Analysis of the Impact of Undergraduate and Master Degree on Salary
Analysis of Variance

Anastasia Nica m20210516| Bruna Ribeiro m20210226 | Maria Machado m20210371
Nova Information Management School

Introduction

The group decided to analyze the effected of the undergraduate's degree and master's specialization in the salary of the recent employees. The purpose of this investigation is to understand how the choices of the students regarding the areas of their studies affect their expected salary. For this, it was used data from a university, and it is intended to give some advices to their students.
Given this goal, we defined our research questions:

- Are there differences in terms of expected salary according to undergraduate's degree? And according the master's specialization?
- Is there evidence of an interaction between both factors?
- If there are differences, what is the proposed path that leads to the higher expected salary?

Materials

First, it was needed to clean the data so, we decided to cut some variables that are not meaningful for our research.

After that, it's important to clarify the description of the variables presents in the model.

Our experimental units are the students of the university and the measurement unit is the unit on which the salary is measured.

Firstly, we have the Factors:

Variable	Type	Description	Levels
Undergraduate degree	Categorical	Undergraduate degree – Field of degree education	Comm&Mgmt; Sci&Tech; Others
Master Specialization	Categorical	Master Specialization	Mkt&Fin; Mkt&HR

Table 1: Factors

Then, we have the response:

Variable	Type	Description	Values
Salary	Numeric	Salary of the former student	[200000,425000]

Table 2: Response variable

Our final database is as follows:

Undergraduate degree	Master Specialization	
	Mkt&Fin	Mkt&HR
Comm&Mgmt	250000	275000
	425000	265000
Sci&Tech	200000	270000
	252000	360000
Others	360000	250000
	240000	252000

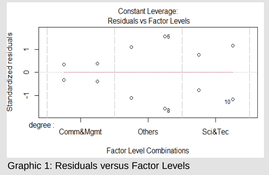
Methodology

The results of the tests were obtained trough codes in the R Studio.

We decided to use the Two Factor Design because in this design all possible combinations of the levels of the factors are investigated.

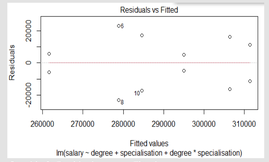
We should guarantee two principles of experimental design:

- Randomization - it is possible to conclude that the data is randomized because:



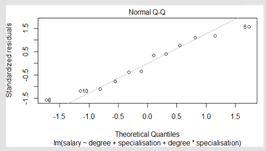
Graphic 1: Residuals versus Factor Levels

The variability of the residuals is constant between the factor levels.



Graphic 2: Residuals versus Fitted

The red line remains constant over the graph. So, the variability remains constant as the fitted values increase.



Graphic 3: Normal Q-Q

Through the Normal Probability Plot of Residuals, it's possible justify the normality assumption. The points are closer to the line so we assume normality.

- Replication – this principle is verified because our data has 2 independent repeats run of each factor combination.

In this case we have 2 factors, A and B, respectively, degree and specialization.
n = 2 replicates
a = 3 levels
b = 2 levels

The effects model is:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1,2,3 \\ j = 1,2 \\ k = 1,2 \end{cases}$$

The formal test of the hypothesis of no differences in the degree treatment, specialisation treatment and in the interaction term is given by:

Degree:

$H_0: \tau_1 = \tau_2 = \tau_3 = 0$

$H_1: \exists\, \tau_i \neq 0$, i = 1, 2, 3

Specialization:

$H_0: \beta_1 = \beta_2 = 0$

$H_1: \exists\, \beta_j \neq 0$, j = 1, 2

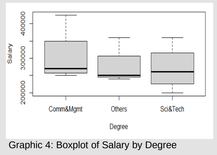
Interaction:

$H_0: \exists\, (\tau\beta)_{ij} = 0$, i = 1, 2, 3 and j = 1, 2

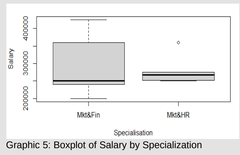
$H_1: \exists\, (\tau\beta)_{ij} \neq 0$, i = 1, 2, 3 and j = 1, 2

Boxplots gives an uncluttered view of the mean, the 25% and 75% quartile and indicates the presence of unusually small or large outlying values.

The analysis of boxplots is important to get an idea of the expected results obtained through the ANOVA.



Graphic 4: Boxplot of Salary by Degree



Graphic 5: Boxplot of Salary by Specialization

As we can see in graphic 4, the mean salary is very close in all courses, which mean that the means salaries don't differ much considering the different degrees.

Taking into account the specialization the mean salary is very close too. But in this case, we have an outlier in Mkt&HR, which means that exists one student in this specialization that stands out from the others.

Analysis/ Results

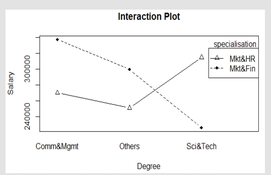
To test what factors have differences in treatment means we made a ANOVA test.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree_t	2	2.5715e+09	1285750000	0.2758	0.7681
specialisation	1	2.5208e+08	252083333	0.0541	0.8238
degree_t:specialisation	2	1.4626e+10	7313083333	1.5690	0.2831
Residuals	6	2.7967e+10	4661083333		

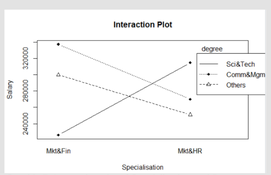
Output 1: ANOVA test

By this output, we can conclude that we do not reject H_0 for any of the factors or the interaction, at a significance level of 5%, and conclude that the degree, the specialization and the interaction term does not influence the salary.

We also made the interaction plots in order to understand the interaction effects.



Graphic 6: Interaction plot of Degree and Specialization



Graphic 7: Interaction plot of Specialization and Degree

Although ANOVA test gives us information that there are no significant differences in the means salary, by these graphics we observe that exists residual differences. The highest difference is between students with a degree in Comm&Mgmt and specialization in Mkt&Fin, and students with a degree in Sci&Tech and specialization in Mkt&Fin. The smallest difference is between students with a degree in Sci&Tech and a specialization in Mkt&HR and those with a degree in Others and a specialization in Mkt&Fin.

By the same rationale, we use the Fisher´s LSD method to compare all pairs of treatment means in order to confirm the conclusions that we observe in the interaction plots.

```
$degree_t
                    diff    lwr.ci    upr.ci   pval
Others-Comm&Mgmt   -28250 -146376.3  89876.34 0.5798
Sci&Tech-Comm&Mgmt -33250 -151376.3  84876.34 0.5167
Sci&Tech-Others     -5000 -123126.3 113126.34 0.9209

$specialisation
                    diff    lwr.ci   upr.ci   pval
Mkt&HR-Mkt&Fin -9166.667 -105616.4 87283.09 0.8238

$`degree_t:specialisation`
                                          diff     lwr.ci    upr.ci   pval
Others:Mkt&Fin-Comm&Mgmt:Mkt&Fin       -37500 -204555.88 129555.88 0.6026
Sci&Tech:Mkt&Fin-Comm&Mgmt:Mkt&Fin    -111500 -278555.88  55555.88 0.1536
Comm&Mgmt:Mkt&HR-Comm&Mgmt:Mkt&Fin     -67500 -234555.88  99555.88 0.3610
Others:Mkt&HR-Comm&Mgmt:Mkt&Fin        -86500 -253555.88  80555.88 0.2521
Sci&Tech:Mkt&HR-Comm&Mgmt:Mkt&Fin      -22500 -189555.88 144555.88 0.7529
Sci&Tech:Mkt&Fin-Others:Mkt&Fin        -74000 -241055.88  93055.88 0.3200
Comm&Mgmt:Mkt&HR-Others:Mkt&Fin        -30000 -197055.88 137055.88 0.6757
Others:Mkt&HR-Others:Mkt&Fin           -49000 -216055.88 118055.88 0.4999
Sci&Tech:Mkt&HR-Others:Mkt&Fin          15000 -152055.88 182055.88 0.8334
Comm&Mgmt:Mkt&HR-Sci&Tech:Mkt&Fin       44000 -123055.88 211055.88 0.5431
Others:Mkt&HR-Sci&Tech:Mkt&Fin          25000 -142055.88 192055.88 0.7268
Sci&Tech:Mkt&HR-Sci&Tech:Mkt&Fin        89000  -78055.88 256055.88 0.2401
Others:Mkt&HR-Comm&Mgmt:Mkt&HR        -19000 -186055.88 148055.88 0.7901
Sci&Tech:Mkt&HR-Comm&Mgmt:Mkt&HR       45000 -122055.88 212055.88 0.5343
Sci&Tech:Mkt&HR-Others:Mkt&HR          64000 -103055.88 231055.88 0.3847
```

Output 2: Fisher LSD test

By this output, we see that the lower p-value is 0.1536, which corresponds to the biggest difference in the mean salary found before.

We also observed that the highest p-value is 0.8334, which matches to the smallest difference, that is almost neglected.

Conclusions

In conclusion, it was used a two-factor design to test whether undergraduate's degree and master's specialization impact the salary of the recent employees.

It was concluded that these factors do not affect the response variable, neither exits evidence of a interaction between them.

Recommendations

According to our main findings we suggest extending the analysis to other factors that can have an impact on the salary.

Nevertheless, at this point we recommend the students to follow the areas that they consider to be best at, because such choice does not lead to significant differences in the salary.

References

Database:https://www.kaggle.com/benroshan/factors-affecting-campus-placement
Book: Montgomery, Douglas C., 2017. Design and analysis of experiments. 9th ed. Wiley.