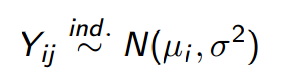
Final STP 531 Lu Wang

**ONE**

**PART I**

(a). Should use the AVONA model. And assumption is



(b) H0: μ.1= μ.2= μ.3= μ.4

Ha: Not all μi are equal

From SAS output, F value for the test is 127.60 with DF=3, 12, pvalue<0.0001<α=0.05. Thus, reject the null. Not all μi are equal. The variety of wheat has an effect on the growth of plants.

(c) Using TUKEY method. The confidence intervals are:

| **i** | **j** | **Difference Between Means** | **Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)** | |
| --- | --- | --- | --- | --- |
| **1** | **2** | -3.400000 | -4.428660 | -2.371340 |
| **1** | **3** | 2.220000 | 1.191340 | 3.248660 |
| **1** | **4** | 4.060000 | 3.031340 | 5.088660 |
| **2** | **3** | 5.620000 | 4.591340 | 6.648660 |
| **2** | **4** | 7.460000 | 6.431340 | 8.488660 |
| **3** | **4** | 1.840000 | 0.811340 | 2.868660 |

This result shows that at 90%family confidence coefficient, we can conclude that the mean of the four levels can be ranked as: 2>1>3>4

(d) From sas output, the interval is: (-4.275420, -2.524580) for μ.1- μ.2. This shows that at 95% confidence level, we can conclude that the mean height for level 1 is lower than level 2.

(e) From SAS output, t value for the test is -8.46, with pvalue<0.0001<α=0.05. Thus, reject the null, conclude there is difference in the mean plant height for the first tow group. This is consistent with the result in d.

(f) The t value for the test is -8.46 for comparing μ.1- μ.2. This means that μ.1 is small than μ.2. And pvalue for two tale test is <0.0001, which means the pvalue for one tale is <0.0002<α=0.025. Thus reject the null, conclude that at α=0.025, μ.1< μ.2.

(g) The block efficiency can be calculated using E=[(nb-1)MSBL+nb(r-1)MSE]/[(nbr-1)MSE]

From SAS output. The related data are:

| **Source** | **DF** | **Type I SS** | | **Mean Square** | **F Value** | | **Pr > F** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **variety** | 3 | 154.4920000 | | 51.4973333 | 127.60 | | <.0001 |
| **block** | 4 | 12.3930000 | | 3.0982500 | 7.68 | | 0.0026 |
| **Error** | 12 | 4.8430000 | 0.4035833 | | |

E=[4\*3.0982500+5\*3\*0.4035833]/[(5\*4-1)\* 0.4035833]= 2.40565343668

Edfb=4\*3=12

Edfr=5\*4-1=19

E’=(12+1)(19+3)\*E/[(12+3)(19+1)]= 2.29338960963

So the blocking variable is not very efficient, but can still reduce half the observations needed if without block.

**PART2**

(a)

(b) Input the dataset into SAS. Use the ANOVA model to test whether the mean of different varieties are equal with sas code:

**proc** **glm** data=wheat alpha=**0.05**;

class variety block;

model growth=variety block;

**run**;

**quit**;

the output AVONA TABLE from SAS is below:

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 7 | 166.8850000 | 23.8407143 | 59.07 | <.0001 |
| **Error** | 12 | 4.8430000 | 0.4035833 |  |  |
| **Corrected Total** | 19 | 171.7280000 |  |  |  |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **variety** | 3 | 154.4920000 | 51.4973333 | 127.60 | <.0001 |
| **block** | 4 | 12.3930000 | 3.0982500 | 7.68 | 0.0026 |

We can get the F value and test statistics from the table and decide to reject the null.

(c) Since we want to get all pairwise comparison, Tukey procedure is better (produce narrower C.I.)

To get all the pairwise comparison of level means, I used this SAS code:

**proc** **glm** data=wheat;

class variety block;

model growth=variety block;

lsmeans variety /adjust=TUKEY cl alpha=**0.1**;

**run**;

**quit**;

From the output, we can get all the confidence intervals. If the difference between two levels are negative, it means the lsmean for first level is smaller than second, vice versa. Thus, from SAS output.

| **i** | **j** | **Difference Between Means** | **Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)** | |
| --- | --- | --- | --- | --- |
| **1** | **2** | -3.400000 | -4.428660 | -2.371340 |
| **1** | **3** | 2.220000 | 1.191340 | 3.248660 |
| **1** | **4** | 4.060000 | 3.031340 | 5.088660 |
| **2** | **3** | 5.620000 | 4.591340 | 6.648660 |
| **2** | **4** | 7.460000 | 6.431340 | 8.488660 |
| **3** | **4** | 1.840000 | 0.811340 | 2.868660 |

We can get that mean of 1< 2, 1>3 1>4 2>3 2>4 3>4. Combine all these information together, we can get the rank of 2>1>3>4

(d) Use the following SAS script to get the confidence intevel:

**proc** **glm** data=wheat alpha=**0.05**;

class variety block;

model growth=variety block;

lsmeans variety/cl pdiff;

**run**;

**quit**;

(e) The following script is used to get the result:

**proc** **glm** data=wheat alpha=**0.05**;

class variety block;

model growth=variety block;

estimate 'μ.1 vs μ.2' variety **1** -**1** **0** **0**;

**run**;

**quit**;

we can get the test result:

| **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **µ.1 vs µ.2** | -3.40000000 | 0.40178767 | -8.46 | <.0001 |

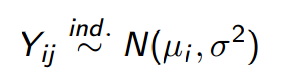
Pvalue<0.05, thus, we reject the null. There is difference in mean in level 1 and 2.

(f) the same script in (e) will be used. The t value for one tale test and two tale test are the same. The pvalue for one tale test is two times the pvalue for two tale test. So pvalue<0.0002 for this problem. Smaller than 0.025, and since the estimate difference is negative. We conclude that μ.1< μ.2.

**TWO**

**PART I**

(a) Should use the AVONA model. And assumption is



(b) The AVONA table is :

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3 | 154.4920000 | 51.4973333 | 47.80 | <.0001 |
| **Error** | 16 | 17.2360000 | 1.0772500 |  |  |
| **Corrected Total** | 19 | 171.7280000 |  |  |  |

(c) H0: μ1= μ2= μ3= μ4

Ha: Not all μi are equal

From SAS output, F value for the test is 47.80 with DF=3, 16, pvalue<0.0001<α=0.05. Thus, reject the null. Not all μi are equal. The variety of wheat has an effect on the growth of plants.

**PART II**

(b) The AVONA table is retrieved using the following code:

**proc** **glm** data=wheat alpha=**0.05**;

class variety;

model growth=variety;

**run**;

**quit**;

(c) We can get the F value and test statistics from the table and decide to reject the null.

**THREE**

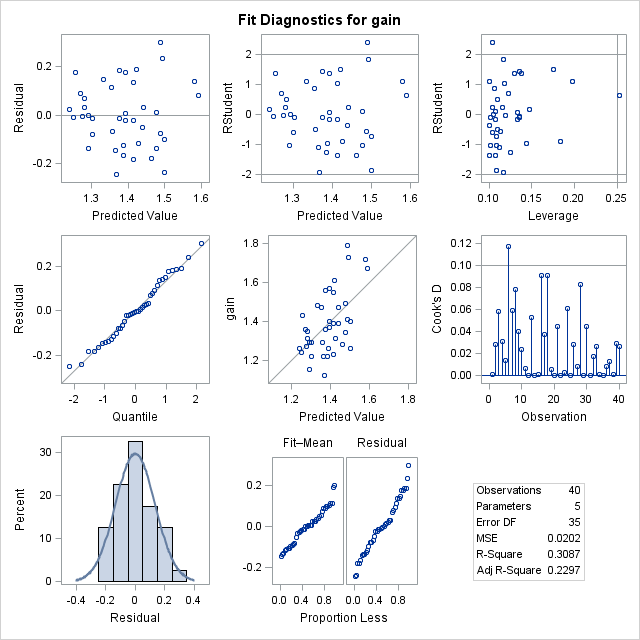
**PART I**

(a)



And the assumption is that εij~ N (0, σ2), i.i.d.

Residual analysis is performed to check the appropriateness of the model. QQ plot is also provided there. From the plot, we can see the variance is constant, the residuals seem to be normally distributed and the residuals seems to be independent.



Correlation coefficient test is also used to test the residuals are normal or not. Based on the output statistics, pvalue<0.001, we conclude the residuals are normally distributed

| **Pearson Correlation Coefficients, N = 40  Prob > |r| under H0: Rho=0** | | |
| --- | --- | --- |
|  | **r** | **rankit** |
| **r** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.99298 | | <.0001 | |
| **rankit** | |  | | --- | | 0.99298 | | <.0001 | | |  | | --- | | 1.00000 | |

Based on all the analysis above, the model is appropriate here,

(b) AVONA table is provided:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | | **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** | | --- | --- | --- | --- | --- | --- | | **Model** | 4 | 0.31577099 | 0.07894275 | 3.91 | 0.0100 | | **Error** | 35 | 0.70700651 | 0.02020019 |  |  | | **Corrected Total** | 39 | 1.02277750 |  |  |  |      | **R-Square** | **Coeff Var** | **Root MSE** | **gain Mean** | | --- | --- | --- | --- | | 0.308739 | 10.23788 | 0.142127 | 1.388250 |      | **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** | | --- | --- | --- | --- | --- | --- | | **diet** | 3 | 0.16231353 | 0.05410451 | 2.68 | 0.0620 | | **weight** | 1 | 0.13822349 | 0.13822349 | 6.84 | 0.0130 | |  |  |  |  |  | |  |  |  |  |  |

From the AVONA table, we can see the overall pvalue is 0100, which is significant at α=0.05 level. As a result, we conclude that there is difference in the mean of weight gain at different level of diets and initial weight. However, when we look into the SSIII table. The main effect of diet has a pvalue of 0.062, which is not significant at α=0.05 level. It seems the initial weight might be a more important factor here.

(c) From the SAS output, we can get the point estimate for the slope is 0.005384841, and the 95% confidence interval is (0.001205780, 0.009563903)

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Intercept** | 1.058450356 | B | 0.11136958 | 9.50 | <.0001 | 0.832358093 | 1.284542618 |
| **diet 1** | 0.112614278 | B | 0.06435668 | 1.75 | 0.0889 | -0.018036722 | 0.243265277 |
| **diet 2** | -0.035616627 | B | 0.06456179 | -0.55 | 0.5847 | -0.166684024 | 0.095450770 |
| **diet 3** | 0.101153055 | B | 0.06396833 | 1.58 | 0.1228 | -0.028709566 | 0.231015677 |
| **diet 4** | 0.000000000 | B | . | . | . | . | . |
| **weight** | 0.005384841 |  | 0.00205854 | 2.62 | 0.0130 | 0.001205780 | 0.009563903 |

(d) The confidence interval as in SAS output is

| **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| --- | --- | --- | --- | --- | --- | --- |
| **estimate new obs** | 1.33515452 | 0.04536691 | 29.43 | <.0001 | 1.24305479 | 1.42725426 |

(1.24305479,1.42725426)

(e) The confidence intervals are in SAS output below:

| **Least Squares Means for Effect diet** | | | | |
| --- | --- | --- | --- | --- |
| **i** | **j** | **Difference Between Means** | **Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)** | |
| **1** | **2** | 0.148231 | -0.011627 | 0.308089 |
| **1** | **3** | 0.011461 | -0.148531 | 0.171453 |
| **1** | **4** | 0.112614 | -0.049213 | 0.274442 |
| **2** | **3** | -0.136770 | -0.296932 | 0.023393 |
| **2** | **4** | -0.035617 | -0.197960 | 0.126727 |
| **3** | **4** | 0.101153 | -0.059698 | 0.262004 |

Bonferroni method is used here. It shows that at 90% family confidence coefficient level, the mean of weight gain for any pairwise comparison is not significant at this level. So we conclude the main effect are not different here at different levels.

**PART II**

(a) In this problem, we are investigating the effect of three factors. So the factor effect model should be appropriate. From the fitting diagnostics, we can see most observations are fitted. There are two possible outliers, but should not influence the model too much, because AVONA model is robust against slight violation of assumptions.

Both (a) and (b) is gained using the following SAS script:

**proc** **glm** data=test plots=diagnostics;

class diet;

model gain=diet weight/ss3 clparm;

output out=out r=r;

**run**;

**quit**;

**%macro** corrTest(data=, residual=r, plot=on);

%\*\*\* data= (name of the data set that contains the residual);

%\*\*\* residual= (variable name of the residual in the data set, default is 'r');

%\*\*\* plot = (on/off), the normal probability will not be created if plot=off;

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\* order the residual;

proc sort data=&data. out=s&data.;

by &residual.;

run;

%\*\*\* use sql to obtain the size of the data and assign the size to a macro variable n;

proc sql noprint;

select count(\*) into :n

from s&data.;

quit;

data s&data.;

set s&data.;

kID = \_n\_; %\*\* create a cloumn indicating the rank of the response;

rankit = quantile('NORMAL', (kID-(**3**/**8**))/(&n.+**.25**)); %\*\* create the rankit;

run;

%\*\*\* normal probability plot, can be turned off by assign plot=off;

%if &plot=on %then

%do;

goptions reset=all;

symbol1 value=dot color=blue height=**2**;

proc gplot;

plot &residual.\*rankit;

run;

goptions reset=all;

%end;

%\*\*\* obtain the correlation coefficient;

proc corr;

var &residual. rankit;

ods select PearsonCorr;

run;

**%mend**;

\*\*\*\* To use the macro function, just enter the data name that contains the residual;

\*\*\*\* You can change the value of one or more parameters, as in the example below;

%***corrTest***(data=out);

(c) the slope should be the parameter estimate of weight. The following SAS code is used to generate the result:

**proc** **glm** data=test plots=diagnostics;

class diet;

model gain=diet weight/ ss3 clparm;

**run**;

**quit**;

(d) We want to estimate new observations, so estimate procedure should be used. As a result, the following SAS script is used to retrieve the result:

**proc** **glm** data=test plots=none;

class diet;

model gain=diet weight/clparm alpha =**0.05**;

estimate 'estimate new obs' intercept **1** diet **0** **1** **0** **0** weight **58**;

**run**;

**quit**;

(e) We want to estimate new observations, so estimate procedure should be used. As a result, the following SAS script is used to retrieve the result:

**proc** **glm** data=test;

class diet;

model gain=diet weight;

lsmeans diet/cl adjust=Bon alpha=**0.1**;

lsmeans diet/cl adjust=SCHEFFE alpha=**0.1**;

**run**;

**quit**;

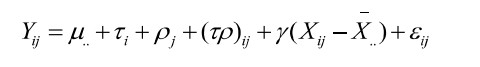
Both Bonferroni adjust and scheffe are tried out. Bonferroni yields narrower confidence interval. So is used in the result. The intervals generated weith schefee is below:

| **Least Squares Means for Effect diet** | | | | |
| --- | --- | --- | --- | --- |
| **i** | **j** | **Difference Between Means** | **Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)** | |
| **1** | **2** | 0.148231 | -0.016840 | 0.313302 |
| **1** | **3** | 0.011461 | -0.153748 | 0.176671 |
| **1** | **4** | 0.112614 | -0.054491 | 0.279719 |
| **2** | **3** | -0.136770 | -0.302155 | 0.028616 |
| **2** | **4** | -0.035617 | -0.203254 | 0.132021 |
| **3** | **4** | 0.101153 | -0.064944 | 0.267250 |

**FOUR**

**PART I**

(a)



And the assumption is that εij~ N(0, σ2), i.i.d.

From the SAS output below, we can see the pvalue for Type III SS is 0.0520for diet main effect. So at α=0.1, we reject the null. Main effect of diet is present.

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **diet** | 3 | 0.17800565 | 0.05933522 | 2.87 | 0.0520 |
| **age** | 1 | 0.03518804 | 0.03518804 | 1.70 | 0.2013 |
| **weight** | 1 | 0.16585630 | 0.16585630 | 8.03 | 0.0080 |
| **diet\*age** | 3 | 0.03512550 | 0.01170850 | 0.57 | 0.6408 |

(b) H0: (7μ11+3μ12)/10= (7μ21+3μ22)/10= (7μ31+3μ32)/10= (7μ41+3μ42)/10

Ha: not all (7μi1+3μi2)/10 are equal

| **Contrast** | **DF** | **Contrast SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **L** | 3 | 0.14403102 | 0.04801034 | 2.33 | 0.0941 |

With pvalue=0.0941, at α=0.1, we reject the null, the main effect of diet is present

(c) The confidence interval as in SAS output is

| **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| --- | --- | --- | --- | --- | --- | --- |
| **estimate new obs** | 1.37268184 | 0.05222376 | 26.28 | <.0001 | 1.26617079 | 1.47919290 |

(1.26617079,1.47919290)

(d) H0: gain>0

Ha: gain<=0

T value-26.28, and pvalue for two tale test<0.0001, which means for one tale test is<0.0002. At α=0.05 level, if pvalue<0.05, reject the null. Else, do not reject. With pvalue<0.0002, we would reject the null. The mean weight gain is positive.

**PART II**

(a) We want to test the main effect of diet. So we should be using type III SS and corresponding F value and p-value. The following SAS code is used to retrieve the table

**proc** **glm** data=test;

class diet age;

model gain=diet age weight age\*diet/ss3;

**run**;

**quit**;

(b)

H0: (7μ11+3μ12)/10= (7μ21+3μ22)/10= (7μ31+3μ32)/10= (7μ41+3μ42)/10

Ha: not all (7μi1+3μi2)/10 are equal

SAS code as below i:

**proc** **glm** data=test plots=diagnostics;

class diet age;

model gain=diet age weight age\*diet;

contrast 'L' diet **1** -**1** **0** **0** age\*diet **0.7** **0.3** -**0.7** -**0.3** **0** **0** **0** **0**,

diet **0** **1** -**1** **0** age\*diet **0** **0** **0.7** **0.3** -**0.7** -**0.3** **0** **0**,

diet **0** **0** **1** -**1** age\*diet **0** **0** **0** **0** **0.7** **0.3** -**0.7** -**0.3**;

**run**;

**quit**;

(c) We want to estimate new observations, so estimate procedure should be used. Weight of the ages are 0.7 for first level and 0.3 for second level. As a result, the following SAS script is used to retrieve the result in c and d:

**proc** **glm** data=test plots=diagnostics;

class diet age;

model gain=diet age weight age\*diet/clparm;

estimate 'estimate new obs' intercept **1** age **0.7** **0.3** diet **0** **1** **0** **0** weight **58** age\*diet **0** **0** **0.7** **0.3** **0** **0** **0** **0** ;

**run**;

**quit**;