

```
> library(Stat5303libs); library(cfcdae); library(lme4)
```

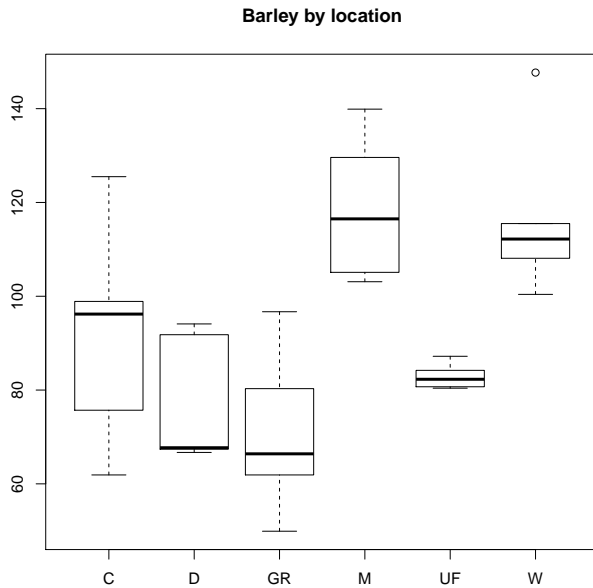
```
> immer
```

One early and famous example of a Randomized Complete Block analyzed by Fisher involves five varieties of barley grown at six locations (which included Crookston, Waseca, ..., data from Minnesota!). The original experiment had two years of data; we will only look at the second year. This data set is included in the MASS package, which Stat5303 automatically loads.

	Loc	Var	Y1	Y2
1	UF	M	81.0	80.7
2	UF	S	105.4	82.3
3	UF	V	119.7	80.4
4	UF	T	109.7	87.2
5	UF	P	98.3	84.2
6	W	M	146.6	100.4
7	W	S	142.0	115.5
8	W	V	150.7	112.2
9	W	T	191.5	147.7
10	W	P	145.7	108.1
11	M	M	82.3	103.1
12	M	S	77.3	105.1
13	M	V	78.4	116.5
14	M	T	131.3	139.9
15	M	P	89.6	129.6
16	C	M	119.8	98.9
17	C	S	121.4	61.9
18	C	V	124.0	96.2
19	C	T	140.8	125.5
20	C	P	124.8	75.7
21	GR	M	98.9	66.4
22	GR	S	89.0	49.9
23	GR	V	69.1	96.7
24	GR	T	89.3	61.9
25	GR	P	104.1	80.3
26	D	M	86.9	67.7
27	D	S	77.1	66.7
28	D	V	78.9	67.4
29	D	T	101.8	91.8
30	D	P	96.0	94.1

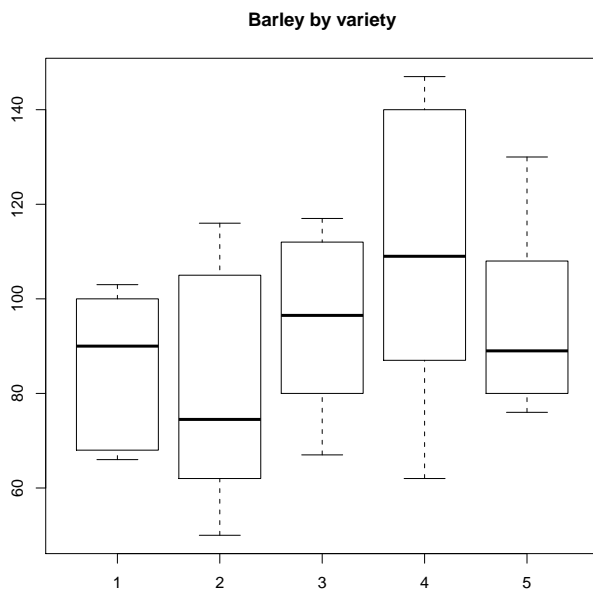
```
> boxplot (Y2~Loc, data=immer, main="Barley by location")
```

Here are the results separately by block. There are large block to block differences, so blocking should be a real help.



```
> boxplot (Y2~Var, data=immer, main="Barley by variety")
```

Here are the results separately by treatment. We can see some treatment differences, but we also see a lot of variation within each treatment. Much of this is block to block variability that would show up in our error if we hadn't blocked.

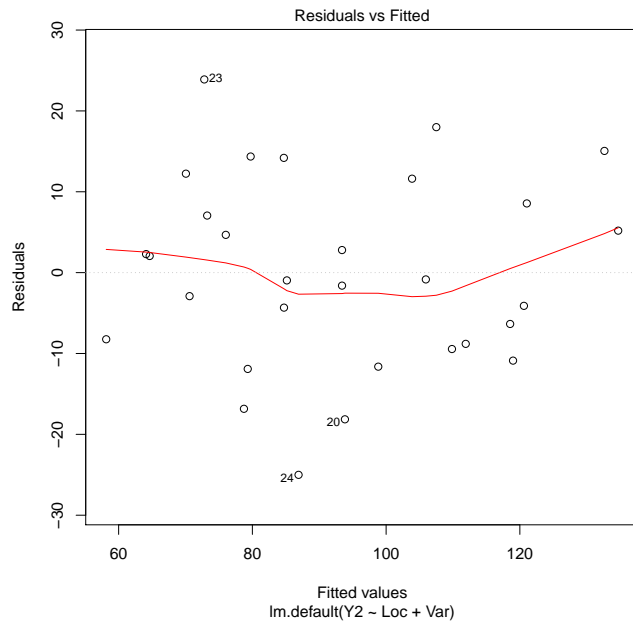


```
> fit1 <- lm(Y2~Loc+Var, data=immer)
```

Get into the habit of treatments after blocks.

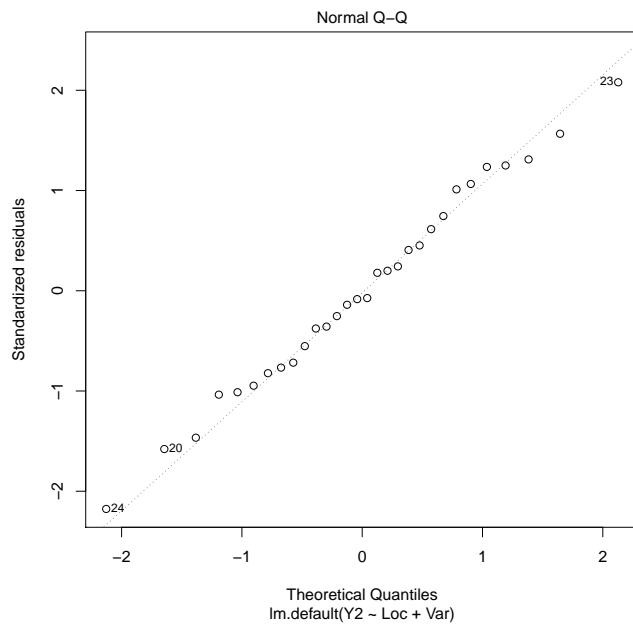
```
> plot(fit1, which=1)
```

Residuals look pretty good, although there is just a hint of a dip in the middle.



```
> plot(fit1, which=2)
```

Normality looks good.



```
> anova(fit1)
```

Here is the ANOVA, we just enter treatments after blocks. Variety effects are moderately significant. We do not test the block effects (even though R prints an F, it doesn't know about blocking factors).

```
Analysis of Variance Table
```

```
Response: Y2
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Loc	5	10285.0	2056.99	10.3901	5.049e-05	***
Var	4	2845.2	711.29	3.5928	0.02306	*
Residuals	20	3959.5	197.98			

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(fit1)
```

Varieties M and S have low yields, variety T has a high yield, and the others are in the middle. Again, ignore "testing" related to block effects.

```
Call:
```

```
lm.default(formula = Y2 ~ Loc + Var, data = immer)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-25.007	-8.665	-0.900	8.185	23.893

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	93.133	2.569	36.254	< 2e-16	***
Loc1	-1.493	5.744	-0.260	0.797543	
Loc2	-15.593	5.744	-2.715	0.013344	*
Loc3	-22.093	5.744	-3.846	0.001008	**
Loc4	25.707	5.744	4.475	0.000232	***
Loc5	-10.173	5.744	-1.771	0.091790	.
Var1	-6.933	5.138	-1.349	0.192261	
Var2	2.200	5.138	0.428	0.673081	
Var3	-12.900	5.138	-2.511	0.020748	*
Var4	15.867	5.138	3.088	0.005797	**

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 14.07 on 20 degrees of freedom
```

```
Multiple R-squared:  0.7683, Adjusted R-squared:  0.664
```

```
F-statistic: 7.369 on 9 and 20 DF,  p-value: 0.0001069
```

> **pairwise (fit1, Var)**

Using HDS, only S and T seem to differ, the others cannot be distinguished.

Pairwise comparisons (hsd) of Var

	estimate	signif	diff	lower	upper
1 - 2	-9.1333333	24.30866	-33.441989	15.175322	
1 - 3	5.9666667	24.30866	-18.341989	30.275322	
1 - 4	-22.8000000	24.30866	-47.108656	1.508656	
1 - 5	-8.7000000	24.30866	-33.008656	15.608656	
2 - 3	15.1000000	24.30866	-9.208656	39.408656	
2 - 4	-13.6666667	24.30866	-37.975322	10.641989	
2 - 5	0.4333333	24.30866	-23.875322	24.741989	
* 3 - 4	-28.7666667	24.30866	-53.075322	-4.458011	
3 - 5	-14.6666667	24.30866	-38.975322	9.641989	
4 - 5	14.1000000	24.30866	-10.208656	38.408656	

> **fit2 <- lmer(Y2~(1|Loc)+Var, data=immer)**

OK, so what if we thought that blocks should be random?

> **summary (fit2)**

We get the same estimates of variety effects and residual error as well as same SE for variety effects.

Linear mixed model fit by REML ['lmerMod']

Formula: Y2 ~ (1 | Loc) + Var

Data: immer

REML criterion at convergence: 227

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.92838	-0.49474	-0.05208	0.72439	1.54701

Random effects:

Groups	Name	Variance	Std.Dev.
Loc	(Intercept)	371.8	19.28
	Residual	198.0	14.07

Number of obs: 30, groups: Loc, 6

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	93.133	8.280	11.247
Var1	-6.933	5.138	-1.349
Var2	2.200	5.138	0.428
Var3	-12.900	5.138	-2.511
Var4	15.867	5.138	3.088

Correlation of Fixed Effects:

(Intr)	Var1	Var2	Var3
Var1	0.000		
Var2	0.000	-0.250	
Var3	0.000	-0.250	-0.250
Var4	0.000	-0.250	-0.250

```
> Anova(fit2, test="F")
```

Kenward and Roger agrees with fit1.

Analysis of Deviance Table (Type II Wald F tests with Kenward–Roger df)

Response: Y2

	F	Df	Df.res	Pr(>F)
Var	3.5928	4	20	0.02306 *

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

```
> PBmodcomp(fit2, lmer(Y2 ~ (1 | Loc), data=immer))
```

Parametric bootstrap agrees, too.

Parametric bootstrap test; time: 42.90 sec; samples: 1000 extremes: 26;

large : Y2 ~ (1 | Loc) + Var

small : Y2 ~ (1 | Loc)

	stat	df	p.value
LRT	13.073	4	0.01093 *
PBtest	13.073		0.02697 *

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

```
> fit2.mcmc <- lmer.mcmc(fit2, 20000)
```

Let's be complete and use our MCMC method.

```
> lmer.mcmc.intervals(fit2.mcmc)
```

Both the intervals and the anova from MCMC match with the simple version we began with.

	lower	median	upper	SE
(Intercept)	67.954600	93.123651	114.730024	11.844891
Var1	-19.360989	-7.316400	5.033749	5.906038
Var2	-8.639146	2.561818	12.676264	5.424877
Var3	-23.692054	-13.253614	-1.523182	5.603044
Var4	5.835096	16.100567	27.635781	5.386200
Loc	55.337759	327.464959	1983.786153	664.435222
sigma2	122.445486	220.660282	487.140982	91.495355

```
> lmer.mcmc.anova(fit2.mcmc)
```

	chisq	Df	MC p-value
(Intercept)	61.82272	1	0.000
Var	13.31483	4	0.013

```
> (2056.99-197.98)/5
```

“Old school” estimate of block variation based on fit1, which agrees with REML.

[1] 371.802

> **(5*2056.99+24*197.98)/29**

This is our estimate of what the error variance would be in a CRD from the same kind of data we used in our RCB.

[1] 518.499

> **(21*28/23/26) * 518.499/197.98**

Here is the estimate of relative efficiency. Mostly it is a ratio of the MSE in the CRD (estimated) to the MSE in the RCB. In addition, there is a degrees of freedom correction term that adjusts for the loss of degrees of freedom in the RCB. This term is usually pretty close to 1. Here it is .98.

[1] 2.575151