Example of Normal Mixed Model

Balanced Multi-Location Trial

Want to compare the difference in weight gain between pigs given one of two pre-feeding treatments (A or B). The study is conducted on 4 different farms, where each farm has 24 pens, with one pig from each treatment in a pen (2 pigs per pen). Data located in **multiloc.dat** on the class website.

Analysis models

Model	Fixed effects	Random effects	Method
1	treatment, farm	---	Ordinary least squares
2	treatment	farm	REML
3	treatment	farm, treatment*farm	REML

Expected mean squares

Source	df	Model 1	Model 2	Model 3
farm	**f-1=3**		$\sigma^2 + pt\sigma^2_f$	$\sigma^2 + pt\sigma^2_f$
pen(farm)	**f(p-1)=92**	$\sigma^2 + t\sigma^2_{pf}$	$\sigma^2 + t\sigma^2_{pf}$	$\sigma^2 + t\sigma^2_{pf}$
treatment	**(t-1)=1**	$\sigma^2 + \varphi_{trt}$	$\sigma^2 + \varphi_{trt}$	$\sigma^2 + p\sigma^2_{ft} + \varphi_{trt}$
farm*treatment	**(f-1)(t-1)=3**	$\sigma^2 + \varphi_{ft}$	---	$\sigma^2 + p\sigma^2_{ft}$
error	**f(p-1)(t-1)=92**	σ^2	σ^2	σ^2

Inferences on Model 1

-first test farm*treatment interaction
-if not significant, then test treatment main effect
-if farm*treatment significant, then need to test farm-specific treatment effects

Inferences on Model 2

-test treatment effect using MS(error).

Requires strong assumption that there is no farm*treatment interaction in order to apply the inferences to the population of farms that these farms were selected from.

Inferences on Model 3

-test treatment effect using MS(farm*treatment) as error term
-*possibly:* estimate farm-specific treatment effects

Because farm*treatment effect is random, treatment effects assumed to vary randomly between farms, therefore results can be related to the population of farms.

Model 1 – Farm treated as fixed effect

```
PROC MIXED DATA=M;
 CLASS FARM PEN TRT;
 MODEL Y=FARM|TRT;
 RANDOM PEN(FARM);
 LSMEANS TRT FARM*TRT/ DIFF SLICE=FARM CL;
RUN;
```

Selected output:

```
                  Covariance Parameter
                        Estimates

                  Cov Parm      Estimate

                  pen(farm)       1.0062
                  Residual        2.6412

            Type 3 Tests of Fixed Effects

                    Num     Den
        Effect       DF      DF    F Value    Pr > F

        farm          3      92     192.43    <.0001
        trt           1      92     150.76    <.0001
        farm*trt      3      92      17.25    <.0001
```

The farm*treatment effect is significant, therefore we need to test farm specific treatment effects.

```
TITLE 'MODEL 1, BY FARM BECAUSE OF SIGNIFICANT FARM*TREATMENT EFFECT';
PROC MIXED DATA=M;
BY FARM;
CLASS PEN TRT;
MODEL Y=TRT;
RANDOM PEN;
LSMEANS TRT/DIFF PDIFF CL;
RUN;
```

Farm	Trt 1	Trt 2	Difference	s.e.(diff)	t	pr>|t|
1	104.34	103.14	1.20	0.51	2.34	0.028
2	111.73	107.39	4.34	0.31	13.95	<0.001
3	114.75	113.55	1.20	0.54	2.25	0.035
4	110.14	105.36	4.78	0.48	9.88	<0.001

Although the treatment differences are all in the same direction, the degree of the difference is not the same among the farms.

The inference applies only to these four farms. If the farm*treatment effect had not been significant, even though the inferences still apply only to these four farms, it would be more reasonable to extrapolate the treatment effect to other farms.

Model 2 – Farm treated as random effect

```
PROC MIXED DATA=M;
CLASS FARM PEN TRT;
MODEL Y=TRT
RANDOM FARM PEN(FARM);
LSMEANS TRT/PDIFF;
RUN;
```

Selected output:

```
                    Covariance Parameter
                          Estimates

                    Cov Parm      Estimate
                    farm           18.5595
                    pen(farm)       0.3287
                    Residual        3.9963

     Type 3 Tests of Fixed Effects

                      Num     Den
         Effect        DF      DF    F Value    Pr > F
         trt            1      95      99.64    <.0001

                    Least Squares Means

                             Standard
 Effect   trt   Estimate       Error     DF   t Value   Pr > |t|
 trt      1       110.24      2.1645     95     50.93     <.0001
 trt      2       107.36      2.1645     95     49.60     <.0001
```

Significant treatment effect, however model assumes treatment effects are not farm specific. Do we believe this?

Model 3 – Farm and farm*treatment treated as random

```
PROC MIXED DATA=M;
CLASS FARM PEN TRT;
MODEL Y=TRT;
RANDOM FARM PEN(FARM) FARM*TRT;
LSMEANS TRT/PDIFF;
ESTIMATE 'TRT 1 BROAD MEAN' INTERCEPT 1 TRT 1 0;
ESTIMATE 'TRT 1 NARROW MEAN' INTERCEPT 4 TRT 4 0 | FARM 1 1 1 1 FARM*TRT 1 0  1 0  1 0  1
0/DIVISOR=4;
ESTIMATE 'TRT EFF AT FARM 1' TRT 1 -1 | FARM*TRT  1 -1  0  0   0  0   0  0;
ESTIMATE 'TRT EFF AT FARM 2' TRT 1 -1 | FARM*TRT  0  0  1 -1   0  0   0  0;
ESTIMATE 'TRT EFF AT FARM 3' TRT 1 -1 | FARM*TRT  0  0  0  0   1 -1   0  0;
ESTIMATE 'TRT EFF AT FARM 4' TRT 1 -1 | FARM*TRT  0  0  0  0   0  0   1 -1;
RUN;
```

Selected output:

```
                    Covariance Parameter
                          Estimates

                     Cov Parm      Estimate
                     farm           17.6655
```

pen(farm)	1.0062
farm*trt	1.7880
Residual	2.6412

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
trt 1 broad mean	110.24	2.2139	3	49.79	<.0001
trt 1 narrow mean	110.24	0.1949	3	565.57	<.0001
trt eff at farm 1	1.2974	0.4588	3	2.83	0.0663
trt eff at farm 2	4.2569	0.4588	3	9.28	0.0027
trt eff at farm 3	1.2974	0.4588	3	2.83	0.0663
trt eff at farm 4	4.6691	0.4588	3	10.18	0.0020

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	110.24	2.2139	3	49.79	<.0001
trt	2	107.36	2.2139	3	48.49	<.0001

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	2	2.8802	0.9742	3	2.96	0.0597

The standard errors for the treatment effect at each farm will be biased downward. SAS PROC Mixed has methods for determining the correct degrees of freedom for the estimates, to correct for the downward bias. Kenward-Rogers is a more general degree-of-freedom procedure. Satterthwaite's approximation procedure is a special case of the Kenward-Rogers. For designs with missing data, the Kenward-Rogers is recommended.

```
PROC MIXED DATA=M;
CLASS FARM PEN TRT;
MODEL Y=TRT/DDFM=KR;
RANDOM FARM PEN(FARM) FARM*TRT;
LSMEANS TRT/PDIFF;
 ESTIMATE 'TRT 1 BROAD MEAN' INTERCEPT 1 TRT 1 0;
 ESTIMATE 'TRT 1 NARROW MEAN' INTERCEPT 4 TRT 4 0 | FARM 1 1 1 1 FARM*TRT 1 0  1 0  1 0  1
0/DIVISOR=4;
 ESTIMATE 'TRT EFF AT FARM 1' TRT 1 -1 | FARM*TRT  1 -1  0  0   0  0   0  0;
 ESTIMATE 'TRT EFF AT FARM 2' TRT 1 -1 | FARM*TRT  0  0  1 -1   0  0   0  0;
 ESTIMATE 'TRT EFF AT FARM 3' TRT 1 -1 | FARM*TRT  0  0  0  0   1 -1   0  0;
 ESTIMATE 'TRT EFF AT FARM 4' TRT 1 -1 | FARM*TRT  0  0  0  0   0  0   1 -1;
RUN;
```

Selected output:

```
                    Cov Parm      Estimate
                    farm           17.6655
                    pen(farm)       1.0062
                    farm*trt        1.7880
                    Residual        2.6412

                                Standard
Label              Estimate        Error    DF  t Value  Pr > |t|
trt 1 broad mean     110.24       2.2139   3.3    49.79    <.0001
trt 1 narrow mean    110.24       0.1949   171   565.57    <.0001
trt eff at farm 1    1.2974       0.4730  94.4     2.74    0.0073
trt eff at farm 2    4.2569       0.4730  94.4     9.00    <.0001
trt eff at farm 3    1.2974       0.4730  94.4     2.74    0.0073
trt eff at farm 4    4.6691       0.4730  94.4     9.87    <.0001
```

Note that the treatment means and differences are not affected. Also note that the s.e. for the treatment effects at each farm is larger than the s.e. from the model without the df correction (0.4588).

```
                     Least Squares Means

                           Standard
Effect   trt   Estimate      Error     DF   t Value   Pr > |t|
trt      1       110.24     2.2139    3.3     49.79     <.0001
trt      2       107.36     2.2139    3.3     48.49     <.0001

              Differences of Least Squares Means

                                Standard
Effect  trt  _trt  Estimate      Error    DF  t Value  Pr > |t|
trt     1    2       2.8802     0.9742     3     2.96    0.0597
```

Unbalanced multiple location with missing data

Go back to the blood pressure data. We'll fit the same three types of models that we did for the above example, except that now we will include a baseline covariate

Model	Fixed effects	Random effects	Method
1	baseline,treatment, center	---	Ordinary least squares

2	baseline,treatment	center	REML
3	baseline,treatment	center, treatment*center	REML

Models 1 and 3 we have fit previously.

Summarize the treatment effects from the 3 models

Model	**baseline**	**A-B**	**A-C**	**B-C**
1	0.22 (0.11)	1.20 (1.24)	2.99 (1.23)	1.79 (1.27)
2	0.22 (0.11)	1.03 (1.22)	2.98 (1.21)	1.95 (1.24)
3	0.28 (0.11)	1.29 (1.39)	2.93 (1.37)	1.64 (1.40)

Variance Components:

Model	**center**	**treatment*center**	**residual**
1	---	---	71.9
2	7.82 (3.99)	---	70.9 (6.2)
3	6.46 (4.35)	4.10 (5.33)	68.4 (6.5)

Because the trial is unbalanced, we would expect there to be bias in the standard errors of the fixed effect estimates. One way to see the differences in the standard error estimates is to look at the 95% confidence limits.

The following mixed model is run first with the default degrees of freedom (containment), then with ddfm=satterth, then with ddfm=kr.

```
proc mixed covtest;
class trt center;
model dbp=trt dbp0; * add /ddfm=satterth or /ddfm=kr for other models
random center center*trt;
estimate 'A-B' trt 1 -1 0/ E CL;
estimate 'A-C' trt 1 0 -1/ E CL;
estimate 'B-C' trt 0 1 -1/ E CL;
run;
```

From the default df.

Label	Estimate	Standard Error	DF	Pr>|t|	Lower	Upper
A-B	1.2859	1.3914	48	0.3600	-1.5116	4.0834
A-C	2.9274	1.3727	48	0.0381	0.1674	5.6874
B-C	1.6415	1.4049	48	0.2484	-1.1832	4.4662

From the Satterthwaite df.

Label	Estimate	Standard Error	DF	Pr>|t|	Lower	Upper
A-B	1.2859	1.3914	23.8	0.3646	-1.5867	4.1585
A-C	2.9274	1.3727	25.6	0.0427	0.1036	5.7512
B-C	1.6415	1.4049	25.7	0.2534	-1.2482	4.5312

From the Kenward-Rogers df.

Label	Estimate	Standard Error	DF	Pr>|t|	Lower	Upper
A-B	1.2859	1.4300	23.8	0.3775	-1.6665	4.2384
A-C	2.9274	1.4109	25.6	0.0482	0.02511	5.8297
B-C	1.6415	1.4453	25.7	0.2666	-1.3314	4.6144

Model Checking

We will carry out model checking on Model 3.

```
proc mixed data=ldbp;
class trt center;
model dbp=trt dbp0/ddfm=satterth outpred=pred outpm=pm;
random center center*trt/solution;
estimate 'A-B' trt 1 -1 0/ E CL;
estimate 'A-C' trt 1 0 -1/ E CL;
estimate 'B-C' trt 0 1 -1/ E CL;
make 'solutionr' out=solut;
run;
```

The outpred=pred outputs the residuals and the outpm=pm outputs the mean predicted values. The make statement outputs the random effects estimates to the solut dataset.

The center variable in the solut dataset is alphanumeric and needs to be numeric. The following creates a new numeric variable centerx.

```
data solut;
set solut;
centerx=center*1;
drop center;
run;
```

Create datasets

C_EST – random effects estimated for center
CT_EST – random effects estimated for center*treatment

```
data c_est(keep=center c_est) ct_est(keep=center trt ct_est);
set solut;
center=centerx;
if Effect='center' then do;
c_est=Estimate;
output c_est;
end;
else do;
ct_est=Estimate;
output ct_est;
end;
proc sort data=ct_est;
by center trt;
proc sort data=pm;
by center trt;
```

Merge these two datasets with the predicted values and calculate the mean predicted values and output these into new dataset.

```
data c_esta;
merge pm c_est;
by center;
proc means noprint; by center; id c_est;
var pred; output out=c_est mean=c_pred n=freq;
run;
data ct_esta;
merge pm ct_est;
by center trt;
proc means noprint; by center trt; id ct_est;
var pred; output out=ct_est mean=ct_pred n=freq;
```

Print the residuals.

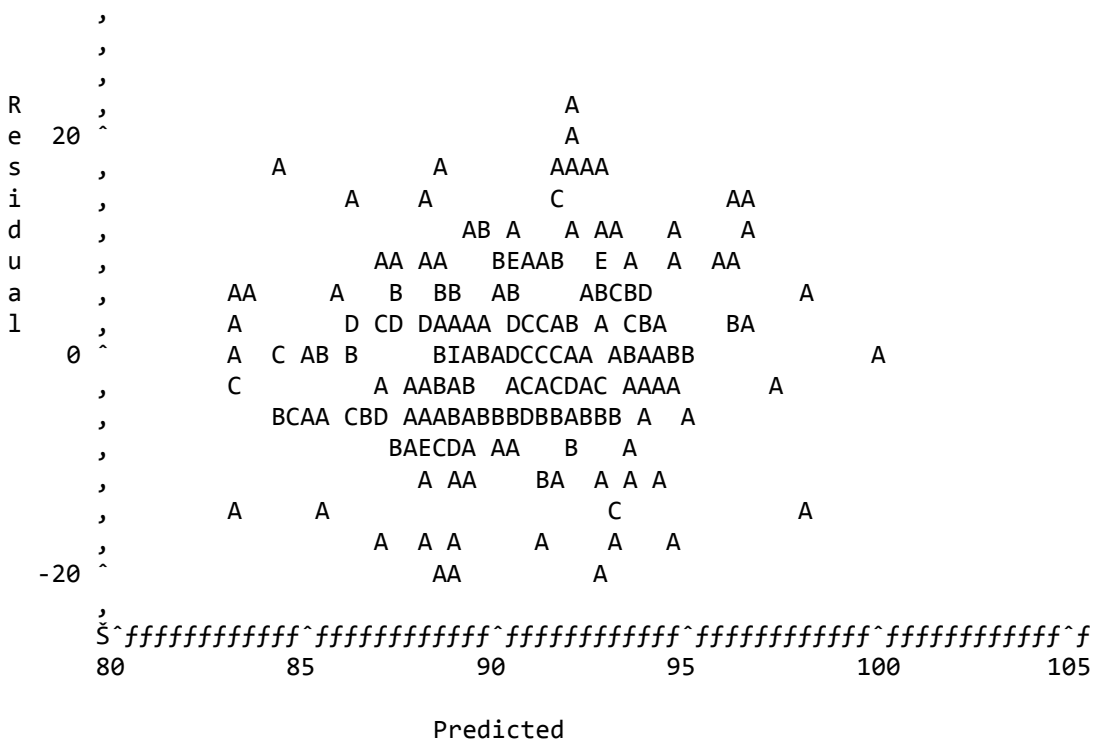
```
proc print noobs data=pred;
var pat trt center Resid Pred;
run;
```

Plot the residuals against their predicted values.

```
title 'Residuals and Predicted Values';
proc plot data=pred; plot resid*pred;
title 'Residuals against their predicted values';
```

```
       Plot of Resid*Pred.  Legend: A = 1 obs, B = 2 obs, etc.

      ,
      ,
   40 ^                                                  A
      ,
      ,
```



Look at normal probability plot (plot ordered residuals against their values expected from the standard normal distribution given their ranks). Should form a straight line

```
proc rank data=pred out=norm normal=tukey; var resid; ranks s_est;
proc plot data=norm;
plot resid*s_est;
title 'resuduals - normal plot';
run;
```

```
Plot of Resid*s_est.  Legend: A = 1 obs, B = 2 obs, etc.

      ,
      ,
   40 ˆ                                                                A
      ,
      ,
      ,
      ,
      ,
R     ,                                                               A
e  20 ˆ                                                             A
s     ,                                                          BAAAA
i     ,                                                       CBB
d     ,                                                    BDC
u     ,                                                GEEEB
a     ,                                            CHGG
l     ,                                        IJIIF
    0 ˆ                                   FKJKJA
      ,                               HJJD
      ,                          DGGHIHB
      ,                     CEDFC
      ,                ACDA
      ,             BBB
      ,         AAAAB
  -20 ˆ  A  A A
      ,
      Šˆƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒˆ
      -3         -2         -1          0          1          2          3

                          Rank for Variable Resid
```

Looking at the plots, we see one possible outlier. The tables below shows what happens when the patient is removed from the dataset.

	Variance Components		
Model	**Center**	**Treatment*Center**	**Residual**
With outlier	6.46	4.10	68.36
Without outlier	6.97	1.50	63.27

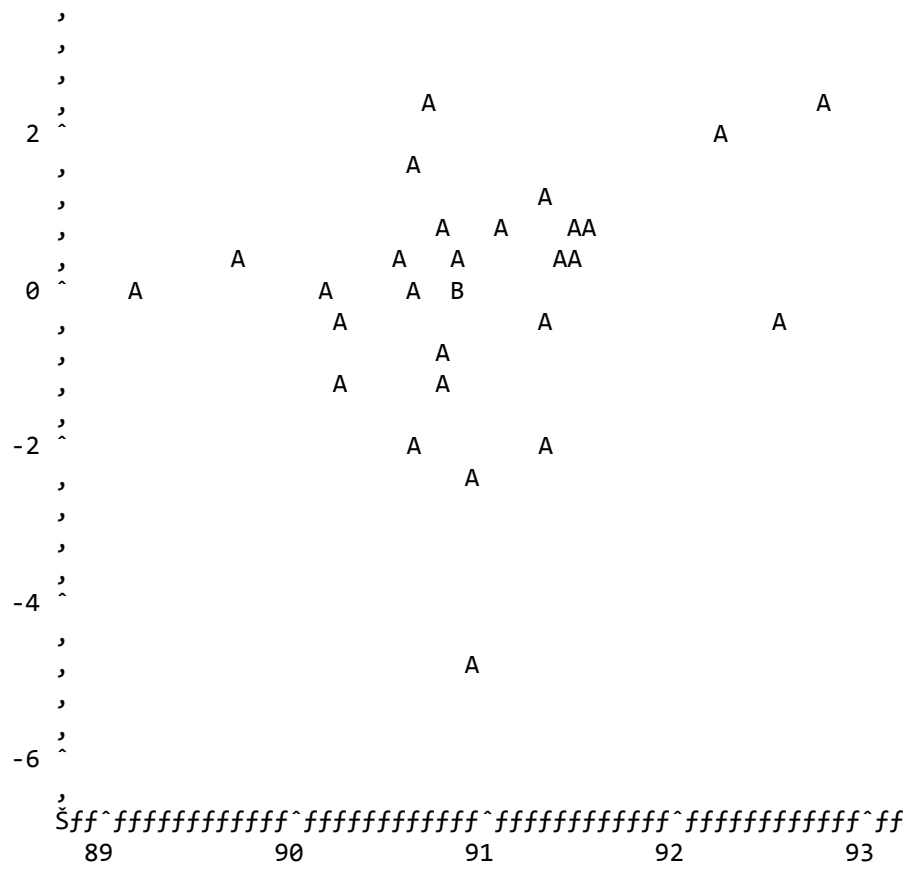
	Treatment effects (Ses)		
Model	**A-B**	**A-C**	**B-C**
With outlier	1.29 (1.39)	2.92 (1.37)	1.64(1.40)
Without outlier	0.74(1.24)	2.49 (1.22)	1.76 (1.25)

The parameter estimates have changed. The outlier value was 140 mmHg, which could be due to a recording error. However, this patient had a high baseline value (113) and dropped out of the study due to "insufficient effect". Would report results both with and without the outlier.

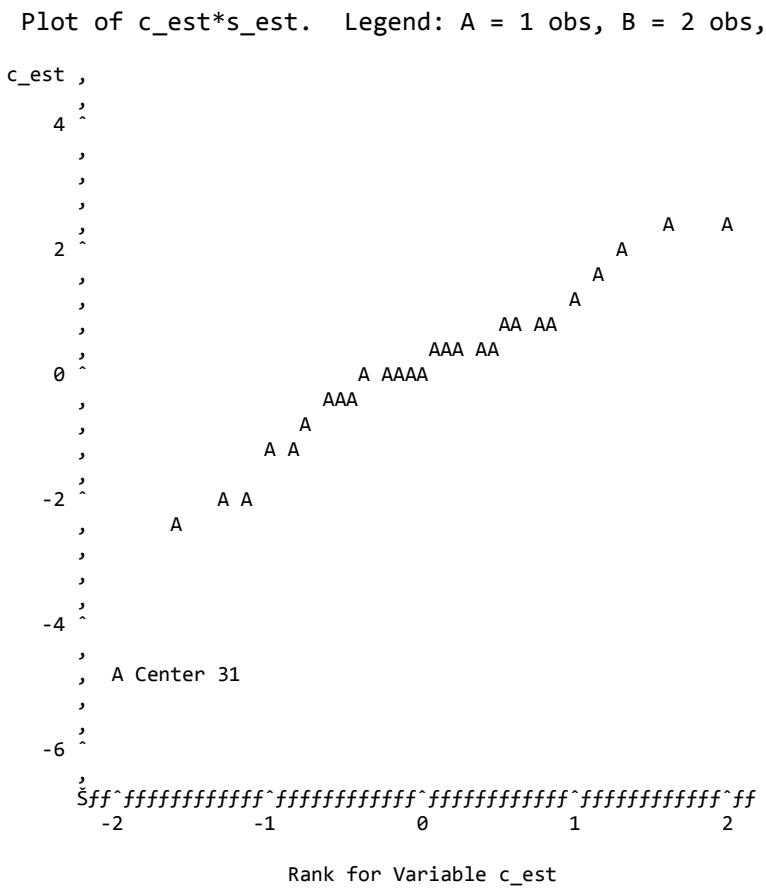
Next the predicted values for the center will be plotted.

```
proc plot data=c_est;
plot c_est*c_pred;
title 'center effects against their predicted values';
Plot of c_est*c_pred.  Legend: A = 1 obs, B = 2 obs, etc.

     c_est ,
           ,
         4 ˆ
```




```
 proc rank out=norm normal=tukey data=c_est; var c_est;
ranks s_est;
proc plot data=norm; plot c_est*s_est;
title 'center effects - normal plot';
run;
```



Looking at the plots, we see one possible outlier Center 31. The tables below shows what happens when the patient is removed from the dataset.

	Variance Components		
Model	**Center**	**Treatment*Center**	**Residual**
With 31	6.46	4.10	68.36
Without 31	0.81	5.37	73.1

	Treatment effects (Ses)		
Model	**A-B**	**A-C**	**B-C**
With 31	1.29 (1.39)	2.92 (1.37)	1.64(1.40)
Without 31	1.33 (1.55)	3.03 (1.52)	1.70 (1.56)

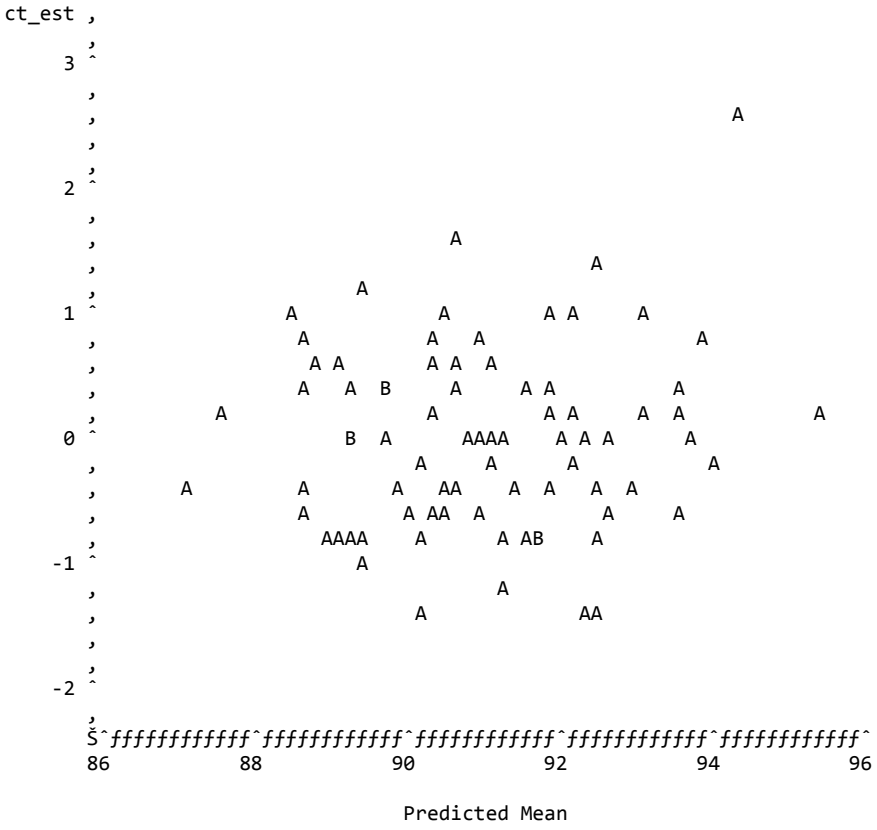
Not much change in treatment estimates. Therefore would keep Center 31 in the analysis.

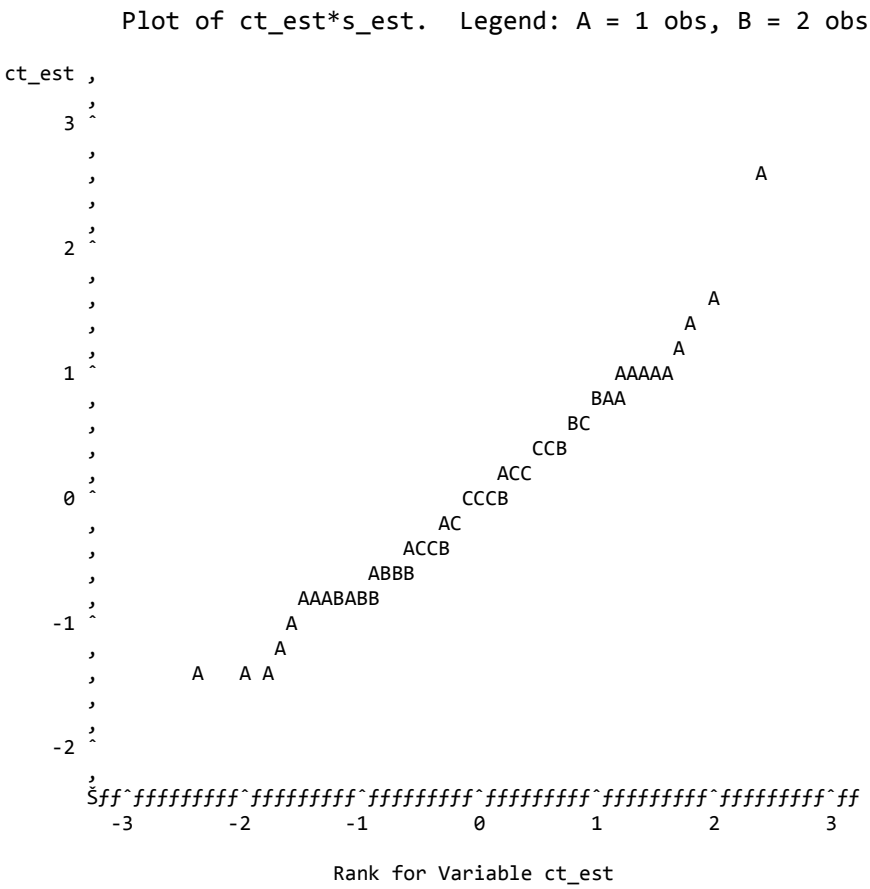
Now investigate center*treatment

```
title 'center*treatment effects and predicted values';
proc plot data=ct_est;
plot ct_est*ct_pred;
title 'center*treatment effects against their predicted values';
proc rank out=norm normal=tukey data=ct_est; var ct_est;
```

```
ranks s_est;
proc plot data=norm; plot ct_est*s_est;
title 'center*treatment effects - normal plot';
run;
```

Plot of ct_est*ct_pred. Legend: A = 1 obs, B = 2 obs, etc.





Neither plots indicate any noticeable outliers.

Now check the homogeneity of treatment variances.

```
proc sort data=pred; by trt;
proc means data=pred; var resid; by trt;
title 'proc means to check residual variance is homogeneous across treatments';

proc sort data=ct_est; by trt;
proc means data=ct_est; var ct_est; by trt;
title 'proc means to check center*treatment effect variance is homogeneous across
treatments';
run;
```

```
proc means to check residual variance is homogeneous across treatment

--------------------------------- trt=A ----------------------------------
                    Analysis Variable : Resid Residual

     N            Mean         Std Dev         Minimum         Maximum
   ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
   100    -3.42482E-14       9.1545453     -17.2589799      39.8192060
--------------------------------- trt=B ----------------------------------
                    Analysis Variable : Resid Residual

     N            Mean         Std Dev         Minimum         Maximum
   ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
    93    -2.68937E-14       6.5308257     -18.7823225      18.2781249
--------------------------------- trt=C ----------------------------------
                    Analysis Variable : Resid Residual

     N            Mean         Std Dev         Minimum         Maximum
   ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
    95    1.017198E-14       7.8623760     -18.7313617      17.3011936

proc means to check center*treatment effect variance is homogeneous across
treatments

--------------------------------- trt=A ----------------------------------
                        Analysis Variable : ct_est

     N            Mean         Std Dev         Minimum         Maximum
    ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
    26    -2.90793E-15       0.8457655      -1.4859312       2.5895978
--------------------------------- trt=B ----------------------------------
                        Analysis Variable : ct_est

     N            Mean         Std Dev         Minimum         Maximum
    ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
    25    -3.33067E-16       0.6983672      -1.3161783       1.6899927
--------------------------------- trt=C ----------------------------------
                        Analysis Variable : ct_est

     N            Mean         Std Dev         Minimum         Maximum
    ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
    28    6.723045E-15       0.6670323      -0.9090217       1.4170861
```

The standard deviations of the residuals and the center*treatment effects were similar among the treatment groups.