

Assignment2_210034046

```
#Importing baldy.csv data.
PROC import out= Najmi050.baldy
dbms=csv replace
file="/home/u59691081/sasuser.v94/SasFolder/New Folder3/Baldy.csv";
getnames=yes;
datarow=2;
run;
DATA najmi050.baldy_mm;
SET najmi050.baldy;
Luxuriant=luxuriant*25.4;
placebo=placebo*25.4;
baldbegone=baldbegone*25.4;
skinheadnomore=skinheadnomore*25.4;
RUN;
#Checking contents.
proc contents data=najmi050.baldy_mm;
run;
#Checking numeric variables.
proc univariate data=najmi050.baldy_mm;
run;

#Task 1: Comparing Luxuriant with placebo.

#Extracting subset of required data.
data najmi050.LuxuriantVsPlacebo;
set najmi050.baldy_mm;
keep luxuriant placebo;
run;

#Double transposing, to get factor variable.

#First transpose converts data into two rows of brands, each having their own 100 columns of hair growth.
proc transpose data=najmi050.luxuriantvsplacebo out=temp;
by rows;;
var luxuriant placebo;
run;
#Second transpose gets data into two columns, brand names as factor variable and 2nd column hair growth.
#Therefore, each row is a subject that has used a particular brand and his hair growth.
proc transpose data=temp out=want;
by _name_;
var col;;
run;
proc print data=want;
run;

#Saving data to permanent library and renaming the columns to brand and growth_mm.
```

```

data najmi050.LuxuriantVsPlacebo;
set work.want;
rename _name_=LUXorPlacebo col1=growth_mm;
run;

proc print data=najmi050.LuxuriantVsPlacebo;
run;

#Conducting a two sample ttest to check difference between the two groups' growth.

proc ttest alpha=0.05 data=najmi050.luxuriantvsplacebo sided=2;
class LUXorPlacebo;
var growth_mm;
run;

#The F-test for equality variance ( $p < 0.0001$ ) shows that there is enough evidence to reject the null hypothesis.
#the variance of the two groups is equal.
#Therefore, options are to either use the un-pooled Satterthwaite method,
#provided the normality assumption holds or go for a non-parametric test.
#Although, the QQplot for placebo shows deviation from the normal straight line,
#to confirm our suspicion we conduct a normality test.
proc univariate data=najmi050.luxuriantvsplacebo normaltest;
run;

#According to Shapiro-wilk test, the assumption that data is normal, doesn't
#satisfy,  $p\text{-value} < 0.0001$ .
#To this end, we use non-parametric tests to test growths are equal.
proc npar1way wilcoxon data=najmi050.luxuriantvsplacebo;
class luxorplacebo;
var growth_mm;
exact;
run;

#The  $p\text{-value} < 0.0001$ , lower than the 0.05 significance level.
#Therefore, we conclude that the placebo and luxuriant have significantly
#different growth rates. It is evident from the mean scores (luxuriant=144.7>placebo=56.3), luxuriant
#is causing more growth, in comparison to placebo.

proc univariate data=najmi050.baldy_mm;
run;

#Task 2: Luxuriant vs all other brands.

#Extracting subset of data for analysis.

Data Najmi050.LUXVSALLBRANDS;
set najmi050.baldy_mm;
keep luxuriant baldbegone skinheadnomore;
run;
#Double transposing, to get it into a factor variable form.

#1st transpose converts the data into 3 rows, each corresponding to brands,
#with each having 100 columns.

```

```

proc transpose data=najmi050.luxvsallbrands out=temp;
by rows;;
var luxuriant baldbegone skinheadnomore;
run;
#Sorting data according to brand.
proc sort data= temp;
by _name_;
run;
#Transposing data again to get 400 rows and 2 columns.
#1st column is a factor variable with brand names and the 2nd column
#contains the values of growth in mm of a subject that used that brand.
proc transpose data=work.temp out=need;
by _name_;
var col;;
run;
#Transferring data to luxvsallbrands in the permanent library.
data najmi050.luxvsallbrands;
set need;
run;
#Renaming columns with brand and growth_mm.
data najmi050.luxvsallbrands;
set najmi050.luxvsallbrands;
rename _name_=Brand col1=Growth_mm;
run;

#Conducting an anova analysis to test whether there is difference in mean
#in any one of the 3 brand's growth, followed by a Tukey's HSD to analyse pairwise differences.
proc anova data=najmi050.luxvsallbrands;
class brand;
model growth_mm=brand;
means brand/tukey cldiff;
run;
#P-value<0.0001 for anova shows that our null hypothesis should be rejected and there is a significant difference
#between the growth rate of atleast one brand compared to the others.
#From the Tukey's HSD it is clear that there are significant differences between all the brands.

#Task3: analysing the effect of age on hair growth
#Dividing into two separate data sets, one for growth data, one for age data.

#Extracting growth data set.
data najmi050.Growths_set;
set najmi050.baldy_mm;
keep luxuriant placebo baldbegone skinheadnomore;
run;

#Double transposing growth set, to get into factor variable form for analysis.
#First transpose converts 4 rows of brands, with 100 columns of growth data.
proc transpose data=najmi050.growths_set out=temp;
by rows;;
var luxuriant placebo baldbegone skinheadnomore;
run;
#Sorting the data by brand name.
proc sort data=temp;

```

```

by _name_;
run;
#2nd transpose to get into factor variable form, 1st column is brand names and 2nd is growth in mm.
proc transpose data=temp out=need;
by _name_;
var col;;
run;
#Saving from temporary to permanent library and renaming columns to brand and growth_mm.
data najmi050.Growths_set;
set need;
rename _name_=brand col1=growth_mm;
run;

#Extracting 2nd set, with ages and agebrand.
data najmi050.age_set;
set najmi050.baldy_mm;
drop luxuriant placebo baldbegone skinheadnomore;
run;

#Renaming the columns to brand names.
data najmi050.age_set;
set najmi050.age_set;
rename ageluxuriant=luxuriant ageplacebo=placebo agebaldbegone=baldbegone ageskinheadnomore=skinheadnomore;
run;

#Double transposing again to get factor variable form.

#1st transpose to get 4 rows with 100 columns of ages.
proc transpose data=najmi050.age_set out=temp;
by rows;;
var luxuriant placebo baldbegone skinheadnomore;
run;

#Sorting the data set by brand-name.
proc sort data=temp;
by _name_;
run;

#2nd transpose to get factor variable form.
proc transpose data=temp out=need;
by _name_;
var col;;
run;

#Saving data to permanent library and renaming columns to brand and ages.
data najmi050.ages_set;
set need;
rename _name_=brand col1=ages;
run;

#Merging the two data sets side by side.
data najmi050.Merged_Growth_Ages;
merge najmi050.growths_set najmi050.ages_set;

```

```
run;

#Fitting a linear model similar to R's: growth~ages.
proc glm data=najmi050.merged_growth_ages;
class brand;
model growth_mm=ages/ solution clparm;
run;
#From this analysis we conclude that, age doesn't effect the growth. As it
#is evident from a very small parameter estimate (-0.067).
#Furthermore, the p-value for the t-test is less than 0.5919 which is indicative
#that no evidence can be found against the null hypothesis:
#hair growth and age are equal.
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.