Biometrics 000, 000–000
000 0000

DOI: 000

Ocular Irritation Study-Rabbits

Tyler Green
University of Guelph

1. Introduction

The dataset (Ocular Irritation Study) chosen is from the Bradstreet datasets. Ocular Irritation Study is a dataset that comes from the testing of two drugs at different doses to the effect it has at creating ocular lesions in rabbits measured by draize scores. The dataset consists of 2 experiments. Experiment one is a 2X1 factorial design and experiment two is a 2X3 nested factorial design, with each experiment conducted in essentially identical fashion. Experiment one was conducted with the two drugs (Drug A and Drug B) on one does (100mg) with 4 repetitions per combination. Experiment two was conducted with the two drugs (Drug A and Drug B) on five doses (50mg, 25mg, 5mg, 12.5mg, 2.5mg). In experiment 2 Drug A was combined with dosages: 50mg, 25mg and, 5mg. While Drug B was combined with dosages: 25mg, 12,5mg and, 2.5mg.

2. Potential Problems And Solutions

With this particular dataset and how it has been compiled there are some particular challenges associated with it.

The first problem that could arise is that this test is conducted by using drugs on rabbits, the rabbits may have some sort of extreme reaction or no to little reaction at all to the drug. This should only be a problem if all rabbits have the same effect of either extreme or no reaction at all when there is supposed to be some standard reaction. This is very unlikely to happen such that there are 32 rabbits and the chances of all 32 being either highly reactive to the drugs or non reactive at all is very minimal and this can be ignored. However it is still good to know whether a rabbit has had some form of reaction and therefore a Tukey test is used. This can tell us what is significantly different. If one rabbit has a large statistical difference we know that, that particular rabbit is having a more or less severe reaction.

The second problem and arguably the worst problem is that, since the information is based on two experiments conducted in nearly identical fashion, there could be a difference in some key section of the experiment. This nearly identical fashion could cause some sort of skew to the data or some unknown effect on the errors. To test if there is homoscedasticity between the two experiments, Levene's test can be used. If the test return an insignificant p-value then the two experiments are good to be used together. If the p-value is significant then the two experiments must be split up and tested.

The third problem is that two sevenths of the data is missing and roughly one third of the data are 0's. The missing data could contain some unusual effect or trend that is not shown in the observed values. The large number of 0 values could act as a strong influence to the model. With both the missing values and the 0's that makes roughly thirteen twenty-firsts of values potentially problematic. A solution to the missing values is that the values are considered missing at random (MAR) values. If the values where missing completely at random (MCAR) or missing not at random (MNAR), the values would need to be estimated to proceed with the analysis. Since the values are MAR, we can simply omit the values and proceed as normal. To deal with the large quantity of values that are 0 we can simply remove all but the first 0 obtained for rabbit. This will help normalize the data and keep the number of 0's from influencing the model. One thing to note about doing this and with the model it will provide is that all negative values produced given the model is used as a predictor or to estimate should be treated as a 0, due to the lower limit of 0 of the draize score. To help determine the influence the 0 values have the tests shall be conducted with only the first 0 in the model and all others removed as well as the complete dataset

3. Analysis

The general model: $Y_{ijkl} = \mu + \tau_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \gamma_l + \epsilon_{ijkl}$ shall be used for the analysis of the dataset. Where:

- i's go from 1-32
- j's go from 1-2
- k's go from 1-5
- l's go from 1-9

The model has the following assumptions:

- τ_i are the Rabbit effect, where $\sum \tau_i = 0$
- α_j are the Drug effect, where $\sum \alpha_j = 0$
- β_k are the Dose effect, where $\sum \beta_k = 0$
- $\alpha\beta_{jk}$ are the Drug:Dose interaction effect, where $\sum \alpha\beta_{jk} = 0$
- γ_l are the time effect, where $\sum \gamma_l = 0$

This paper has been submitted for consideration for marking in *STAT 3210*

- ϵ_{ijkl} are iid N(0,σ^2)

The following tests shall be performed and analyzed:

(1) Error Term Comparison
(2) Rabbit Reaction
(3) Drug Comparison
(4) Dose Comparison
(5) Drug-Dose Interaction
(6) Effects of Time

3.1 *Error Term Comparison*

To check if there is homoscedasticity among the different experiments a Levene's test is conducted. The null hypothesis of there is homoscedasticity is tested against the alternative hypothesis of there is heteroscedasticity. Conducting this test consits of joining the two experiments as one large dataset and then using Levene's test using all treatments as factors as a group. If the p-value is significant then the two experiments are heteroscedasticity.

3.1.1 *Full Dataset.* Table One was obtained when conducting the test upon the two experiments treated as one on all the data.
See Table One
Since the p value is 0.9998 which is greater than α of 0.05 we cannot reject the null hypothesis that the two models have homoscedasticity.

3.1.2 *Reduced Dataset.* Table Two was obtained when conducting the test upon the two experiments treated as one on the reduced dataset.
See Table Two
Since the p value is 0.9989 which is greater than α of 0.05 we cannot reject the null hypothesis that the two models have homoscedasticity.

3.1.3 *Section Conclusion.* The very large p values helps to determine that the data is very homoscedasticitic whether it is the full data or not. This shows that among all the data collected over the two experiments were done in nearly identical fashion with very little to no changes. This shows there is no change in error over the two experiments. This means that the datasets do not need to be split up to be examined further, they can be left as one larger set and examined.

3.2 *Rabbit Reaction*

In order to tell if there is a statistical difference between the rabbits reactions a Tukey test is needed. If there is a significant p value then the two rabbits have statistically different results. Furthermore if in the general model with two factor interactions is not significant as well then each rabbit produces statistically the same effect.

3.2.1 *Full Dataset.* For the full dataset the general model was created. An anova table was created to test the significance of the animal parameter (τ_i) for this dataset.
See Table Three
Based on the anova table Animal has a p value of nearly 0.8.

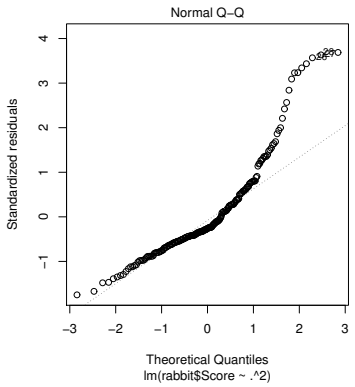

Figure 1. The Full Datasets QQ Plot.

This is not significant at most level of α. This shows that animal is not significant in predicting the score. Although animal is not significant in predicting score, it is still important to check whether an animal has a reaction effect. Based on the Tukey test (omitted due to space) all animals are not statistically significant. This means that not only are the animals not significant to predict the draize scores but that they are all produce statistically the same results.

3.2.2 *Reduced Dataset.* For the reduced dataset the general model was created. An anova table was created to test the significance of the animal parameter (τ_i).
See Table Four
Based on the anova table Animal has a p value of nearly 1. This is not significant at most level of α. This shows that animal is not significant in predicting the score. Although animal is not significant in predicting score again, it is still important to check whether an animal has a reaction effect. Based on the Tukey test (omitted due to space) all animals are not statistically significant. This means that not only are the animals not significant to predict the draize scores but that they are all produce statistically the same results.

3.2.3 *Section Conclusions.* Both datasets got identical conclusions for animal parameter (τ_i). This shows that across both datasets the animal parameter is not statistically significant. Also in both datasets all animals are not statistically different. This means that the animals are producing statistically equal reactions. This makes the analysis of the other effects easier as there are no unusual effects happening within the rabbits.

3.3 *Model Adequacy*

For both versions of the dataset we can run residual analysis to ensure that the models are following the proper assumptions. The main test that will be used is the test of the qqplot. Figure one was obtained from the full dataset and figure two was obtained from the reduced dataset. Both qqplots indicate the data follows a normal distribution with a right skew to it. The skew can be turned removed by square rooting the draize score. This makes the models follow the qqplot with a normal distribution.

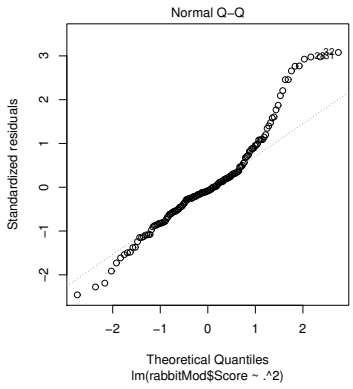

Figure 2. The Reduced Datasets QQ Plot.

3.4 *Parameters*

From the anova tables for the full dataset (table 3) and the reduced dataset (table 4) we can determine was factors are statistically important and what factors are not. We will be looking at the null hypotheses drug has no effect on the draize score, dose has no effect on draize score, and time has no effect. These will be versus the alternate hypotheses that these factors do have an effect.

3.4.1 *Full Dataset.* For the full datatset the anova table (table 3) is obtained. The p values for drug, dose and, time are 0.03778, 1.688e-09, 2.2e-16 respectivly. That means that all three factors are important. Ignoring the effects of two plus factor interactions. However, what effects do these factors have? To see the effects we look at the effects table.

See Table 5

From this table we can see the effect the drug has for the full model. Drug A produces an effect of causing a score that is 0.4 less than that of drug B. The effect of dose ranges from differences of 1.6 to 0.1. The effect time has is parabolic, the more time that goes on the larger the decrease there is. From this we can see that smaller the dose smaller the irritation. This should not be surprising as the less irritant there is the less irritating it can do. We can also see that drug A causes less lesion than drug B. This can be very important to know as most companies want to minimize the adverse effects their product has on consumers.

3.4.2 *Partial Dataset.* For the partial dataset the the anova table (table 4) is obtained. The p values for drug, dose and, time are 0.298827, 0.001764, 2.2e-16 respectively. For this model only dose and time are significant at the 5 percent level. This could be caused by the lack of 0s extending the dataset and acting as a marker for how quickly the drug gets out of the system. To see the effects of this we need another effects table.

However upon attempting to get this effects table R was only producing effects for rabbit. This leads me to believe that the addition of all the 0 values that even out the dataset and give all treatment combinations a value has effects for the way tests are conducted. I believe that because they are uneven the tests become skewed. This in tern makes the standard tests useless.

3.4.3 *Section Conclusions.* Based on these tests, for the full model all basic interactions minus the block effect (rabbit) are all significant and that there are, although not tested, interactions between parameters. For the reduced dataset only dose and time were significant. The effects of which were unable to be analyzed yet it still brought forward useful information. That information is that the extra 0's in the model do not hinder the analysis, they actually improve the results and help balance the number of observations for each treatment combination.

4. Conclusions

Overall, the reduced dataset contained less information and thus provided less results and was in tern less useful. However for the full dataset the information was able to be found out through several analyses that were conducted. This information included that drug A caused less ocular lesions than drug B resulting in a lower draize score. Time had a effect on reducing the lesions that increased the more time had passed. On a more personal note of what information the dataset contained, was the information that even though there may be what appears to be lots of useless additional information like an abundance of 0 values as time goes on, still contains a lot of information and can act as a regulatory system to even out the data. This effect also helps conduct analyses by providing further information as to what happens after the bounds of the experiment, rather than having a possible drop below 0 values when you have a minimum value.

Table 1
Levenes Test Table (Full Dataset)

	Df	F Value	P (>F)
Group	31	0.3245	0.9998
	192		

Table 2
Levenes Test Table (Reduced Dataset)

	Df	F Value	P (>F)
Group	31	0.3753	0.9989
	135		

Table 3
General Model Anova Table (Full Dataset)

Source	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Drug	1	10.04	10.04	4.3876	0.03778
Dose	1	93.56	93.56	40.8927	1.688e-9
Animal	29	51.62	1.78	0.7780	0.78370
Time	1	550.86	550.86	240.7599	2.2e-16
Drug:Time	1	12.89	12.89	5.6347	0.01879
Dose:Time	1	3.16	3.16	1.3812	0.24165
Animal:Time	29	8.71	0.30	0.1313	1.00000
Residuals	160	366.08	2.29		

Table 4
General Model Anova Table (Reduced Dataset)

Source	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Drug	1	2.092	2.092	1.0904	0.298827
Dose	1	19.783	19.783	10.3116	0.001764
Animal	29	24.831	0.856	0.4463	0.992623
Time	1	309.342	309.342	161.2353	2.2e-16
Drug:Time	1	8.751	8.781	4.5766	0.034774
Dose:Time	1	18.051	18.051	9.4087	0.002759
Animal:Time	29	82.596	2.848	1.4845	0.076956
Residuals	103	197.613	1.919		

Table 5
Effects Table (Full Dataset)

Drug	A	B					
Effect	-0.2117	0.2117					
Dose	2.5mg	5mg	12.5mg	25mg	50mg	100mg	
Effect	-0.5681	-0.6992	-0.3933	-0.2622	0.0874	1.0487	
Time	15	120	1440	2880	4320	5760	8640	10080
Effect	1.6504	1.6016	0.9884	0.3195	-0.3494	-1.0422	-2.1889	-3.0250

Dataset: http://webserv.jcu.edu/math//faculty/TShort/Bradstreet/part6/part6-table3.html