

STAT 514 Homework 3

1. A factor with three levels was studied in an experiment. The data is given as follows, in which the first column includes the treatments and the second column includes the responses. You can download the data, hw4.dat, from the Blackboard Learning:

```
1 2.23
1 3.04
. ...
3 8.12
```

- (a) Test the hypothesis that there is no difference across the treatments (use $\alpha = .05$).

solution:

Usual ANOVA shows that $F_0 = 21.31$ and P-value is less than 0.001. We reject null hypothesis and conclude that difference exists among the treatment effects.

- (b) Use proper plots to check whether the constant variance assumption is valid. Can you use a formal test to support your conclusion?

solution:

Residual plots indicate that constant variance assumption might be invalid. Both Levene's test and bartlett's test report P-value less than 5%, which implies that some remedy is in orde.

- (c) Generate the $\log s_i$ vs. $\log \bar{y}_i$. plot (template code "trans.sas" is available on Blackboard) and estimate the possible transformation for variance stabilization.

solution:

Using the SAS file for approximate Box-Cox transformation, one has

$$\log s_i = -0.714 + .835 \log \bar{y}_i,$$

with $\hat{\beta} = 0.835$, which implies the possible power transformation will be

$$Y' = Y^{1-\hat{\beta}} = Y^{.165}.$$

Since $\hat{\beta}$ is sort of close to one, it is also appropriate to use

$$Y' = \log(Y)$$

- (d) Use the formal Box-Cox procedure to identify the optimal transformation. You need use template code "trans1.sas" for this data set and generate proper output and plot to make the choice.

solution:

Use the sas file for the exact Box-Cox transformation, SS_E is minimized at **Approximately** $\lambda = 0.25$. (A more accurate result is $\lambda \approx 0.35$) The transformation is

$$Y' = Y^{0.25}$$

In fact, $\lambda = .25$ and $\lambda = 0$ might not be different statistically, so both transformations can be used.

- (e) Repeat (a) and (b) for the transformed response. You may need use some sas function in the data step to generate the new responses.

solution:

Apply ANOVA to the transformed responses. Residual plots and formal tests show that the violation of constant variance assumption has been corrected.

2. Four different designs for a digital computer circuit are being studied to compare the amount of defects. The following data have been obtained (“defects.dat” on the Blackboard):

```
design defect
1 7
1 2
1 4
1 7
1 2
. .
4 2
2 7
```

- (a) Is the amount of defects present the same for all four designs? (use $\alpha = 0.05$).

solution:

Apply ANOVA first to the data. It appears that the p value is significant.

- (b) Analyze the residuals from a). In particular, how do you think about the normality assumption? Can you use any formal test to support your conclusion?

solution:

QQ plot reveals some departure from normality, formal tests (Shapiro-Wilk’s test) report p-values less than 5%. This implies that the normality assumption is not valid and the result from ANOVA are questionable

- (c) Use the Kruskal-Wallis test for the data and compare the results with a).

solution:

Use *PROC NPAR1WAY* to perform the Kruskal-Wallis test.

```
Kruskal-Wallis Test
Chi-Square      13.3467
DF              3
Pr > Chi-Square 0.0039
```

The conclusion from Kruskal-Wallis is consistent with that from ANOVA in this problem.

3. In a study of the effect of glucose on insulin release, identical specimens of pancreatic tissue were equally and randomly assigned to three different levels of glucose concentration (low, medium, high). The amount of insulin produced by each tissue after treatment was recorded. The data set, “insulin.dat”, can be downloaded from the Blackboard Learning. In “insulin.dat”, the first column contains the amounts at the low concentration, the second column the amounts at the medium concentration, and the third column the amounts at the high concentration. To read a data set like this, do the following in the data step to create a data set suitable for the glm procedure.

```
data insulin;
infile 'H:\dataset\insulin.dat';
input t1 t2 t3;
y=t1; trt=1; output;
y=t2; trt=2; output;
y=t3; trt=3; output;
drop t1 t2 t3;
```

This creates a treatment variable (trt) and a response variable (y).

- (a) Test the hypothesis that there is no difference across treatments in the amount of insulin produced (use $\alpha = 0.01$).

solution:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.28516000	1.64258000	74.11	<.0001
Error	57	1.26340000	0.02216491		
Corrected Total	59	4.54856000			

Significant!

- (b) Diagnose whether the assumptions are valid?

solution Residual plots and QQ plot are generated to check model assumptions. Two formal tests for constant assumption are Levene's test and Bartlett's test. For normality, Kolmogorove-Smirnov test and Anderson-Darling test can be used.

- (c) Construct 99% CIs for the average insulin amounts at the low, medium and high glucose concentrations separately (not simultaneously). The formula can be found in Montgomery Section 3-3.3 Equation (3-12). Based on each confidence interval, does it appear the average amount of insulin is significantly different than 3.5?

solution

There is no significant violation to the model assumptions.

The GLM Procedure

t Confidence Intervals for y

Alpha	0.01
Error Degrees of Freedom	57
Error Mean Square	0.022165
Critical Value of t	2.66487
Half Width of Confidence Interval	0.088714

trt	N	Mean	99% Confidence Limits
3	20	4.05100	3.96229 4.13971
2	20	3.90500	3.81629 3.99371
1	20	3.49800	3.40929 3.58671

For treatment 3 and 2, there are significant difference between 3.5 and treatment means. For treatment 1, there is no significant difference between 3.5 and treatment mean.

Code:

```
proc glm;
class trt;
model y = trt;
means trt / hovtest=bartlett hovtest=levene hovtest=bf;
means trt / alpha=0.01 lsd clm;
output out=diag p=pred r=res;

proc sort; by pred;
symbol1 v=circle i=sm50; title1 "Residual Plot";
proc gplot; plot res*pred/frame;

proc univariate data=diag normal noprint;
var res; qqplot res / normal (L=1 mu=est sigma=est);
histogram res / normal;

run;
```