MT5763: Software for Data Analysis - Assignment III

Alexander Ross

20-11-2022

Abstract

This report concerns the outcomes of a hypothetical clinical trial investigating the effect of a new drug to treat male-pattern baldness. Data was collected on patients' ages, their treatment product, and their resulting hair growth. We present the methods and findings from our analysis undertaken to assess the relative efficacy of the new drug. We describe the data wrangling process as well as the statistical tests, namely Wilcoxon Rank Sum Tests, Kruskal-Wallis Test, and Analysis of Covariance, used to quantify the effects of the variables treatment and age on the response. It was found that the new drug was effective beyond the placebo but that it was not the most effective drug on the market. Patient age was found to have had no significant effect on their measured hair growth.

Introduction

Androgenetic alopecia is the most common type of hair loss in men, in whom the condition is referred to as male-pattern baldness. More than half of men over the age of fifty will be affected by the condition (NLM, 2015). As such, there is high demand for products that can treat male-pattern baldness. Analysis of proposed drugs' effects in patients is essential to ensuring future availability of safe and effective treatments on the market.

The trial which this report concerns investigated the effect of a new drug, *Luxuriant*, on hair growth. Its effect was compared alongside two other drugs, *BaldBeGone* and *Hairy Goodness*, and a placebo. Patients in the study group consisted of bald men in middle-age to early elderhood from thirty to seventy-three years old. Patients were randomised to a treatment, had their head shaved, and after a month had hair growth measured. Growths for each treatment are shown in Figure 1.

Our analysis sought to quantitatively compare the results between treatments to determine whether *Luxuriant* had an effect beyond the placebo, whether *Luxuriant* was more effective than the two exising treatments, and to determine whether age had any bearing on hair growth. Analysis was conducted using the SAS Studio platform (SAS Institute Inc).

Methods

We first needed to wrangle the data to render it amenable to further analysis. The raw data consisted of four columns describing the hair growth (one for each treatment) and four corresponding columns for the patients' ages. We converted all lengths from inches by multiplying them by 25.4, as millimetres were desired in reporting. We then restructured the data to simplify analysis. Four tables ('LUXF', 'BBGF', 'HGF', and 'PLACF' in the appendix) were created from the main dataset by taking the hair growth and corresponding age columns for each treatment. To each of these two-column tables, a 'treatment' column was added. The ages were sorted to give the tables structure. These tables were combined to create one ('STACK') containing hair growth (mm), age (years), and treatment columns.

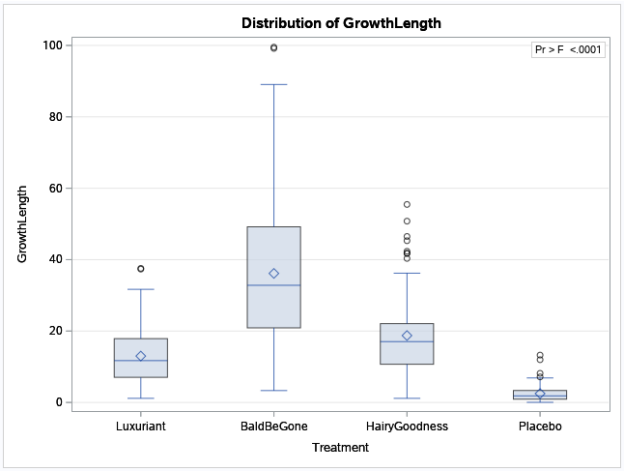

Figure 1: Exploratory plot showing the distribution of the growth lengths for each treatment with boxplots

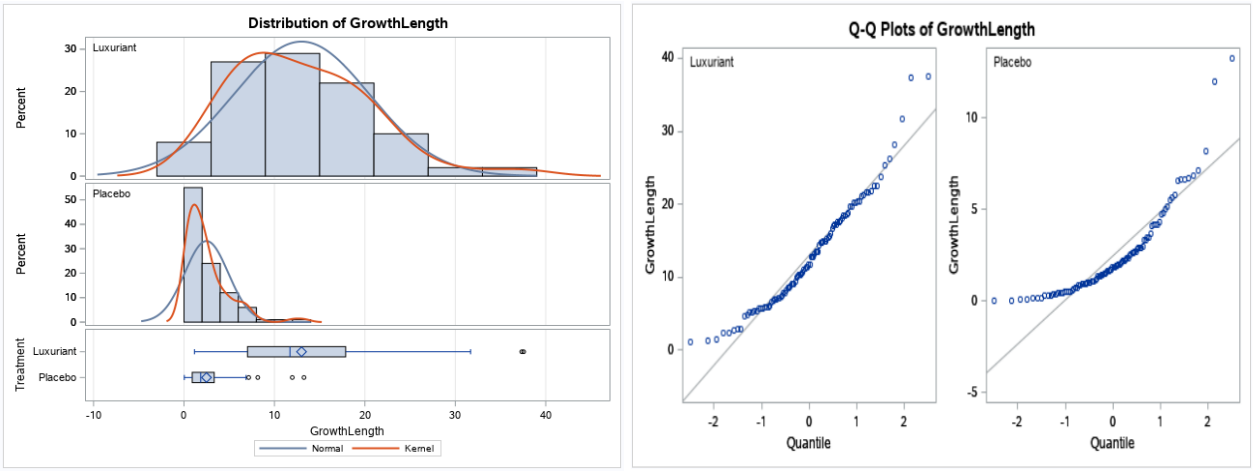

Figure 2: Normality assessment of the response for treatments *Luxuriant* and placebo. Although the distribution for *Luxuriant* is approximately Normal, it is far from Normal for the placebo; the QQ-plot shows that the response for placebo is not fitted well by the line representing Normality.

A two-sample t-test of hair growth between the placebo and *Luxuriant* was considered in answer to the first question. However, it was found that the response for the placebo was non-Normal as shown in Figure 2.

We therefore conducted a non-parametric test - the Wilcoxon Rank Sum Test - as this does not assume response Normality (SAS Institute Inc.). We were able to assume independence of the data based on the randomised study design and implicitly assumed equal sample variance.

To determine whether *Luxuriant* was more effective than the other two existing treatments, a Kruskal-Wallis Test was performed. This non-parametric test was used as we could not assume Normality of the response, but we wished to compare multiple samples. The test is an extension of the Wilcoxon Rank Sum Test for multiple samples. Thus we were able to compare the effect of *Luxuriant* to the effect of each other treatment separately by the Wilcoxon rank sums.

Finally, Analysis of Covariance (ANCOVA) was employed to ascertain whether patient age had any effect on hair growth, and to determine whether this effect differed between treatment groups. ANCOVA combines ANOVA and regression by testing for difference in the means of hair growth across levels of the categorical treatment variable, while controlling for the age effect (SAS Institute Inc.). The trial's study design meant this method was appropriate, and the assumption of independent observations could be made.

Results

Our Wilcoxon test resulted in a test-statistic of $z = 10.80$, with p-value $p =< 0.0001$, significant at the 1% level. Thus it can be concluded to reject the null hypothesis of the two samples coming from the same distribution.

From our Kruskal-Wallis Test we obtained a test-statistic of $\chi^2 = 258.3$ on three degrees of freedom, associated with a p-value of $p =< 0.0001$. We therefore conclude at the 1% level that the medians of the responses for each treatment are not equal. The Wilcoxon scores for the treatments reveal that *Luxuriant* is ranked third, with mean score 191.5. This suggests that *Luxuriant* is the third most effective treatment.

Wilcoxon Scores (Rank Sums) for Variable GrowthLength Classified by Variable Treatment

Treatment	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Luxuriant	100	19153.0	20050.0	1001.24922	191.530
BaldBeGone	100	31519.0	20050.0	1001.24922	315.190
HairyGoodness	100	23626.0	20050.0	1001.24922	236.260
Placebo	100	5902.0	20050.0	1001.24922	59.020

Figure 3: Wilcoxon scores for each treatment

Our ANCOVA test produced F-values for the treatment and age coefficients, as well as for the coefficient of its interaction with treatment:

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Treatment	3	59510.68149	19836.89383	137.92	<.0001
Age	1	117.82362	117.82362	0.82	0.3660
Age*Treatment	3	303.64415	101.21472	0.70	0.5503

Figure 4: Table summarising the ANCOVA results for the coefficients of treatment, age and age-treatment interaction. p-values for both age-related coefficients are insignificant.

Since the p-values associated with both F statistics are greater than 0.05, we conclude at the 5% level that age does not affect patients' hair growth, and that the effect of age does not differ between treatments. This is in line with our expectation from Figure 5.

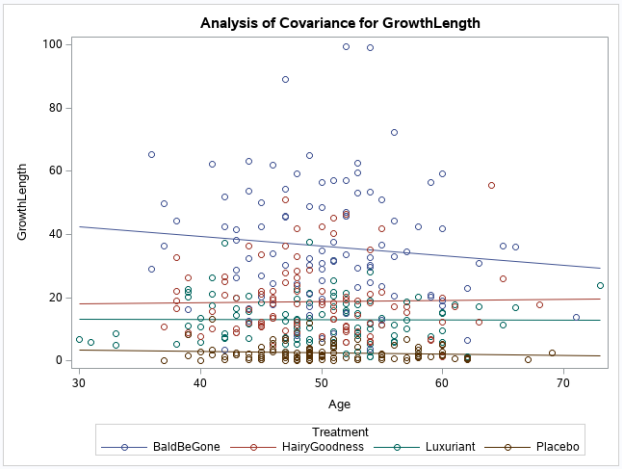

Figure 5: Plot comparing the effect of age on the response between the four treatments - the slopes for *Hairy Goodness*, *Luxuriant*, and the placebo all look to be about zero. *BaldBeGone* shows a slight negative relationship between age and hair growth, but this is statistically insignificant.

Discussion

Since our Wilcoxon test statistic was significant and positive, it may be deduced that *Luxuriant* had an effect significantly *greater* than the placebo, answering our first question. In answer to the second, we concluded that *Luxuriant* was not more effective than the existing treatments, as both other drugs resulted in significantly greater hair growth in patients. Age was found to not be relevant to any effect on the response, since both the age and age-treatment coefficients were insignificant. When considering future research on this topic, larger sample sizes may be considered to increase the accuracy of the conclusions. To improve the generalisability of the study's results, subjects of sundry ethnicities could be included in future efforts.

References

MedlinePlus [Internet],
Bethesda (MD): National Library of Medicine (US) [updated 2020 Jun 24],
Androgenetic Alopecia [updated 2015 Aug 1; cited 2022 Nov 22],
Available at https://medlineplus.gov/genetics/condition/androgenetic-alopecia/

SAS Institute Inc.,
SAS 9.1.3 Help and Documentation,
Cary, NC: SAS Institute Inc., 2002-2004

SAS Institute Inc.,
SAS 9.1.3 Help and Documentation,
Cary, NC: SAS Institute Inc., 2002-2004
Available at https://documentation.sas.com/doc/en/statcdc/14.2/statug/statug_npar1way_examples03.htm

SAS Institute Inc.,
SAS 9.1.3 Help and Documentation,
Cary, NC: SAS Institute Inc., 2002-2004,
Available at https://documentation.sas.com/doc/en/statcdc/14.2/statug/statug_glm_examples04.htm

Code Appendix

```
%web_drop_table(WORK.IMPORT);

FILENAME REFFILE '/home/u62758472/Software Assignment III/Baldy.csv';

/*import data*/
proc import DATAFILE=REFFILE
    DBMS=CSV
    OUT=BALDY;
    GETNAMES=YES;
run;

proc contents DATA=BALDY;
run;
%web_open_table(BALDY);

/*transform lengths from inches to mm*/
data MMData;
set BALDY;
Luxuriant=Luxuriant*25.4;
Placebo=Placebo*25.4;
BaldBeGone=BaldBeGone*25.4;
HairyGoodness=HairyGoodness*25.4;
run;

/*create table for Luxuriant data sorted by age*/
proc sort DATA = MMData;
by AgeLuxuriant;
run;

data Lux;
set MMData;
keep AgeLuxuriant Luxuriant;
run;

data LuxF;
set Lux;
Product="Luxuriant";
rename Luxuriant=GrowthLength;
rename AgeLuxuriant=Age;
run;

/*create table for BaldBeGone data sorted by age*/
proc sort DATA = MMData;
by AgeBaldBeGone;
run;

data BBG;
set MMData;
keep AgeBaldBeGone BaldBeGone;
run;

data BBGF;
```

```
set BBG;
Product="BaldBeGone";
rename BaldBeGone=GrowthLength;
rename AgeBaldBeGone=Age;
run;

/*create table for HairyGoodness data sorted by age*/
proc sort DATA = MMData;
by AgeHairyGoodness;
run;

data HG;
set MMData;
keep AgeHairyGoodness HairyGoodness;
run;

data HGF;
set HG;
Product="HairyGoodness";
rename HairyGoodness=GrowthLength;
rename AgeHairyGoodness=Age;
run;

/*create table for Placebo data sorted by age*/
proc sort DATA = MMData;
by AgePlacebo;
run;

data PLAC;
set MMData;
keep AgePlacebo Placebo;
RUN;

data PLACF;
set PLAC;
Product="Placebo";
rename Placebo=GrowthLength;
rename AgePlacebo=Age;
run;

/*create combined table of the four treatments*/
data STACK;
length Product $ 15;
set LUXF BBGF HGF PLACF;
run;

/*create table including only treatments to be compared by t-test*/
data TSTACK;
length Product $ 15;
set LUXF PLACF;
run;

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

```
/* Test for Normality */
proc univariate data=WORK.TSTACK normal mu0=0;
    ods select TestsForNormality;
    class Product;
    var GrowthLength;
run;

/*1 - t-test */
proc ttest data=WORK.TSTACK sides=2 h0=0 plots(showh0);
    class Product;
    var GrowthLength;
run;

/*Wilcoxon Rank Sum Test*/
proc npar1way data=WORK.TSTACK WILCOXON;
    class Product;
    var GrowthLength;
run;

/*2 - non-parametric ANOVA to compare treatment pairs*/
proc npar1way data=WORK.STACK;
    class Product;
    var GrowthLength;
run;

/*3 - ANCOVA to compare effect of age on response across treatments*/
proc glm data=WORK.STACK;
    class Product;
    model GrowthLength=Product Age Age * Product;
    lsmeans Product / adjust=tukey pdiff alpha=.05;
quit;
```