MT5763 Assignment 2

210032607 | https://github.com/williamlogan77/MT5763_2_210032607

Abstract

While `Luxuriant` does not provide the greatest hair regrowth results, it provides an alternative to other brands on the market. `Luxuriant` allows for a more predictable result and can be argued that this makes it the best product out there. Similar to competing brands, `Luxuriant` works well for all ages without concern.

Introduction

This report will explore how the use of hair regrowth products can impact the length of hair regrown (in millimetres) after one month. The main focus will be to explore the efficacy of `Luxuriant` by comparing it to a placebo. Further probing into how the data is distributed will spotlight its effectiveness in comparison to the other products on the market. Along with `Luxuriant`, all brands will be considered when exploring how the age of the participant can effect the results of the hair growth to further reinforce reliability of results.

Analysis Description

The initial factor to consider is whether or not `Luxuriant` works. This can be verified by comparing the sample to the placebo (the null hypothesis being that there is no difference in the two samples). Performing a two sample t-test with the two samples can lead to mathematical proof of a difference in distribution. The two sample t-test can allow for comparison of the distribution by considering the confidence intervals, means and standard deviation for each and therefore provide evidence that the `Luxuriant` sample is not random.

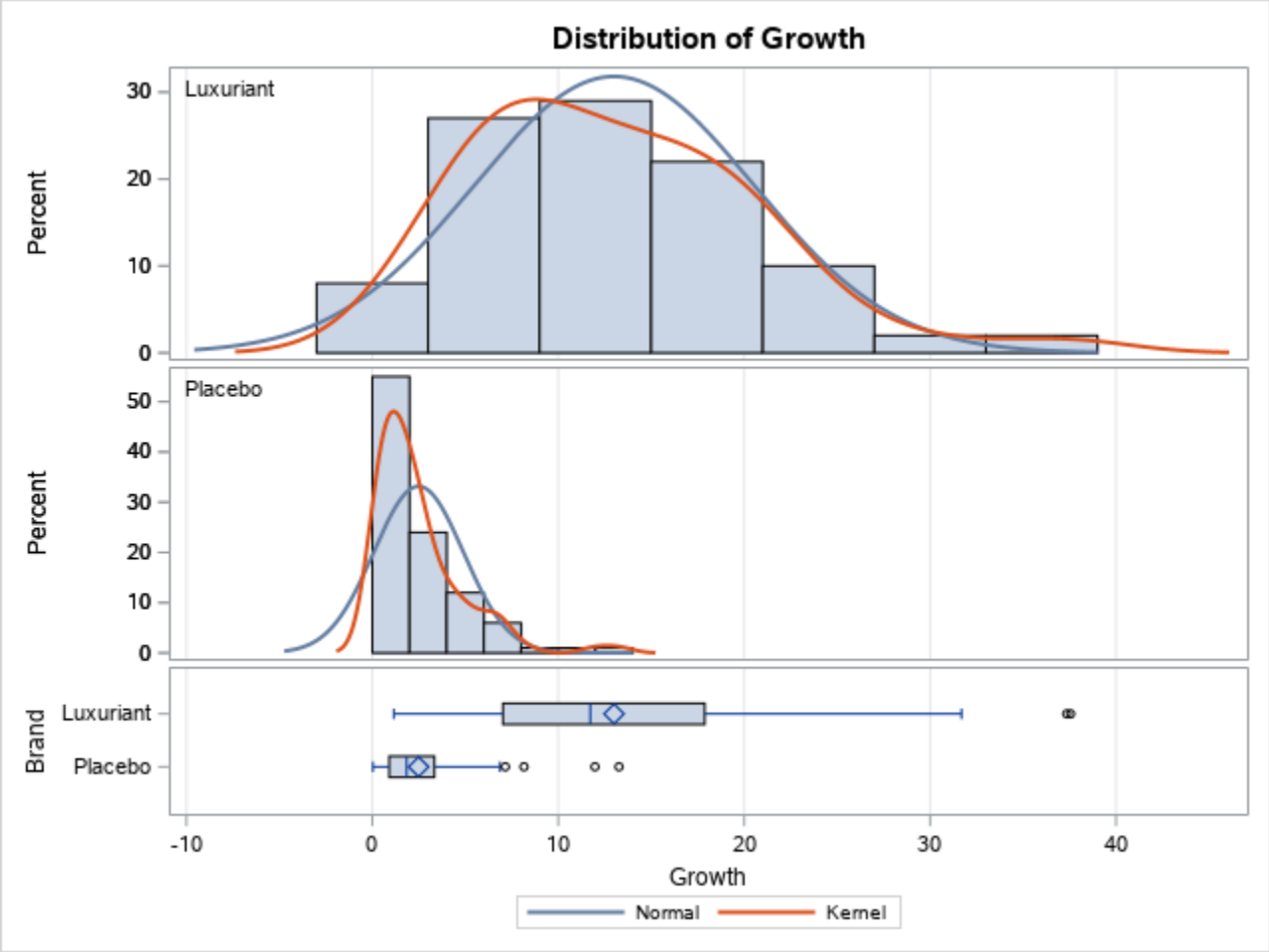
The next step is to compare how well `Luxuriant` works in comparison to the other competing products. This can be done using analysis of variance (ANOVA) , where each distribution is compared against one another and effectiveness of each brand can be discussed. Similar to determining if `Luxuriant` works, the mean and standard deviations will be analysed and weighed up against each brand.

Finally, to establish whether the age of the participant carries any effect, calculating Pearson's correlation coefficient can highlight how much of an impact one variable has on another. This should be done for all brands to highlight the effect multiple times. This will increase the reliability of the results as the conclusion does not depend solely on one sample which may be biased.

Results

Does `Luxuriant` work?

As discussed, this can be explored using a two sample t-test. The plotted histogram depicts a large discrepancy in the distributions, with `Luxuriant` looking normally distributed and a visually shifted mean in the positive direction. This provides strong evidence for `Luxuriant` having an effect on hair growth but needs to be consolidated quantitatively.

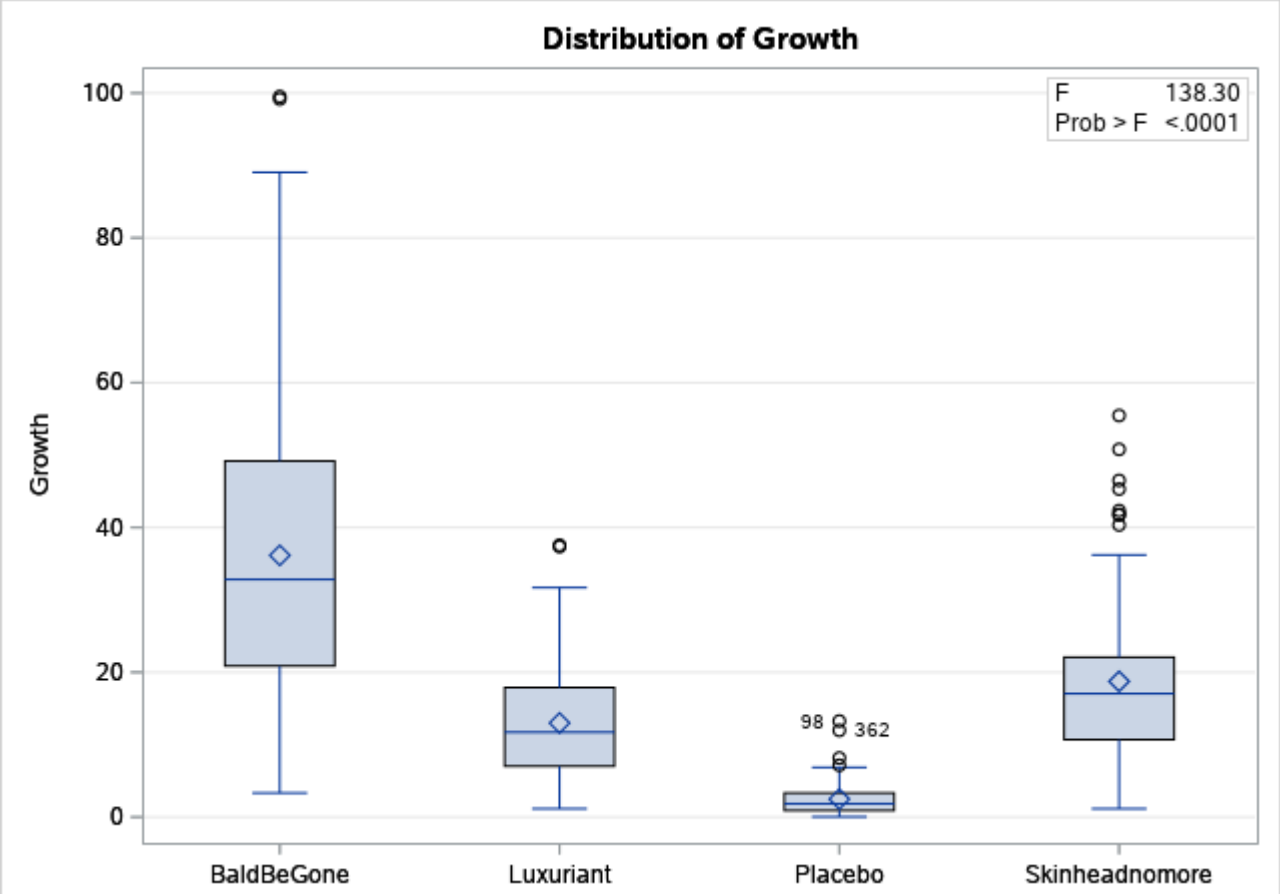


Brand	Mean	Std Dev	Minimum	Maximum	2.5% CL Mean	97.5% CL Mean
`Luxuriant`	12.9995	7.5258	1.1551	37.5585	11.5062	14.4928
Placebo	2.4851	2.4068	0.0353	13.2652	2.0075	2.9626
Diff (1-2)	10.5144	5.5871				

The table highlights that there is a difference of 10.51 in the means and there is no overlap in confidence intervals whatsoever. This shows that there is there must be an outside factor making the first sample (`Luxuriant`) deviate from the placebo - obviously this is the hair product. Therefore confirming that `Luxuriant` has an effect on hair regrowth.

How good is `Luxuriant`?

Performing an ANOVA test highlights each distribution against one another. In these results the placebo sample was kept to highlight the effectiveness of each brand. Again, analysis of the visual plot (in this case a box plot) of the distributions can illuminate the effectiveness of each brand.



Here it is much harder to draw conclusions from solely looking at a plot, however first impressions show that `BaldBeGone` is the provides the most regrowth out of the three products analysed, with `Luxuriant` and `Skinheadnomore` looking very similar in distribution.

Note The placebo distribution is so much smaller than the other distributions reinforcing the fact that these products **work**.

Extracting the means and standard deviations from the plot allows for more in-depth interpretation of the results.

Brand	Mean	Std Dev
`BaldBeGone`	36.1521295	19.7447168
`Luxuriant`	12.9994956	7.5258279
`Skinheadnomore`	18.7400940	11.0209323

While it is clear that `BaldBeGone` on average out-performs Luxuriant, there is a wider spread of results, as shown by it having the highest standard deviation. This is similar for `Skinheadnomore` where the mean is higher than `Luxuriant`'s, but the distribution has a larger standard deviation.

Finding the difference in the values relative to Luxuriant proves helpful.

Brand	Mean	Std Dev
`BaldBeGone`	+ 23.15263	+ 12.21889
`Skinheadnomore`	+ 5.740598	+ 3.495104

This reinforces the point of the other products have a wider spread of data, proving unreliable in comparison to `Luxuriant`. While the distribution is smaller, and there is less growth, `Luxuriant` is more consistent and accurate. Depending on the outlook, this could be interpreted as positive due to being able to give accurate directions and results to participants.

How does age effect hair regrowth?

Age is a very viable influencer on hair regrowth and to determine whether it has an impact on this sample, the Pearson correlation coefficient is calculated.

Brand	Pearson Correlation Coefficient
`BaldBeGone`	-0.10295
`Luxuriant`	-0.01074
`Skinheadnomore`	0.01697

With each coefficient being so small (and approximately 0), it is evident there is no impact of age on hair growth. This is reinforced by each sample having such a small coefficient and if one of the samples coefficient was large, more investigation would have to be done to confirm whether this was an outlier or not.

Discussion

To further reinforce the findings in this paper, the data must be more reliable. This would be achieved by obtaining multiple samples, potentially bigger samples and also a wider range of competing brands. Moreover, an initial clear objective of the company's vision for the product would have aided in verifying the goal that they set out to achieve.

To conclude, while `Luxuriant` does not provide as much potential hair regrowth as other brands, it provides hair regrowth that is consistent across a wide sample. One could argue that `Luxuriant` is the best brand on the market due to being accurate and reproducible as affirmed by the smallest standard deviation from all products. As discussed in this paper, the product works for all age ranges and there is no evidence to argue against this.

Appendix

Below is all of the scripts used in SAS to provide plots and data for this report.

```
/* This script reads in the data and modifies it accordingly for use across all tests */

/* First read in the data to be manipulated and store it as `REFILE` */
FILENAME REFFILE '/home/u59728190/Assignment 2/Data/Baldy.csv';

/* Convert the csv file into a table in SAS ready for manipulation */
PROC IMPORT DATAFILE=REFFILE DBMS=CSV OUT=WORK.IMPORT replace;
    GETNAMES=YES;
RUN;

/* Update the values for each hair regrowth to convert from inches to mm by multiplying by 25.4 */
proc sql NOPRINT;
    update work.import set Luxuriant=Luxuriant * 25.4;
    update work.import set Placebo=Placebo * 25.4;
    update work.import set BaldBeGone=BaldBeGone * 25.4;
    update work.import set Skinheadnomore=Skinheadnomore * 25.4;
    run;
```

```
/* The objective of this script is to create two separate tables that can later on
   be combined into one table with column names `brand`, `age` and `growth` */

/* Create a temporary table and add a unique number to each row */

data work.__tmp__;
    set WORK.IMPORT;
    ID=_n_;
run;

/* Transpose the data, renaming the columns the `brand` and `growth` and ignore
   all age variables for now */

proc transpose data=work.__tmp__ out=work.Stacked1(drop=_Label_
        rename=(col1=Growth)) name=Brand;
    var `Luxuriant` Placebo BaldBeGone Skinheadnomore;
    by ID;
run;

/* Delete the temporary table since it is no longer needed */

proc delete data=WORK.__tmp__;
run;

/* Recreate the temporary table with unique indexs for the next modification */

data work.__tmp__;
    set WORK.IMPORT;
    ID=_n_;
run;

/* Transpose the data again, but instead of putting the values back into the original
   columns, using the index to assign rows. This avoids putting the brand names back into
   columns and instead is now a new column. */

/* Similar to the first transpose, create a table with brand names and ages, ready
   for combining with the first table */

proc transpose data=work.__tmp__ out=work.Stacked2(drop=_Label_
        rename=(col1=Age)) name=Brand;
    var AgeLuxuriant AgePlacebo AgeBaldBeGone AgeSkinheadnomore;
    by ID;
run;

/* Delete the index table again */

proc delete data=WORK.__tmp__;
run;

/* End of transpose */

/* The objective of this script is to combine the two seperate tables into one */
```

```
/* First, sort both tables by the index so the same entries match and store it in
   two temporary tables */

proc sort data=WORK.STACKED1 out=work._tmpsort1_;
    by ID;
run;

proc sort data=WORK.STACKED2 out=work._tmpsort2_;
    by ID;
run;

/* Combine the two tables keeping `brand`, `growth` and `age` as columns and
   combine by the unique index number given to them */

data work.combined;
    merge _tmpsort1_ (keep=Brand Growth ID) _tmpsort2_(keep=Age ID);
    by ID;
run;

/* Delete the two temporary tables created */

proc delete data=work._tmpsort1_ work._tmpsort2_;
run;

/* Delete the ID column as it is no longer needed */

data work.combined(DROP= ID);
set work.combined;
run;

/* End of combination script */

/* For one of the tests, the placebo and Luxuriant data is just needed on its own.
   This is where we create a table that removes all other columns other than the
   ones needed */

proc sql noprint;
    create table work.placebo as select * from WORK.COMBINED where(Brand EQ
        'Luxuriant' OR Brand EQ 'Placebo');
quit;
```

```
/* This script perfoms a two sided t-test on the Luxuriant and Placebo data sets */

/* Initial "housekeeping" for SAS */
ods noproctitle;
ods graphics / imagemap=on;


/* Perform the two sided t-test with specified data sets and output a summary plot
   and a quantile-quantile plot */
proc ttest data=WORK.PLACEBO sides=2 h0=0 plots(only showh0)=(summaryPlot
        qqplot);
    class Brand;
    var Growth;
```

```
run;
```

```
/* This script performs an analysis of variance test */

/* Inital "housekeeping" for SAS */

Title;
ods noproctitle;
ods graphics / imagemap=on;

/* Call the glm function to perform the ANOVA test, using each brand as individual
   tests and the growth variable as the variable of interest. Output a range of
   statistical tests that were later cherry picked to add into the report */

proc glm data=WORK.COMBINED;
    class Brand;
    model Growth=Brand;
    means Brand / hovtest=levene welch plots=none;
    lsmeans Brand / adjust=tukey pdiff alpha=.05;
    run;
quit;
```

```
/* This script performs the correlation analysis */

/* Inital "housekeeping" for SAS */

ods noproctitle;
ods graphics / imagemap=on;

/* Sort the table by brand so there is no confusion on which data belongs to which brand */
proc sort data=WORK.COMBINED out=Work.SortTempTableSorted;
    by Brand;
run;

/* Perform the pearson correlation coefficient test of growth and age and plot a
   scatter graph of the variables. I decided the graphs were irrelevant and added
   nothing to report so they were omitted */

proc corr data=Work.SortTempTableSorted pearson nosimple noprob
        plots(maxpoints=none)=scatter(noinset ellipse=none nvar=10 nwith=10);
    var Growth;
    with Age;
    by Brand;
run;

/* Delete the temporary sorted table */
proc delete data=Work.SortTempTableSorted;
run;
```