Brian Holliday

Professor Yuce

Computational Statistics

22 May 2020

Project 3: Two-Way ANOVA Test

An ANOVA or Analysis of Variance Test is typically used to check if there is a difference in the mean value of a variable with respect to some perimeter. This is like the T-Test, but it is different in that with an ANOVA test you can test for two or more classifications for a parameter. For example, if you were to check sales for company by month, you could test it for 12 separate months. You could check the means of 12 separate months at the same time with ANOVA. If this was a T-Test you might have to test for two 6 months periods.

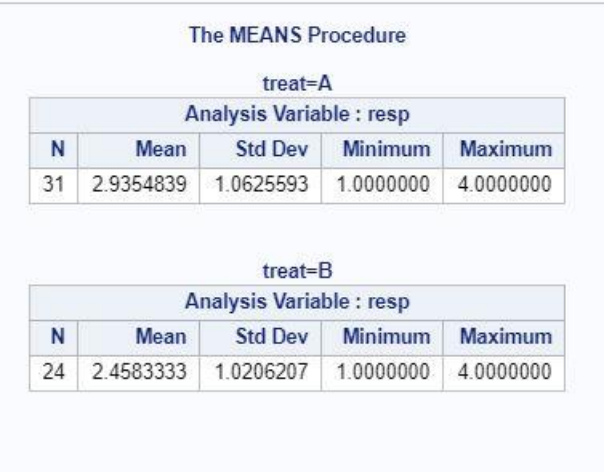
The ANOVA test works by running an F-test on the variation between the groups over the variation within the group. If the variation between the groups is high and the variation within the group is low, then typically would reject the null hypothesis. If the variation between the groups is low and the variation is high, then typical we would fail to reject the null hypothesis. In this case the null hypothesis would be that the mean values of a test parameter with respect to the groups that are being tested are the same. In this case we used the Two-Way ANOVA test to test cholesterol data.

Patients in this study were participating in a study with the aim to lower gastroparesis levels for the people in involved. There was one group A that was given an experimental treatment and another group B that was given dietary changes. The history of the participants with respect to gastroparesis episodes was also recorded as well. The categories for this were, no prior episodes, one prior episode, and two or more prior episode. I chose to do two separate ANOVA tests for this data and another Two-Way ANOVA test. A Two-Way ANOVA test just like a One-Way ANOVA, but the test is done to the mean of a parameter with respect to two difference categories. To go back to our sales example, we could run the ANOVA test on sales for the month and days of the week. Therefore, you might see a mean difference in sales on Saturdays in December compared to Mondays in January. This works for our gastroparesis data because we can test not only test for the treatment and history, but we can test for combination of the history and treatment. How well the treatment works in ranked on a response level from 1-4, with 4 being the best and 1 being the worst.

From running our ANOVA test, we can see that there was a clear difference in the means for those that are no prior episodes, but for the treatment we could not rule out that the means are the same although A did do better in the study. So, we could walk away saying the better treatment is A, but it is not significant. The cross product of seemed to have no difference in the means. Our estimate test shows that the mean of Hist A, no prior episodes is biggest indicator for response, but if you are suffering with episode regardless of the severity, go with treatment A.

Figures:

Figure 1: Means by Treatment



The MEANS Procedure

treat=A

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
31	2.9354839	1.0625593	1.0000000	4.0000000

treat=B

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
24	2.4583333	1.0206207	1.0000000	4.0000000

In this figure we can see that mean value of response of A is about 0.5 bigger than B.

Figure 2: Means by Hist

The MEANS Procedure

hist=0

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
9	3.4444444	0.7264832	2.0000000	4.0000000

hist=1

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
28	2.7500000	0.9670497	1.0000000	4.0000000

hist=2

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
18	2.3333333	1.1881771	1.0000000	4.0000000

The mean value of the resp 0, no prior incidents is much higher than the other histories.

Figure 3: QQ Response Treatment A

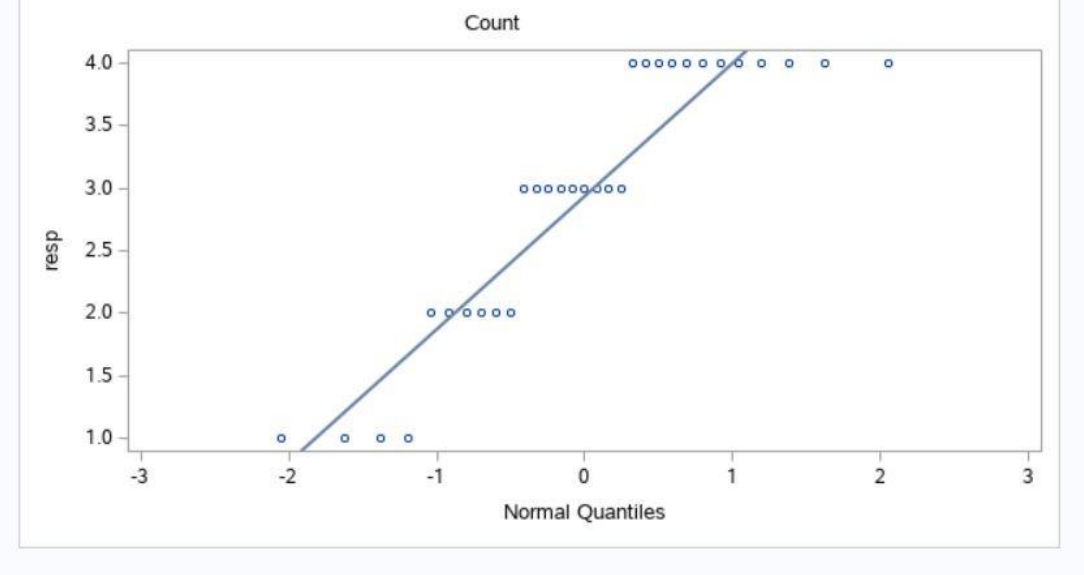
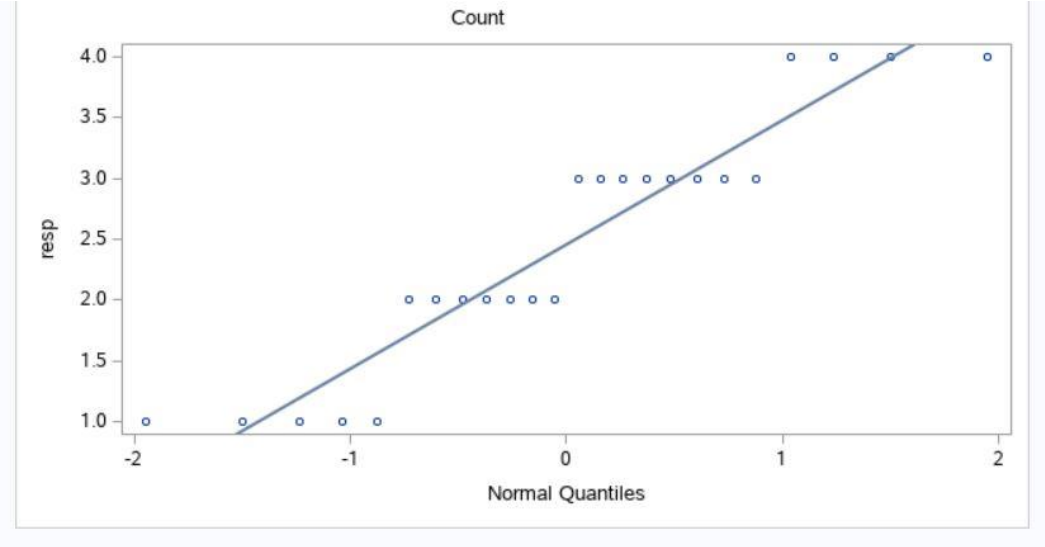


Figure 4: QQ Response Treatment B



We can see that values from the response variable are not normal. The are not close to normal with only four different values that are equidistant from each other.

Figure 5: Two Way Means

The MEANS Procedure

treat=A hist=0

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
5	3.6000000	0.5477226	3.0000000	4.0000000

treat=A hist=1

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
16	3.0000000	0.9660918	1.0000000	4.0000000

treat=A hist=2

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
10	2.5000000	1.2692955	1.0000000	4.0000000

treat=B hist=0

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
4	3.2500000	0.9574271	2.0000000	4.0000000

treat=B hist=1

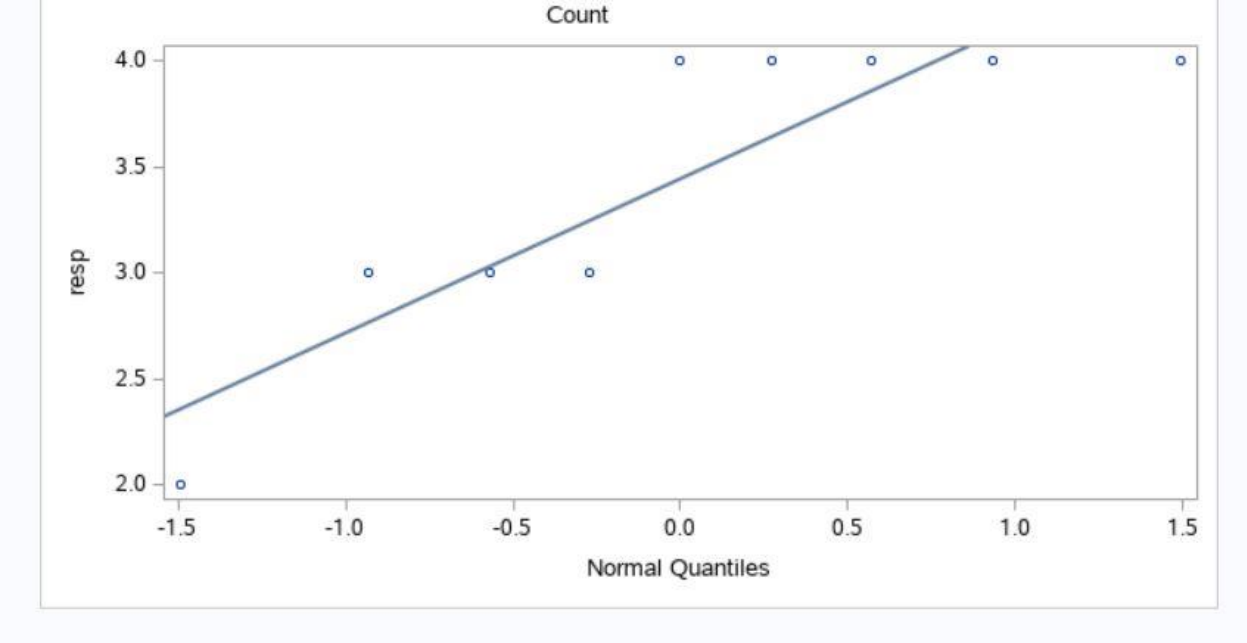
Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
12	2.4166667	0.9003366	1.0000000	4.0000000

treat=B hist=2

Analysis Variable : resp				
N	Mean	Std Dev	Minimum	Maximum
8	2.1250000	1.1259916	1.0000000	4.0000000

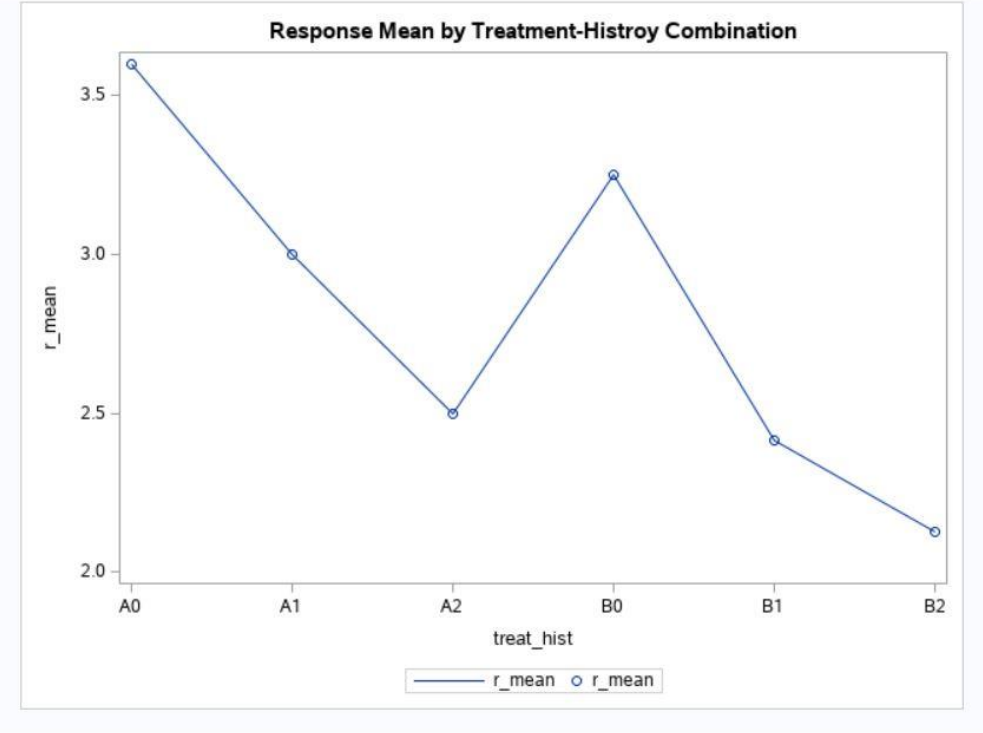
Two way means for response

Figure 6: QQ Response Hist 0



We can see that with respect to history the responses are not coming for a normal distribution.

Figure 7: Treatment-History vs Response Mean



We could see that response mean is higher for treatment A for the same history but is it significant.

Figure 8: GLM Response

The GLM Procedure

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	10.66742424	2.13348485	2.08	0.0837
Error	49	50.24166667	1.02534014		
Corrected Total	54	60.90909091			

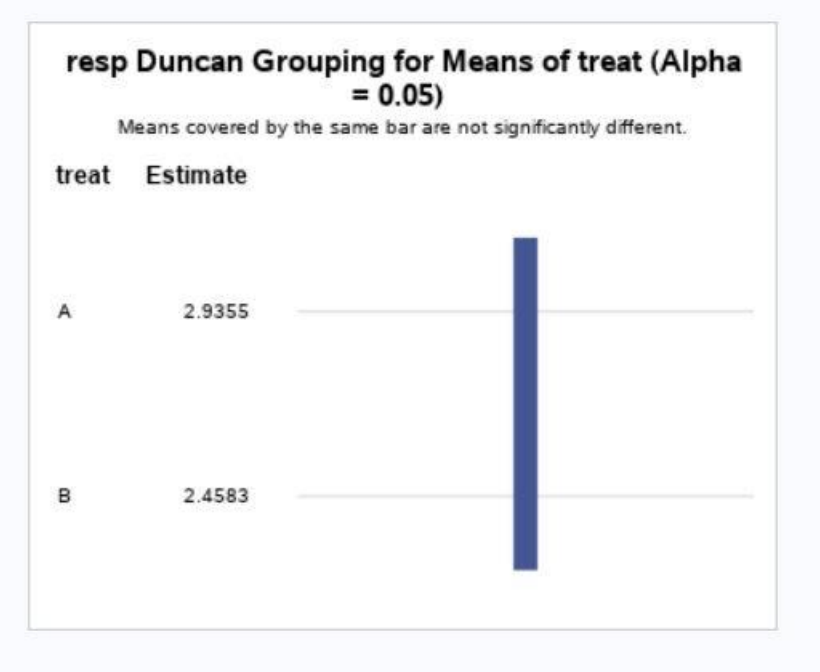
R-Square	Coeff Var	Root MSE	resp Mean
0.175137	37.12833	1.012591	2.727273

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	1	3.07978983	3.07978983	3.00	0.0894
hist	2	7.42802304	3.71401152	3.62	0.0341
treat*hist	2	0.15961137	0.07980569	0.08	0.9252

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	1	2.08536379	2.08536379	2.03	0.1602
hist	2	7.34270996	3.67135498	3.58	0.0354
treat*hist	2	0.15961137	0.07980569	0.08	0.9252

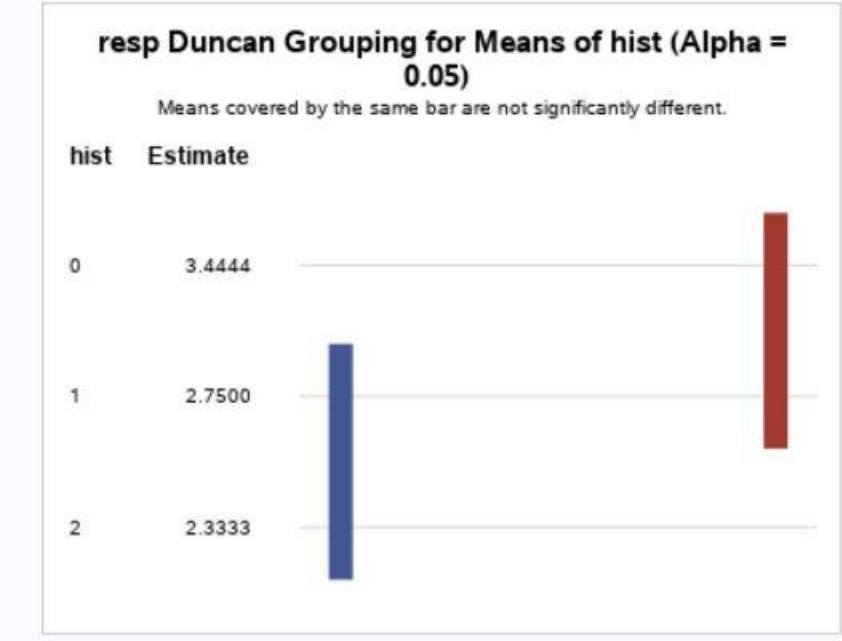
We can see or our model, we would fail to reject the null hypothesis. Which means that we cannot rule our that that the means are the same for our model with respect to response. For the treatment we fail to reject the null hypothesis, we cannot rule out that the means are the same for our model for treatment with respect to response. We can say that the means are not the same between people with different histories. For the cross-product treat*hist, there is strong evidence for the null hypothesis.

Figure 9: Duncan A-B



Duncan Test shows that the means are the same between treatment A and B

Figure 10: Duncan History



We have and association rule here. History 1 and 2 are the same, 0 and 1 are the same. We cannot rule out that they are all the same.

Figure 11: Estimate

The GLM Procedure

Dependent Variable: resp

Parameter	Estimate	Standard Error	t Value	Pr > |t|
hist 0 vs and hist1 and hist 2	0.91458333	0.37298145	2.45	0.0178

Our estimate shows that history 0 is different from 1 and 2.

Code:

```
DATA gastro ;
        infile '/home/u45153057/gastro.dat' expandtabs;
        input num treat $ hist resp ;
run ;

PROC print data = gastro ;
run ;

PROC sort data = gastro out = sort_gastro ;
        by treat ;
run ;

PROC print data = sort_gastro ;
run ;

PROC sort data = gastro out = sort_gastro2 ;
        by hist ;
run ;

PROC print data = sort_gastro2 ;
run;

PROC means data = sort_gastro ;
        var resp ;
        by treat ;
run ;

PROC means data = sort_gastro2 ;
```

```
        var resp ;
        by hist ;
run ;

PROC sort data = gastro out = sort_gastro3 ;
        by treat hist ;
run ;

PROC print data = sort_gastro3 ;
run;



PROC univariate data = sort_gastro3 normal plot ;
        var resp ;
        by treat ;
run ;

PROC univariate data = sort_gastro2 normal plot ;
        var resp ;
        by hist ;
run ;

PROC freq data = sort_gastro ;
        by treat ;
run ;

PROC sort data = gastro out = sort_gastro4 ;
        by treat hist ;
run ;
```

```
PROC print data = sort_gastro4 ;
run ;


PROC means data = sort_gastro4 ;
        var resp ;
        by treat hist ;
        output out = gastro_m mean = r_mean ;
run ;

PROC print data = gastro_m ;
run ;


/* Combine the treatment and history columns to
find the mean by the treatment and history combination */

DATA gastro_m2 ;
        set gastro_m ;
        treat_hist = catt(treat, hist) ;
run ;

PROC print data = gastro_m2 ;
run ;

PROC sgplot data = gastro_m2 ;
        title 'Response Mean by Treatment-Histroy Combination' ;
        series x = treat_hist y = r_mean ;
        scatter x = treat_hist y = r_mean ;
run ;
```

```
PROC glm data = sort_gastro ;
        class treat hist ;
        model resp = treat hist treat*hist ;
        means treat hist treat*hist / duncan ;
        estimate 'hist 0 vs and hist1 and hist 2' hist 1 -.5 -.5 ;
        estimate '.60A vs .40B ' treat 0.6 -.4
run ;
```