Session 10

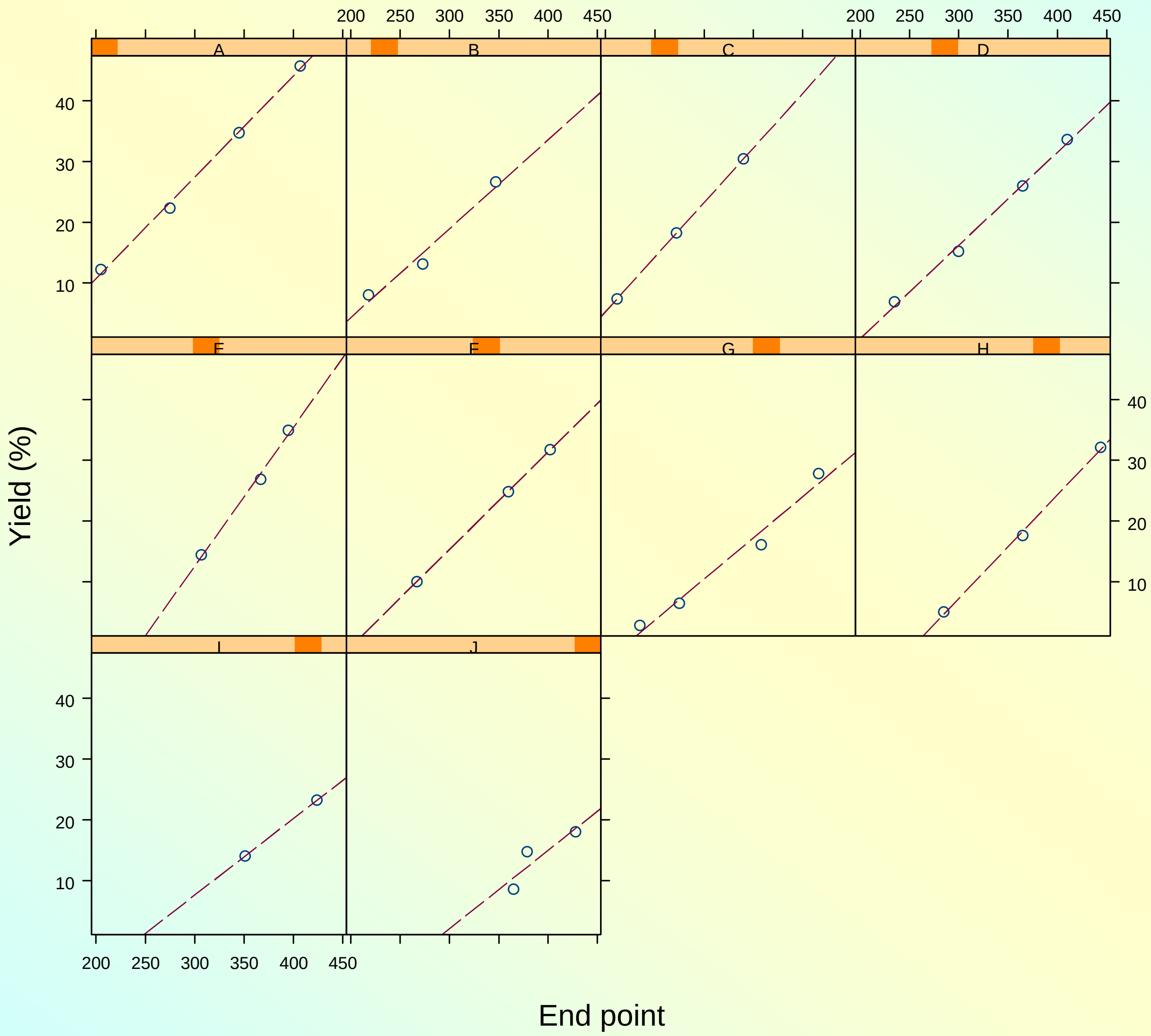
Linear Mixed Effects Models

Example: The petroleum data of N H Prater

- 10 samples of crude oil
- Partitioned and tested for yield at different "end points"
- The samples themselves are classified by specific gravity (SG), vapour pressure (VP) and volatility as measured by the ASTM 10% point (V10)
- Problems:
  - Estimate the rate of rise in yield with end point
  - Can we explain differences between samples in terms of external measures?

The petrol data displayed

```
xyplot(Y ~ EP | No, petrol,
  panel = function(x, y, ...) {
     panel.xyplot(x, y, ...)
     panel.lmline(x, y, col = 3, lty = 4)
  },
  as.table = T,
  aspect = "xy",
  xlab = "End point",
  ylab = "Yield (%)")
```


A
B
C
D
E
F
G
H
I
J
200 250 300 350 400 450
10 20 30 40
Yield (%)
End point

Can the slopes be regarded as parallel?

```
petrol.lm1 <- aov(Y ~ No/EP, petrol)
petrol.lm2 <- aov(Y ~ No + EP, petrol)
anova(petrol.lm2, petrol.lm1)
```

```
Analysis of Variance Table

Response: Y

    Terms Resid. Df      RSS    Test Df Sum of Sq  F Value     Pr(F)
1 No + EP        21 74.13199
2 No/EP          12 30.32915 1 vs. 2  9  43.80284 1.925665 0.1439048
```

Yes?

Can we explain differences in intercepts?

```
petrol.lm3 <- aov(Y ~ . - No, petrol)
anova(petrol.lm3, petrol.lm2)
```

```
Analysis of Variance Table

Response: Y

    Terms Resid. Df     RSS    Test Df Sum of Sq  F Value      Pr(F)
1 SG + VP +
  V10 + EP       27 134.804
2 No + EP        21  74.132 1 vs. 2  6  60.67197 2.864511 0.03368118
```

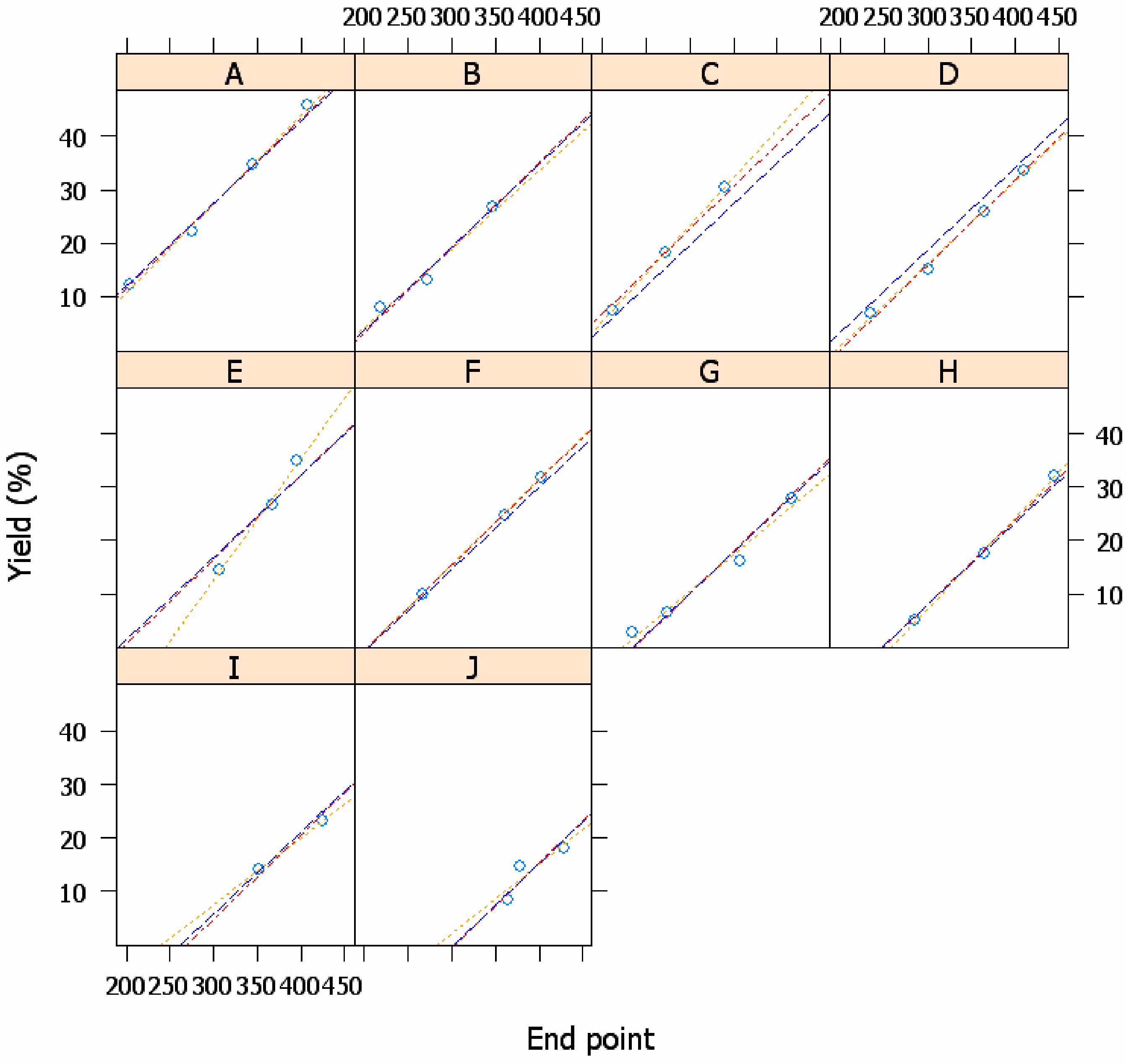
- Looks doubtful, but how close is it?

Looking at all three models

```
tmp <- update(petrol.lm2, .~.-1)
a0 <- predict(tmp, type="terms")[, "No"]
b0 <- coef(tmp)["EP"]
a <- cbind(1, petrol$SG, petrol$VP, petrol$V10) %*%
             coef(petrol.lm3)[1:4]
b <- coef(petrol.lm3)["EP"]

palette(c("red", "blue", "orange"))

xyplot(Y ~ EP | No, petrol, subscripts = T,
  panel = function(x, y, subscripts, ...) {
  panel.xyplot(x, y, ...)
  panel.lmline(x, y, col = 3, lty = 3)
  panel.abline(a0[subscripts][1], b0, col = 4, lty = 4)
  panel.abline(a[subscripts][1], b, col = 5, lty = 5)
}, as.table = T, aspect = "xy", xlab = "End point",
  ylab = "Yield (%)")
```


A
B
C
D
E
F
G
H
I
J
Yield (%)
End point
200 250 300 350 400 450
10 20 30 40

Adding a random term

- The external (sample-specific) variables explain much of the differences between the intercepts, but there is still significant variation between samples unexplained.
- One natural way to model this is to assume that, in addition to the systematic variation between intercepts explained by regression on the external variables, there is a random term as well.
- We assume this term to be normal with zero mean. Its variance measures the extent of this additional variation

```
petrol.lme <- lme(Y ~ SG + VP + V10 + EP, data =
   petrol, random = ~1 | No)

summary(petrol.lme)

Linear mixed-effects model fit by REML
 …

Random effects:
 Formula: ~1 | No
        (Intercept) Residual
StdDev:    1.444741 1.872208

Fixed effects: Y ~ SG + VP + V10 + EP
                Value Std.Error DF   t-value p-value
(Intercept) -6.134791 14.554171 21 -0.421514  0.6777
SG           0.219398  0.146938  6  1.493136  0.1860
VP           0.545863  0.520528  6  1.048673  0.3347
V10         -0.154243  0.039962  6 -3.859697  0.0084
EP           0.157177  0.005588 21 28.127561  0.0000
 …
```

Comparison with previous model

```
round(summary.lm(petrol.lm3)$coef, 5)
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) -6.82077   10.12315 -0.67378  0.50618
SG           0.22725    0.09994  2.27390  0.03114
VP           0.55373    0.36975  1.49756  0.14585
V10         -0.14954    0.02923 -5.11597  0.00002
EP           0.15465    0.00645 23.99221  0.00000
```

```
Fixed effects: Y ~ SG + VP + V10 + EP
                Value Std.Error DF   t-value p-value
(Intercept) -6.134795  14.55411 21  -0.42152  0.6777
         SG  0.219398   0.14694  6   1.49314  0.1860
         VP  0.545863   0.52053  6   1.04868  0.3347
        V10 -0.154243   0.03996  6  -3.85971  0.0084
         EP  0.157177   0.00559 21  28.12753  <.0001
```

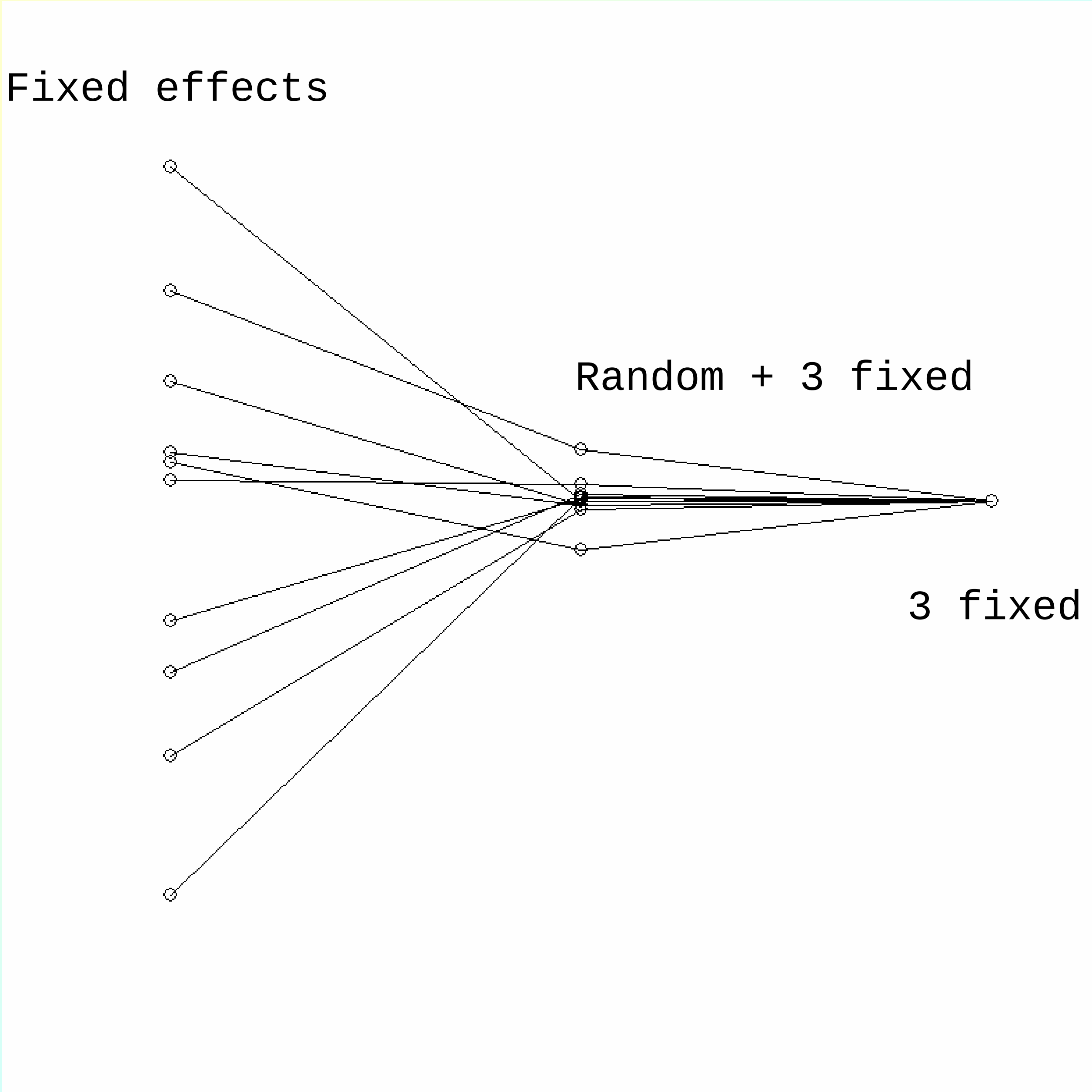
Shrinkage

```
dat <- with(petrol,
  expand.grid(No = levels(No), SG = mean(SG),
      VP = mean(VP), V10=mean(V10), EP = mean(EP)))

fixM <- predict(petrol.lm2, dat)
ranM <- predict(petrol.lme, dat)
conM <- predict(petrol.lm3, dat)

X <- rbind(cbind(1,fixM), cbind(2,ranM), cbind(3,conM))
plot(X, axes = F, ann = F)

segments(1,fixM, 2,ranM)
segments(2,ranM, 3,conM)
```



Fixed effects
Random + 3 fixed
3 fixed

Adding random slopes as well

- The deviation from parallel regressions was not significant, but still somewhat suspicious.
- We might consider making both the intercept and the slope have a random component.

```
petrol.lme2 <- lme(Y ~ SG + VP + V10 + EP, data =
  petrol,random = ~ 1 + EP | No)

summary(petrol.lme2)
```

…

```
Random effects:
 Formula: ~1 + EP | No
 Structure: General positive-definite, Log-Cholesky
  parametrization
            StdDev      Corr
(Intercept) 1.725640232 (Intr)
EP          0.003283688 -0.535
Residual    1.854119617

Fixed effects: Y ~ SG + VP + V10 + EP
                Value Std.Error DF   t-value p-value
(Intercept) -6.229186 14.632337 21 -0.425714  0.6746
SG           0.219109  0.147941  6  1.481050  0.1891
VP           0.548118  0.523123  6  1.047781  0.3351
V10         -0.153940  0.040213  6 -3.828130  0.0087
EP           0.157248  0.005666 21 27.753481  0.0000
```

…