Luxuriant: The Newest Treatment Ready for the Hair Growth Market

Joshua Arrabaca - 220029955

2023-11-17

1 Abstract

This paper demonstrates statistical analysis performed in Python on a dataset containing growth performance of 3 hair treatment brands, plus a Placebo.

The analysis showed that age did not have an effect on hair growth, and that Luxuriant performed significantly well against the Placebo, with an average growth of 13.0mm. However, it was was outperformed by the established brands HairyGoodness and BaldBeGone.

Still, Luxuriant's performance over the Placebo by 10.51mm indicates that it can be released to market, although future trials should separate placebo testing and existing treatment testing.

2 Introduction

Analysis was conducted on hair growth and age data for four hair treatments: Luxuriant, a new product; HairyGoodness and BaldBeGone, both established brands; and a Placebo. The means and 95% confidence intervals were computed, and ANOVA and Ordinary Least Squares modeling performed to answer the questions:

1. Is there an effect of Luxuriant above and beyond the placebo?
2. Is Luxuriant more effective than the existing treatments on the market?
3. Is age relevant to any effect?

The results show that Luxuriant works significantly better than the Placebo, with an average growth of 13.0mm versus 2.49mm. However, HairyGoodness and BaldBeGone outperformed Luxuriant with averages of 36.15mm and 18.74mm respectively. Age was not found to have a significant effect.

3 Methods

A dataset containing 100 rows of hair growth data was prepared in Python. Growth was converted from inches to millimeters (via `pandas.apply()` and a lambda function `x: x*25.4`), and age rounded off with `.round()`, and casted `.astype(int)`. This preliminary dataset (Table 9) was used for ANOVA and t-tests.

This was further tidied to have one row per participant with columns "Brand", "Growth" and "Age". Temporary datasets were created for each brand, then appended together with `pd.concat()`. The full dataset previewed in Table 1 has 400 rows and 3 columns, and was used for visualisations and OLS linear modeling.

Table 1: Head of Tidy Dataframe

	Brand	Growth	Age
0	Luxuriant	20.122736	59
1	Luxuriant	15.377107	61
2	Luxuriant	17.638841	54
3	Luxuriant	1.379751	63
4	Luxuriant	37.558535	50

Plots were created using `sns.scatterplot`, `.boxplot` and `.histplot` with the parameter `(hue=brand)` to visualise the groupings.

To determine differences between groups, the means were calculated along with 95% confidence intervals using `statsmodels.stats.api`. ANOVA tests were also conducted for all brands, while t-tests were conducted between Luxuriant and Placebo using `scipy.stats`. Participants' age was checked first (i.e., if they had the same age, then it's suitable to compare their hair growth).

To determine if age was relevant, a linear model was created using `statsmodels.formula.api`, and the p-values were evaluated.

For this analysis, the following are assumed: the samples are independent; the samples come from a normally distributed population; and the standard deviations are the same (Heumann et al., 2022).

4 Results

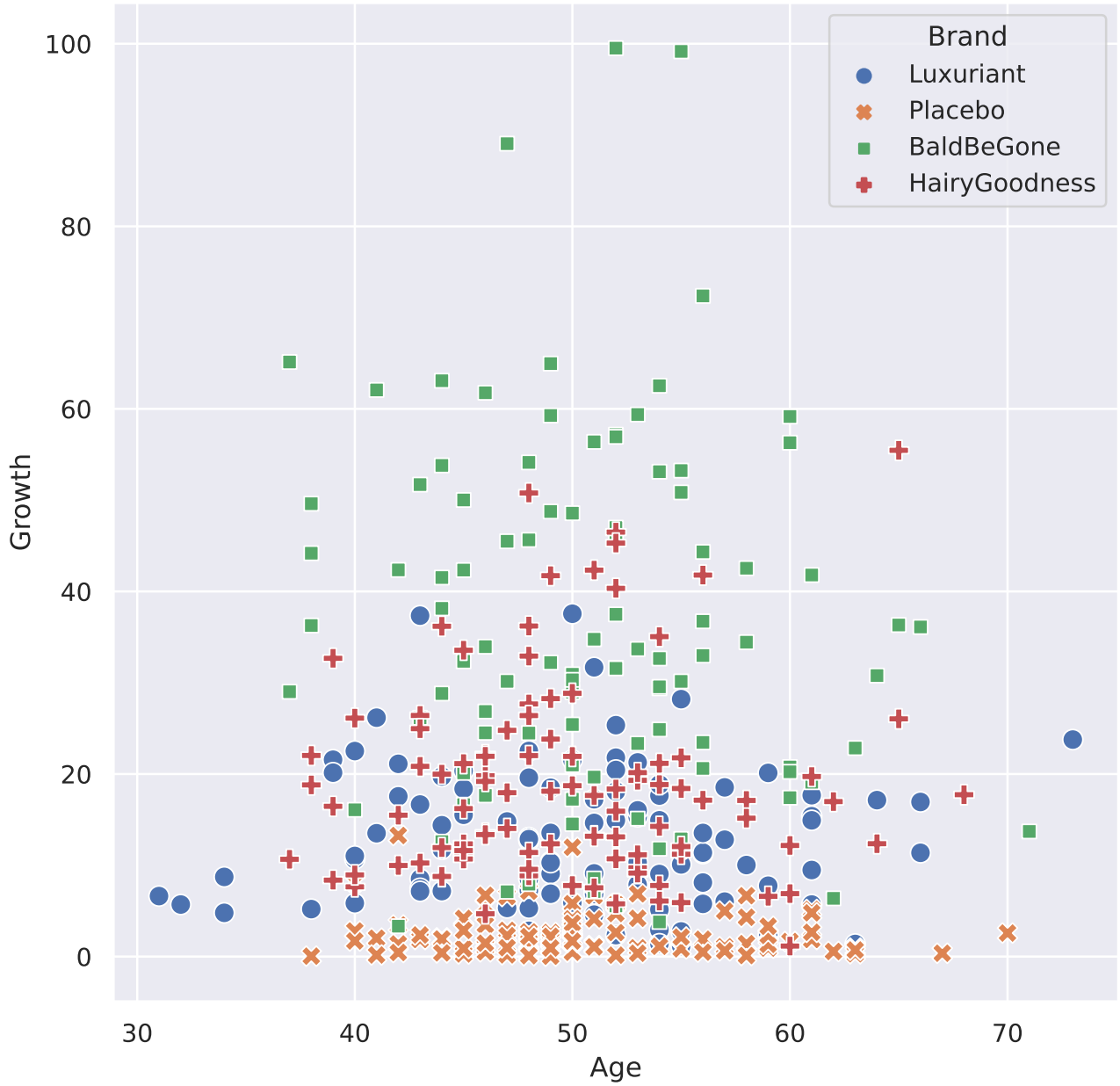

Figure 1: Scatter Plot of Age vs Hair Growth

A visual assessment in Figure 1 does not show any particular pattern for age vs growth. However, the groupings show that Placebo had the lowest effect on growth at up to 20mm, and BaldBeGone had the highest from 20mm to above 60mm. Luxuriant performed averagely from 20mm to 40mm.

The mean ages were compared in Figure 2 and showed similar quantiles (i.e., the 25^{th} to 75^{th} quantiles cover ages 45 to 55). Table 2 shows the mean age is around 50, and the 95% confidence intervals in Table 3 exhibit overlapping ranges (from 49 to 51). Figure 4 (in the Appendix) shows similar results.

The visualisations, similar mean ages and overlapping confidence intervals show that group ages are not statistically different. This makes it suitable to compare hair growth between groups.

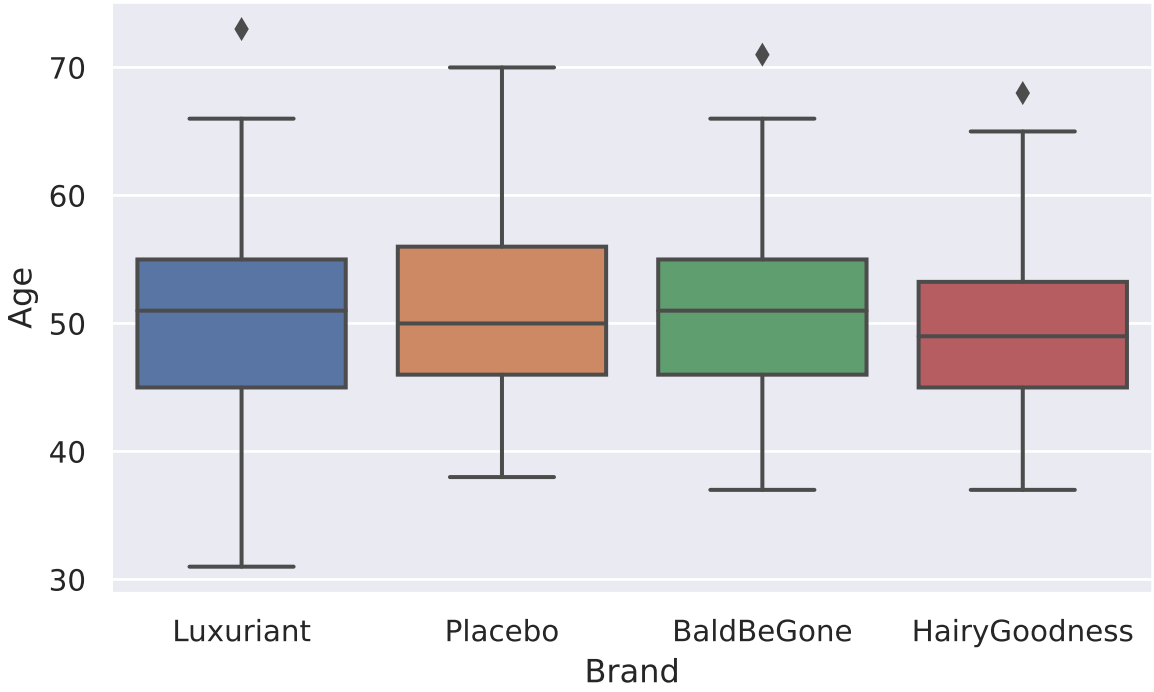

Figure 2: Box Plot of Participants' Age

Table 2: Mean Age by Brand

Brand	Age
Luxuriant	50.29
Placebo	50.96
BaldBeGone	50.93
HairyGoodness	49.73

Table 3: 95% Confidence Intervals for Age

Names	Lower Bound	Upper Bound
Luxuriant	48.78	51.8
Placebo	49.64	52.28
BaldBeGone	49.6	52.26
HairyGoodness	48.45	51.01

Table 4 shows varying average hair growth by brand, with Placebo having the smallest growth at 2.49mm and BaldBeGone the highest at 36.15mm. The boxplot in Figure 3 also shows the 50^{th} quantitles are different for all groups, with Luxuriant outperforming Placebo which has the lowest and narrowest growth, although BaldBeGone has the widest range. Figure 5 histogram also shows different distributions.

The Confidence Intervals in Table 5 show no overlaps between any group for hair growth. All these strongly suggest that there is a significant difference in growth among brands.

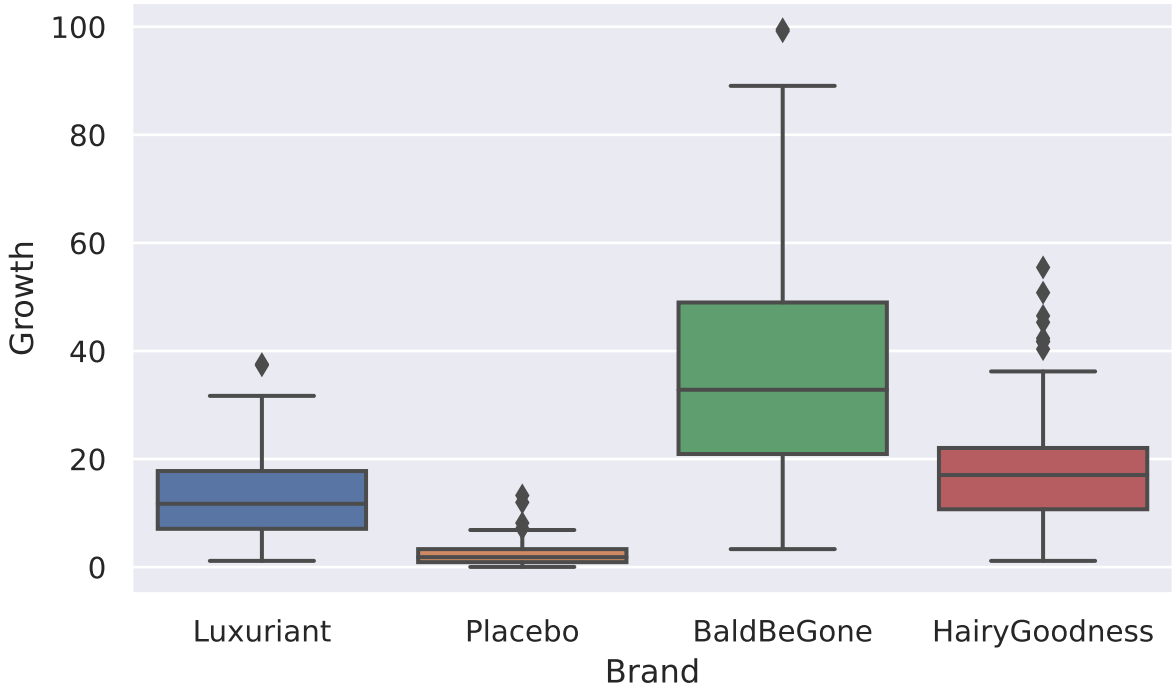

Figure 3: Box Plot of Average Hair Growth

Table 4: Means Hair Growth by Brand

Brand	Growth
Luxuriant	13.0
Placebo	2.49
BaldBeGone	36.15
HairyGoodness	18.74

Table 5: 95% Confidence Intervals for Growth

Names	Lower Bound	Upper Bound
Luxuriant	11.51	14.49
Placebo	2.01	2.96
BaldBeGone	32.23	40.07
HairyGoodness	16.55	20.93

To verify the null hypothesis (H_0: There is no difference in growth between groups), Table 6 shows that growth has a significant p-value of almost zero (unlike age which is not significant at the 0.1 level).

For the other null hypothesis (H_0: There is no difference in growth between Luxuriant and Placebo), Table 7 also shows a significant p-value of close to zero.

Table 6: ANOVA Results for All Brands

Test Variable	Statistic	P-Value
Age	0.730105	0.534507
Growth	138.297504	0.000000

Table 7: T-Test Between Luxuriant and Placebo

Test Variable	Statistic	P-Value
Age	-0.662467	0.508442
Growth	13.307183	0.000000

These results show that both null hypotheses can be rejected. Therefore, there is a significant difference in hair growth between groups, and between Luxuriant and Placebo.

The OLS linear model can be described as:

$$\hat{Growth} = 39.89 + Brand - 0.07Age \tag{1}$$

The resulting p-value for age (Table 10) is non-significant. This implies that age is not a determining factor in hair growth.

Moreover, the model's adjusted R-squared value is only 0.5. There are further evidence of non-normality; the residuals plot (Figure 6) looks non-normally distributed, and the QQ Plot (Figure 7) shows skewed data. The Shapiro-Wilk test also yields W=0.91 and p-value=9.93e-15, confirming non-normal distribution (Navarro, 2015).

Although the model is underpredicting, it can still be used to illustrate growth by brand. Table 8 shows that Placebo has the least growth while BaldBeGone has the highest, with Luxuriant in the middle.

Table 8: Predicted Growth For Age=50

	Brand	Predicted Growth
0	Luxuriant	13.02
1	Placebo	2.56
2	BaldBeGone	36.22
3	HairyGoodness	18.72

5 Conclusion and Recommendation

The analysis shows that Luxuriant has a significant growth effect over the Placebo by 10.51mm (13.0mm for Luxuriant versus 2.49mm for Placebo). However, Luxuriant is less effective than BaldBeGone and HairyGoodness which average 36.15mm and 18.74mm respectively. Age was not found to be a relevant effect.

While Luxuriant's performance is not as large as the existing brands, it still has a significant effect over the Placebo. It can therefore be released with the tested claim that it improves hair growth regardless of age.

To improve clinical trials, there should be separate test phases (Hulley et al., 2013); one phase for testing against the Placebo (phase 2), another to test against existing hair treatments (phase 3), and an optional study after Luxuriant's introduction to the public (phase 4).

6 References

Heumann, C., Schomaker, M., & Shalabh. (2022). *Introduction to statistics and Data Analysis: With exercises, solutions and applications in R* (Second Edition). Springer International.

Hulley, S. B., Cummings, S. R., Browner, W. B., Grady, D. G., & Newman, T. B. (2013). *Designing Clinical Research* (Fourth edition). Lippincott, Williams & Wilkins.

McKinney, W. (2022). *Python for Data Analysis. Python for Data Analysis*, Third Edition. https://wesmckinney.com/book/

Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners*. Open Textbook Library. https://open.umn.edu/opentextbooks/textbooks/559

Seaborn User Guide and Tutorial. User guide and tutorial - seaborn 0.13.0 documentation. (2023). https://seaborn.pydata.org/tutorial.html

SciPy Documentation. SciPy documentation - SciPy v1.11.3 Manual. (2023, September 27). https://docs.scipy.org/doc/scipy/

7 Appendix

Table 9: Head of Dataframe with Conve

	Luxuriant	Placebo	BaldBeGone	HairyGoodness	AgeLuxuriant	AgePlacebo	AgeBaldBeGone	AgeHairyGoodness
0	20.122736	2.047563	7.125638	13.230100	59	44	51	51
1	15.377107	0.957729	57.176273	10.711428	61	53	52	52
2	17.638841	2.467823	36.108605	46.497552	54	48	66	52
3	1.379751	5.521011	17.651599	21.165904	63	46	46	54
4	37.558535	4.992816	19.100094	7.640771	50	57	61	40

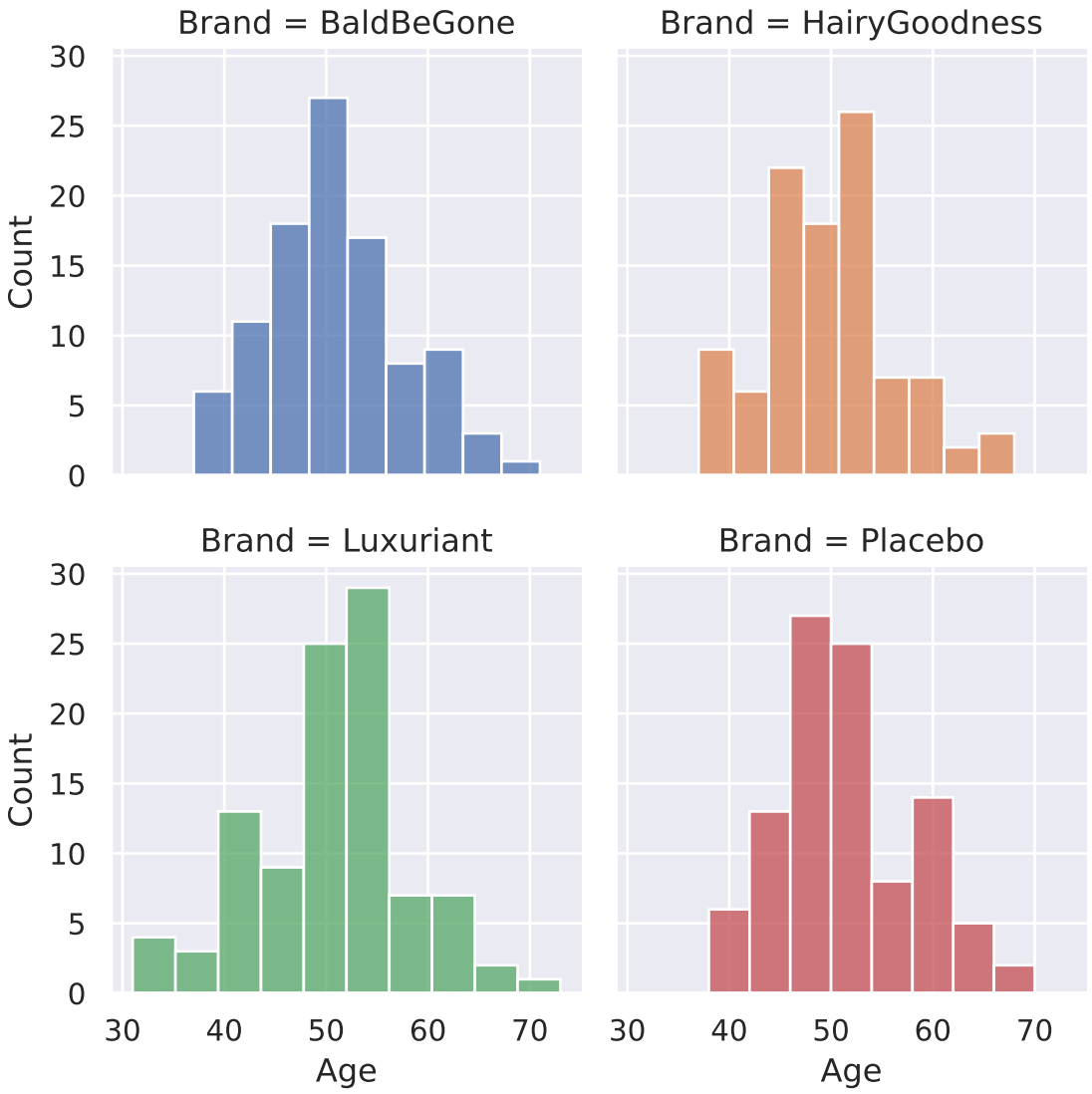

Figure 4: Histograms of participants' age by brand, showing fairly similar distributions

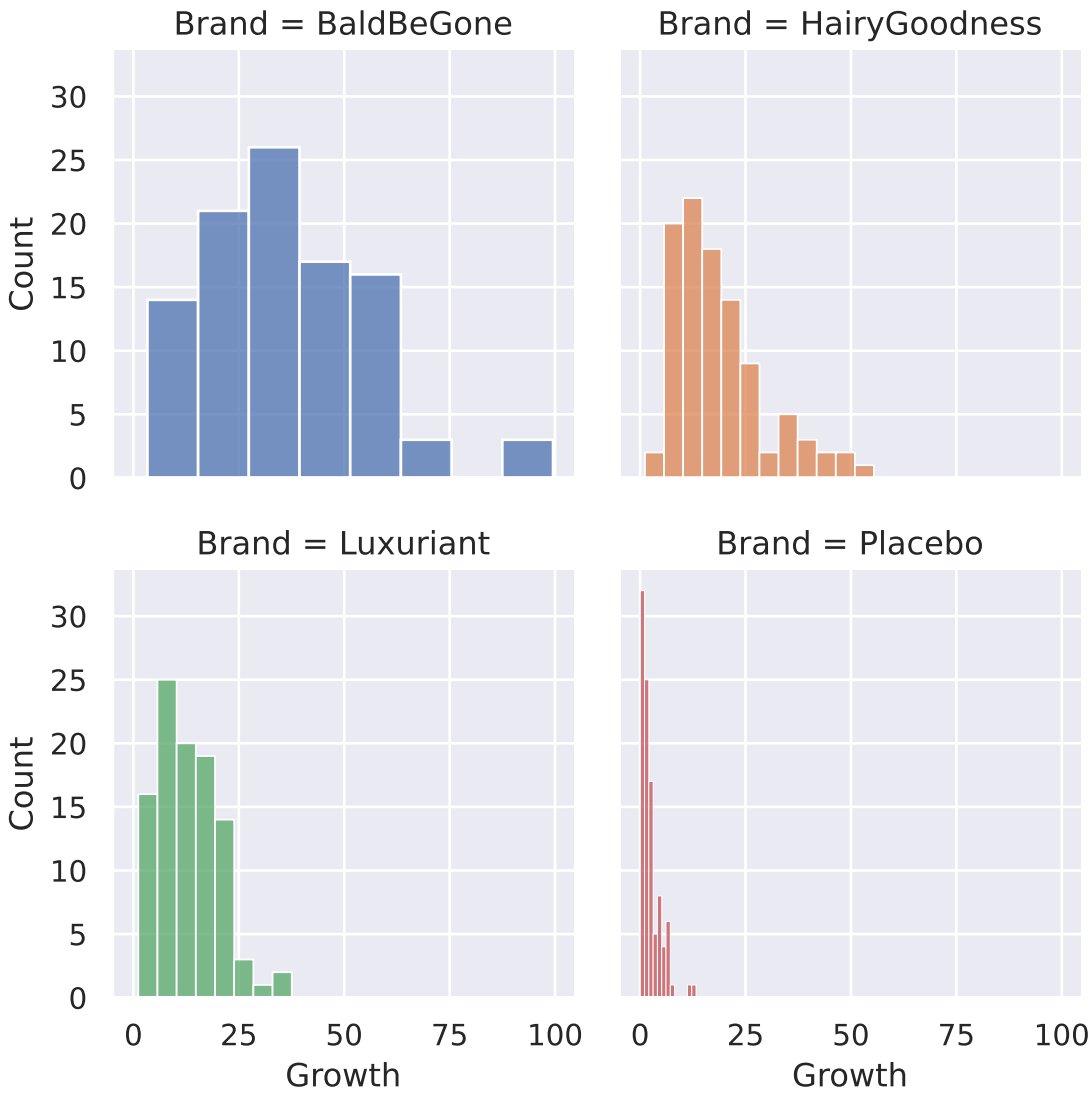

Figure 5: Histograms of hair growth by brand, showing signficantly varying distributions

Table 10: OLS Summary Table showing the Adj R-squared value and Age p-value

Model:	OLS	Adj. R-squared:	0.508
Dependent Variable:	Growth	AIC:	3126.7791
Date:	2023-11-17 07:52	BIC:	3146.7365
No. Observations:	400	Log-Likelihood:	-1558.4
Df Model:	4	F-statistic:	103.8
Df Residuals:	395	Prob (F-statistic):	2.41e-60
R-squared:	0.513	Scale:	143.55

	Coef.	Std.Err.	t	P> |t|	[0.025	0.975]
Intercept	39.8905	4.6262	8.6228	0.0000	30.7956	48.9855
Brand[T.HairyGoodness]	-17.5001	1.6976	-10.3085	0.0000	-20.8377	-14.1626
Brand[T.Luxuriant]	-23.1996	1.6953	-13.6846	0.0000	-26.5326	-19.8667
Brand[T.Placebo]	-33.6649	1.6944	-19.8686	0.0000	-36.9960	-30.3337
Age	-0.0734	0.0877	-0.8366	0.4033	-0.2459	0.0991

Omnibus:	111.514	Durbin-Watson:	1.977
Prob(Omnibus):	0.000	Jarque-Bera (JB):	495.130
Skew:	1.137	Prob(JB):	0.000
Kurtosis:	7.953	Condition No.:	398

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

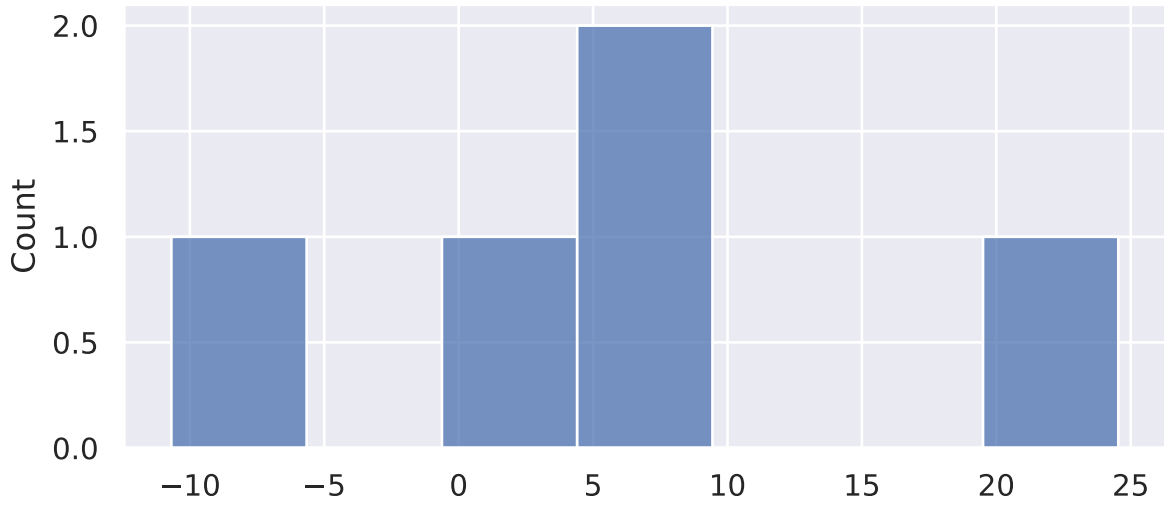

Figure 6: Histograms of OLS model residuals showing non-normal distribution

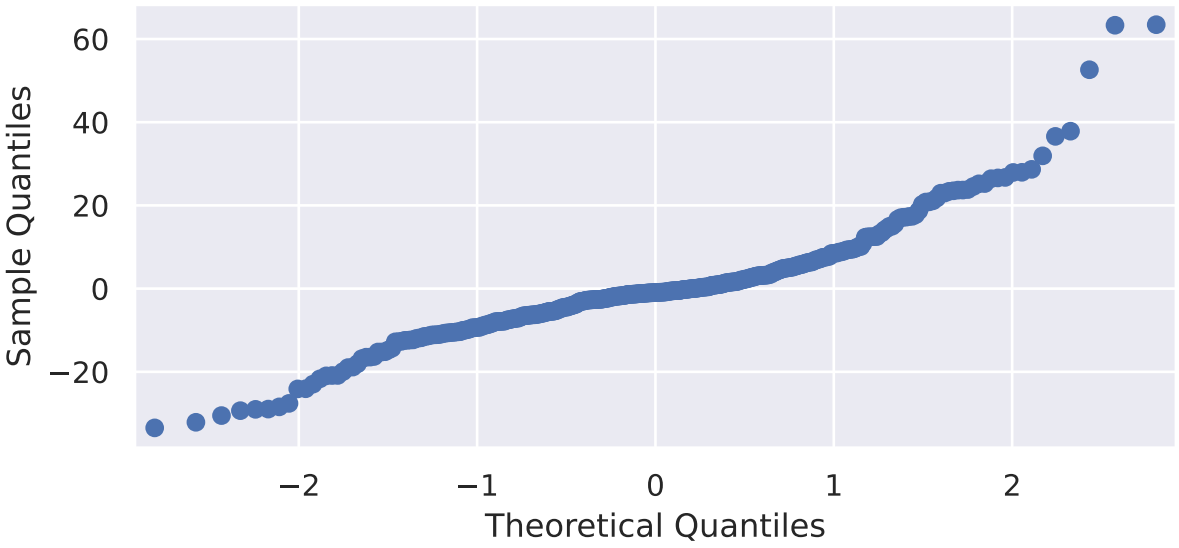

Figure 7: QQ Plot showing slightly downward skewed data

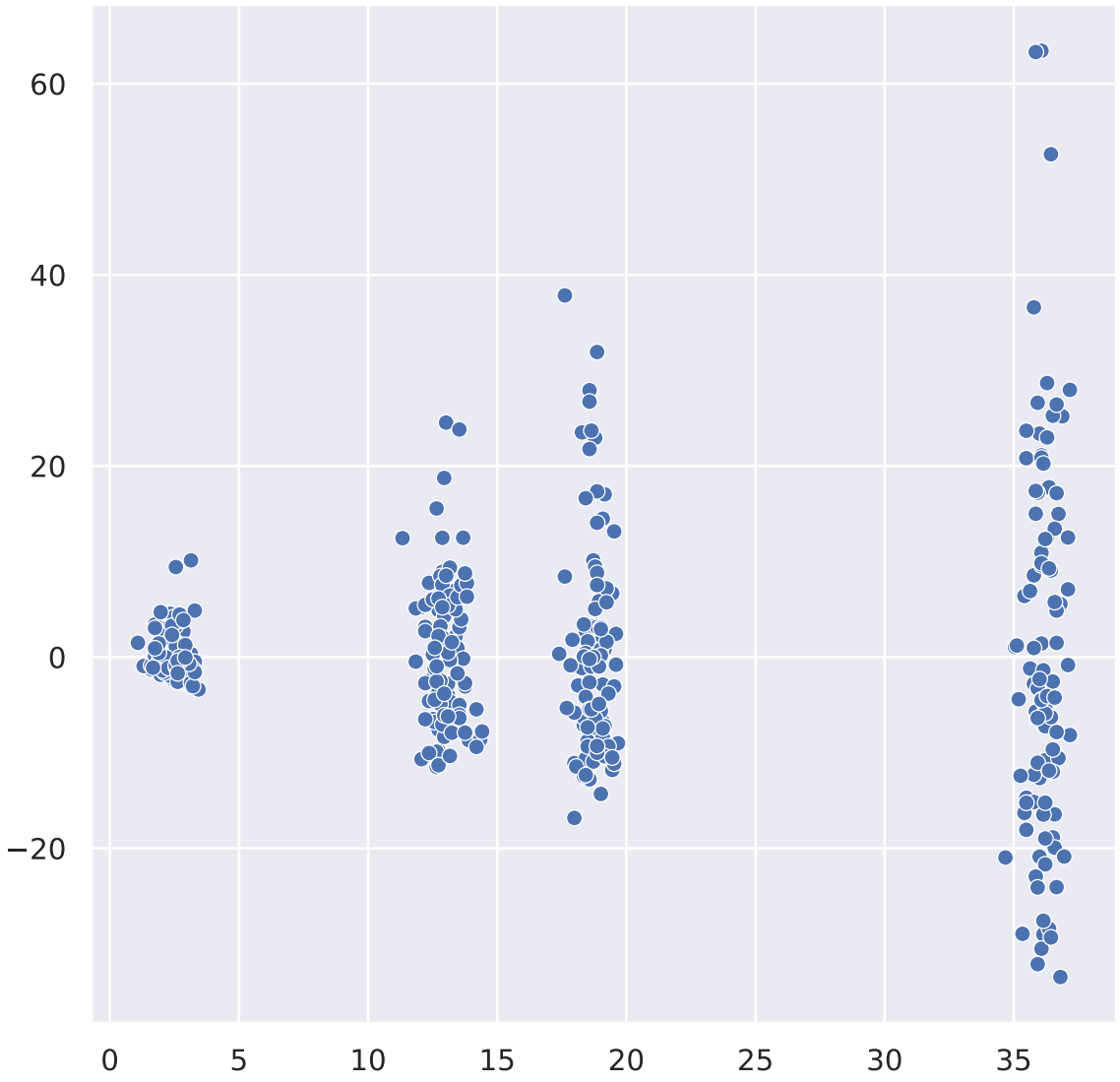

Figure 8: Residuals vs Fitted Values showing dense points in the lower residual values