

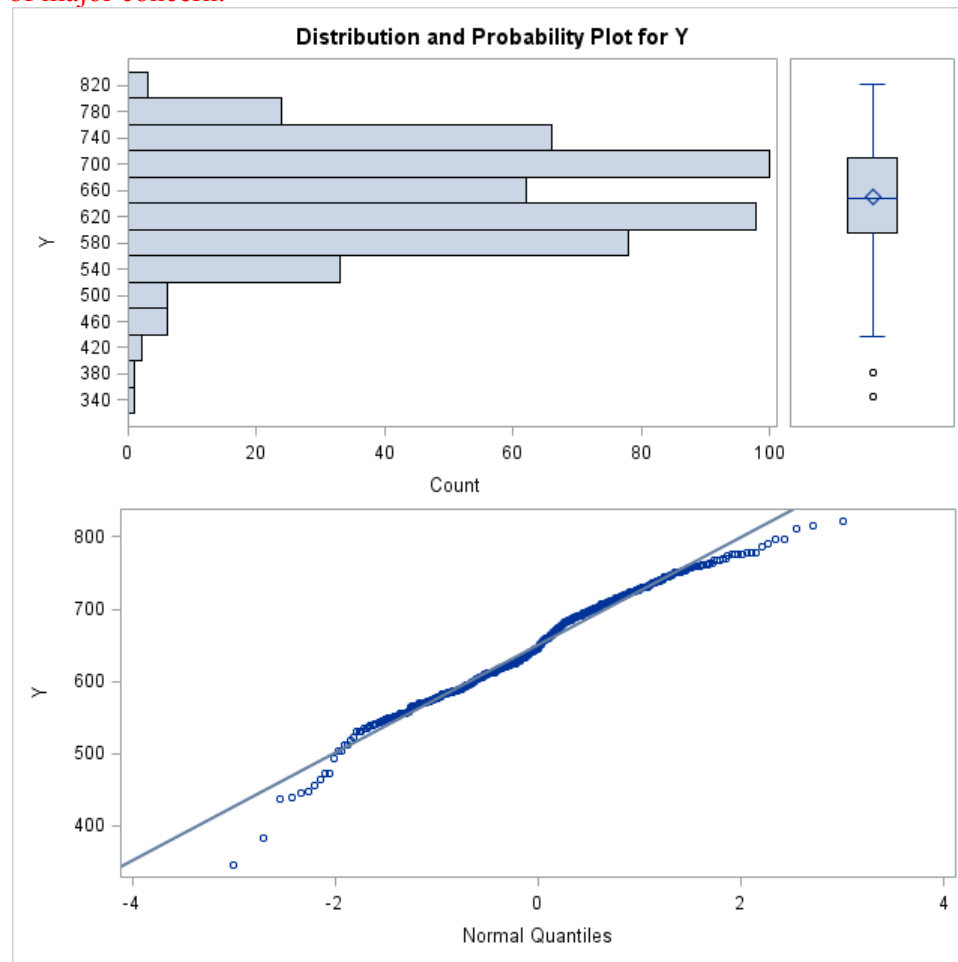
HW 5

Let's look through some data generated by the National Institute of Standards and Technology (NIST) that investigates the **strength** of ceramics. Use the code `5_homeworkCeramic.sas` for the data set in SAS form. For our purposes, we are interested in the three production variables: **tableSpeed**, **feedRate**, and **gritSize**. Additionally, the ceramics are produced in two different batches in **eight different laboratories**. The **batches are coded 1 and 2** to indicate the first batch produced in a given lab and the second batch, respectively.

Let's work through analyzing this data set.

- a) A good first step is to go looking for anomalous observations. In this case, that will mainly be in terms of the response. One way of doing this is to look at summary statistics or histogram.

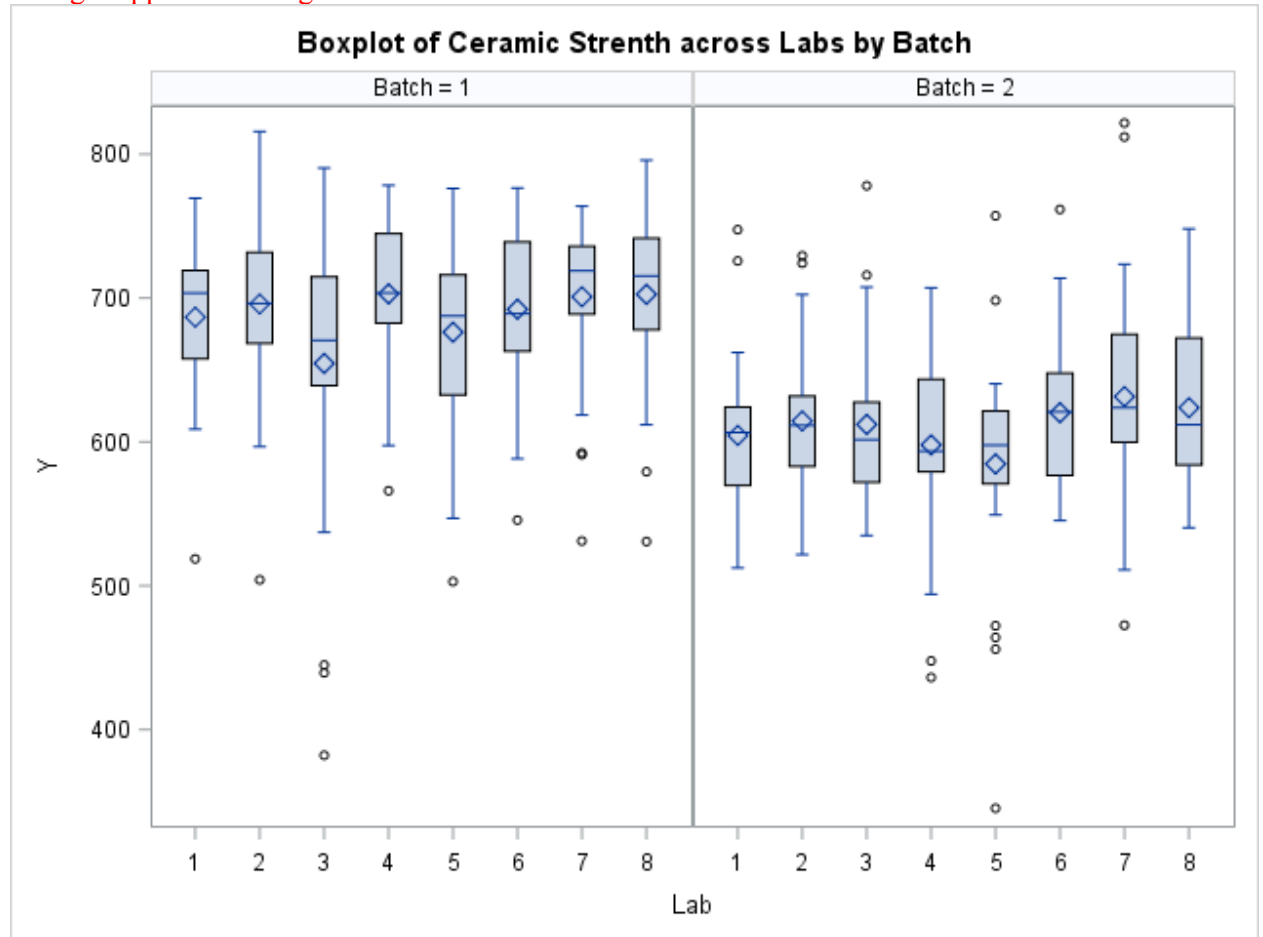
There are 2 observations with extremely low strength, but the sample size is large so that they are not of major concern.



Obs	Id	Lab	Num	Test	Y	tableSpeed	feedRate	gritSize	direction	Trt	Set	Llab	Rep	Batch	Sblab	Set2
151	301	3	1	3	382.133	1	1	-1	-1	4	1	2.75	1	1	2.5	11
270	570	5	30	3	345.294	-1	-1	-1	-1	1	2	5.25	2	2	5.5	18

- b) Next, let's visually investigate the possibility of a batch or lab effect. A useful tool for this is the boxplot. What do your findings suggest?

From the boxplots, mean strength across labs do not appear to be too different, while mean strength appear to be higher in batch 1 for each lab. There could be a batch effect.



- c) Now we need to decide what type of study this was (ideally, we would have designed the study ourselves).

This is a nested study, because the batches are different for different labs thus batch is nested in lab.

Note: You might notice that lab is partially nested with tableSpeed and feedRate. For practice purpose, just ignore it.

- d) Write down a model for this study with pairwise interactions for the main explanatory variables of interest. Include batch and lab, but not as interactions.

$$Y_{ijklmn} = \mu + \mu_j + \mu_k + \mu_l + \mu_{jk} + \mu_{jl} + \mu_{kl} + \mu_m + \mu_{n(m)} + \varepsilon_{ijklmn}$$

μ_j - fixed main effect of j-th level of tableSpeed;

μ_k - fixed main effect of k-th level of feedRate;

μ_l - fixed main effect of l-th level of gritSize;

μ_{jk} - fixed interaction effect of tableSpeed and feedRate;

μ_{jl} - fixed interaction effect of tableSpeed and gritSize;

μ_{kl} - fixed interaction effect of feedRate and gritSize;

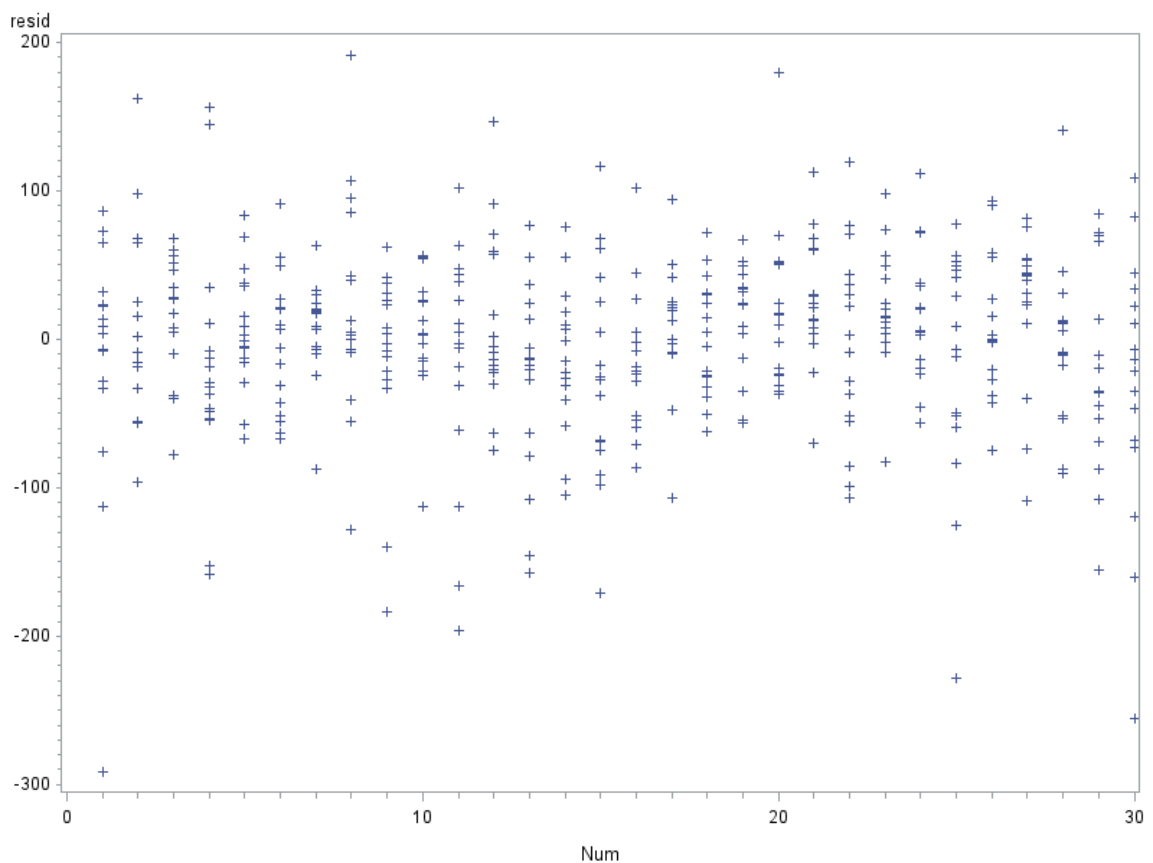
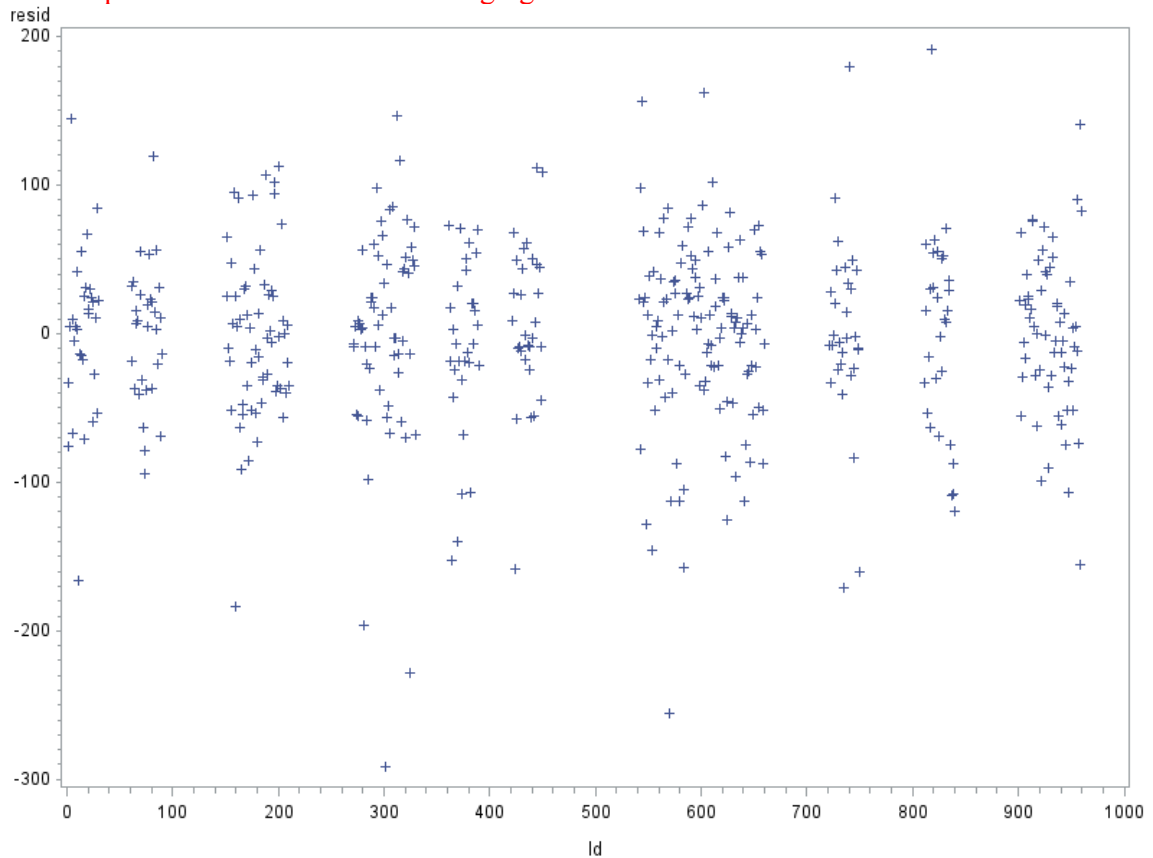
μ_m - fixed main effect of m-th lab;

$\mu_{n(m)}$ - fixed main effect of n-th batch, nested with Lab;

ε_{ijklmn} - random experiment error.

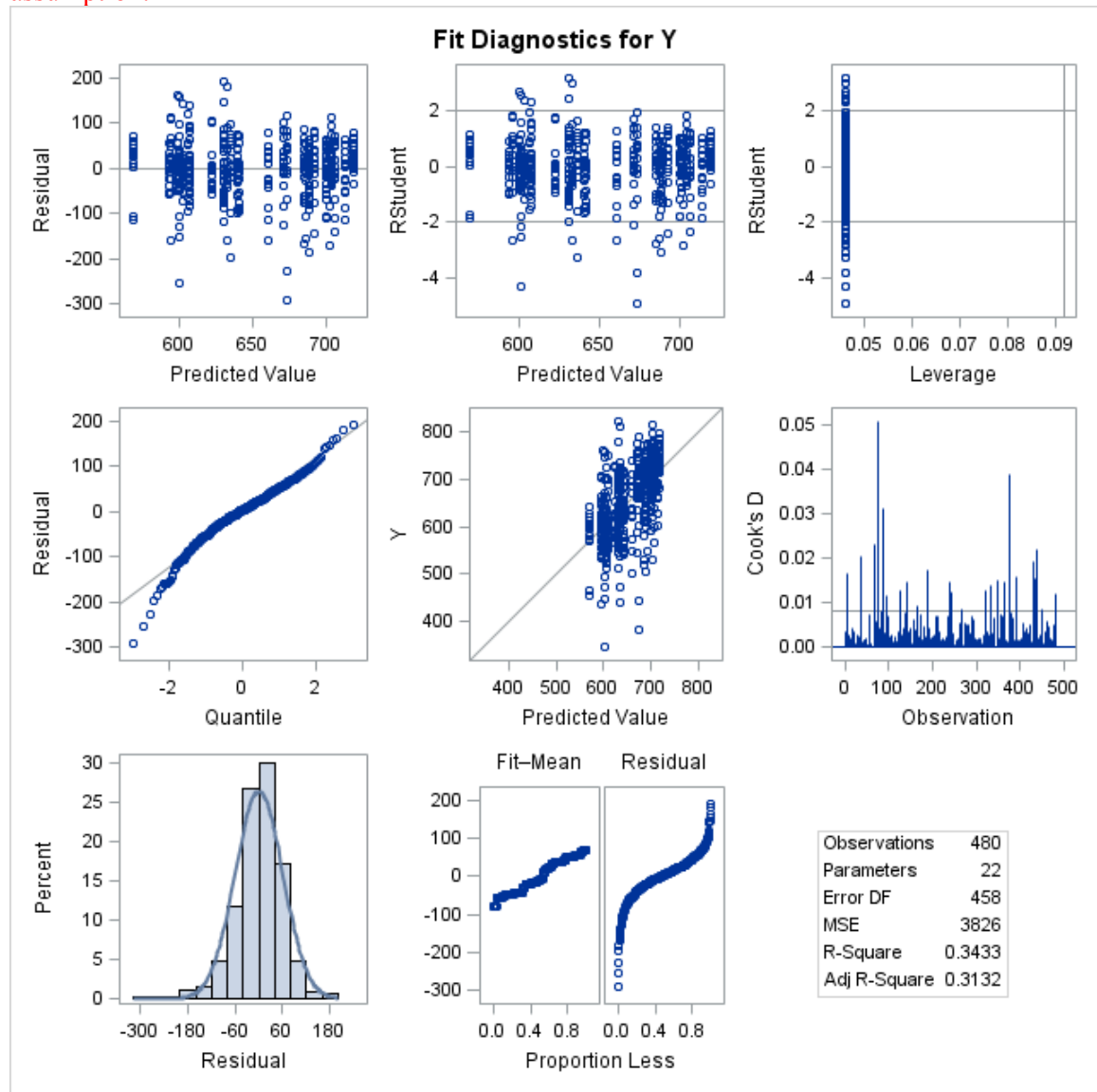
- e) A helpful diagnostic plot in many instances is called a “run plot”. A run plot plots a variable of interest versus measurements that might exhibit correlation between observations. Here, make run plots of the residuals versus id and versus num.

The run plots below do not reveal changing variance.



- f) Check the assumptions for the model you wrote down in d)

The residual vs. predicted value do not reveal changing variance. The qq plot shows the residuals are slightly skewed to the left, but this is not too much a departure from normality assumption.



- g) Continue the analysis reporting your conclusions (not necessarily with the model specified in d). I'm just not explicitly outlining the remaining steps.)

By type III sum of squares test, the interaction effect of table speed and feed rate is not significant ($p\text{-value} = 0.2837$), nor is the interaction effect of feed rate and grit size ($p\text{-value} = 0.1630$). After leaving them out of the model, the residual plots still do not reveal serious violation to assumption about error term. Even though main effect of feed rate in the reduced model is not significant ($p\text{-value} = 0.5$), it is better to keep it in the model as it is of interest for the research.

Conclusion

There is sufficient evidence ($p\text{-value} = 0.0505$) evidence that table speed and grit size jointly affect mean ceramic strength. On the other hand, there is little evidence ($p\text{-value} = 0.53$) feed rate affect mean ceramic strength. Setting table speed at level -1 and grit size at 1 yields the highest average strength, which is only significantly higher than setting both of them at 1 after Tukey's adjustment for multiple comparison.

Scope

This is an experimental study in which table speed, feed rate, and grit size are randomly assigned, therefore causal relation between them and the ceramic strength can be concluded.

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	594.7561875	B	16.72262670	35.57	<.0001
tableSpeed -1	41.7583167	B	11.30655928	3.69	0.0002
tableSpeed 1	0.0000000	B	.	.	.
feedRate -1	-5.2054833	B	7.99494474	-0.65	0.5153
feedRate 1	0.0000000	B	.	.	.
gritSize -1	26.6240083	B	9.79176757	2.72	0.0068
gritSize 1	0.0000000	B	.	.	.
tableSpeed*gritSize -1 -1	-31.3545333	B	15.98988948	-1.96	0.0505
tableSpeed*gritSize -1 1	0.0000000	B	.	.	.
tableSpeed*gritSize 1 -1	0.0000000	B	.	.	.
tableSpeed*gritSize 1 1	0.0000000	B	.	.	.
Lab 1	-27.0395000	B	17.42457809	-1.55	0.1214

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

tableSpeed	gritSize	Y LSMEAN	LSMEAN Number
-1	-1	660.752571	1
-1	1	665.483096	2
1	-1	650.348788	3
1	1	623.724779	4

Least Squares Means for effect tableSpeed*gritSize
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Y

i/j	1	2	3	4
1		0.9628	0.7941	0.0010
2	0.9628		0.4111	0.0014
3	0.7941	0.4111		0.0342
4	0.0010	0.0014	0.0342	

Appendix: SAS Code

```
DATA ceramic;
    INPUT Id Lab Num Test Y tableSpeed feedRate gritSize
direction Trt Set Llab Rep Batch Sblab Set2;
    DATALINES;
    .....
;
RUN;

/*(a). Check anomalous response*/
proc univariate data = ceramic plots;
    var Y;
run;

proc print data = ceramic;
    where Y <= 400;
run;

/*(b). Visual examination of lab/batch effect*/
proc sort data = ceramic;
    by Batch Lab;
run;

proc sgpanel data = ceramic;
    title 'Boxplot of Ceramic Strenth across Labs by Batch';
    panelby Batch;
    vbox Y / category = Lab;
run;
title;

/* (c). */
proc sql;
select
    Lab
    ,Batch
    ,tableSpeed
    ,feedRate
    ,gritSize
    ,sum(1) as rep
from ceramic
group by Lab
    ,Batch
    ,tableSpeed
    ,feedRate
    ,gritSize
order by Lab
    ,Batch
    ,tableSpeed
    ,feedRate
    ,gritSize
;

proc glmmod data = ceramic noprint outparm = parm;
```

```

        class tableSpeed feedRate gritSize Lab Batch;
        model Y = tableSpeed feedRate gritSize tableSpeed*feedRate
tableSpeed*gritSize feedRate*gritSize Lab Batch(Lab)/;
run;

proc print data = parm;
run;

/* (d). */
proc glm data = ceramic plots = all;
    class tableSpeed feedRate gritSize Lab Batch;
    model Y = tableSpeed feedRate gritSize tableSpeed*feedRate
tableSpeed*gritSize feedRate*gritSize Lab Batch(Lab)/ss3
solution;
    output out = resout r = resid;
run;

/* (f). Run plot */
proc gplot data = resout;
    plot resid*id;

proc gplot data = resout;
    plot resid*num;
run;

/* (g). */

proc glm data = ceramic plots = all;
    class tableSpeed feedRate gritSize Lab Batch;
    model Y = tableSpeed feedRate gritSize tableSpeed*gritSize
Lab Batch(Lab)/ss3 solution;
    lsmeans tableSpeed*gritSize / pdiff adjust = tukey
plots=diffplot(center) ;
run;

```