**Anova Analysis**

Ajay

June 18 2018

**help**("type")

## starting httpd help server ... done

## R Markdown  
  
*# =====setting working directory=====*  
  
**setwd**("E:/ML/Greatlakes/Advanced statistics/Project/Anova")  
**getwd**()

## [1] "E:/ML/Greatlakes/Advanced statistics/Project/Anova"

*# Install necessary packages*  
*# install.packages("ggplot2")*  
*# library(ggplot2)*  
*# install.packages("nortest")*  
  
  
**library**(ggplot2)  
**library**(nortest)  
**library**(car)

## Loading required package: carData

**library**(psych)

##   
## Attaching package: 'psych'  
##   
## The following object is masked from 'package:car':  
##   
## logit  
##   
## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

**library**(foreign)  
**library**(MASS)  
  
  
*# =====Reading the file=====*   
  
promo <- **read.csv**("PL\_X\_SELL.csv")  
**attach**(promo)  
  
*# Checking the total number of rows and columns*  
**dim**(promo)

## [1] 20000 10

*# So we have 20000 rows and 10 columns in our dataset.*  
  
*# ===== Names of the Columns in the data set =====*  
**names**(promo)

## [1] "Cust\_ID" "Target" "Age" "Gender"   
## [5] "Balance" "Occupation" "No\_OF\_CR\_TXNS" "AGE\_BKT"   
## [9] "SCR" "Holding\_Period"

*# ===== Find out Class of each Feature, along with internal structure =====*  
**str**(promo)

## 'data.frame': 20000 obs. of 10 variables:  
## $ Cust\_ID : Factor w/ 20000 levels "C1","C10","C100",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Target : int 0 1 0 0 0 0 0 0 0 0 ...  
## $ Age : int 30 41 49 49 43 30 43 53 45 37 ...  
## $ Gender : Factor w/ 3 levels "F","M","O": 2 2 1 2 2 2 2 2 2 2 ...  
## $ Balance : num 160379 84371 60849 10559 97100 ...  
## $ Occupation : Factor w/ 4 levels "PROF","SAL","SELF-EMP",..: 2 3 1 2 4 2 1 2 1 1 ...  
## $ No\_OF\_CR\_TXNS : int 2 14 49 23 3 2 23 45 3 33 ...  
## $ AGE\_BKT : Factor w/ 7 levels "<25",">50","26-30",..: 3 6 7 7 6 3 6 2 6 5 ...  
## $ SCR : int 826 843 328 619 397 781 354 239 339 535 ...  
## $ Holding\_Period: int 9 9 26 19 8 11 12 5 13 9 ...

*# #Out of the 10 Variables - Cust\_ID(Unique values),Target,Age,Balance,No\_OF\_CR\_TXNS,SCR and Holding period are integer variables. Gender and Occupation are categorical.*  
  
  
*# ===== Descriptive stats and check if there are any missing values =====*  
**summary**(promo)

## Cust\_ID Target Age Gender   
## C1 : 1 Min. :0.00000 Min. :21.0 F: 5525   
## C10 : 1 1st Qu.:0.00000 1st Qu.:30.0 M:14279   
## C100 : 1 Median :0.00000 Median :38.0 O: 196   
## C1000 : 1 Mean :0.08665 Mean :38.4   
## C10000 : 1 3rd Qu.:0.00000 3rd Qu.:47.0   
## C10001 : 1 Max. :1.00000 Max. :55.0   
## (Other):19994   
## Balance Occupation No\_OF\_CR\_TXNS AGE\_BKT   
## Min. : 0 PROF :5463 Min. : 0.00 <25 :1784   
## 1st Qu.: 23737 SAL :5839 1st Qu.: 7.00 >50 :3020   
## Median : 79756 SELF-EMP:3366 Median :13.00 26-30:3404   
## Mean : 146181 SENP :5332 Mean :16.65 31-35:3488   
## 3rd Qu.: 217311 3rd Qu.:22.00 36-40:2756   
## Max. :1246967 Max. :50.00 41-45:3016   
## 46-50:2532   
## SCR Holding\_Period   
## Min. :100.0 Min. : 1.00   
## 1st Qu.:333.0 1st Qu.: 8.00   
## Median :560.0 Median :16.00   
## Mean :557.1 Mean :15.34   
## 3rd Qu.:784.0 3rd Qu.:23.00   
## Max. :999.0 Max. :31.00   
##

**describe**(promo)

## vars n mean sd median trimmed mad  
## Cust\_ID\* 1 20000 10000.50 5773.65 10000.50 10000.50 7413.00  
## Target 2 20000 0.09 0.28 0.00 0.00 0.00  
## Age 3 20000 38.40 9.60 38.00 38.30 11.86  
## Gender\* 4 20000 1.73 0.46 2.00 1.78 0.00  
## Balance 5 20000 146181.31 169812.53 79755.74 113711.68 101211.66  
## Occupation\* 6 20000 2.43 1.15 2.00 2.41 1.48  
## No\_OF\_CR\_TXNS 7 20000 16.65 12.98 13.00 15.11 10.38  
## AGE\_BKT\* 8 20000 4.08 1.86 4.00 4.09 2.97  
## SCR 9 20000 557.14 260.52 560.00 558.42 335.07  
## Holding\_Period 10 20000 15.34 8.95 16.00 15.27 11.86  
## min max range skew kurtosis se  
## Cust\_ID\* 1 20000 19999 0.00 -1.20 40.83  
## Target 0 1 1 2.94 6.63 0.00  
## Age 21 55 34 0.10 -1.16 0.07  
## Gender\* 1 3 2 -0.76 -0.65 0.00  
## Balance 0 1246967 1246967 1.80 3.57 1200.76  
## Occupation\* 1 4 3 0.16 -1.41 0.01  
## No\_OF\_CR\_TXNS 0 50 50 0.95 0.01 0.09  
## AGE\_BKT\* 1 7 6 0.02 -1.12 0.01  
## SCR 100 999 899 -0.03 -1.21 1.84  
## Holding\_Period 1 31 30 0.03 -1.14 0.06

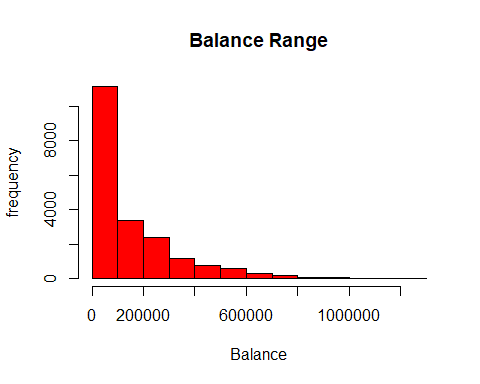
*# Observation:*  
  
*# We have unique coustomer Id's for each customer.*  
*# Target is either 0 or 1. 0- non respondent variable and 1 for respondent variable.*  
*# Age varies from 21 to 55.*  
*# We have less females i.e 5525 and more males i.e 14,279 in our data set, also 196 - companies or firms.*  
*# Balance(Double type) varying from 0 to 1246967*  
*# Occupation - Prof,Salaried, Delf employed and Senp*  
*# No\_OF\_CR\_TXNS - Number of credit transactions.*  
*# Age bracket - ages distributed in buckets.*  
*# SCR - Generic marketic scores for customer.*  
*# Holding\_Period - Ability of a customer to hold money in the account( no. of days) varies from 0 to 31.*  
  
*# Lets check if there are any NA values.*  
  
**colSums**(**is.na**(promo))

## Cust\_ID Target Age Gender Balance   
## 0 0 0 0 0   
## Occupation No\_OF\_CR\_TXNS AGE\_BKT SCR Holding\_Period   
## 0 0 0 0 0

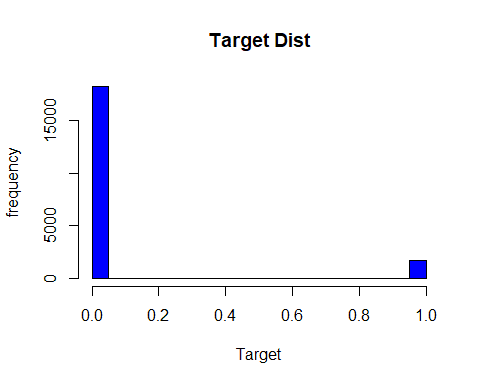
*# we see there are no NA values in our data set.*  
  
  
*# ===== Display first 6 rows in our data set using head function =====*  
**head**(promo)

## Cust\_ID Target Age Gender Balance Occupation No\_OF\_CR\_TXNS AGE\_BKT SCR  
## 1 C1 0 30 M 160378.60 SAL 2 26-30 826  
## 2 C10 1 41 M 84370.59 SELF-EMP 14 41-45 843  
## 3 C100 0 49 F 60849.26 PROF 49 46-50 328  
## 4 C1000 0 49 M 10558.81 SAL 23 46-50 619  
## 5 C10000 0 43 M 97100.48 SENP 3 41-45 397  
## 6 C10001 0 30 M 160378.60 SAL 2 26-30 781  
## Holding\_Period  
## 1 9  
## 2 9  
## 3 26  
## 4 19  
## 5 8  
## 6 11

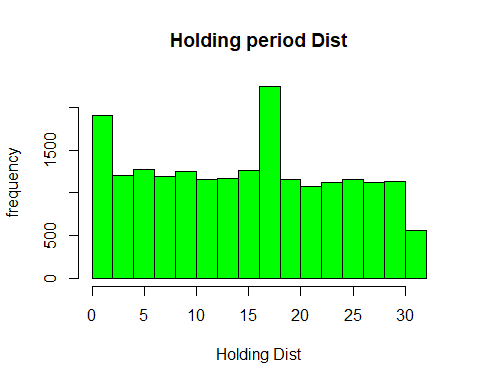
*# Lets start with Univariate Analysis, first Balance column*  
  
  
**hist**(Balance,main="Balance Range",xlab="Balance",ylab="frequency",col="Red")



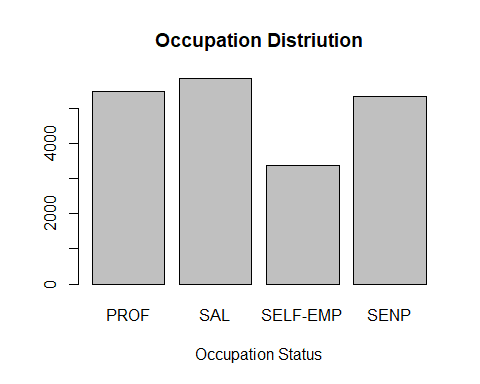
*# Almost 60-65% of customers are in the range of 0-1 Lakh range.*  
  
*# Target distribution*  
  
**hist**(Target,main="Target Dist",xlab="Target",ylab="frequency",col="Blue")



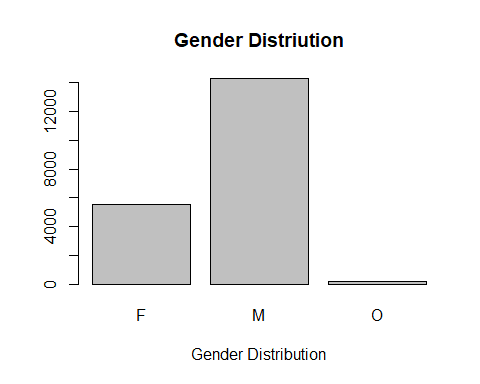
*# Around 90% are non-respondent customers.*  
  
*# Holding period*  
  
**hist**(Holding\_Period,main="Holding period Dist",xlab="Holding Dist",ylab="frequency",col="Green")



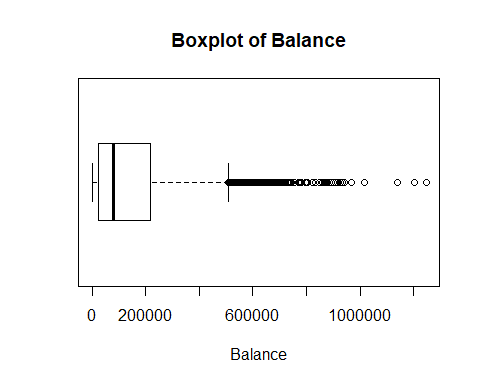
*# # lets see the Occupation column*   
count <- **table**(promo**$**Occupation)  
**barplot**(count,main="Occupation Distriution",xlab= "Occupation Status")



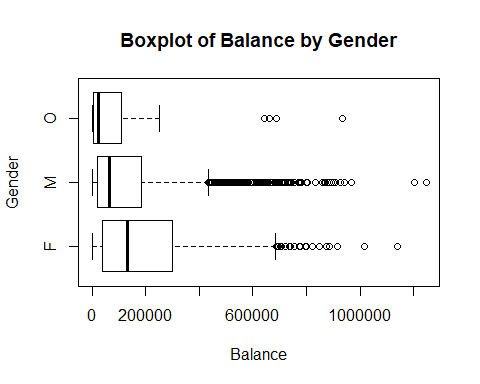
*# we see less number of self employeed people in our dataset.*  
  
*# gender*  
count2 <- **table**(promo**$**Gender)  
**barplot**(count2,main="Gender Distriution",xlab= "Gender Distribution")



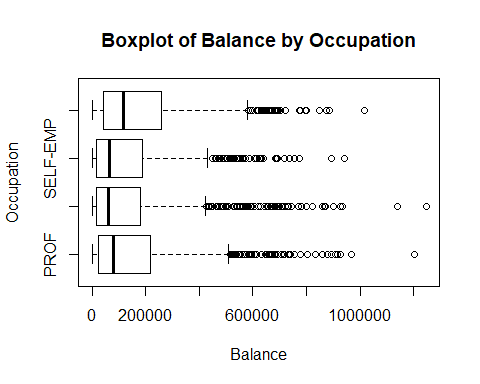
*# We can say there are high no. i.e around 14k males, less females and very few ompanies.*  
  
*# Lets draw boxplot and see if there are any outliers*  
  
**boxplot**(Balance,xlab="Balance",horizontal = TRUE,main="Boxplot of Balance")



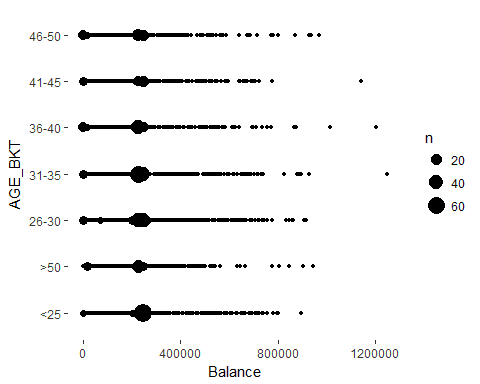
*# We see there are many outliers higher than 5L.*  
  
*# Now lets check the outliers again for the balance but with other variables like Gender and occupation.*  
  
**boxplot**(Balance**~**Gender,xlab="Balance",ylab="Gender",horizontal = TRUE,main="Boxplot of Balance by Gender")



*# we see outliers for all 3 types of genders. We see more no. of outlers for males (obviosly there are more males in our data set)*   
  
*# Occupation*  
  
  
**boxplot**(Balance**~**Occupation,xlab="Balance",ylab="Occupation",horizontal = TRUE,main="Boxplot of Balance by Occupation")



*# All 4 types has outliers. Also most of the occupation type have soe variation in mean but difference is not much.*  
  
*# Lets take age\_bkt column*  
  
g <- **ggplot**(promo, **aes**(Balance,AGE\_BKT))  
g **+** **geom\_count**()



*# So we can say all age\_bkt consists of outliers.*  
  
  
*# Thats all about EDA. Now lets start Anova analysis.*  
  
*# 1. Assuptions -*  
  
*# 1.1 - Normality test*  
  
*# Ho : data is distributed normally around each occupation*  
*# H1 : data is not distributed normally around each occupation*  
  
*#Note that we have 20k values in our dataset. Shapiro has an limitation that it works only on 5K values.*   
*# so we use another test which is anderson - darling normality test*  
  
**cat**("Normality p-values for Occupation:")

## Normality p-values for Occupation:

**for**(i **in** **unique**(**factor**(promo**$**Occupation))){  
 **cat**(**ad.test**(promo[promo**$**Occupation**==**i, ]**$**Balance)**$**p.value," ")  
}

## 3.7e-24 3.7e-24 3.7e-24 3.7e-24

*# Now we get p values way less than 0.05 . So we go for robost methods.*  
  
*# Anderson darling test for entire occupation column*  
  
**ad.test**(promo**$**No\_OF\_CR\_TXNS)**$**p.value

## [1] 3.7e-24

*# but before that lets check for Homogeneous in variance test.*  
  
**leveneTest**(promo**$**Balance**~**promo**$**Occupation)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 3 54.545 < 2.2e-16 \*\*\*  
## 19996   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*# H0 : Means of average balance across all 4 occupation is same.*  
*# H1 : Means of average balance e across all 4 occupation is different.*  
  
*# Here p value is lesss than 0.05. Null hypothesis is rejected.*  
  
*# Hence we might need to check robost method.*  
  
*# lets see bartlett test for variance test.*  
  
**bartlett.test**(promo**$**Balance**~**promo**$**Occupation)

##   
## Bartlett test of homogeneity of variances  
##   
## data: promo$Balance by promo$Occupation  
## Bartlett's K-squared = 96.401, df = 3, p-value < 2.2e-16

*# even bartletts test tells the same. p value < 0.05*  
  
*# Normality tests is violated. Hence there is difference in variance.*  
  
*# Even though Normality assumption is violated, anova test is given below -*  
  
aov2 <- **aov**(Balance**~**Occupation)  
**summary**(aov2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Occupation 3 1.052e+13 3.506e+12 123.8 <2e-16 \*\*\*  
## Residuals 19996 5.662e+14 2.831e+10   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

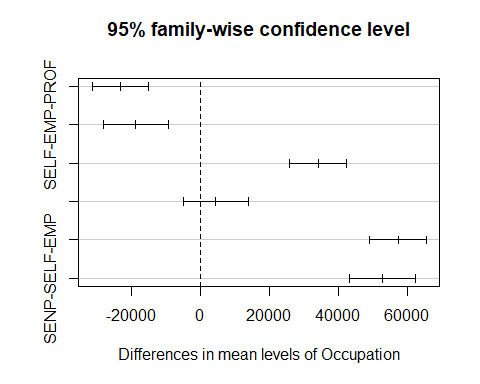
*# As p value< .05,we reject null hypothesis and thus concludethere is significant difference in balance amount for*  
*# different occupation types*   
  
*# Now as part of Robust method, we conduct Huber M-Estimator Test.*  
  
**oneway.test**(promo**$**Balance **~** promo**$**Occupation,var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: promo$Balance and promo$Occupation  
## F = 120.18, num df = 3, denom df = 10305, p-value < 2.2e-16

*# So, this test tells us that even though we have assumed that variances are not the same across the groups, we get p value < 0.05.*  
  
*# So we can conclude that there is significant difference in balance amount for different occupation types.*  
  
*# Now to know where the difference is, we need to run post hoc i.e tukey test.*  
  
*# POST HOC test :*  
  
**TukeyHSD**(aov2)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Balance ~ Occupation)  
##   
## $Occupation  
## diff lwr upr p adj  
## SAL-PROF -23151.230 -31288.977 -15013.482 0.0000000  
## SELF-EMP-PROF -18592.178 -28065.330 -9119.026 0.0000028  
## SENP-PROF 34199.915 25877.257 42522.573 0.0000000  
## SELF-EMP-SAL 4559.052 -4797.095 13915.198 0.5936835  
## SENP-SAL 57351.145 49161.914 65540.376 0.0000000  
## SENP-SELF-EMP 52792.093 43274.678 62309.508 0.0000000

**plot**(**TukeyHSD**(aov2))



*# AS per tukey plot, we can say except for self emp and salaried, there is significant difference across occupations.*  
  
*# Final Conclusion -*  
  
*# 1. assuptions violated for both normality and homogeneos ofvariance.*  
*# 2. Robust method shows similar result as of anova analysis.*  
*# 3. SO we conclude there is significant differnce in balance across different occupation.*  
*# 4. Even Tukey test suggests same as above point(point no.3)*  
  
*# ===========================*END *OF ONE WAY ANOVA======================================*  
  
  
*# ==========================Two way Anova========================*  
  
*# loading required libraries*  
  
**library**(psych)  
**library**(car)  
**library**(foreign)  
**library**(MASS)  
  
**library**(rcompanion)  
**library**(robustHD)

## Loading required package: perry  
## Loading required package: parallel  
## Loading required package: robustbase  
##   
## Attaching package: 'robustbase'  
##   
## The following object is masked from 'package:psych':  
##   
## cushny

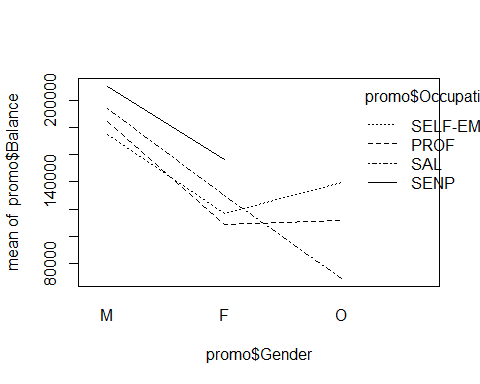
**library**(WRS2)  
  
  
*# Lets see how the visualization looks like.*  
  
*# Before that Lets assign the categorical column as factors*  
  
promo**$**Occupation <- **factor**(promo**$**Occupation,labels=**c**("SAL","SELF-EMP","PROF","SENP"))  
  
promo**$**Gender <- **factor**(promo**$**Gender,labels = **c**("M","F","O"))  
  
**tapply**(promo**$**Balance,**list**(promo**$**Occupation , promo**$**Gender),mean)

## M F O  
## SAL 194154.4 129761.6 69081.36  
## SELF-EMP 174860.4 116906.7 139669.79  
## PROF 184217.4 108727.4 111960.16  
## SENP 210156.0 156289.2 NA

*# Here we can say below :*  
  
*# Mean Balance for salaried Males is more 1.94 L, whereas for females its almost 1.3 L and for firms its less ~ 70k.*  
*# similarly Mean balance for self employed Males is 1.74 L, whereas for females its almost 1.2 L and for firms this # time its more which is ~ 1.4L.*  
*# similarly Mean balance for professional Males is 1.84 L, whereas for females its almost 1.08 L and for firms this # time its ~ 1.2L.*  
*# similarly Mean balance for SENP Males is 2.1 L, whereas for females its almost 1.56 L and we dont have any values # for thefirms.*  
  
*# Similarly apply for standard deviation*  
  
**tapply**(promo**$**Balance,**list**(promo**$**Occupation , promo**$**Gender),sd)

## M F O  
## SAL 186560.2 164560.0 147598.3  
## SELF-EMP 198516.3 152463.1 224833.8  
## PROF 188517.8 138033.2 225062.5  
## SENP 193518.9 158528.5 NA

*# Now lets draw interaction plot*   
  
**interaction.plot**(promo**$**Gender,promo**$**Occupation,promo**$**Balance)



*# We could clearly say there exists a same pattern in all 3 occupation types for males gender.*  
*# i.e the balance is high for males, decreases for females and increases only for self emp and prof occupation type.*  
*# note that there are no entries for firms in SENP.*  
  
*# Now lets check for the normality tests before conducting anova test.*  
*# same as one way anova, we use anderson - darling normality test as datset has 20k values.*  
*# first occupation -*   
  
**cat**("Normality p-values for Occupation:")

## Normality p-values for Occupation:

**for**(i **in** **unique**(**factor**(promo**$**Occupation))){  
 **cat**(**ad.test**(promo[promo**$**Occupation**==**i, ]**$**Balance)**$**p.value," ")  
  
}

## 3.7e-24 3.7e-24 3.7e-24 3.7e-24

*# we get very less p values. less than 0.05*  
  
*# now lets check for Gender -*  
  
**cat**("Normality p-values for gender:")

## Normality p-values for gender:

**for** (z **in** **unique**(**factor**(promo**$**Gender))){  
 **cat**(**ad.test**(promo[promo**$**Gender**==**z, ]**$**Balance)**$**p.value," ")  
}

## 3.7e-24 3.7e-24 3.7e-24

*# Even here we get values less than 0.05. so normality assumption is violated.*  
  
*# Next lets check for levene's Homogeneous test.*  
  
**leveneTest**(promo**$**Balance**~**promo**$**Occupation**\***promo**$**Gender)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 10 52.553 < 2.2e-16 \*\*\*  
## 19989   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

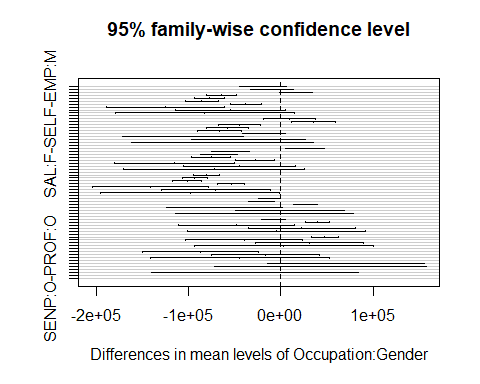
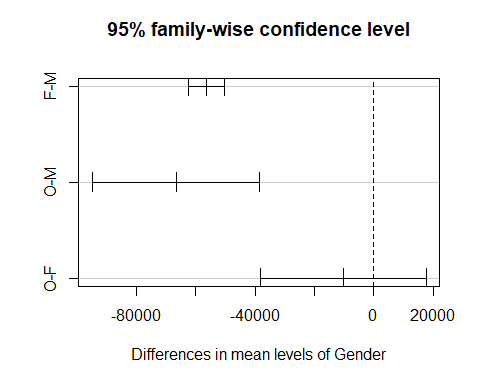
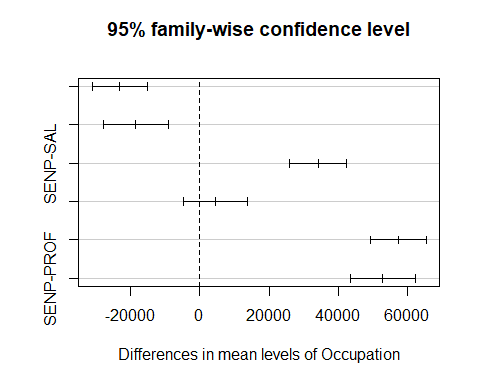
*# H0 : Variance across all occupation and gender is same.*  
*# H1 : variance across all occupation and gender is different.*  
  
*# We see p value is less than 0.05. Null hypothesis of equal variance is rejected.*  
  
*# so Both Normality and Homogeneous test's are violated. Thus we may have to check the robust method at the later*   
*# point of time.*  
  
*# Lets check 2 way Anova.*  
  
anova1 <- **aov**(Balance**~**Occupation**+**Gender**+**Occupation**:** Gender,data=promo)  
  
**summary**(anova1)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Occupation 3 1.052e+13 3.506e+12 127.035 < 2e-16 \*\*\*  
## Gender 2 1.401e+13 7.004e+12 253.797 < 2e-16 \*\*\*  
## Occupation:Gender 5 5.145e+11 1.029e+11 3.729 0.00225 \*\*   
## Residuals 19989 5.517e+14 2.760e+10   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*# Observation :*  
  
*# 1. Balance across occupation is different.*  
*# 2. Balance amoung gender is different.*   
*# 3. Balance across occupation and Gender (interaction effect) is also different or significant.*  
  
*# So in this model all 3 effects are significant.*  
  
*# Lets check Tukey test :*  
  
**TukeyHSD**(anova1)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Balance ~ Occupation + Gender + Occupation:Gender, data = promo)  
##   
## $Occupation  
## diff lwr upr p adj  
## SELF-EMP-SAL -23151.230 -31185.335 -15117.124 0.0000000  
## PROF-SAL -18592.178 -27944.680 -9239.676 0.0000020  
## SENP-SAL 34199.915 25983.254 42416.576 0.0000000  
## PROF-SELF-EMP 4559.052 -4677.935 13796.039 0.5833428  
## SENP-SELF-EMP 57351.145 49266.212 65436.078 0.0000000  
## SENP-PROF 52792.093 43395.892 62188.294 0.0000000  
##   
## $Gender  
## diff lwr upr p adj  
## F-M -56408.97 -62578.25 -50239.69 0.0000000  
## O-M -66673.63 -94975.49 -38371.76 0.0000001  
## O-F -10264.66 -38267.73 17738.41 0.6660915  
##   
## $`Occupation:Gender`  
## diff lwr upr p adj  
## SELF-EMP:M-SAL:M -19294.052 -44579.905 5991.801 0.3436234  
## PROF:M-SAL:M -9937.073 -32958.860 13084.714 0.9617493  
## SENP:M-SAL:M 16001.539 -1675.625 33678.702 0.1210509  
## SAL:F-SAL:M -64392.833 -80797.831 -47987.835 0.0000000  
## SELF-EMP:F-SAL:M -77247.743 -93077.993 -61417.493 0.0000000  
## PROF:F-SAL:M -85427.058 -103105.599 -67748.518 0.0000000  
## SENP:F-SAL:M -37865.200 -55051.667 -20678.733 0.0000000  
## SAL:O-SAL:M -125073.083 -188883.781 -61262.386 0.0000000  
## SELF-EMP:O-SAL:M -54484.648 -114006.364 5037.067 0.1108627  
## PROF:O-SAL:M -82194.280 -179176.705 14788.146 0.1929317  
## SENP:O-SAL:M NA NA NA NA  
## PROF:M-SELF-EMP:M 9356.979 -18643.034 37356.992 0.9950594  
## SENP:M-SELF-EMP:M 35295.590 11494.747 59096.434 0.0000805  
## SAL:F-SELF-EMP:M -45098.782 -67970.642 -22226.921 0.0000000  
## SELF-EMP:F-SELF-EMP:M -57953.691 -80416.880 -35490.503 0.0000000  
## PROF:F-SELF-EMP:M -66133.007 -89934.873 -42331.140 0.0000000  
## SENP:F-SELF-EMP:M -18571.148 -42009.849 4867.552 0.2854352  
## SAL:O-SELF-EMP:M -105779.031 -171549.870 -40008.193 0.0000096  
## SELF-EMP:O-SELF-EMP:M -35190.596 -96809.040 26427.848 0.7801076  
## PROF:O-SELF-EMP:M -62900.228 -161183.436 35382.980 0.6288024  
## SENP:O-SELF-EMP:M NA NA NA NA  
## SENP:M-PROF:M 25938.611 4558.517 47318.706 0.0042222  
## SAL:F-PROF:M -54455.761 -74796.614 -34114.907 0.0000000  
## SELF-EMP:F-PROF:M -67310.670 -87190.890 -47430.450 0.0000000  
## PROF:F-PROF:M -75489.986 -96871.219 -54108.752 0.0000000  
## SENP:F-PROF:M -27928.127 -48904.328 -6951.926 0.0008288  
## SAL:O-PROF:M -115136.011 -180070.065 -50201.956 0.0000004  
## SELF-EMP:O-PROF:M -44547.575 -105272.043 16176.892 0.4074842  
## PROF:O-PROF:M -72257.207 -169982.420 25468.006 0.3945523  
## SENP:O-PROF:M NA NA NA NA  
## SAL:F-SENP:M -80394.372 -94402.575 -66386.169 0.0000000  
## SELF-EMP:F-SENP:M -93249.282 -106579.795 -79918.768 0.0000000  
## PROF:F-SENP:M -101428.597 -116908.797 -85948.397 0.0000000  
## SENP:F-SENP:M -53866.739 -68782.516 -38950.961 0.0000000  
## SAL:O-SENP:M -141074.622 -204311.564 -77837.680 0.0000000  
## SELF-EMP:O-SENP:M -70486.187 -129392.387 -11579.987 0.0052347  
## PROF:O-SENP:M -98195.818 -194801.702 -1589.935 0.0422783  
## SENP:O-SENP:M NA NA NA NA  
## SELF-EMP:F-SAL:F -12854.910 -24445.615 -1264.205 0.0152781  
## PROF:F-SAL:F -21034.225 -35044.166 -7024.284 0.0000596  
## SENP:F-SAL:F 26527.633 13143.976 39911.291 0.0000000  
## SAL:O-SAL:F -60680.250 -123573.435 2212.935 0.0705829  
## SELF-EMP:O-SAL:F 9908.185 -48628.831 68445.201 0.9999931  
## PROF:O-SAL:F -17801.446 -114182.661 78579.768 0.9999831  
## SENP:O-SAL:F NA NA NA NA  
## PROF:F-SELF-EMP:F -8179.315 -21511.655 5153.024 0.6901015  
## SENP:F-SELF-EMP:F 39382.543 26709.930 52055.156 0.0000000  
## SAL:O-SELF-EMP:F -47825.340 -110571.062 14920.381 0.3453523  
## SELF-EMP:O-SELF-EMP:F 22763.095 -35615.455 81141.645 0.9822955  
## PROF:O-SELF-EMP:F -4946.537 -101231.589 91338.516 1.0000000  
## SENP:O-SELF-EMP:F NA NA NA NA  
## SENP:F-PROF:F 47561.858 32644.448 62479.268 0.0000000  
## SAL:O-PROF:F -39646.025 -102883.352 23591.302 0.6593423  
## SELF-EMP:O-PROF:F 30942.410 -27964.203 89849.023 0.8608972  
## PROF:O-PROF:F 3232.779 -93373.357 99838.914 1.0000000  
## SENP:O-PROF:F NA NA NA NA  
## SAL:O-SENP:F -87207.883 -150309.416 -24106.351 0.0003919  
## SELF-EMP:O-SENP:F -16619.448 -75380.260 42141.363 0.9988935  
## PROF:O-SENP:F -44329.080 -140846.380 52188.220 0.9407042  
## SENP:O-SENP:F NA NA NA NA  
## SELF-EMP:O-SAL:O 70588.435 -14436.640 155613.510 0.2194253  
## PROF:O-SAL:O 42878.804 -71541.560 157299.168 0.9871046  
## SENP:O-SAL:O NA NA NA NA  
## PROF:O-SELF-EMP:O -27709.631 -139794.640 84375.377 0.9996908  
## SENP:O-SELF-EMP:O NA NA NA NA  
## SENP:O-PROF:O NA NA NA NA

**plot**(**TukeyHSD**(anova1))



*# We can say there is difference between and within groups.*  
  
*# Final Conclusion -*  
  
*# 1. assuptions violated for both normality and homogeneos ofvariance.*  
*# 2. Anova test conducted shows significant difference amoung gender and balance.*  
*# 3. Post hoc shows significaant difference bw grp and within grp.*  
  
  
*# Finally lets check for robost method for 2 way anova (as assumptions got violated)*  
  
  
  
*#pbad2way(formula = Balance ~ Occupation + Gender +*   
*#Occupation:Gender,*  
 *# data = promo, est = "mom", nboot = 5000)*