Effect of Anti-Anxiety Drugs on Memory Recall

De Broux, Evan (50%)

Singh, Abhimanyu (50%)

MATH 483-01

December 1, 2019

Contents

1 Abstract **2**

2 Introduction **2**

3 Data and Sources **2**
3.1 Data Sources 2
3.2 Preliminary Data Analysis 3

4 Analysis **5**
4.1 Building of Model to Estimate Treatment Effects 5
4.1.1 Variable Selection and Definition 5
4.1.2 Regressed Model and Results 6
4.2 Tukey Test for Differences in Treatment Means 8
4.3 Model Diagnostics 9
4.3.1 Analysis of Residuals 9
4.3.2 Breusch-Pagan Test for Constant Variance 10
4.4 Power 11

5 Conclusion **12**

6 Proposed Future Work **13**

7 Appendix **15**
7.1 Plots 15
7.2 Model Summaries 16

1 Abstract

Major depressive disorder is a chronic, recurring, and debilitating mental illness that is the most common mood disorder in the United States. When people are overwhelmed by heart-pounding panic, paralyzed by fear, or exhausted by another worrying sleepless night, people do just about anything to get relief. There is evidence that when anxiety is disabling, medication may help, but are drugs always the best answer?

2 Introduction

The positive effects of long-term potentiation between cells on memory recall and acquired associations have been impaired by benzodiazepines. By differentiating Alprazolam (long-term) and Triazolam (short-term) adverse long-term effects, patients may be given a better diagnosis to mitigate any damage to the metacognition and memory recall ability of the brain. Additional study has also shown that merely remembering specific memories associated with strong emotions in the present time can cause these emotions to materialize and influence future thoughts for a short period of time (about 10 minutes). The presence of happy emotions and sad emotions is of interest and is known to have a significant impact on memory recall, thus driving the question, what are the effects on memory recall performance of benzodiazepines after being primed with a happy or sad memory? This clinical study will show if the type of memories which support or impede memory recall is independent of other variables if the efficacy of benzodiazepines depends not only on the sensitivity but also on the mood of a participant, and finally on the possibility of strengthening or deteriorating memory recall performance when presented in conjunction with the use of benzodiazepines or memories related to mood alone beyond the known response 7, 8.

3 Data and Sources

3.1 Data Sources

We found the dataset on Kaggle:

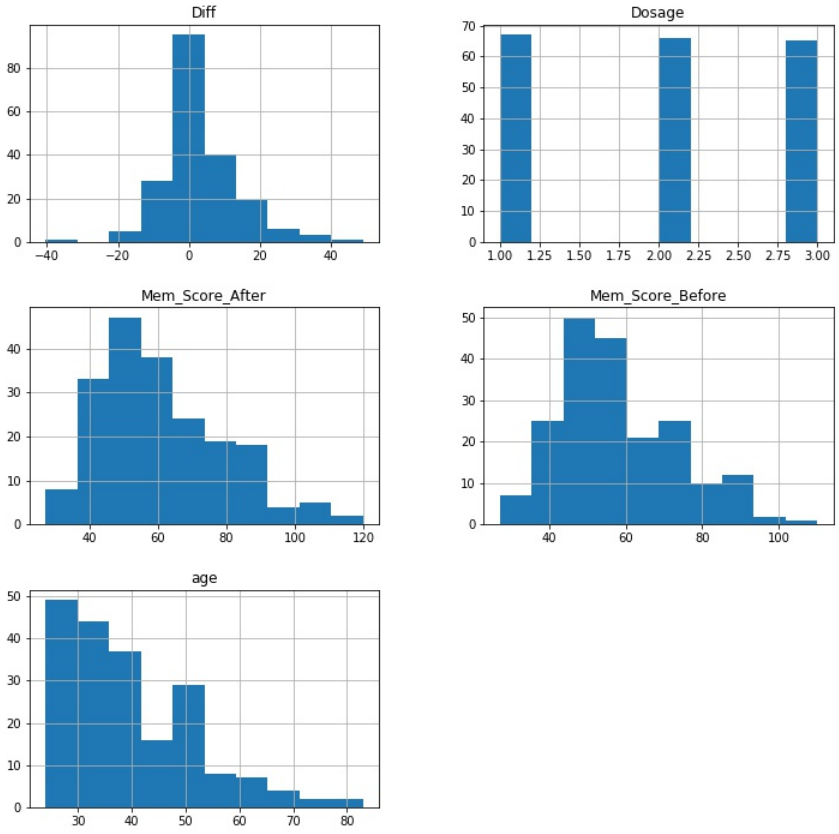
Memory Test	Memory Test on Drugged Islanders Data: Anti-Anxiety Medicine on Novel Islanders grouped by Happy or Sad Memory Priming [1]	Kaggle/UCLA

Each of the 9 drug treatment combinations had equal amounts of participants, and each group had equal amounts of participants placed into blocks primed with happy memories or sad memories. All genders were used and all participants of islanders were aged 24 and older to ensure that the prefrontal cortex of participants was fully developed, thus ensuring that anomalies in the observed response variable are not caused by an underdeveloped mind. Each islander is administered their drug once-a-day for a week to mimic addiction and the happy or sad memories are primed 10 minutes prior to the memory test.

The dataset can be obtained on Kaggle at:
https://www.kaggle.com/steveahn/memory-test-on-drugged-islanders-data
Or our GitHub repository:
github.com/edebroux/final_project_483 _f2019

3.2 Preliminary Data Analysis

We first looked at the distributions of the data we have been given.We see that there is evidence that the islanders' differences in memory score, kept in seconds, is normally distributed. We see that the memory scores before and after are similarly unimodal, but the data is skewed to the right for both variables. Also, the distribution for age may not entirely reflect the age distribution of islanders, as this is also skewed to the left. However, without knowing more information about where these islanders come from, it is impossible to say if this sample truly reflects the age distribution of islanders. We will just have to trust the experiment organizer's sampling methods in this case. We cannot control for this variable either, so we are assuming that assignment to treatment blocks was completely randomized to try to balance out the effects of age on the difference in memory scores.



Caption: Distributions of all variables containing real numbered values. Additionally, we can confirm that the dosages were evenly distributed, by looking at the upper right histogram. We also can refer to the appendices to confirm this for the drug type and blocks of happy or sad primed memories, 7, 8.

4 Analysis

4.1 Building of Model to Estimate Treatment Effects

4.1.1 Variable Selection and Definition

We start by defining our variables for the model. In truth, we could regress a model of predictor variables with the memory score after as the response. However, since we are only interested in figuring out if the drugs have any effect on the change in memory score, we will not perform this sort of regression.

Now, when defining the variables, it appears that we may have some sort of 3^2 factorial experiment on our hands. We have three different drugs and three different dosage levels. However, we cannot truly rank the drugs effect on memory here. Sure, we may be able to speculate that the effects of the sugar tablet, Triazolam, and Alprazolam on memory scores may be ordinal in some sense. However, until more strenuous testing is carried out and confirmed, it seems inappropriate to us to do so. We thus define our variables for drug administered categorically with two separate binary categorical variables.

Similarly, the dosage amounts, even though administered in one, two, and three doses, are not delievered in the same amount of units (mgs and pure number of tablets). Hence, it could again be inappropriate to define the dosage variable as a continuous variable. We again define the dosage amount as two separate categorical variables. We include the block effect for the happy or sad primed memories before testing. This leads to a model matrix:

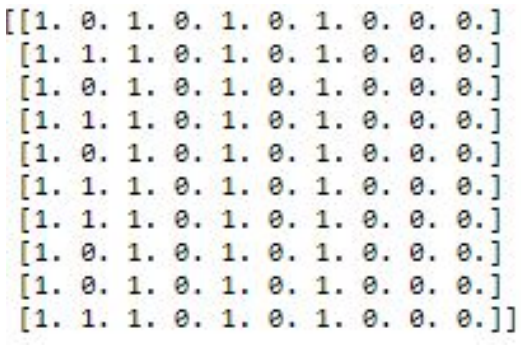


Figure 1: First 10 rows of the model matrix used for the regression model.

4.1.2 Regressed Model and Results

Using 'statsmodels.formula.api.ols', we were able to fit a regression model with the difference as the response, the first order terms of the drug administered and the dosage given, and then their interaction terms. The model summary is displayed in the appendix 9.

The $R^2_{adjusted}$ value is 0.444. The regression model is significant at an $\alpha = 0.05$ significance level, as the corresponding F-statistic, $MS_R/MS_E = 18.45$, is larger than the critical value for $F_{\alpha=0.05,9,188} = 1.93$. We can also account for what amount of the total sum of squares can be attributed to the drug type and dosage amount using another F-test. The F-test would be carried out as follows in a general sense, where A is the drug type and B is the dosage with a type I error probability of $\alpha = 0.05$:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_n \tag{1}$$

$$H_1 : \mu_i \neq \mu_j, i \neq j \in \{1, 2, \ldots, n\} \tag{2}$$

Test statistic: $F^* = \frac{MS_A}{MS_E} = \frac{SS_A abc(n-1)}{SS_E(a-1)}$ or $F^* = \frac{MS_B}{MS_E}$

Where the rejection region's critical value is defined by:
$|F^*| > F_{\alpha=0.05,a-1=3-1=2,188} = 3.04$

for both A and B. The block effect's F-test would be carried out similarly, but instead the critical value for the rejection region would be defined by:

$|F^*| > F_{\alpha=0.05, c-1=2-1=1, 188} = 3.89.$

For the first order interactions between the drug and dosage type, we would have an F-test defined as follows:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_{\binom{ab}{2}}$
$H_1 : \mu_i \neq \mu_j$, for some $i \neq j \in \{1, 2, \ldots, \binom{ab}{2}\}$

Test statistic: $F^* = \frac{MS_{AB}}{MS_E}$
Where the rejection region critical value is defined by:
$|F^*| > F_{\alpha=0.05, (a-1)(b-1)=2*2=4, 188} = 2.42$

The output of the regression model is shown below here:

	df	sum_sq	mean_sq	F	PR(>F)
Happy_Sad_group	1.0	9.689091	9.689091	0.150544	6.984550e-01
C(Dosage, Sum)	2.0	1224.603642	612.301821	9.513623	1.159451e-04
C(Drug, Sum)	2.0	4309.257305	2154.628652	33.477483	3.658230e-13
C(Dosage, Sum):C(Drug, Sum)	4.0	5141.980736	1285.495184	19.973348	1.012301e-13
Residual	188.0	12089.780135	64.360533	NaN	NaN

Figure 2: ANOVA table of the effects for drug type, drug dosage, and their interaction effects.

Based on the outputs from our model, we see that the block effect is not significant at an $\alpha = 0.05$ significance level. Hence, there is not nearly enough evidence to suggest that the type of memories primed before testing significantly affects the memory score.

Yet, based on the output of the ANOVA table, the drug type, the amount of drug administered, and the interactions between the drug type and dosage are all significant factors in determining the difference in memory scores before and after the drugs were given to the participants. It obviously would

be useful to figure out which differences in drug type, drug dosage, and their interactions are significant, but we can figure this out using a Tukey test for pairwise differences in treatment means.

4.2 Tukey Test for Differences in Treatment Means

A Tukey test for pairwise differences in treatment means is carried out in the general case as follows:

$H_0 : \mu_i = \mu_j$
$H_1 : \mu_i \neq \mu_j$
It uses the studentized range statistic:
$q = \frac{\bar{y_{max}} - \bar{y_{min}}}{\sqrt{MS_{error}/n}}$

Two means are different if the absolute value of their sample mean differences exceeds:
$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_{error}}{n}}$

where a is the number of parameters and $f = N - a$. This is how the test for the drug dosage and drug type are run. The Tukey test for the interactions is the same, but now we have $a * b$ different treatment combinations and $f = N - ab$ [3].

We show the output below for the drug and dosage tests. The Tukey test for the interactions are shown in the appendix 10.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	1.6118	0.6389	-2.6956	5.9193	False
1	3	5.8994	0.0042	1.5753	10.2235	True
2	3	4.2876	0.0537	-0.0527	8.6278	False

Figure 3: Pairwise Tukey test for differences amongst dosage treatment groups.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	S	-9.6414	0.001	-13.629	-5.6537	True
A	T	-10.0578	0.001	-14.0609	-6.0547	True
S	T	-0.4165	0.9	-4.4345	3.6015	False

Figure 4: Pairwise Tukey test for differences amongst drug treatment groups.

Based on the results of the Tukey tests, we can see that there is a significant difference between those who receive 1 dose and 3 doses. Additionally, there are significant differences between those administered Alprazolam and the placebo and the Alprazolam and Triazolam groups. Based on the interactions between the treatment groups, we see that there are 9 different pairwise treatment interactions groups that have significant mean differences at an $\alpha = 0.05$ significance level. For the sake of space, refer to 10 in the appendix.

4.3 Model Diagnostics

We must look at the residual diagnostics for this model. If there are any issues with the assumption that $\epsilon_i \sim N(0, \sigma^2)\ \forall i$, i.e. the errors are distributed with mean zero and constant variance, then many of the assumptions and tests that have been made in this model are quite useless.

4.3.1 Analysis of Residuals

We start by plotting the normal probability plots of the residuals, the residuals against each level of each treatment, and the residuals against the predicted values the model suggests based on the predictor variables.

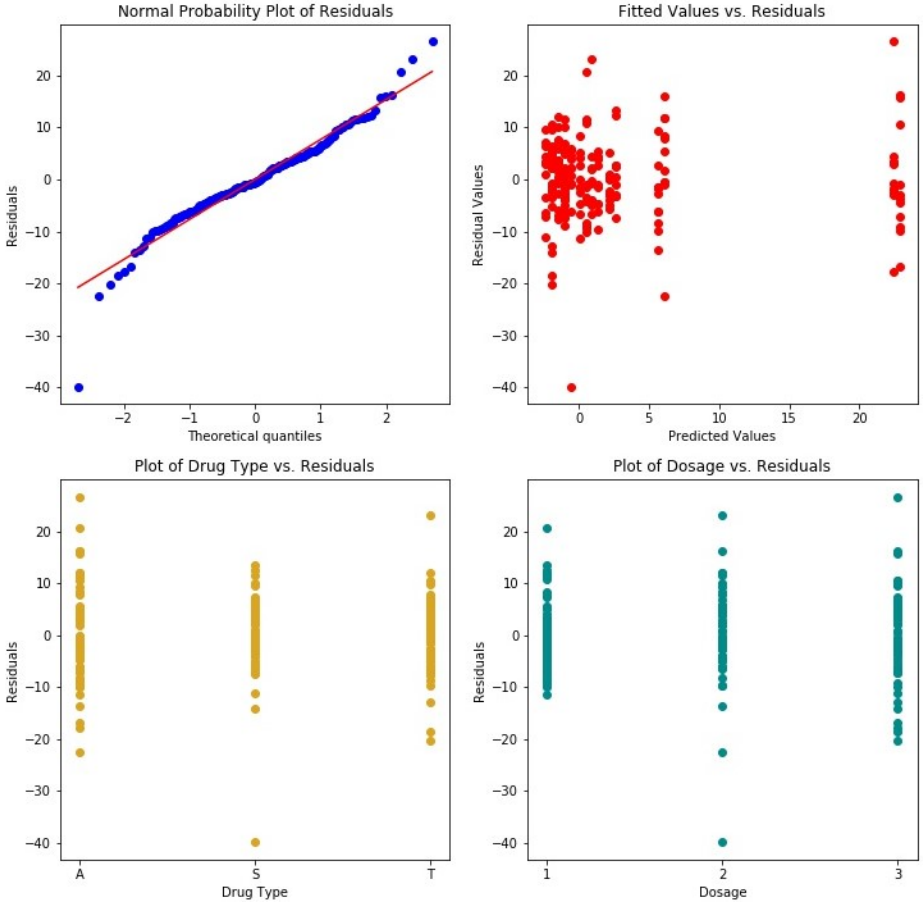


Figure 5: Residual plots

Based on the residuals plots, there is a slight hint of non-constant variance, and there appears to be some divergence from normality near the tails of the distribution based on the normal probability plot. However, besides a few potential outliers, we see that the residuals are reasonably normally distributed. We will formally test to see if the variance is constant.

4.3.2 Breusch-Pagan Test for Constant Variance

In order to test for constant variance, we will carry out a Breusch-Pagan test. The procedure for the test is carried out as follows:

We regress the function: $log(\sigma_i^2) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i1} x_{i2}$

$H_0 : \gamma = 0$

$H_1 : \gamma \neq 0$

Test statistic: $\frac{SSR^*/4}{(SSE/n)^2}$

Where SSR^* is the regression sum of squares for the log regression of the residuals. Critical value for the rejection region would be $\chi^2_{\alpha/2;3}$, and we will test at an $\alpha = 0.05$ level [2].

The output of the test is given in our notebook. The corresponding test statistic p-value is 0.588, which means there is not enough evidence to suggest that the variance is not constant. Therefore, it is unnecessary to make some sort of Box-Cox transformation of the response variable in order to stabilize the variance.

Based on these diagnostics, we would have to investigate the potential outlying observations. In our view there are two, potentially three outliers that require investigation. 31 year old Eva Takakashi had a significantly lower memory score (almost 40 seconds less), despite being administered the placebo in two doses. 35 year olds Miki Carrasco and Naoto Lopez had increases in recall times after being administered 3 doses of Alprazolam and 2 doses of Triazolam, respectively. In the case of Carrasco and Lopez, there may be some other factor that is unknown to us that caused significant increases in recall times. This would require further investigation. In the case of Takakashi, this appears to potentially be a bad trial, as the drug administered was technically the control. Either discarding the trial from the analysis or re-running the trial may be the right move, as the Takakashi outlier may be leveraging the model.

4.4 Power

The last thing we would like to calculate is the statistical power of the experiment that was carried out. Recall that the power calculates the probability of not making a type II error, the probability that the null hypothesis is not rejected given it is not true. Based on the parameter estimate with the largest degree of freedom (4), the number of samples (198), the probability of a type I error defined in our tests ($\alpha = 0.05$), and a fairly low estimation

of maximum effect size, the standardized difference between the greatest and least means divided by the standard deviation, which we put at 0.25, we calculated that the power of the test we carried out was approximately 0.8444.

5 Conclusion

Based on the analysis of the results, we can conclude that as our model assumptions withstand the diagnostic analysis, we can conclude that anxiety medicine does have a significant effect on recall. Both the type of medicine administered and the amount given to participants was found to significantly effect the memory recall scores of participants in this experiment.

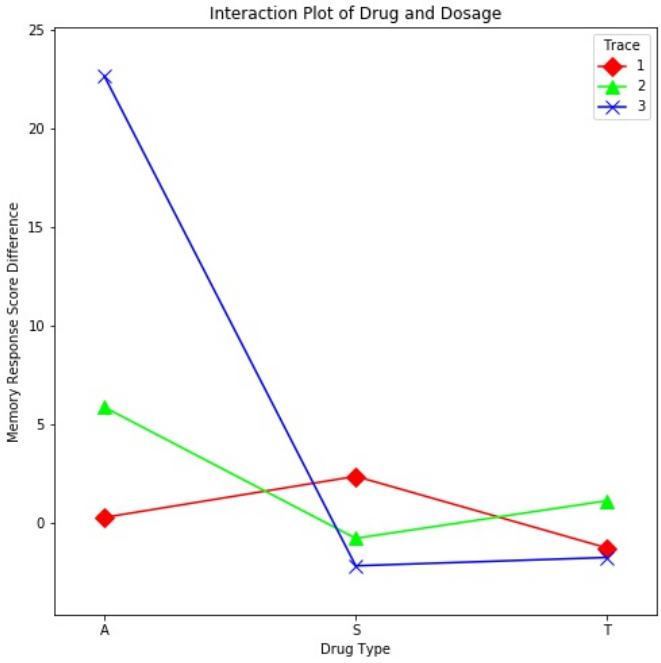


Figure 6: Plot of the interaction effect between the drug administered and the dosage of each drug given to patients.

If one wished to significantly effect the memory of a patient using an anti-anxiety drug, based on the results of the experiment they should give three doses of Alprazolam to the patients.

6 Proposed Future Work

There is a fair amount of subject matter that we did not really cover in this project. There are obvious variables that we did not block for, such as gender, mostly because the gender of the participants was not given. It is true that we could have ventured a guess for what gender each participant was, but that can be difficult to determine only going off of their names. Additionally, there may be some familial traits that should be investigated that could be effecting the response variable here.

We look at the power calculations here too. We estimate the maximum effect size to be approximately 0.385 after running this experiment. This would change the power of the test we ran to 0.998, which makes sense since as effect size increases the power of a test increases as well. Given the prior assumptions, we would not know this beforehand. Since we assumed an effect size of 0.25, and say we want the probability of a type II error to at least match the probability of a type I error that we defined in our statistical tests above, we would need to increase the sample size to at least 279 participants. Additionally, if one wants to retain equal amounts of participants based on the type of drug administered, dosage amount, and type of memory primed before testing, this would need to be increased to 288 participants to ensure all 18 different treatment/block effect combinations get 16 participants per combination. Otherwise, one would need to use some sort of balanced incomplete block design to find the optimal assignment of treatments. This ensures that the probability of a type II error, β, is less than 0.05.

Our final discussion point regards optimality of design. If we want to make a D optimal design, in which the goal is to maximize the determinant of $|\mathbf{X}^T\mathbf{X}|$, which concurrently minimizes the volume of the joint confidence region for the parameter estimates, we would need to find an matrix that maximizes that determinant. Our current model matrix gives us a determinant of $|\mathbf{X}^T\mathbf{X}| \approx 3 * 10^{19}$. This matrix gives an extremely large determinant value, but we would need to do a further investigation and a lot of algorithmic optimization to conclude that the current algorithm is D-optimal. Additionally, if we wanted to make our design I optimal, we would try to

minimize the prediction variance. In our current model (without defining any of the variables explicitly here), we regress a model that looks like so:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_2 x_4 + \beta_7 x_3 x_4 + \beta_8 x_2 x_5 + \beta_9 x_3 x_5 \tag{3}$$

In order to calculate the maximum prediction variance for this model, we would need to calculate:

$$V(\mathbf{X}\beta) = \sigma^2 \sum_{i=1}^{10} (\mathbf{X}^T \mathbf{X}_{ii})^{-1} (1 + x_1 + x_2 + x_3 + x_4 + x_5 + x_2 x_4 + x_3 x_4 + x_2 x_5 + x_3 x_5) \tag{4}$$

This is maximized when $x_1 = 1$ and then one of x_2, x_3 and one of x_4, x_5 are equal to one as well. We found this value to be $V(\mathbf{X}\beta) = 0.75445\sigma^2 = 48.56$. It seems plausible that this is the minimum possible maximum variance of this experimental design, but again further investigations would be needed to check if this is the case.

References

[1] Akram Almohalwas.

[2] Michael et al Kutner. *Applied Linear Statistical Models*. India, 2005.

[3] Douglas Montgomery. *Design and Analysis of Experiments*. John Wiley Sons, Singapore, 2013.

7 Appendix

7.1 Plots

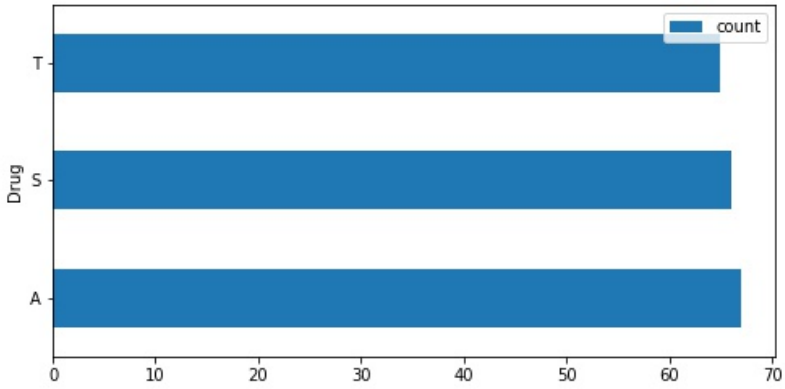


Figure 7: Bar graph of participants given each type of drug, a sugar tablet placebo (S), Triazolam (T), and Alprazolam (A)

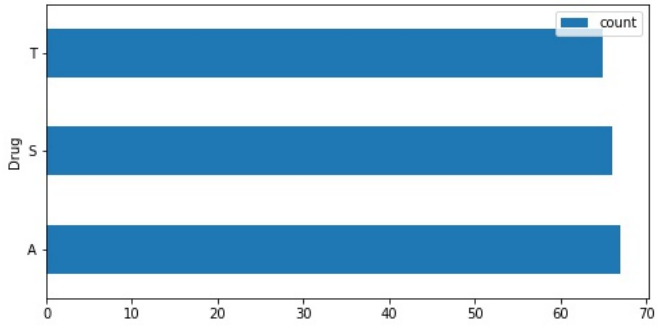


Figure 8: Bar graph of participants primed with happy (H) or sad (S) memories.

7.2 Model Summaries

OLS Regression Results

Dep. Variable:	Diff	R-squared:	0.469
Model:	OLS	Adj. R-squared:	0.444
Method:	Least Squares	F-statistic:	18.45
Date:	Sat, 30 Nov 2019	Prob (F-statistic):	8.03e-22
Time:	21:39:32	Log-Likelihood:	-688.10
No. Observations:	198	AIC:	1396.
Df Residuals:	188	BIC:	1429.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.7333	0.806	3.390	0.001	1.143	4.324
Happy_Sad_group[T.S]	0.4221	1.141	0.370	0.712	-1.828	2.672
C(Dosage, Sum)[S.1]	-2.4642	0.804	-3.067	0.002	-4.049	-0.879
C(Dosage, Sum)[S.2]	-0.8520	0.806	-1.057	0.292	-2.443	0.739
C(Drug, Sum)[S.A]	6.6616	0.804	8.291	0.000	5.077	8.247
C(Drug, Sum)[S.S]	-3.1156	0.806	-3.864	0.000	-4.706	-1.525
C(Dosage, Sum)[S.1]:C(Drug, Sum)[S.A]	-6.8466	1.130	-6.059	0.000	-9.076	-4.618
C(Dosage, Sum)[S.2]:C(Drug, Sum)[S.A]	-2.8722	1.138	-2.523	0.012	-5.118	-0.627
C(Dosage, Sum)[S.1]:C(Drug, Sum)[S.S]	5.0218	1.138	4.412	0.000	2.776	7.267
C(Dosage, Sum)[S.2]:C(Drug, Sum)[S.S]	0.2686	1.140	0.236	0.814	-1.981	2.518

Omnibus:	29.694	Durbin-Watson:	2.112
Prob(Omnibus):	0.000	Jarque-Bera (JB):	119.731
Skew:	-0.451	Prob(JB):	1.00e-26
Kurtosis:	6.701	Cond. No.	3.42

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

Figure 9: Model summary for linear regression model.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A1	A2	5.5775	0.3256	-1.913	13.0679	False
A1	A3	22.3366	0.001	14.8461	29.827	True
A1	S1	2.082	0.9	-5.4084	9.5724	False
A1	S2	-1.0589	0.9	-8.5493	6.4315	False
A1	S3	-2.4498	0.9	-9.9402	5.0406	False
A1	T1	-1.5453	0.9	-9.0357	5.9452	False
A1	T2	0.8457	0.9	-6.6448	8.3361	False
A1	T3	-2.0282	0.9	-9.6092	5.5529	False
A2	A3	16.7591	0.001	9.1859	24.3323	True
A2	S1	-3.4955	0.8691	-11.0687	4.0777	False
A2	S2	-6.6364	0.1375	-14.2096	0.9368	False
A2	S3	-8.0273	0.0287	-15.6005	-0.4541	True
A2	T1	-7.1227	0.0837	-14.6959	0.4505	False
A2	T2	-4.7318	0.5624	-12.305	2.8414	False
A2	T3	-7.6056	0.0535	-15.2685	0.0572	False
A3	S1	-20.2545	0.001	-27.8277	-12.6813	True
A3	S2	-23.3955	0.001	-30.9687	-15.8223	True
A3	S3	-24.7864	0.001	-32.3596	-17.2132	True
A3	T1	-23.8818	0.001	-31.455	-16.3086	True
A3	T2	-21.4909	0.001	-29.0641	-13.9177	True
A3	T3	-24.3647	0.001	-32.0275	-16.7019	True
S1	S2	-3.1409	0.9	-10.7141	4.4323	False
S1	S3	-4.5318	0.6121	-12.105	3.0414	False
S1	T1	-3.6273	0.8364	-11.2005	3.9459	False
S1	T2	-1.2364	0.9	-8.8096	6.3368	False
S1	T3	-4.1102	0.7285	-11.773	3.5527	False
S2	S3	-1.3909	0.9	-8.9641	6.1823	False
S2	T1	-0.4864	0.9	-8.0596	7.0868	False
S2	T2	1.9045	0.9	-5.6687	9.4777	False
S2	T3	-0.9693	0.9	-8.6321	6.6936	False
S3	T1	0.9045	0.9	-6.6687	8.4777	False
S3	T2	3.2955	0.9	-4.2777	10.8687	False
S3	T3	0.4216	0.9	-7.2412	8.0845	False
T1	T2	2.3909	0.9	-5.1823	9.9641	False
T1	T3	-0.4829	0.9	-8.1457	7.1799	False
T2	T3	-2.8738	0.9	-10.5366	4.789	False

Figure 10: Tukey tests for the interaction groups.