 4 Interaction plot ………………………………………………………………………………….5

Observations ………………………………………………………………………………..6

Que1.6: Perform a two-way ANOVA based on the Education and Occupation (along with their interaction Education\*Occupation) with the variable ‘Salary’. State the null and alternative hypotheses and state your results. How will you interpret this result? …………………………6

Null Hypothesis ………………………………………………………………………………6 Alternate Hypothesis ………………………………………………………………………………..6

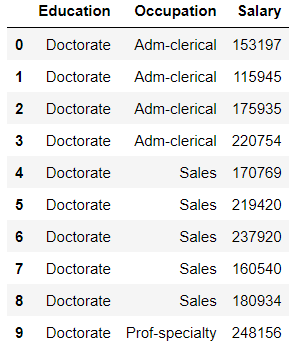
Table 4 …………………………………………………………………………………………

**Problem 1:**

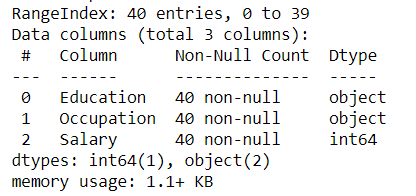
Salary is hypothesized to depend on educational qualification and occupation. To understand the dependency, the salaries of 40 individuals [[SalaryData.csv](https://olympus.greatlearning.in/courses/49115/files/3505183/download?verifier=sezP94lkiZL92J3w4IijaGARLdok1amDbp7LowAC&wrap=1)] are collected and each person’s educational qualification and occupation are noted. Educational qualification is at three levels, High school graduate, Bachelor, and Doctorate. Occupation is at four levels, Administrative and clerical, Sales, Professional or specialty, and Executive or managerial. A different number of observations are in each level of education – occupation combination.

 [Assume that the data follows a normal distribution. In reality, the normality assumption may not always hold if the sample size is small.]

Let us have a look at our **sample data set:**



Also getting some **information about the data:**



From the above information we can see that the dataset contains of total 40 entries with 3 columns. Dataset has variables with 1 integer and 2 objects.

**Que1.1: State the null and the alternate hypothesis for conducting one-way ANOVA for both Education and Occupation individually.**

**Hypothesis for Education:**

H0 (Null Hypothesis) - Mean Salary is equal for all types of Educational Qualification.

HA (Alternate Hypothesis) - Mean salary is different for at least one Educational qualification.

**Hypothesis for Occupation:**

H0 (Null Hypothesis) - Mean Salary is equal for all types of Occupations.

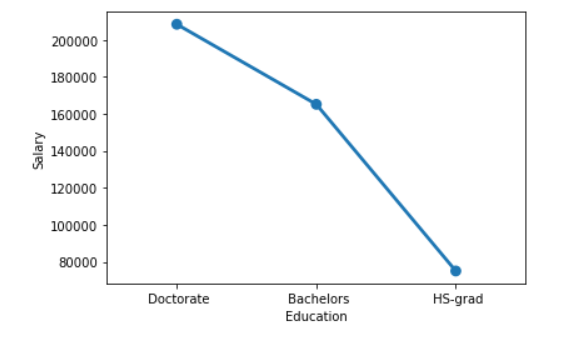
HA (Alternate Hypothesis) - Mean salary is different for at least for any one Occupation.

**Que1.2: Perform one-way ANOVA for Education with respect to the variable ‘Salary’. State whether the null hypothesis is accepted or rejected based on the ANOVA results.**

|  |
| --- |
|  |

**Table 1**

We can also have plot for the same:



**Fig.1**

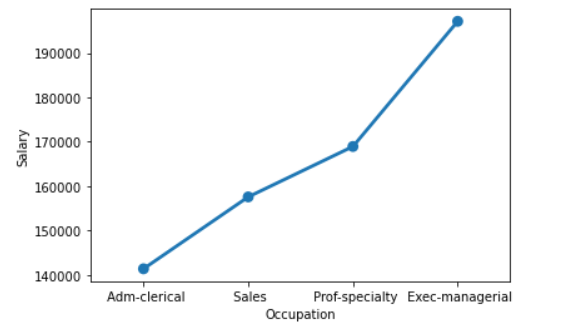
Since we can see that the **p-value i.e.PR (>F) is less than the significance level (0.05), we can reject the null hypothesis** and state that there is a difference in the mean salary at least for one any Educational qualification.

**Que1.3: Perform one-way ANOVA for variable Occupation with respect to the variable ‘Salary’. State whether the null hypothesis is accepted or rejected based on the ANOVA results.**

|  |
| --- |
|  |

**Table 2**

We can also have plot for the same:



**Fig. 2**

Since the **p-value i.e. PR (>F) is greater than the significance level (0.05)**, we have no evidence that mean salary is different for at least one of the Occupations. That means we have to accept the null Hypothesis.

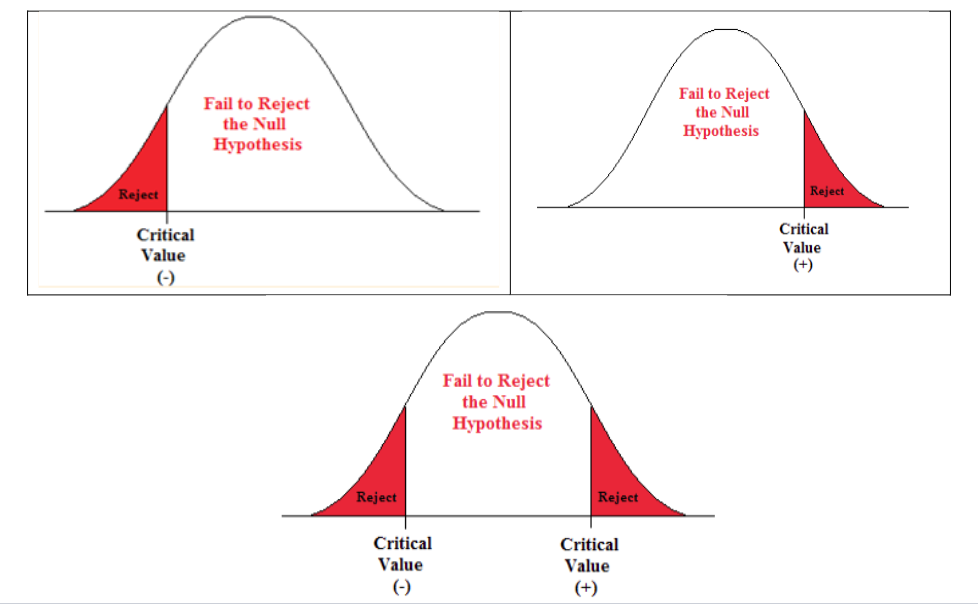
**Que1.4: If the null hypothesis is rejected in either (1.2) or in (1.3), find out which class means are significantly different. Interpret the result.**

Null Hypothesis is rejected in the above problem (1.2) so we can conclude that the alternative hypothesis is true and there is **not enough statistical evidence to infer that the null hypothesis is true. The t-statistic falls in the rejection region.**

And in problem (1.3) we have no evidence to reject null hypothesis, so we conclude that that there is **no enough statistical evidence to infer that null hypothesis is rejected** and that the alternate is true**. Here, the t-statistic falls in the acceptance region.**

Education means is significantly different. This is because the P-value here if too very small then the significance level (i.e. 0.05)

Following figures can make it more clear on when to reject the null hypothesis and when to accept the null hypothesis.



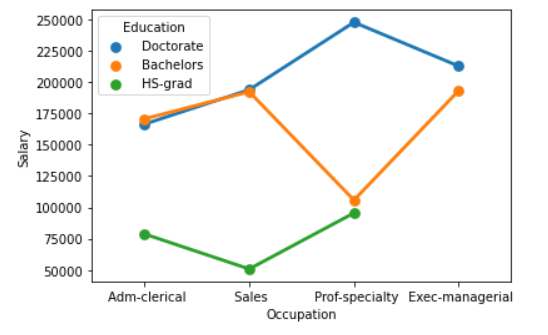
**Fig.** **3**

**Que1.5: What is the interaction between the two treatments? Analyze the effects of one variable on the other (Education and Occupation) with the help of an interaction plot.**

|  |
| --- |
|  |

**Table 3**

**Interaction Plot:**



**Fig. 4**

**Observations:**

From the above interaction plot between Educational Qualification and Occupation with respect to the Salary we can conclude following:

\* Salary even for the occupation Adm-Clerical is higher for the Doctorate Education. Though we can see it decreasing for Exec-managerial.

\* Salary for the Bachelors also can be seen on higher side for Adm\_clerical but can see considerable drop in salary for Prof-specialty and again on increasing side for Exec-managerial.

\* While salary for HS-grad is seen to be quite low for Adm-clerical, sales and increasing for Prof-specialty. Also we can see that there are no jobs for Hs-grad in Exec-managerial.

**Que1.6: Perform a two-way ANOVA based on the Education and Occupation (along with their interaction Education\*Occupation) with the variable ‘Salary’. State the null and alternative hypotheses and state your results. How will you interpret this result?**

**H (0): Null Hypothesis -** Means of salary are same for all types of educational qualifications with respect to the occupation.

**H (A): Alternate Hypothesis -** At least one type of educational qualification with respect to the occupation does not have the similar salary mean.

|  |
| --- |
|  |

**Table 4**