Does Wearing Make-Up Impact How Much a Waiter Receives in Tips? A Shocking Discovery

Barinder Thind[1] and Kevin Multani[2]

Abstract— **This work is too important for an abstract.**

I. Introduction

In this short report, I highlight an atrocious problem that is tireless in its efforts to plague society; that is, the impact of "make-up" on the psyche of regular restaurant-goers is significant enough to alter the trajectory of entire lives. In fact, make-up serves as an invaluable tool in the arsenal of any waiter (although, I suppose maybe that is not true) when the ultimate goal is to make some damn money. Through an *extensive* statistical analysis, I - essentially - **prove** that society seems to only care about aesthetics and what is underneath is brushed away as if it had the most minimal conceivable value comprehend-able.

In Section II, I first highlight the super secret data used in this groundbreaking analysis; next, plots are presented - some of which which paint a deceiving picture of the status-quo. Then, a quick overview of the statistical approaches taken is given. In section III, the terrifying results are given, complete with ridiculous commentary complementary of yours truly. Lastly, in Section IV, the conclusion is driven home under comedic pretenses..

II. Methodology

A. Data Description

In a world where heroes are seldom seen, one woman of war ventured out into the wild fully understanding of the risks: her goal? to collect concrete information on this hideous reality. Like any fearsome warrior, she escaped relatively unscathed and provided, what is now, priceless data.

There are 22 observations and a number of variables (12, to be exact) that provide information on the amount of money made (in tips) before and after tip-out[1], the time at which the shift began and ended, and whether or not she wore make-up. A couple of features were also engineered; namely, the length of the shifts and the overall payout for the day.

B. Exploratory Data Analysis

The obvious initial approach when comparing a factor with a continuous response is to look at the box-plots for the data[2]. First, we explore whether there is some relationship between the variables related to tipping and the given state (make-up or no make-up).

*This work was not supported by any organization

[1]B. Thind is a Masters of Science student in the Department of Statistics at Simon Fraser University, British Columbia

[2]K. Multani is a PhD Student in the Department of Physics at Stanford University, California. K. Multani also contributed very little to the analysis done here.

[1]The net difference being how much she took home

[2]Albeit, when there is this little data, you may even be better off just plotting all the data

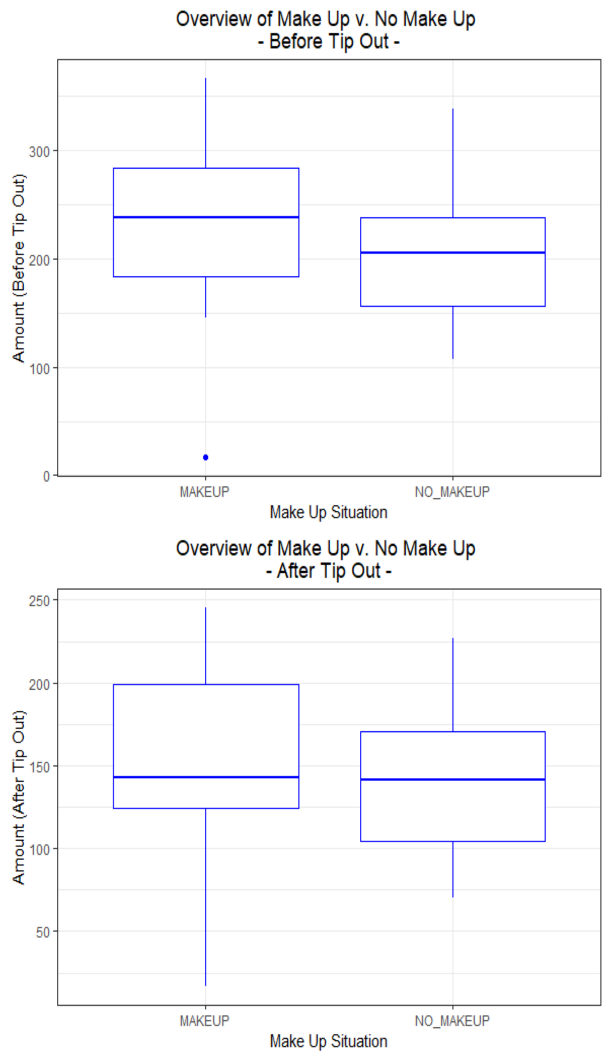

Fig. 1. Boxplots comparing tipping outcomes between different states

Looking at these plots, there doesn't seem to be any real indicator that there is a statistically significant difference between the two states and how much the waiter makes however, the larger concern has to do with the variances involved - it seems that when using make-up, there is almost surely a skew towards greater monetary gains[3]. Also, note that there seems to be one extreme outlier (or perhaps not) in the make-up group - whether this says something about the make-up ability, the type of customer, or the number of customers during that shift is left to the interpretation of the reader.

Next, we can look to make some initial impressions on the effect of shift lengths on the final tipout and whether or not this differed by state:

Fig. 2. The linear curves were fit using regular multiple linear regression models

Wow, what a result. This plot seems to indicate that when you work longer, there is a much steeper increase in tips for someone wearing make-up vs. not and this is DESPITE the high leverage point referred to earlier[4]. As an aside, it is important to note that the regression lines here serve as visual aids much more so than they reveal some underlying pattern about the provided data.

We can also look to explore the relationship between shift lengths and the amount of money made - in particular, the violin plot in Fig 3. will display how much is made in tips and how it is spread across some shift lengths.

The way to interpret this plot is to look at the 'width' of the 'violins' and

[3]This is further exemplified in the bootstrapping results

[4]The 'outlier' from before which can be seen in the bottom left hand corner of Fig 2. Note also that this data point only affects the red (make-up) regression line

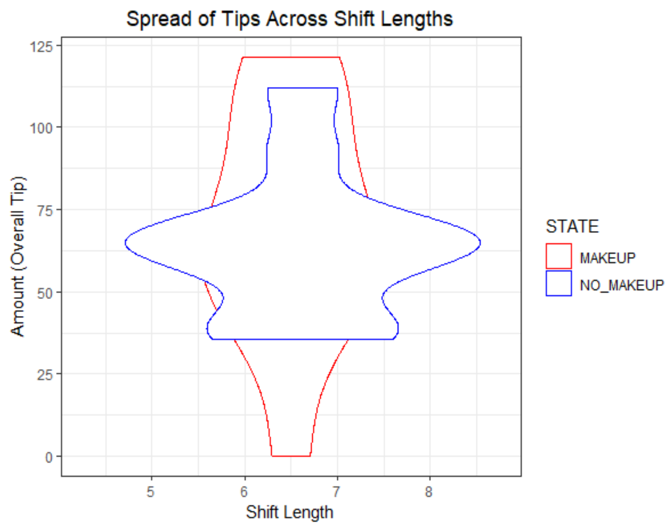

Fig. 3. Violin Plot highlighting the relationship between shift length and tip amount separated by the state of the waiter

see how they spread at a particular tip amount. In other words, think about this plot as a mirror image and only look at one half of it - this is a combination of a kernel density plot and a boxplot. For example, the protruding point for the 'no make-up' state around the 60$ tip mark indicates that there is a higher probability of that amount (60$) occurring across the shift lengths. The key takeaway is that when we consider to look at the larger tip amounts (moving up the y axis), there is a higher probability of them occurring (as shifts are longer) for the make-up state v. the non make-up state.

C. Analysis of Variance with Baseline Constraints

The Analysis of Variance (ANOVA) model is the classic scientific approach for the given problem. Generally the hypothesis are written as:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_p$$
$$H_1 : \text{At least one } \mu \text{ is different.}$$

For some set of p means. This statistical test follows an F-distribution as the mathematics would indicate that it is the ratio of the sum of squared normally distributed random variables.

We can restate the above model in its regression form as:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Where the response is the tip amount, the μ refers to the baseline constraint on the model[5] or the mean of the 'null' model and α is an indicator variable that highlights which groups' mean we are referring to; and of course, ϵ is the classic error term. Lastly, we need to consider the assumptions of a one-factor ANOVA model - they are all the same as the least squares regression (which makes sense because we have written it as a regression model) in that we expect normal residuals, constant variance, and good data collecting services[6]

If none of this makes sense or you don't care to understand what all this means, don't worry about it.

D. Bootstrapping

In this section, I provide a quick overview of the non-parametric bootstrapping algorithm and why it is useful. In short, we can summarize the algorithm as follows: Let $\hat{F}_n$ be the empirical distribution of your collected data. Then,

1) Draw $X_1^*, ..., X_n^*$ from $\hat{F}_n$
2) Compute $T_n^* = g(X_1^*, ..., X_n^*)$
3) Repeat steps 1 and 2, B times, to get $T_1^*, ..., T_n^*$

Where T is the particular statistic (for example, the mean) that you are looking to analyze. And, from this point, you can check how your set of B T's varies, etc. The main idea is

[5]This means that for one of the groups, we treat α to be 0 so just μ alone is the mean response for that group (or state)

[6]This last point is a given, given that there is no chance at all that something may have gone terribly wrong with this data

that you assume your sample is a good representative of the population and that sampling from your sample is like sampling from the population. Doing so provides you with a way to see how your statistic varies on repeat samples.

Again, if this doesn’t make sense, just ignore this as the important part is to know that it works and that it has revealed some terrible emet.

III. Results

A. ANOVA Results

This model provided results that were bullshit 😇. They can be seen in Fig 4.

```
Analysis of Variance Table

Response: FINAL_TIP
          Df  Sum Sq Mean Sq F value Pr(>F)
STATE      1   209.3  209.26  0.2269  0.639
Residuals 20 18443.1  922.15
```

Fig. 4. ANOVA Results

In this case, our μ’s were the averages in the two make-up states. The probability that a value is as or more extreme than the one observed is nearly 64% - i.e. it is fairly likely. This implies that there is probably no difference between make-up and not make-up however, as the results in the next section prove, this is definitely **not** the case.

Fun side note, when you are only considering the two groups (like in our case here, we have make-up and not make-up), the square of the t-test is equivalent to the F-test; that is, if you look at the F-value you see and square root it, you will get the t-value (ambiguous is of course whether or not it is non-negative but moot due to symmetry). To demonstrate, consider the regression model which provides us with t-statistics on our variables:

```
Call:
lm(formula = FINAL_TIP ~ STATE, data = tips)

Residuals:
    Min      1Q  Median      3Q     Max
-70.629 -20.628  -1.525  17.575  50.721

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     70.629      9.156   7.714 2.04e-07 ***
STATENO_MAKEUP  -6.168     12.949  -0.476    0.639
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.37 on 20 degrees of freedom
Multiple R-squared:  0.01122,   Adjusted R-squared:  -0.03822
F-statistic: 0.2269 on 1 and 20 DF,  p-value: 0.639
```

Fig. 5. Regression Results

Notice that the t-value is -0.476 and squaring this, we get 0.2266 $\approx$ 0.2269 - which is the F-value garnered from the ANOVA model. You could also look at the p-values. Ignore this part if you don’t care.

Lastly, let’s consider the diagnostic plots to confirm the legitimacy of using such a model:

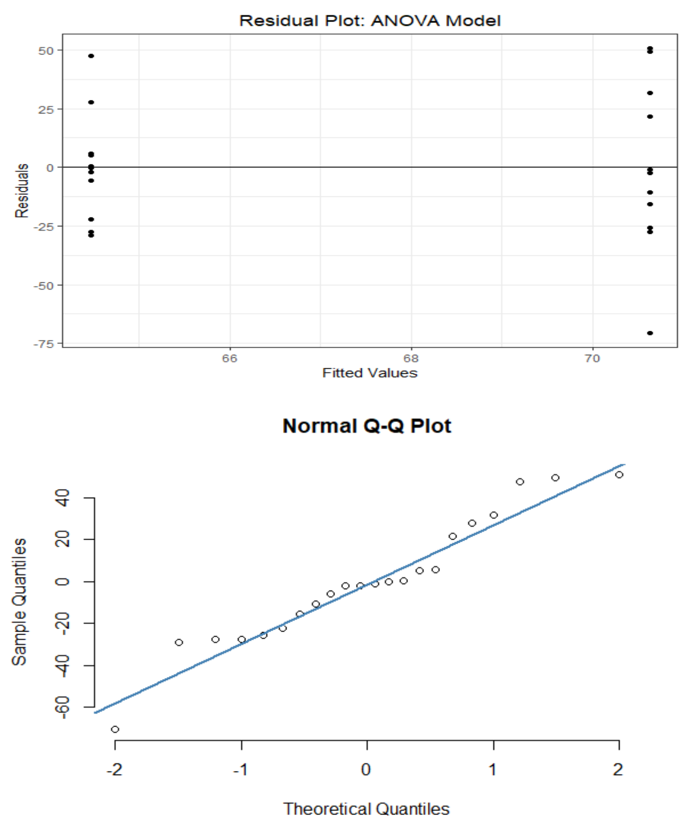

Fig. 6. Regression Residual Plots

Loosely speaking, the assumptions do

not seem to be violated indicating that there is some validity to be performing an ANOVA (among other statistical tests) on this data.

B. Bootstrapping Results

Here, we look to see whether, on repeated samples from the provided data, there is a difference between the amount of money made (on average) between the two states. After running some simulations, the following densities were realized:

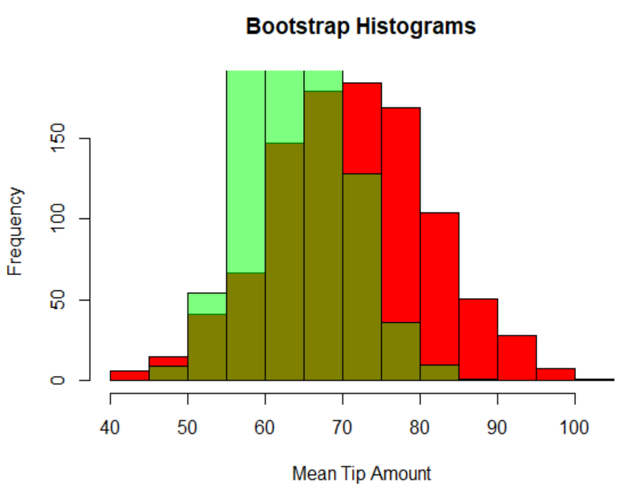

Fig. 7. Results of Bootstrapping: 10000 Simulations. The red histogram corresponds to the make-up state whereas the green one is the non-make-up state.

What we notice here is that, over the simulations, there is an undeniable difference between the average tip amounts across the two states. For the non-make-up state, the mean tip amount taken home at the end of the day seems to bunch up on to the left end of the make-up density. Moreover, the high leverage point referred to earlier was not removed but, despite this, the make-up histogram still indicates that the *mean* amount is greater than it would be if no make-up was involved.

This is in contrast with the statistical model results from before however, the results here are more concrete. The data set is fairly small and so, there isn't much *power* for us to get a significant result regardless. It is much more fruitful to just look at how the data varies on repeat samples to get a good read on whether or not make-up significantly affects tips.

C. Other Results

To further drive this point home, a more detailed look was given to the data. The proposed question to be considered was:

Do shorter shifts with make-up result in the same amount or more money than longer shifts without make up?

In order to answer this question, I considered shifts that were less than 5.5 hours long to be "short" shifts. After generating this data via the bootstrapping strategy and looking at the results, it was clear that make-up shifts, despite being shorter, resulted in a greater average tip amount (73$) whereas longer shifts without make-up had a lower average tip ($\approx$ 70$). The variation in these numbers, despite the observed overlap, is likely more of a function of the sparse data than it is indicative of no relationship between state and tips earned given the conditions.

Regression models were also fit to confirm whether some kind of relationship even exists between shift length and the amount of money made and, as observed in Fig 2., this relationship is indeed statistically significant and is even more pronounced in the case of the make-up state (which is to say, a larger magnitude in difference is seen).

Lastly, let's consider the actual difference in pay. Purely looking at averages and by removing the outlier (a data point whose value is the result of a statutory holiday), we see that on days which the the brave waitress wore make-up, she made 13$ more. Averaging over the total shift length, we see that, per hour, you can expect to make about 2$ more in tips (13$ v. 11$). The variation is also greater for the make-up days (as seen earlier, this is skewed to the left).

D. Future Analysis

Whether it be a result of subpar memory (a plight that is all too common to me) or because of the strain hoisted onto to the data collector, the information on hair quality (for a given shift) is incomplete. Hair quality is extremely importantly (maybe) with respect to how much you generate in tips so this is an area left to be explored. In particular, future analysis could be done on the effect of "bombing" hair on monetary gains - surely an avenue for endless research.

In addition, there seems to be an opportunity to see how the time of day effects tips gained - do you make more later on in the day? Are people more shallow in the evening? What kind of customers come in on the weekends?

IV. Conclusion

Sometimes, events occur by chance. In greek mythology, Hades gets treated as the Greek equivalent of the devil. But, in actuality, he got the job because he and his brothers drew straws and he got stuck with the underworld. Zeus received the sky, Poseidon the seas, and Hades was just unfortunate (or not?). Was that just bad luck? The ancient Greeks were not equipped to answer such questions and perhaps, even if the statistical tools were readily available, the outcomes may still have been a consequence of chance - to think otherwise is a still commonly occurring error; however, the conclusions here are different.

The analysis done shows, beyond a shadow of a doubt (probably), that there is a significant effect associated with make-up on the amount of tips garnered by a waiter. While the initial results seemed to show otherwise, a further, more careful analysis has left little to the imagination. A number of avenues for future research have also been opened and serves as an aid for scientists across the world.

While these conclusions are devastating, they also mark a step in the right direction: identifying and conceding that a problem exists is a powerful first stride in solving it. Perhaps we will never get to some realization from the set of idealizations detailing what ***our*** world ought to be but it is important to remember that the power of our imagination is limited in scope - we are slaves to our experiences and in thinking not, you further slip into the jaws of self-chicanery. There is always so much more progress to be made in our little realms despite all the constraints plastered on to us whether by some irreversible personal choices or by the hands of god. In dark times, when the persistent conclusion associated with self-perception is a gravely wrong one, it is of utmost urgency to *focus* on where you focus your gaze; as Nietsche famously stated, *if you stare into the abyss, the abyss stares back at you*. This is commentary on the interaction between something and nothing - the hope is clear, you *should* concede and from there you can find answers - but just not to the questions you would like; those have no answers. The goal is to live as if the moments leading up to[7] some great Epiphany are all that there is and that while you may arrive at nothing more than a *mirage* of a personal heaven, you are rich in wisdom enough to have already built your own.

APPENDIX

There is no appendix.

ACKNOWLEDGMENT

I acknowledge nothing.

REFERENCES

[1] I did it myself.

[7]This distance is unknown