Which is more effective treatment for male pattern baldness and does age influence the hair growth?

MT5763 - Individual Courserwork 2

Student ID number: 200028225

October 29, 2020

Github Link of private repository: ***Click here***

Abstract

Significant research has been undertaken to find out which male pattern baldness treatments are more effective. Along with which factor matters in hair growth values after the treatments. This analytic paper uses the data of hair growth on randomly selected male groups after receiving different male pattern baldness (MPB) treatments over a month. Data wrangling has been done to optimize the dataset in a way that can be easier to read and operate the statistical operations. Fits Liner model to determine the effects of a person's age over hair growth measurement after receiving different MPB treatments for a month. This study finds that luxuriant treatment is quite effective than other existing treatments. Besides, there is not much correlation between age and hair growth values according to our dataset. However, some researchers s justify that there is a relationship between hair growth and age in male pattern baldness treatments (Ellis, J. A., & Sinclair, R. D. (2008))

Introduction

The male pattern baldness (MPB) has been a major concern in the global community. The demand for new products and treatments has increased a lot. The MPB is the most common way to lose hair among men at age of 20-30 and it increased 10% per decade (Liu, F., Hamer, M. A., Heilmann, S., Herold, C., Moebus, S., Hofman, A., Uitterlinden, A. G., Noethen, M. M., van Duijn, C. M., Nijsten, T. E. C., & Kayser, M. (n.d.)) Thus, scientists and researchers are using an advanced statistical approach to analyze and examine the relationship between hair growth on men of different MPB treatments over their ages. This study helps to find which treatment can be more effective for which group of male population which are more or less bald. This paper concludes that luxuriant treatment is more efficient than placebo treatment for MPB. However, this study finds no relation between age and hair growth values in any treatment.

Analysis Description

We have data of newly invented drugs called luxuriant and other 3 existing drugs effect over randomly allocated groups of men. Data wrangling to reshape the data in a form suitable for analysis. The given dataset neither has an extra variable nor any null values required to remove, therefore, not a single column or row was being removed from the original dataset. For further process, a copy of the dataset was made, to avoid any unintentionally change in the original dataset. The wide-to-long transformation has been

implemented to get the optimized version of the data, as it would be clearer to see each row of data as one person. To obtain that, a treatment column has been added to the data. As result, the newly created dataset has 3 columns, named treatment, hair growth values, and age (figure 1). As result, the newly created dataset has 400 rows and 3 columns.

Obs	Treatment	HairGrowthValue	Age
1	Luxuriant	20.122735828	58
2	Luxuriant	15.377107143	60
3	Luxuriant	17.638840643	53
4	Luxuriant	1.3797510632	62
5	Luxuriant	37.55853484	49

Figure 1: First few rows of newly craeted dataset (wide to long transformation)

The hair growth values have been changed from inches to millimeters to determine more human-understandable language. Statistical summary and frequency tables for numeric values have been calculated for the analysis purpose of the dataset after each operation. The liner models has been made to find the relationship between age and hair growth values of different treatments. The results of the analysis have been discussed in the below section.

Results

To determine, if luxuriant treatment is above or beyond the placebo, the series plot has generated. A new dataset with only treatments luxuriant and placebo was made to avoid any accidental manipulation of data with the other 2 treatments.
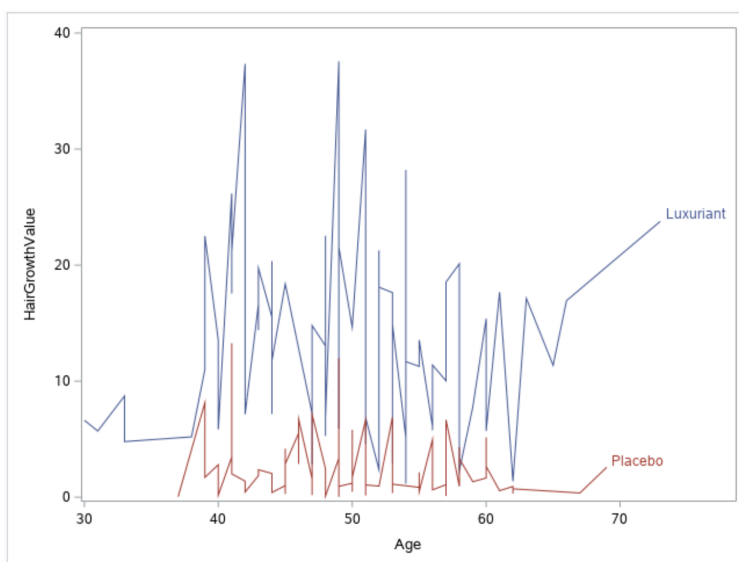

Figure 2: Effect of luxuriant treatment over the placbo treatment

As per Figure 2, we can determine that the newly created drug is more effective than an existing drug placebo for all age groups. Thus, to be more specific, the statistical summary for both treatments has been discussed below.

Treatment	Mean	Std Deviation	Range	Variance	Std Error Mean	p Value (Student's t)
luxuriant	12.99950	7.52583	36.40346	56.6380857	0.75258279	<.0001
placbo	2.485074	2.40679	13.22991	5.79266	0.24067941	<.0001
BaldBeGone	36.1521295	19.7447168	96.19019	389.85384	1.97447168	<.0001
SkinHeadNoMore	18.74009	11.02093	54.30731	121.46095	1.10209323	<.0001

Table 1: Statistical summary of Hair growth variable for different treatments

The mean of hair growth variable for luxuriant treatment is 12.99 whereas for placebo treatment it is only 2.48, which implies that the average hair growth value(mm) is higher in luxuriant treatment. However, we can see the variance of both treatments which are respectively 56.63 and 5.79 implies that data is more spread out around the mean for luxuriant treatment data. The value of standard error for luxuriant treatment is 0.75 which concludes that the sample mean is more likely precise estimate of the population mean. (Refer to table 1)
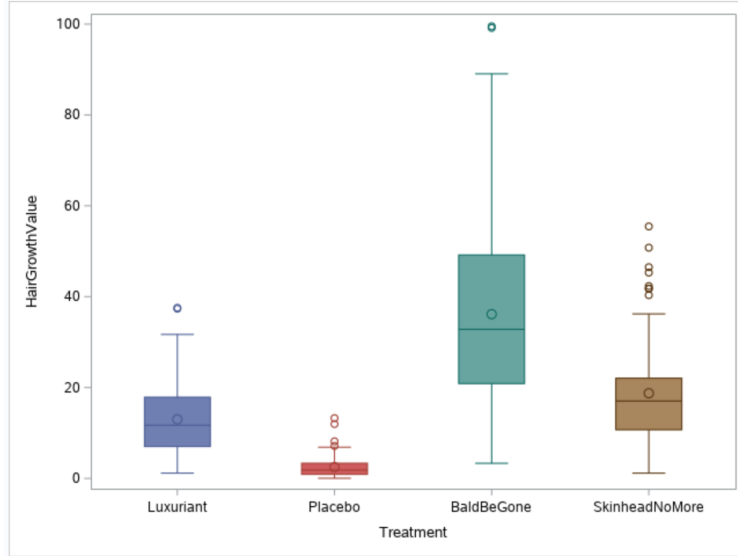

Figure 3: Hair growth effects of different treatments

To find out the more effective treatment, the box plot of hair growth values was generated. (Figure 3) The boxplot shows negative result for BladBeGnone treatment as the values are spread around a lot from sample mean.Placebo treatment does have a less success rate as compared to other treatments. whereas newly created drug(Luxuriant) and skinhead more does have similar results from the experiment. But the variance of shinHeadNoMore treatment is 121.46, which says the data points has spread out a lot around the mean. (Refer to Table 1). Thus we can conclude that luxuriant and SkinHeadNoMore treatments could be more effective than the other 2 treatments.

Treatment	Whole dataset	luxuriant	placbo	BaldBeGone	SkinHeadNoMore
R-Square	0.087737	0.279639	0.270189	0.160687	0.335339
Root MSE	17.14422	7.651073	2.378187	21.36000	10.53586

Table 2: Statistical summary of Linear Models between age and hair growth variables for all treatments

Linear Models were generated to find out the relationship between hair growth and age variables. In figure 4, we can see that data was not evenly distributed, which implies that data is not normally distributed

for the whole dataset. Plus the R-Square value for the whole data set is 0.087, which indicates there is no relationship between age and hair growth in our dataset. The high value of root means square error (17.14422) indicates that data values are widely around its center mean which is not good. (Refer to Table 2)
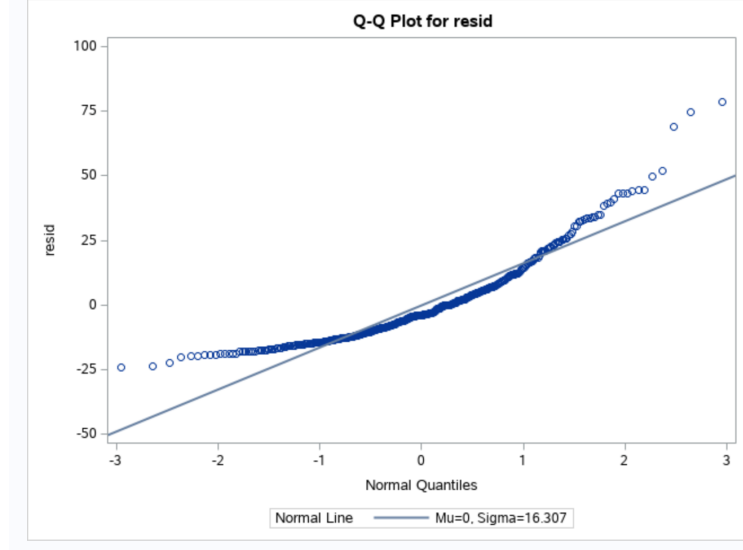

Figure 4: Q-Q plot of whole dataset

The R-square values for luxuriant treatment are also 0.27, which implies that there is no relationship between age and hair growth values for luxuriant treatment. The same logic applies for placebo, bladBeGone, and skinHeadNoMore treatments, R-squares values are 0.270, 0.160, and 0.335 respectively (Refer to Table 2), which justify that there is no relation between age and hair growth values for each individual treatments too. The lower root means square error reflects on the distribution of data and implies that data is centralized around the central moment (mean). (Moody, J. (2019, September 5)). Although our results do not match with theory, we have quite a high root MSE which refers that data is not centralized for any individual treatment.

Assumption Check

Hypothesis for Shapiro-Wilk test (Normality check):

- H0(null hypothesis): The population is normally distributed.
- H1: The population is not normally distributed.

Treatment	Whole dataset	luxuriant	placbo	BaldBeGone	SkinHeadNo- More
Shapiro-Wilk(P Value)	<0.0001	0.0002	<0.0001	0.0203	<0.0001

Table 3: P Value of Shapiro-Wilk test for normality

The P Value <0.0001 (which is <0.05)of Shapiro-Wilk test concludes that we reject the null hypotheses, which mean data is not normally distributed for whole datsset.(Refer to table 3) Similar, for each individual treatment P values are less than 0.05, means data is not normally distributed for any treatments individually.

Software use

The SAS version 3.8 (Basic Edition) was used to run the analysis of data using the university SAS edition. Also, the Apps Anywhere SAS edition was used to generate the graphs. The graph code can be found in the appendices section of this report. R Studio was being used to produce this PDF report.

Further research

The provided dataset information could have more variables in it. The hair growth variable could depend on genetic factors (Ellis, J. A., & Sinclair, R. D. (2008). The quality of hair does have a bit of effect on hair growth for MPB treatments in elderly men (Rexbye, H., Petersen, I., Iachina, M., Mortensen, J., McGue, M., Vaupel, J., & Christensen, K. (n.d.)). By adding a few more variables to the dataset could give more accurate results.

Conclusion

This study does not find any relationship between age to hair growth values for any MPB treatments. But reaches does found some relationship between age to hair growth (Ellis, J. A., & Sinclair, R. D. (2008)). However, the newly created drug (luxuriant) performs better than existing treatments for male pattern baldness problem. Assumptions are being violated and the lower p-value(<0.0001) for Shapiro-Wil test implies that we reject the null hypothesis. thus, the given dataset is not normally distributed. The other variables such as genetic history, hair quality, mental health for the treatment period, and physical health history could affect hair growth in MPB treatments.

References

Ellis, J. A., & Sinclair, R. D. (2008). Male pattern baldness: current treatments, future prospects. Drug Discovery Today, 13(17), 791–797. https://doi.org/10.1016/j.drudis.2008.05.010

Swanson, N., Abramovits, W., Berman, B., Kulp, J., Rigel, D. S., & Levy, S. (2010). Imiquimod 2.5% and 3.75% for the treatment of actinic keratoses: results of two placebo-controlled studies of daily application to the face and balding scalp for two 2-week cycles. Journal of the American Academy of Dermatology, 62(4), 582-590.

Liu, F., Hamer, M. A., Heilmann, S., Herold, C., Moebus, S., Hofman, A., Uitterlinden, A. G., Noethen, M. M., van Duijn, C. M., Nijsten, T. E. C., & Kayser, M. (n.d.). Prediction of male-pattern baldness from genotypes. EUROPEAN JOURNAL OF HUMAN GENETICS, 24(6), 895–902. https://doi.org/10.1038/ejhg.2015.220

Rexbye, H., Petersen, I., Iachina, M., Mortensen, J., McGue, M., Vaupel, J., & Christensen, K. (n.d.). Hair loss among elderly men: Etiology and impact on perceived age. JOURNALS OF GERONTOLOGY SERIES A-BIOLOGICAL SCIENCES AND MEDICAL SCIENCES, 60(8), 1077–1082.

Moody, J. (2019, September 5). What does RMSE really mean? Https://Towardsdatascience.Com/. https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e

Appendices - SAS Code

Code can be reproduce by changing file location and craete library called DATA.

```
/* Generated Code (IMPORT) */
/* Source File: Baldy.csv */
/* Source Path: /folders/myfolders/MT5763_2_200028225 */
/* Code generated on: 10/22/20, 4:48 PM */

%web_drop_table(DATA.BALDY);


FILENAME REFFILE '/folders/myfolders/Baldy.csv';

PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV
    OUT=DATA.BALDY;
    GETNAMES=YES;
RUN;

/* View the basic details of dataset */
PROC CONTENTS DATA=DATA.BALDY;
RUN;

/* Disply the dataset in table format */
%web_open_table(DATA.BALDY);


/* Creating the copy of existing dataset*/
DATA DATA.COPY_DATASET;
SET DATA.BALDY;
RUN;

/*--------------------- SUMMIRINZING DATASET ------------------------------*/

/* Summirizing the numeric variables of different treatments*/
PROC UNIVARIATE DATA= DATA.COPY_DATASET;
VAR LUXURIANT PLACEBO BALDBEGONE SKINHEADNOMORE;
RUN;



/* Summirizing the numeric variables of age when treatment was done*/
PROC UNIVARIATE DATA= DATA.COPY_DATASET;
VAR AGELUXURIANT AGEPLACEBO AGEBALDBEGONE AGESKINHEADNOMORE;
RUN;



/*---------------- CONVERTING THE DATASET TO THREE COLUMNS AGE, HAIR GROWTH VALUE, TREATMENT*/

/* CHANGE DATASET TRANSPOSE WIDE TO LONG  */

/* Creating the seprate dataset of Luxuriant treatment*/
DATA DATA.DATA_LUXURIANT;
    SET DATA.COPY_DATASET ;
    Treatment = "Luxuriant";
```

```
    RENAME AgeLuxuriant = Age;
    RENAME LUXURIANT = HairGrowthValue;
    KEEP AgeLuxuriant LUXURIANT Treatment;
RUN;

PROC UNIVARIATE DATA= DATA.DATA_LUXURIANT;
VAR AGE HAIRGROWTHVALUE;
RUN;

/* Creating the seprate dataset of Placebo treatment*/
DATA DATA.DATA_PLACEBO;
    SET DATA.COPY_DATASET ;
    Treatment = "Placebo";
    RENAME AGEPLACEBO = Age;
    RENAME PLACEBO = HairGrowthValue;
    KEEP AGEPLACEBO PLACEBO Treatment;
RUN;
PROC UNIVARIATE DATA= DATA.DATA_PLACEBO;
VAR AGE HAIRGROWTHVALUE;
RUN;


/* Creating the seprate dataset of BaldBeGone treatment*/
DATA DATA.DATA_BALDBEGONE;
    SET DATA.COPY_DATASET ;
    Treatment = "BaldBeGone";
    length Treatment $ 20;
    RENAME AGEBALDBEGONE = Age;
    RENAME BALDBEGONE = HairGrowthValue;
    KEEP AGEBALDBEGONE BALDBEGONE Treatment;
RUN;

PROC UNIVARIATE DATA= DATA.DATA_BALDBEGONE;
VAR AGE HAIRGROWTHVALUE;
RUN;

/* Creating the seprate dataset of SkinheadNoMore treatment*/
DATA DATA.DATA_SKINHEADNOMORE;
    SET DATA.COPY_DATASET ;
    Treatment = "SkinheadNoMore";
    length Treatment $ 20;
    RENAME AGESKINHEADNOMORE = Age;
    RENAME SKINHEADNOMORE = HairGrowthValue;
    KEEP AGESKINHEADNOMORE SKINHEADNOMORE Treatment;
RUN;

PROC UNIVARIATE DATA= DATA.DATA_SKINHEADNOMORE;
VAR AGE HAIRGROWTHVALUE;
RUN;

/* Merging all above 4 datasets*/
DATA DATA.TIDY_DATASET;
    length Treatment $ 20;
    SET DATA.DATA_LUXURIANT DATA.DATA_PLACEBO DATA.DATA_BALDBEGONE DATA.DATA_SKINHEADNOMORE ;
```

```
RUN;

/* CHANGE HAIR GROTH VALUES(INCHES) TO MILIMIITER*/
DATA DATA.TIDY_DATASET;
    SET DATA.TIDY_DATASET;
    HairGrowthValue = HairGrowthValue * 25.4;
RUN;

/* FREQUENCY TABLES of summery for tidy dataset*/
PROC FREQ DATA=DATA.TIDY_DATASET;
    TABLE Age*HairGrowthValue / CHISQ;
RUN;

/* summerizing tidy dataset */
PROC CONTENTS DATA=DATA.TIDY_DATASET;
RUN;

/* Basic statistical summery for numeric data */
PROC UNIVARIATE DATA= DATA.TIDY_DATASET;
VAR Age HairGrowthValue;
RUN;

/* ------------- Question 1 ------------- */
/* creating a new dataset for Luxuriant and Placebo Treatments*/
DATA DATA.DATA_q1;
    SET DATA.TIDY_DATASET;
    WHERE Treatment = "Luxuriant" OR Treatment = "Placebo";
RUN;

/* statistical summery of above dataset */
PROC UNIVARIATE DATA= DATA.DATA_q1;
VAR AGE HAIRGROWTHVALUE;
RUN;

/*Creating a series plot of hair growth for luxuriant treatment and placbo treatment*/
proc template;
define statgraph sgdesign;
dynamic _AGE _HAIRGROWTHVALUE _TREATMENT _TREATMENT2;
begingraph;
  /* entrytitle halign=center 'Graph: 1 Effect of luxuriant treatment over the placbo treatment';*/
   layout lattice / rowdatarange=data columndatarange=data rowgutter=10 columngutter=10;
      layout overlay;
         seriesplot x=_AGE y=_HAIRGROWTHVALUE / group=_TREATMENT curvelabel=_TREATMENT2 name='series' c
      endlayout;
   endlayout;
endgraph;
end;
run;

/* Displaying the above graph */
proc sgrender data=DATA.DATA_q1 template=sgdesign;
dynamic _AGE="AGE" _HAIRGROWTHVALUE="HAIRGROWTHVALUE" _TREATMENT="TREATMENT" _TREATMENT2="TREATMENT";
run;
```

```
/* ------------- Question 2 ------------- */

/*Creating a box plot of hair growth for all treatments*/
proc template;
define statgraph sgdesign;
dynamic _HAIRGROWTHVALUE _TREATMENT _TREATMENT2;
begingraph;
  /* entrytitle halign=center 'Graph: 2 Hair growth effects of different treatments';*/
   layout lattice / rowdatarange=data columndatarange=data rowgutter=10 columngutter=10;
      layout overlay / xaxisopts=( discreteopts=( tickvaluefitpolicy=splitrotate));
         boxplot x=_TREATMENT y=_HAIRGROWTHVALUE / group=_TREATMENT2 name='box' groupdisplay=Cluster;
      endlayout;
   endlayout;
endgraph;
end;
run;


/* Displying above graph */
proc sgrender data=DATA.TIDY_DATASET template=sgdesign;
dynamic _HAIRGROWTHVALUE="HAIRGROWTHVALUE" _TREATMENT="TREATMENT" _TREATMENT2="TREATMENT";
run;


/* ------------- Question 3 ------------- */
/* graph of age vs different traetments */
proc template;
define statgraph sgdesign;
dynamic _AGE _TREATMENT;
begingraph;
   entrytitle halign=center 'Graph 3: Age over different treatments ';
   layout lattice / rowdatarange=data columndatarange=data rowgutter=10 columngutter=10;
      layout overlay / xaxisopts=( discreteopts=( tickvaluefitpolicy=splitrotate));
         boxplot x=_TREATMENT y=_AGE / name='box' groupdisplay=Cluster;
      endlayout;
   endlayout;
endgraph;
end;
run;


/* Displying above graph */
proc sgrender data=DATA.TIDY_DATASET template=sgdesign;
dynamic _AGE="AGE" _TREATMENT="TREATMENT";
run;

/* fit the model with whole dataset - General Linear model */
PROC GLM data=DATA.TIDY_DATASET outstat=outStat plots=diagnostics;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age / solution clparm;
  ODS output fFitStatistics = modelFit ParameterEstimates =parEsts;
RUN;
Quit;

/* for above dataset: scatter plot with fitted linear regression with prediction interval(CLI) and  con
PROC sgplot data=DATA.TIDY_DATASET;
reg x= HAIRGROWTHVALUE y = age/CLM CLI;
```

```sas
scatter x= HAIRGROWTHVALUE y = age ;
RUN;


/* Assumption for whole dataset*/

PROC GLM data=DATA.TIDY_DATASET;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age;
  OUTPUT out=DATA.obsOut p=yhat r=resid lclm= lowerCI uclm = upperCI;
RUN;
Quit;


PROC UNIVARIATE data=DATA.obsOut normaltest;
var resid;
QQPLOT resid/normal (mu=est sigma=est);
RUN;



/*------------Fit the model with Luxuriant treatment------------*/
/* creating the dataset for Luxuriant treatment*/
DATA DATA.DATA_Luxuriant;
    SET DATA.TIDY_DATASET;
    WHERE Treatment = "Luxuriant";
RUN;

/* fit the model with Luxuriant dataset - General Linear model */
PROC GLM data=DATA.DATA_Luxuriant outstat=outStat_Luxuriant plots=diagnostics;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age / solution clparm;
  ODS output fFitStatistics = modelFit ParameterEstimates =parEsts;
RUN;
Quit;

/* for above dataset: scatter plot with fitted linear regression with prediction interval(CLI) and  con
PROC sgplot data=DATA.DATA_Luxuriant;;
reg x= HAIRGROWTHVALUE y = age/CLM CLI;
scatter x= HAIRGROWTHVALUE y = age ;
RUN;

/* Assumption for Luxuriant treatment*/
PROC GLM data=DATA.DATA_Luxuriant;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age;
  OUTPUT out=DATA.obsOut_Luxuriant p=yhat_l r=resid_l lclm= lowerCI_l uclm = upperCI_l;
RUN;
Quit;

PROC UNIVARIATE data=DATA.obsOut_Luxuriant normaltest;
var resid_l;
QQPLOT resid_l/normal (mu=est sigma=est);
RUN;


/*------------Fit the model with Placebo treatment------------*/
```

```
/* creating the dataset for Placebo treatment*/
DATA DATA.DATA_Placebo;
    SET DATA.TIDY_DATASET;
    WHERE Treatment = "Placebo";
RUN;

/* fit the model with Placebo dataset - General Linear model */
PROC GLM data=DATA.DATA_Placebo outstat=outStat_Placebo plots=diagnostics;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age / solution clparm;
  ODS output fFitStatistics = modelFit ParameterEstimates =parEsts;
RUN;
Quit;

/* for above dataset: scatter plot with fitted linear regression with prediction interval(CLI) and  con:
PROC sgplot data=DATA.DATA_Placebo;
reg x= HAIRGROWTHVALUE y = age/CLM CLI;
scatter x= HAIRGROWTHVALUE y = age ;
RUN;

/* Assumption for Placebo treatment*/
PROC GLM data=DATA.DATA_Placebo;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age;
  OUTPUT out=DATA.obsOut_Placebo p=yhat_p r=resid_p lclm= lowerCI_p uclm = upperCI_p;
RUN;
Quit;

PROC UNIVARIATE data=DATA.obsOut_Placebo normaltest;
var resid_p;
QQPLOT resid_p/normal (mu=est sigma=est);
RUN;


/*-------------Fit the model with BaldBeGone treatment-------------*/
/* creating the dataset for BaldBeGone treatment*/
DATA DATA.DATA_BaldBeGone;
    SET DATA.TIDY_DATASET;
    WHERE Treatment = "BaldBeGone";
RUN;

/* fit the model with BaldBeGone dataset - General Linear model */
PROC GLM data=DATA.DATA_BaldBeGone outstat=outStat_BaldBeGone plots=diagnostics;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age / solution clparm;
  ODS output fFitStatistics = modelFit ParameterEstimates =parEsts;
RUN;
Quit;

/* for above dataset: scatter plot with fitted linear regression with prediction interval(CLI) and  con:
PROC sgplot data=DATA.DATA_BaldBeGone;
reg x= HAIRGROWTHVALUE y = age/CLM CLI;
scatter x= HAIRGROWTHVALUE y = age ;
RUN;
```

```
/* Assumption for BaldBeGone treatment*/
PROC GLM data=DATA.DATA_BaldBeGone;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age;
  OUTPUT out=DATA.obsOut_BaldBeGone p=yhat_b r=resid_b lclm= lowerCI_b uclm = upperCI_b;
RUN;
Quit;

PROC UNIVARIATE data=DATA.obsOut_BaldBeGone normaltest;
var resid_b;
QQPLOT resid_b/normal (mu=est sigma=est);
RUN;

/*-------------Fit the model with SkinheadNoMore treatment-------------*/
/* creating the dataset for SkinheadNoMore treatment*/
DATA DATA.DATA_SkinheadNoMore;
    SET DATA.TIDY_DATASET;
    WHERE Treatment = "SkinheadNoMore";
RUN;

/* fit the model with SkinheadNoMore dataset - General Linear model */
PROC GLM data=DATA.DATA_SkinheadNoMore outstat=outStat_SkinheadNoMore plots=diagnostics;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age / solution clparm;
  ODS output fFitStatistics = modelFit ParameterEstimates =parEsts;
RUN;
Quit;

/* for above dataset: scatter plot with fitted linear regression with prediction interval(CLI) and  con
PROC sgplot data=DATA.DATA_SkinheadNoMore ;
reg x= HAIRGROWTHVALUE y = age/CLM CLI;
scatter x= HAIRGROWTHVALUE y = age ;
RUN;

/* Assumption for SkinheadNoMore treatment*/
PROC GLM data=DATA.DATA_SkinheadNoMore;
  CLASS age;
  MODEL HAIRGROWTHVALUE = age;
  OUTPUT out=DATA.obsOut_SkinheadNoMore p=yhat_s r=resid_s lclm= lowerCI_s uclm = upperCI;_s
RUN;
Quit;

PROC UNIVARIATE data=DATA.obsOut_SkinheadNoMore normaltest;
var resid_s;
QQPLOT resid_s/normal (mu=est sigma=est);
RUN;
```