

1. (1.5pt) Consider the situation described in Problem 13.7, page 602. Provide estimates of the variance components.

```
DATA IN1;
  DO PART = 1 TO 10;
    DO INSPECTOR = 1 to 3;
      DO TEST = 1 TO 3;
        INPUT IMPEDENCE @@; OUTPUT;
      END; END; END;
LINES;
37 38 37 41 41 40 41 42 41
42 41 43 42 42 42 43 42 43
30 31 31 31 31 31 29 30 28
42 43 42 43 43 43 42 42 42
28 30 29 29 30 29 31 29 29
42 42 43 45 45 45 44 46 45
25 26 27 28 28 30 29 27 27
40 40 40 43 42 42 43 43 41
25 25 25 27 29 28 26 26 26
35 34 34 35 35 34 35 34 35
;
proc glm data=in1 plots=all;
class part inspector;
model impedance=part inspector / ss3 solution;
means part inspector;
random part inspector / test;
run;
proc varcomp data=in1 method=reml;
class part inspector;
model impedance=part inspector;
title 'variance components';
run;
```

- (a) (4.5pt) Answer part (a), page 602. Include checking the model assumptions. – Analyze the data from this experiment, assuming that both parts and operators are random effects.

**prob didn't specify whether interaction – I excluded – check with john**

$$H_0: \sigma_{part}^2=0; H_A: \sigma_{part}^2 >0$$

There is strong evidence of part to part variability after accounting for variation between inspectors ( $F_{9,78} = 430.82, P < 0.0001$ ).

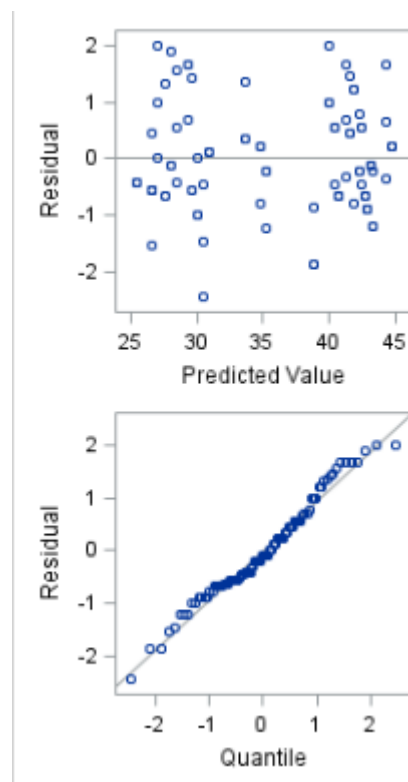
$$H_0: \sigma_{inspector}^2=0; H_A: \sigma_{inspector}^2 >0$$

There is strong evidence of variability between inspectors after accounting for part to part variability ( $F_{2,78} = 19.34, P < 0.0001$ ).

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PART	9	3935.955556	437.328395	430.82	<.0001
INSPECTOR	2	39.266667	19.633333	19.34	<.0001
Error: MS(Error)	78	79.177778	1.015100		

I am not familiar with the subject area. This sounds like an observational study, but if inspectors and parts were randomly assigned to induction motor starters, independence between replicates would be reasonable after accounting for inspector and part variability. It is also assumed parts, inspectors, and the residuals are all independent, and all parts are independent, all inspectors are independent, and all residuals are independent. These will be reasonable if, for the part, the effect of part 1 does not influence the effect of part 2, and likewise for the residuals and inspectors. Inspectors are human, and so may not be independent if trained in subgroups.

Normality is reasonable as all the residuals fall very close to the normal q-q line. There is one outlier residual near the predicted value of 30. However, it is not too extreme in magnitude and given the number of other residuals, we might expect at least one to be a bit larger by chance. Homogeneity of variance seems reasonable as there is random scatter in the residuals vs. predicted plot.



(b) (1.5pt) Provide estimates of the variance components.

ask john what this second plot is

REML Estimates	
Variance Component	Estimate
Var(PART)	48.47926
Var(INSPECTOR)	0.62061
Var(Error)	1.01510

Asymptotic Covariance Matrix of Estimates			
	Var(PART)	Var(INSPECTOR)	Var(Error)
Var(PART)	524.70849	0.00009786	-0.0029357
Var(INSPECTOR)	0.00009786	0.42833	-0.0008807
Var(Error)	-0.0029357	-0.0008807	0.02642

- (c) (1.5pt) Provide a practical interpretation of the variance component estimates in the context of the study.

The variation in impedance within **or is it between?** parts is estimated to be 48.47926. The variation in impedance within inspectors is estimated to be 0.62061 and the variation in impedance after accounting for part and inspector is estimated to be 1.0151.

fix this up and make sure i understand.

2. (1.5pt) Consider the situation described in Problem 13.7, page 602. Provide estimates of the variance components.

by variance components does he mean mse or all the sums of squares... here it says assume operator is fixed, so really only one variance component...

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TEMP	1	258.7812500	258.7812500	37.64	<.0001
OPER	2	130.6408333	65.3204167	9.50	0.0003
TEMP*OPER	2	21.0625000	10.5312500	1.53	0.2254
CYCLE	2	676.6633333	338.3316667	49.21	<.0001
TEMP*CYLE	2	128.4033333	64.2016667	9.34	0.0003
OPER*CYLE	4	14.8633333	3.7158333	0.54	0.7067
TEMP*OPER*CYLE	4	7.7366667	1.9341667	0.28	0.8888

3. (a)

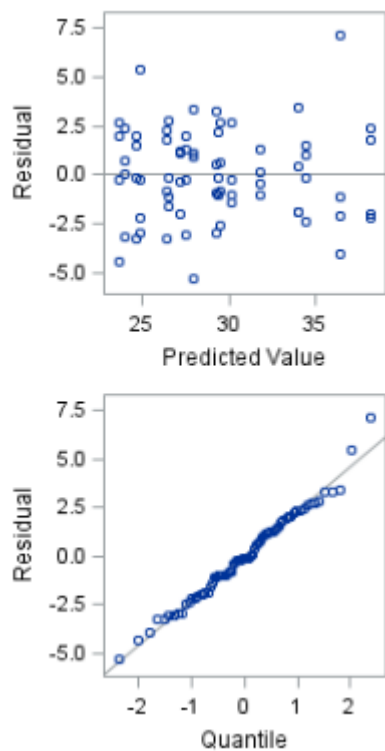
- (b)

- (c)
- (d) The largest effect occurs at 300 degrees, operator 2, and cycle 50 and so that is my recommendation.
- (e) (.5pt) *If a target score of 30 is desirable, what set of factor conditions would you recommend?*

Want sum of effects to be 0, but not sure how to evaluate.

- (f) (1pt) *Is there any serious problem with the model assumptions? If yes, what is the problem?*

Randomization would make independence of the residuals reasonable. The residuals vs. predicted plot shows random scatter, which suggests constant variance is reasonable. The Normal Q-Q plot shows two residuals that are quite large on the standardized scale and deviate from the normal q-q line, suggesting the normality assumption of the residuals may be violated. Looking a bit closer at the residuals vs. predicted plot we do see two residuals greater than 5 in absolute value near predicted values of 28 and 35.



4. Suppose that the data in Problem 3 was not collected as a completely randomized design. Suppose that two replicates of the three-factor factorial were collected on Monday and the remaining two replicates of the three-factor factorial were collected on Friday, and that within each day the 36 experimental runs were completely randomized.
- (a) (1.5pt) *Describe the model that incorporates the model effects from Problem 2 and additional effects involving Days (Monday, Friday) and all two-factor interactions with Days.*

**Make sure John doesn't want more than main fixed effects.**

$$y_{ijklm} = \mu + \beta_i + \gamma_j + \phi_k + \tau_l + \beta\tau_{il} + \gamma\tau_{jl} + \phi\tau_{kl} + \epsilon_{ijklm}$$

where i = level of temperature (1,2)

j = operator level (1,2,3)

k = cycle level (1,2,3)

l = day (1,2)

m = replicate (1,2)

- (b) (1.5pt) Set up a partial ANOVA table with a column for the Source of Variation. This will include rows that corresponding to the model effects in (a) as well as rows for Error and Total. Include a second column with the associated degrees of freedom. No data analysis is needed.

Source	df
temp	1
oper	2
cycle	2
day	1
temp*day	1
oper*day	2
cycle*day	2
error	60
total	71

- (c) (1pt) Typically, the MSE represents an estimate of  $\sigma^2$ . For this to be true, we will need to assume all other potential model effects are negligible and can be ignored. What potential model effects did we assume to be negligible and exclude from the model?

The model effects we assumed to be negligible and excluded from the model include  $\beta\gamma_{ij}$ ,  $\beta\phi_{ik}$ ,  $\gamma\phi_{jk}$ ,  $\beta\gamma\phi_{ijk}$ ,  $\tau\beta\gamma_{ijl}$ ,  $\tau\beta\phi_{ikl}$ ,  $\tau\gamma\phi_{jkl}$ ,  $\tau\beta\gamma\phi_{ijkl}$ .

5. Consider the information on pages 192-194 for Example 5.1 for the battery design experiment. Suppose the researcher is planning another 33 factorial experiment. The goal is to determine the sample size required so that the power of each F-test (for Material Type, Temperature, and their interaction) is  $\geq .90$ .

- (a) (.5pt) Based on this experiment, what is the estimate of  $\sigma$ ?

MSE=675.1 and RMSE= $\hat{\sigma}$ =25.9826865.

(1.5pt) Determine the desired sample size assuming you use the sample mean  $y_{ij}$  as an estimate of  $\mu_{ij}$  for  $i = 1, 2, 3$  and  $j = 1, 2, 3$ . You can assume  $\alpha = .05$ .

The computed total sample size is 18, meaning two replicates per each material and temperature combination.

Computed N Total					
Index	Source	Test DF	Error DF	Actual Power	N Total
1	mat	2	9	>.999	18
2	temp	2	9	>.999	18
3	mat*temp	4	9	>.999	18

```
DATA p5;
  DO mat = 1 TO 3;
    DO temp = 1 to 3;
      INPUT life @@; OUTPUT;
    END; END;
  LINES;
  539 229 230
  623 479 198
  576 583 342
;
PROC GLMPOWER DATA = p5;
CLASS mat temp;
MODEL life=mat|temp;
POWER
STDDEV = 25.98269
ALPHA = 0.05
NTOTAL = .
POWER = 0.9;
RUN;
```

6. Stat 541 students: Suppose that the data in Problem 3 was not collected as a completely randomized design. Suppose that one replicate of the three-factor factorial was collected on Monday, Tuesday, Thursday, and Friday, and that within each of these four days, the 18 experimental runs were completely randomized.
- (a) (3pt) For a model that incorporates model effects involving Days, Temperature, Operator, and Cycle Time, as well as all two-factor and three-factor interactions involving Temperature, Operator, Cycle Time, and Days, set up a partial ANOVA table that includes these model effects and the associated degrees of freedom.

Source	df
temp	1
oper	2
cycle	2
day	3
temp*day	3
oper*day	6
cycle*day	6
temp*oper	2
temp*cycle	2
oper*cycle	4
temp*oper*cycle	4
temp*day*oper	6
temp*day*cycle	6
day*oper*cycle	12
error	12
total	71

(b) (1pt) What does the MSE actually estimate in the context of this study?

In this study, the MSE is actually the four way interaction between temperature, day, operator, and cycle time because there is no replication.