```
/*  This is a program for analyzing the
    penicillan data from Box, Hunter,
    and Hunter.  It is posteded in the
    file
            penclln.sas        */

/*  First enter the data  */


data set1;
  infile 'c:\stat504\penclln.dat';
  input batch process $ yield;
  run;


/* Compute the ANOVA table, formulas for
   expectations of mean squares, process
   means and their standard errors */

proc glm data=set1;
  class batch process;
  model yield = batch process / e e3;
  random batch / q test;
  lsmeans process / stderr pdiff tdiff;
  output out=set2 r=resid p=yhat;
  run;
```

```
The GLM Procedure

    Class Level Information

Class        Levels    Values

batch             5    1 2 3 4 5

process           4    A B C D


Number of Observations Read        20
Number of Observations Used        20
```

The GLM Procedure

General Form of Estimable Functions

Effect		Coefficients
Intercept		L1
batch	1	L2
batch	2	L3
batch	3	L4
batch	4	L5
batch	5	L1-L2-L3-L4-L5
process	A	L7
process	B	L8
process	C	L9
process	D	L1-L7-L8-L9

Type III Estimable Functions

		------Coefficients-------	
Effect		batch	process
Intercept		0	0
batch	1	L2	0
batch	2	L3	0
batch	3	L4	0
batch	4	L5	0
batch	5	-L2-L3-L4-L5	0
process	A	0	L7
process	B	0	L8
process	C	0	L9
process	D	0	-L7-L8-L9

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	334.0000000	47.7142857	2.53	0.0754
Error	12	226.0000000	18.8333333		
Corrected Total	19	560.0000000			

R-Square	Coeff Var	Root MSE	yield Mean
0.596429	5.046208	4.339739	86.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
batch	4	264.0000000	66.0000000	3.50	0.0407
process	3	70.0000000	23.3333333	1.24	0.3387

Quadratic Forms of Fixed Effects in the Expected Mean Squares

Source: Type III Mean Square for process

	process A	process B	process C	process D
process A	3.75000000	-1.25000000	-1.25000000	-1.25000000
process B	-1.25000000	3.75000000	-1.25000000	-1.25000000
process C	-1.25000000	-1.25000000	3.75000000	-1.25000000
process D	-1.25000000	-1.25000000	-1.25000000	3.75000000

Source	Type III Expected Mean Square
batch	Var(Error) + 4 Var(batch)
process	Var(Error) + Q(process)

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: yield

Source	DF	Type III SS	Mean Square	F Value	Pr > F
batch	4	264.000000	66.000000	3.50	0.0407
process	3	70.000000	23.333333	1.24	0.3387
Error: MS(Error)	12	226.000000	18.833333		

Least Squares Means

process	yield LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
A	84.0000000	1.9407902	<.0001	1
B	85.0000000	1.9407902	<.0001	2
C	89.0000000	1.9407902	<.0001	3
D	86.0000000	1.9407902	<.0001	4

Least Squares Means for Effect process
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: yield

i/j	1	2	3	4
1		-0.36434 0.7219	-1.8217 0.0935	-0.72868 0.4802
2	0.36434 0.7219		-1.45736 0.1707	-0.36434 0.7219
3	1.821698 0.0935	1.457359 0.1707		1.093019 0.2958
4	0.728679 0.4802	0.36434 0.7219	-1.09302 0.2958	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

```
/* Compute a normal probability plot for
   the residuals and the Shapiro-Wilk
   test for normality  */

proc rank data=set2 normal=blom out=set2;
  var resid;  ranks q;
  run;

proc univariate data=set2 normal plot;
  var resid;
  run;

  axis1 label=(h=2.5 r=0 a=90 f=swiss 'Residuals')
       value=(f=swiss h=2.0) w=3.0 ;

  axis2 label=(h=2.3 f=swiss 'Standard Normal Quantiles')
       value=(f=swiss h=2.0) w=3.0 ;

  axis3 label=(h=2.3 f=swiss 'Production Process')
       value=(f=swiss h=2.0) w=3.0 ;

  symbol1 v=circle i=none h=2 w=3 c=black;

proc gplot data=set2;
  plot resid*q / vaxis=axis1 haxis=axis2;
  title h=3.0 ls=1.0in f=swiss
       c=black 'Normal Probability Plot';
  footnote ls=0.6in '  ';
  run;

proc gplot data=set2;
  plot resid*process / vaxis=axis1 haxis=axis3;
  title h=3.0 ls=1.0in f=swiss
       c=black 'Residual Plot';
  footnote ls=0.6in '   ';
  run;
```

CHECK THE RESULTS BY YOURSELF

```
/* Fit the same model using PROC MIXED.  Compute
   REML estimates of variance components.  Note
   that PROC MIXED provides appropriate standard
   errors for process means.  When block effects
   are random. PROC GLM does not provide correct
```

```
   standard errors for process means  */

proc mixed data=set1;
  class process batch;
  model yield = process / ddfm=satterth solution;
  random batch / type=vc G solution cl alpha=.05;
  lsmeans process / pdiff tdiff;
run;
```

The Mixed Procedure

Model Information

Data Set	WORK.SET1
Dependent Variable	yield
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
process	4	A B C D
batch	5	1 2 3 4 5

Dimensions

Covariance Parameters	2
Columns in X	5
Columns in Z	5
Subjects	1
Max Obs Per Subject	20

Number of Observations

Number of Observations Read	20
Number of Observations Used	20
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	106.59285141	
1	1	103.82994387	0.00000000

Convergence criteria met.

Estimated G Matrix

Row	Effect	batch	Col1	Col2	Col3	Col4	Col5
1	batch	1	11.7917				
2	batch	2		11.7917			
3	batch	3			11.7917		
4	batch	4				11.7917	
5	batch	5					11.7917

Covariance Parameter Estimates

Cov Parm	Estimate
batch	11.7917
Residual	18.8333

Fit Statistics

-2 Res Log Likelihood	103.8
AIC (smaller is better)	107.8
AICC (smaller is better)	108.8
BIC (smaller is better)	107.0

The Mixed Procedure

Solution for Fixed Effects

Effect	process	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		86.0000	2.4749	11.1	34.75	<.0001
process	A	-2.0000	2.7447	12	-0.73	0.4802
process	B	-1.0000	2.7447	12	-0.36	0.7219
process	C	3.0000	2.7447	12	1.09	0.2958
process	D	0	.	.	.	.

Solution for Random Effects

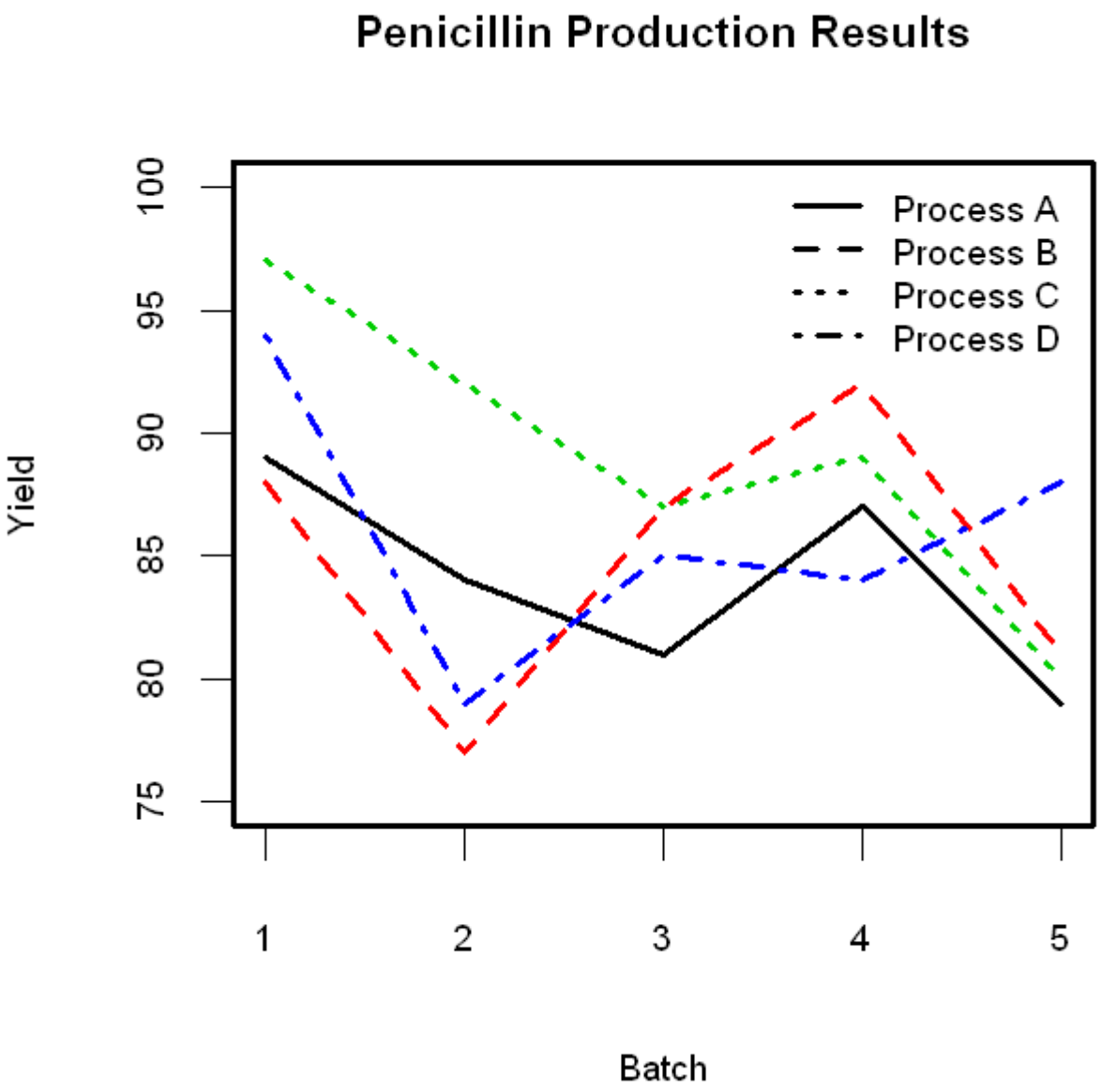
Effect	batch	Estimate	Std Err Pred	DF	t Value	Pr > |t|	Alpha	Lower	Upper
batch	1	4.2879	2.2473	5.29	1.91	0.1115	0.05	-1.3954	9.9712
batch	2	-2.1439	2.2473	5.29	-0.95	0.3816	0.05	-7.8273	3.5394
batch	3	-0.7146	2.2473	5.29	-0.32	0.7627	0.05	-6.3980	4.9687
batch	4	1.4293	2.2473	5.29	0.64	0.5513	0.05	-4.2540	7.1126
batch	5	-2.8586	2.2473	5.29	-1.27	0.2564	0.05	-8.5419	2.8247

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F

```
process        3     12      1.24    0.3387


                      Least Squares Means

                                Standard
Effect    process    Estimate      Error     DF    t Value    Pr > |t|

process    A          84.0000     2.4749   11.1      33.94      <.0001
process    B          85.0000     2.4749   11.1      34.35      <.0001
process    C          89.0000     2.4749   11.1      35.96      <.0001
process    D          86.0000     2.4749   11.1      34.75      <.0001


                  Differences of Least Squares Means

                                          Standard
Effect    process    _process    Estimate      Error     DF    t Value    Pr > |t|

process    A          B          -1.0000     2.7447     12      -0.36      0.7219
process    A          C          -5.0000     2.7447     12      -1.82      0.0935
process    A          D          -2.0000     2.7447     12      -0.73      0.4802
process    B          C          -4.0000     2.7447     12      -1.46      0.1707
process    B          D          -1.0000     2.7447     12      -0.36      0.7219
process    C          D           3.0000     2.7447     12       1.09      0.2958
> # Analyze the penicillin data from Box,
> # Hunter, and Hunter.  This code is
> # posted as  penclln.r
>
> # Enter the data into a data frame and
> # change the Batch and Process variables
> # into factors
>
> penclln <- read.table("penclln.dat",
+     col.names=c("Batch","Process","Yield"))
> penclln$Batch <- as.factor(penclln$Batch)
> penclln$Process <- as.factor(penclln$Process)
> penclln
   Batch Process Yield
1      1       A    89
2      1       B    88
3      1       C    97
4      1       D    94
5      2       A    84
6      2       B    77
7      2       C    92
8      2       D    79
9      3       A    81
10     3       B    87
11     3       C    87
12     3       D    85
13     4       A    87
14     4       B    92
15     4       C    89
16     4       D    84
17     5       A    79
18     5       B    81
19     5       C    80
20     5       D    88
```

```
> # Construct a profile plot.
>
>  means <- tapply(penclln$Yield,list(penclln$Batch,penclln$Process),mean)
>
>  par(cex=1.2,lwd=3,mex=1.5)
>  x.axis <- unique(penclln$Batch)
>  matplot(c(1,5), c(75,100), type="n",
+      xlab="Batch",  ylab="Yield",
+      main= "Penicillin Production Results")
>  matlines(x.axis,means,type='l',lty=c(1,2,3,4),lwd=3)
>
>  legend(3.5,101, legend=c('Process A','Process B',
+     'Process C','Process D'), lty=c(1,2,3,4),bty='n')
```

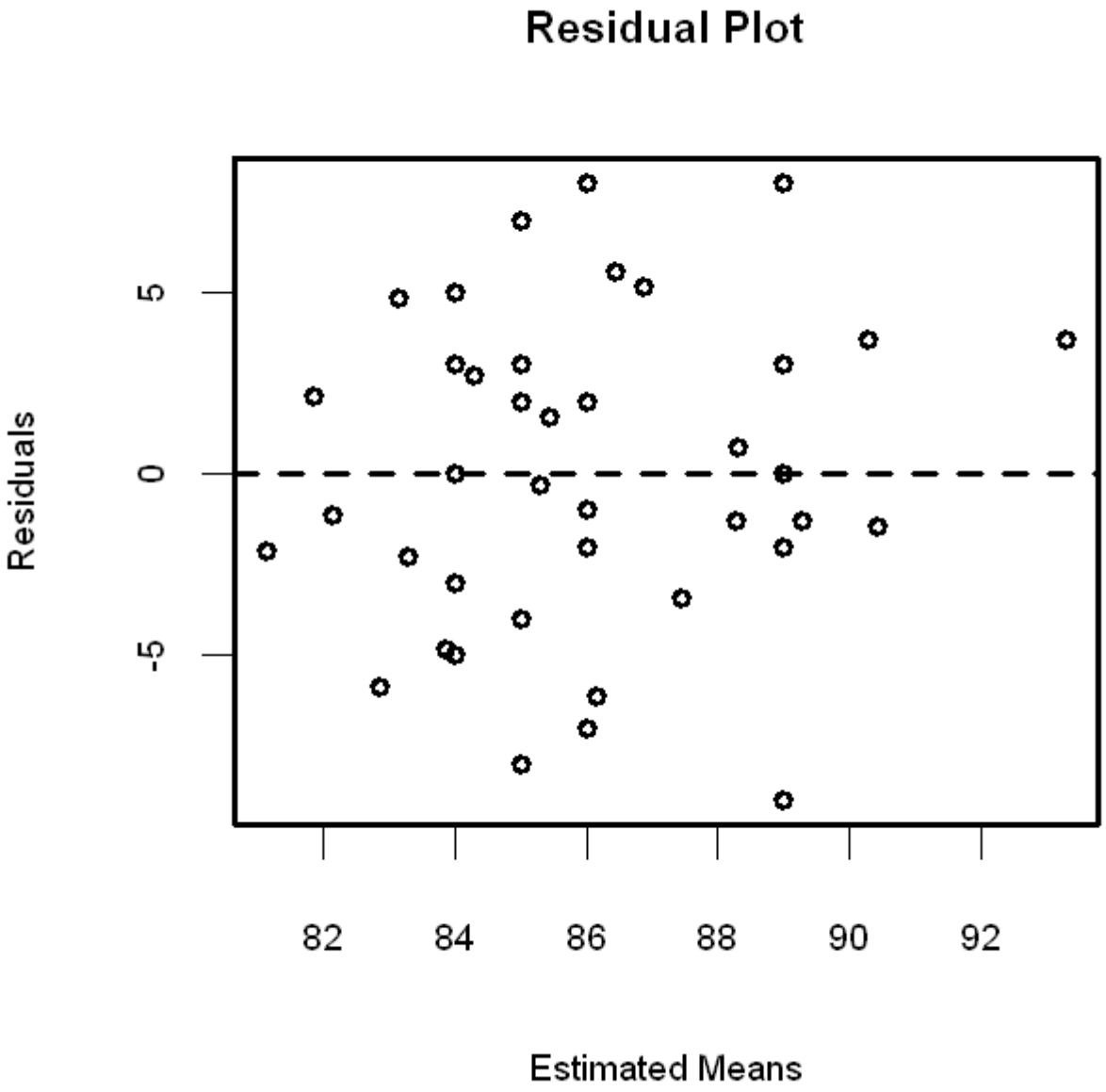



```
> #  Use the  lme( )  function to fit a model
> #  with additive batch (random) and process
> #  (fixed) effects and create diagnostic plots.
>
> library(nlme)
>
> options(contrasts=c("contr.treatment",
+    "contr.poly"))
>
> penclln.lme <- lme(Yield ~ Process,
+    random= ~ 1|Batch, data=penclln,
+    method=c("REML"))
>
> summary(penclln.lme)
Linear mixed-effects model fit by REML
 Data: penclln
      AIC      BIC    logLik
```

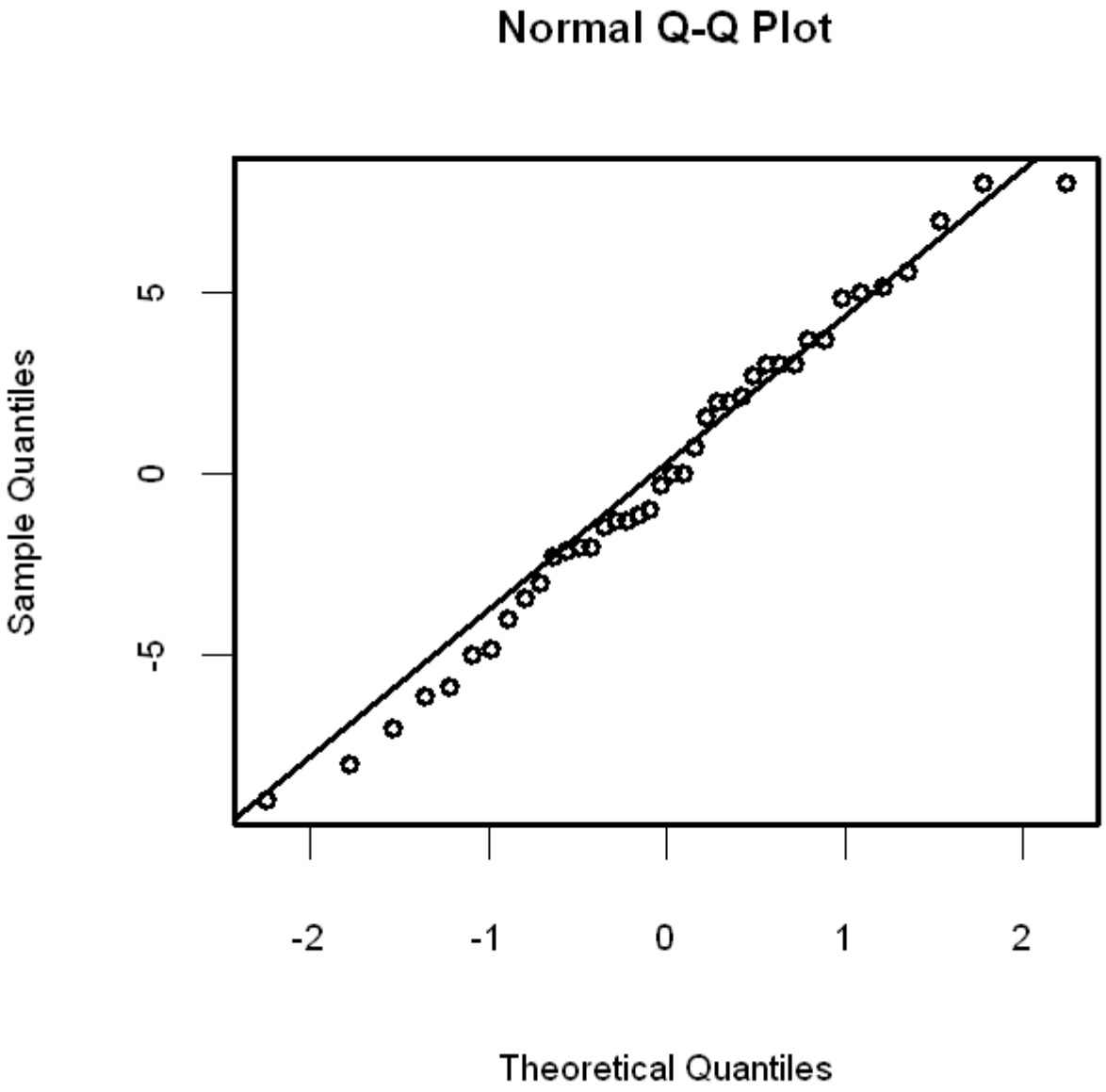
```
  115.8299 120.4655 -51.91497

Random effects:
 Formula: ~1 | Batch
       (Intercept) Residual
StdDev:    3.433900 4.339739

Fixed effects: Yield ~ Process
            Value Std.Error DF  t-value p-value
(Intercept)    84  2.474874 12 33.94113  0.0000
ProcessB        1  2.744692 12  0.36434  0.7219
ProcessC        5  2.744692 12  1.82170  0.0935
ProcessD        2  2.744692 12  0.72868  0.4802
 Correlation:
        (Intr) PrcssB PrcssC
ProcessB -0.555
ProcessC -0.555  0.500
ProcessD -0.555  0.500  0.500

Standardized Within-Group Residuals:
      Min        Q1       Med        Q3       Max
-1.4151575 -0.5017350 -0.1643840  0.6829939  1.2836503

Number of Observations: 20
Number of Groups: 5
> names(penclln.lme)
 [1] "modelStruct"  "dims"         "contrasts"    "coefficients"
 [5] "varFix"       "sigma"        "apVar"        "logLik"
 [9] "numIter"      "groups"       "call"         "terms"
[13] "method"       "fitted"       "residuals"    "fixDF"
[17] "na.action"    "data"
>
> #  Contruct ANOVA table for fixed effects
> anova(penclln.lme)
            numDF denDF   F-value p-value
(Intercept)     1    12 2241.2121  <.0001
Process         3    12    1.2389  0.3387
>
> #  Estimated parameters for fixed effects
> coef(penclln.lme)
  (Intercept) ProcessB ProcessC ProcessD
1    88.28788        1        5        2
2    81.85606        1        5        2
3    83.28535        1        5        2
4    85.42929        1        5        2
5    81.14141        1        5        2
>
> # BLUP's for random effects
> ranef(penclln.lme)
  (Intercept)
1   4.2878788
2  -2.1439394
3  -0.7146465
4   1.4292929
5  -2.8585859
>
> # Confidence intervals for fixed effects
> # and estimated standard deviations
> intervals(penclln.lme)
Approximate 95% confidence intervals
```

```
 Fixed effects:
              lower est.    upper
(Intercept) 78.6077134   84 89.39229
ProcessB    -4.9801698    1  6.98017
ProcessC    -0.9801698    5 10.98017
ProcessD    -3.9801698    2  7.98017
attr(,"label")
[1] "Fixed effects:"



 Random Effects:
  Level: Batch
                lower    est.    upper
sd((Intercept)) 1.285359 3.433900 9.173834

 Within-group standard error:
   lower    est.    upper
2.908802 4.339739 6.474602
>
> # Create a listing of the original data
> # residuals and predicted values
>
> data.frame(penclln$Process,penclln$Batch,
+            penclln$Yield,
+            Pred=penclln.lme$fitted,
+            Resid=round(penclln.lme$resid,3))
   penclln.Process penclln.Batch penclln.Yield Pred.fixed Pred.Batch
1                A             1           89        84   88.28788
2                B             1           88        85   89.28788
3                C             1           97        89   93.28788
4                D             1           94        86   90.28788
5                A             2           84        84   81.85606
6                B             2           77        85   82.85606
7                C             2           92        89   86.85606
8                D             2           79        86   83.85606
9                A             3           81        84   83.28535
10               B             3           87        85   84.28535
11               C             3           87        89   88.28535
12               D             3           85        86   85.28535
13               A             4           87        84   85.42929
14               B             4           92        85   86.42929
15               C             4           89        89   90.42929
16               D             4           84        86   87.42929
17               A             5           79        84   81.14141
18               B             5           81        85   82.14141
19               C             5           80        89   86.14141
20               D             5           88        86   83.14141
   Resid.fixed Resid.Batch
1            5      0.712
2            3     -1.288
3            8      3.712
4            8      3.712
5            0      2.144
6           -8     -5.856
7            3      5.144
8           -7     -4.856
9           -3     -2.285
10           2      2.715
11          -2     -1.285
```

```
12         -1     -0.285
13          3      1.571
14          7      5.571
15          0     -1.429
16         -2     -3.429
17         -5     -2.141
18         -4     -1.141
19         -9     -6.141
20          2      4.859
>
>
> # Create residual plots
> frame( )
> par(cex=1.2,lwd=3,mex=1.5)
>   plot(penclln.lme$fitted, penclln.lme$resid,
+       xlab="Estimated Means",
+       ylab="Residuals",
+       main="Residual Plot")
>   abline(h=0, lty=2, lwd=3)
>
```




```
>   qqnorm(penclln.lme$resid)
>   qqline(penclln.lme$resid)
```


Normal Q-Q Plot
Sample Quantiles
5
0
-5
-2
-1
0
1
2
Theoretical Quantiles