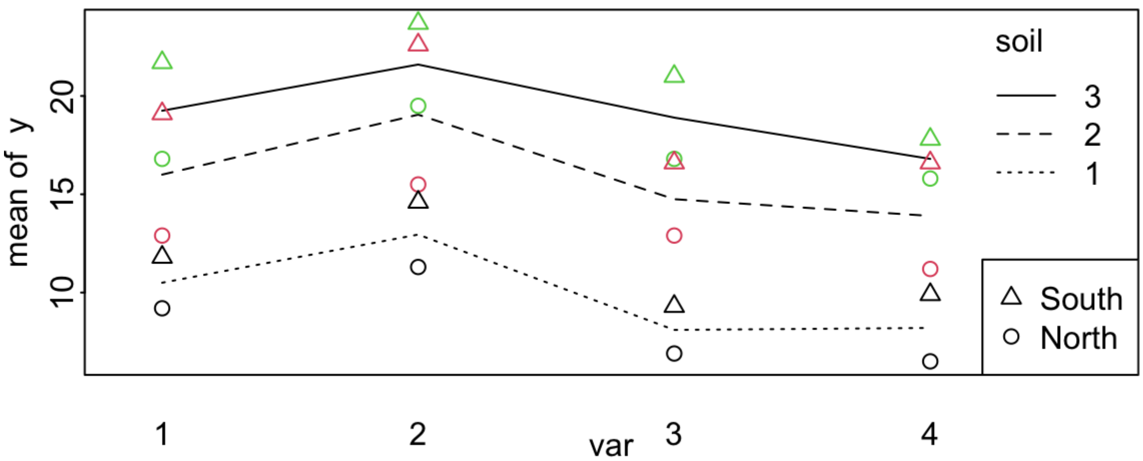
Split Plot analysis with R

```
dat = read.table("splitplot.txt", header=T,
                 colClasses=c("factor","factor","factor","numeric"))
dat$pan = with(dat, block:soil) # or call it "bench"
with(dat,interaction.plot(var, soil, y, ylim=range(y)))
with(dat, points(var, y, col=soil, pch=as.numeric(block)))
legend("bottomright", pch=2:1, legend=levels(dat$block)[2:1])
```




```
library(lme4); library(lmerTest)
fit = lmer(y ~ block + soil + var + soil:var + (1 | pan), dat)
print(VarCorr(fit), comp=c("Variance","Std.Dev.")) # summary(fit)
```

```
 Groups   Name        Variance Std.Dev.
 pan      (Intercept) 0.76281  0.87339
 Residual             0.65417  0.80881
```

```
anova(fit)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
block    17.951 17.9512     1     2 27.4414   0.03456
soil     61.583 30.7916     2     2 47.0699   0.02080
var      81.530 27.1767     3     9 41.5440 1.343e-05
soil:var  3.743  0.6238     6     9  0.9535   0.50444
```

```
ls_means(fit, which="soil")
```

```
Least Squares Means table:

      Estimate Std. Error df t value    lower    upper Pr(>|t|)
soil1  9.93750    0.68057  2  14.602  7.00924 12.86576 0.004657
soil2 15.92500    0.68057  2  23.399 12.99674 18.85326 0.001821
soil3 19.13750    0.68057  2  28.120 16.20924 22.06576 0.001262

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

```
lsm = ls_means(fit, which="soil", pairwise=TRUE); lsm
```

```
Least Squares Means table:

               Estimate Std. Error df t value     lower     upper Pr(>|t|)
soil1 - soil2  -5.98750    0.96247  2 -6.2210 -10.12869  -1.84631  0.02488
soil1 - soil3  -9.20000    0.96247  2 -9.5587 -13.34119  -5.05881  0.01077
soil2 - soil3  -3.21250    0.96247  2 -3.3378  -7.35369   0.92869  0.07924

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

```
show_tests(lsm) # shows which contrasts were used above
```

```
$soil
              (Intercept) blockSouth soil2 soil3 var2 var3 var4 soil2:var2 soil3:var2
soil1 - soil2           0          0    -1     0    0    0    0      -0.25       0.00
soil1 - soil3           0          0     0    -1    0    0    0       0.00      -0.25
soil2 - soil3           0          0     1    -1    0    0    0       0.25      -0.25
              soil2:var3 soil3:var3 soil2:var4 soil3:var4
soil1 - soil2      -0.25       0.00      -0.25       0.00
soil1 - soil3       0.00      -0.25       0.00      -0.25
soil2 - soil3       0.25      -0.25       0.25      -0.25
```

```
ls_means(fit, which="var")
```

```
Least Squares Means table:

     Estimate Std. Error  df t value    lower    upper  Pr(>|t|)
var1 15.25000    0.48597 4.4  31.381 13.94828 16.55172 2.385e-06
var2 17.86667    0.48597 4.4  36.765 16.56495 19.16839 1.190e-06
var3 13.91667    0.48597 4.4  28.637 12.61495 15.21839 3.563e-06
var4 12.96667    0.48597 4.4  26.682 11.66495 14.26839 4.857e-06

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

Why this standard error, and why this degree of freedom of 4.4?
Mean for variety 1 (say): from average over 2 blocks and 3 soils. It's standard error is then:

```
sqrt((0.7628 + 0.6542)/6) # sqrt of (s2_epsilon + s2_delta) * 1/(2*3)
```

```
[1] 0.4859698
```

This estimated variance is a rescaled version of
`MSWPError + (4-1) MSSPError = 4(s2_epsilon + s2_delta)`, so the Satterthwaite approximation will use:

```
MSWPE = 0.6542 + 4*0.7628 # df = 2
MSSPE = 0.6542 # df = 9
dfsat = (MSWPE + 3*MSSPE)^2 / ( MSWPE^2/2 + (3*MSSPE)^2/9)
dfsat # 4.4 degrees of freedom to approximate by a chi-square distribution
```

```
[1] 4.405093
```

```
head(ls_means(fit, which = "soil:var"), n=2)
```

```
Least Squares Means table:

           Estimate Std. Error  df t value    lower    upper  Pr(>|t|)
soil1:var1 10.50000    0.84172 4.4  12.475  8.24535 12.75465 0.0001328
soil2:var1 16.00000    0.84172 4.4  19.009 13.74535 18.25465 2.139e-05

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

```
head(ls_means(fit, which = "soil:var", pairwise=TRUE), n=4)
```

```
Least Squares Means table:

                         Estimate Std. Error  df t value     lower     upper Pr(>|t|)
soil1:var1 - soil2:var1  -5.50000    1.19037 4.4 -4.6204  -8.68855  -2.31145 0.007833
soil1:var1 - soil3:var1  -8.75000    1.19037 4.4 -7.3507 -11.93855  -5.56145 0.001243
soil1:var1 - soil1:var2  -2.45000    0.80881 9.0 -3.0292  -4.27965  -0.62035 0.014266
soil1:var1 - soil2:var2  -8.55000    1.19037 4.4 -7.1826 -11.73855  -5.36145 0.001367

  Confidence level: 95%
  Degrees of freedom method: Satterthwaite
```

Warning: an alternative is the `multcomp` package. It uses a normal approximation to test contrasts and pairwise comparisons. This is very liberal for `soil` comparisons, for which df should be taken to be 2 (not infinite).

```
library(multcomp)
summary(glht(fit, linfct=mcp(soil="Tukey", interaction_average=T)))
```

```


     Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lmer(formula = y ~ block + soil + var + soil:var + (1 | pan), 
    data = dat)

Linear Hypotheses:
           Estimate Std. Error z value Pr(>|z|)    
2 - 1 == 0   5.9875     0.9625   6.221   <1e-04 ***
3 - 1 == 0   9.2000     0.9625   9.559   <1e-04 ***
3 - 2 == 0   3.2125     0.9625   3.338   0.0024 ** 
(Adjusted p values reported -- single-step method)
```