Two-Way ANOVA

Srikar

2021-08-18

Objective:

To check whether there is equal gender pay and equal pay depending on the subjects chosen

Data Description:

The dataset contains two categorical factors: Gender and College major. The three majors are statistics, psychology, and political science and the two genders are male and female. The combination of these two factors (2 genders X 3 majors) produces the following six groups. Each group contains 20 observations.
The data represents the salary earned (in Rupees) with repect to the gender and college major taken. The independent variables are gender and college major with dependt variable (Y) as salary

Important terminologies:

1) Two-way Analysis- A two-way ANOVA test is a statistical test used to determine the effect of two nominal predictor variables on a continuous outcome variable. A two-way ANOVA tests the effect of two independent variables on a dependent variable

2) Post Hoc Test- Post hoc tests are an integral part of ANOVA. When you use ANOVA to test the equality of at least three group means, statistically significant results indicate that not all of the group means are equal. However, ANOVA results do not identify which particular differences between pairs of means are significant. Use post hoc tests to explore differences between multiple group means while controlling the experiment-wise error rate.

```
#Data Summary
library(readxl)
dat <- read_excel("data.xlsx", sheet = "2 way for Sub-Sal")
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
head(dat)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 6 x 3
##   Gender Subject Salary
##   <chr>  <chr>    <dbl>
## 1 Male   Stats   78505.
## 2 Male   Stats   76269.
## 3 Male   Stats   66658.
## 4 Male   Stats   78026.
## 5 Male   Stats   83485.
## 6 Male   Stats   70941.
```

```
summary(dat)
```

```
##     Gender            Subject             Salary
##  Length:40          Length:40          Min.   :66658
##  Class :character   Class :character   1st Qu.:71631
##  Mode  :character   Mode  :character   Median :76119
##                                        Mean   :75909
##                                        3rd Qu.:79772
##                                        Max.   :86084
```

From the table we can observe that Gender and Subject are character type and salary is integer. The minimum salary of all students is Rs.66,658 and maximum salary is Rs. 86084. The average salary is Rs. 75909

Hypothesis Testing

1) Hypothesis with respect to gender

Null Hypothesis (Ho): *μ 1 =μ 2 (#Gender pay is equal)*
Alternative Hypothesis (H1)*: μ 1 ≠ μ 2 (Unequal pay between genders)*

2) Hypothesis with respect to subject

Null Hypothesis (Ho)*: μ 1 =μ 2 =μ 3 (Average pay accross all fields are same)*
Alternative Hypothesis (H1): *μ 1 ≠μ 2 ≠μ 3 (Average pay across all fields are not same)*

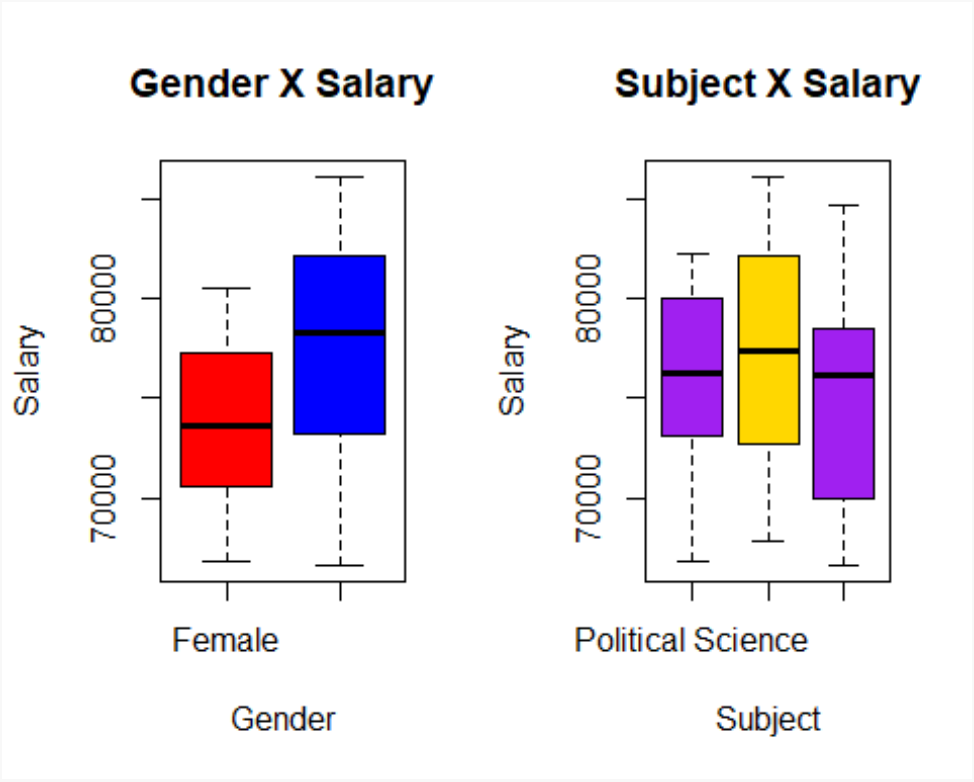
Procedure

1) Converting the factor variables as factors

```
gen=as.factor(dat$Gender)
sub=as.factor(dat$Subject)
```

#2) Visualizing the data

```
par(mfrow=c(1,2))
boxplot(dat$Salary~dat$Gender,xlab = 'Gender',ylab =
'Salary',col=c('red','blue'),main="Gender X Salary")
boxplot(dat$Salary~dat$Subject,xlab = 'Subject',ylab =
'Salary',col=c('purple','gold'),main="Subject X Salary")
```



Using the boxplot, we can visualize our summary statistics and have a general overview of our data

3)Creating the ANOVA model

```
model=aov(dat$Salary~dat$Subject+dat$Gender)
summary(model)
```

```
##             Df    Sum Sq   Mean Sq F value P-Value
## dat$Subject  2  30420618  15210309   0.624 0.5413
## dat$Gender   1 134592790 134592790   5.525 0.0243 *
## Residuals   36 877020315  24361675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of subject w.r.t to salary is 0.5413
The p-value of gender w.r.t to salary is 0.2043

#4) Considering the factor,'gender' as it shows significance

```
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 3.6.3
```

```
## Loading required package: emmeans
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```
lm1=lm(dat$Salary~dat$Subject+dat$Gender)
lsm1=lsmeans(lm1,"Gender")
lsm1
```

```
##  Gender lsmean   SE df lower.CL upper.CL
##  Female  74206 1115 36    71946    76467
##  Male    77875 1115 36    75615    80135
##
## Results are averaged over the levels of: Subject
## Confidence level used: 0.95
```

This shows the average predictions of the salary along with their confidence level.
Female Salary has a 95 % confidence level of (71946,76467)
Male Salary has a 95 % confidence level of (75615,80135)

```
pairs(lsm1)
```

```
##  contrast      estimate   SE df t.ratio p.value
##  Female - Male    -3669 1561 36  -2.350  0.0243
##
## Results are averaged over the levels of: Subject
```

We find an estimate of average Rs. 3669 difference between the gender pay.

Analysis

1)From the above ANOVA table, it can be observed that :

i) The p-value of subject w.r.t to salary is 0.5413
ii) The p-value of gender w.r.t to salary is 0.2043

This means that, since p-value of subject is greater than significance value (5%),we accept the null-hypothesis. Hence we say that there is no #significance of pay between the subjects.

As the p-value of gender is lesser than significance value (0.05), we reject the null hypothesis. There is a significant difference in the salary paid to both the genders

2) With 95% confidence we can say that, males receive a salary between Rs.(75615,80135)
and with 95% confidence we can say that, males receive a salary between Rs.(71946,76467)

3) The average difference between the salary of the 2 genders amounts to Rs 3669.

Conclusion

1) Students who picked any of the courses (Statistics, Political Science or Psychology) get paid the same amount

2) There is a gender pay difference amongst men and women. Men get paid more on an average