STAT-S632

Assignment 6

John Koo

```
import::from(magrittr, `%>%`, `%<>%`)
library(ggplot2)
import::from(lme4, lmer, VarCorr, ranef)
import::from(RLRsim, exactRLRT)

theme_set(theme_bw())
```

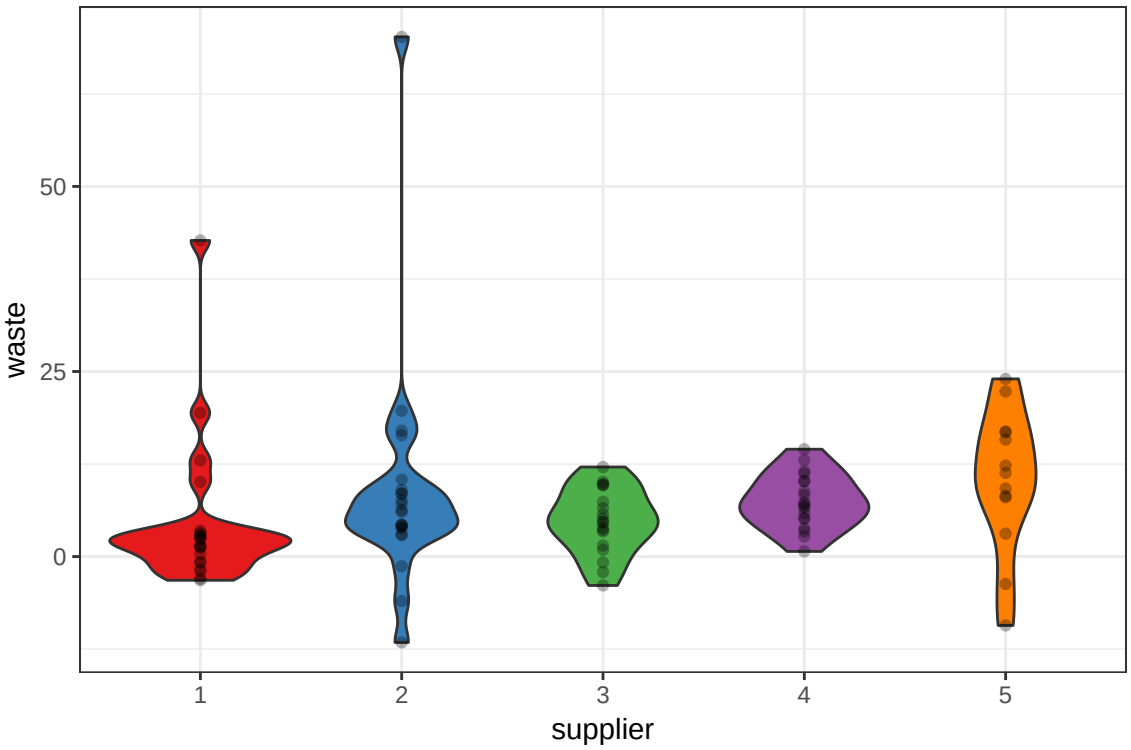
Problem 1

```
denim.df <- faraway::denim
summary(denim.df)
```

```
     waste          supplier
 Min.   :-11.600   1:22
 1st Qu.:  2.550   2:22
 Median :  5.200   3:19
 Mean   :  6.977   4:19
 3rd Qu.:  9.950   5:13
 Max.   : 70.200
```

Part a

```
ggplot(denim.df) +
  geom_violin(aes(x = supplier, y = waste, fill = supplier)) +
  geom_point(aes(x = supplier, y = waste), alpha = .3) +
  scale_fill_brewer(palette = 'Set1') +
  guides(fill = FALSE)
```



Part b

```
fixed.mod <- lm(waste ~ supplier, data = denim.df)
summary(fixed.mod)
```

```
Call:
lm(formula = waste ~ supplier, data = denim.df)

Residuals:
    Min      1Q  Median      3Q     Max
-20.432  -4.377  -1.323   2.639  61.368

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.5227     2.1021   2.152   0.0341 *
supplier2     4.3091     2.9728   1.450   0.1507
supplier3     0.3089     3.0879   0.100   0.9206
supplier4     2.9667     3.0879   0.961   0.3392
supplier5     5.8542     3.4491   1.697   0.0931 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.86 on 90 degrees of freedom
Multiple R-squared:  0.04901,   Adjusted R-squared:  0.006747
F-statistic:  1.16 on 4 and 90 DF,  p-value: 0.334
```

```
drop1(fixed.mod, test = 'F')
```

```
Single term deletions

Model:
waste ~ supplier
         Df Sum of Sq    RSS    AIC F value Pr(>F)
<none>                8749.1 439.67
supplier  4    450.92 9200.0 436.44  1.1596  0.334
```

`supplier` is not significant.

Part c

```
rand.eff.mod <- lmer(waste ~ 1 + (1 | supplier), data = denim.df)
summary(rand.eff.mod)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: waste ~ 1 + (1 | supplier)
   Data: denim.df

REML criterion at convergence: 702.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9095 -0.4363 -0.1669  0.3142  6.3817

Random effects:
 Groups   Name        Variance Std.Dev.
 supplier (Intercept)  0.6711  0.8192
 Residual              97.3350  9.8658
Number of obs: 95, groups:  supplier, 5

Fixed effects:
            Estimate Std. Error t value
(Intercept)    6.997      1.078    6.49
```

$\beta = \mu \in \mathbb{R}^1$

$X \in \mathbb{R}^{95}$ (`nrow(denim.df)`)

$y, \epsilon \in \mathbb{R}^{95}$ as well

$\gamma \in \mathbb{R}^5$, or the number of levels of `supplier`

$Z = \begin{bmatrix} z_1 & & & & \\ & z_2 & & & \\ & & z_3 & & \\ & & & z_4 & \\ & & & & z_5 \end{bmatrix}$ where $\dim(z_i) = |\{x : x = \alpha_i\}|$, or the number of elements in level i. So $Z \in \mathbb{R}^{95 \times 5}$.

Part d

Parametric bootstrapping

```
confint(rand.eff.mod, method = 'boot')
```

```
                2.5 %    97.5 %
.sig01      0.000000  3.788285
.sigma      8.325332 11.300242
(Intercept) 4.866398  8.751813
```

Since the lower bound of the interval is 0, we cannot say that the variance of `supplier` is significant. We can also see this from the p-value.

```
exactRLRT(rand.eff.mod)
```

```
    simulated finite sample distribution of RLRT.

    (p-value based on 10000 simulated values)

data:
RLRT = 0.029383, p-value = 0.3491
```

LR test

```
mle.mod <- lmer(waste ~ 1 + (1 | supplier), data = denim.df, REML = FALSE)
pchisq(2 * (logLik(mle.mod) - logLik(fixed.mod)), 1, lower.tail = FALSE)
```

```
'log Lik.' 1 (df=3)
```

We get a p-value of 1.

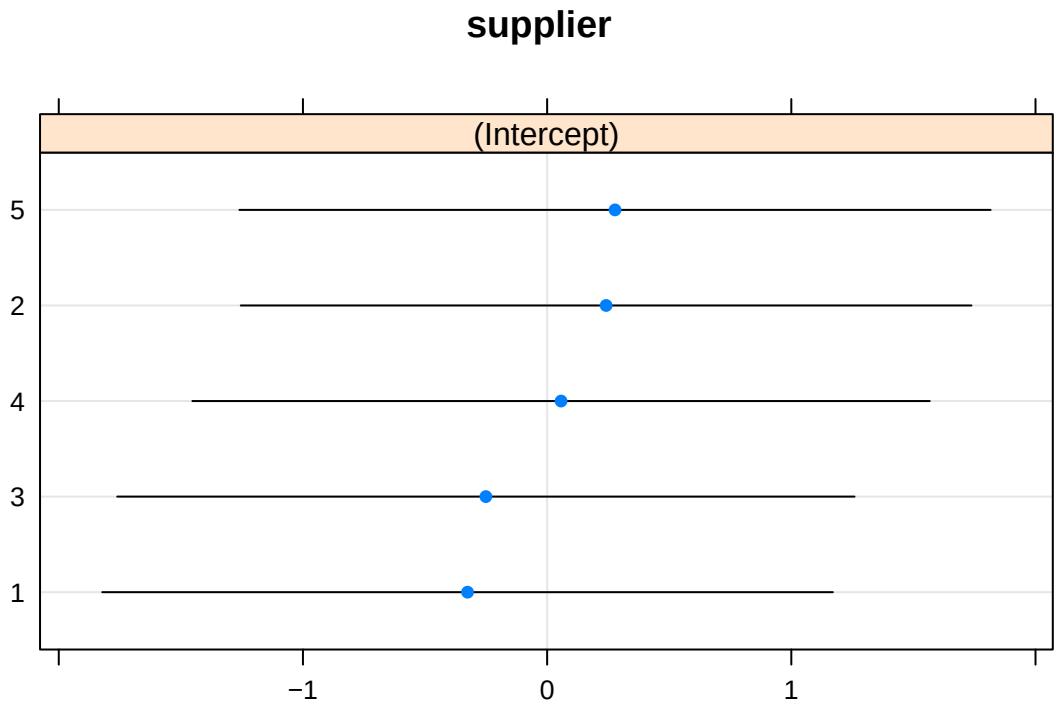
Part e

```
ranef(rand.eff.mod)
```

```
$supplier
  (Intercept)
1 -0.32586973
2  0.24163762
3 -0.25080816
4  0.05703177
5  0.27800850
```

```
lattice::dotplot(ranef(rand.eff.mod, condVar = TRUE))
```

```
$supplier
```



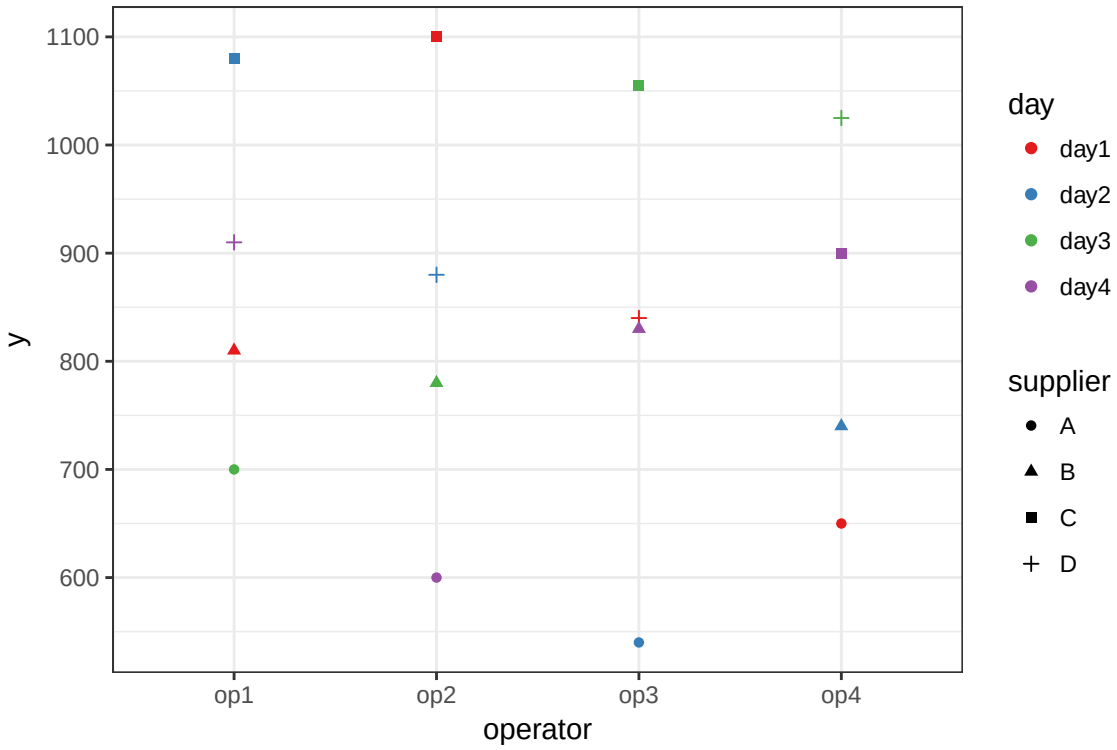
Problem 2

```
breaking.df <- faraway::breaking
summary(breaking.df)
```

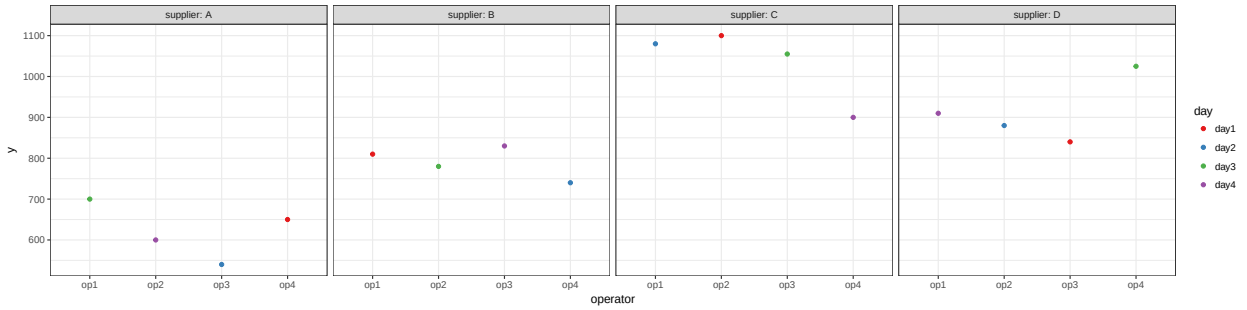
```
       y          operator   day     supplier
 Min.   : 540.0   op1:4    day1:4   A:4
 1st Qu.: 730.0   op2:4    day2:4   B:4
 Median : 835.0   op3:4    day3:4   C:4
 Mean   : 840.0   op4:4    day4:4   D:4
 3rd Qu.: 938.8
 Max.   :1100.0
```

Part a

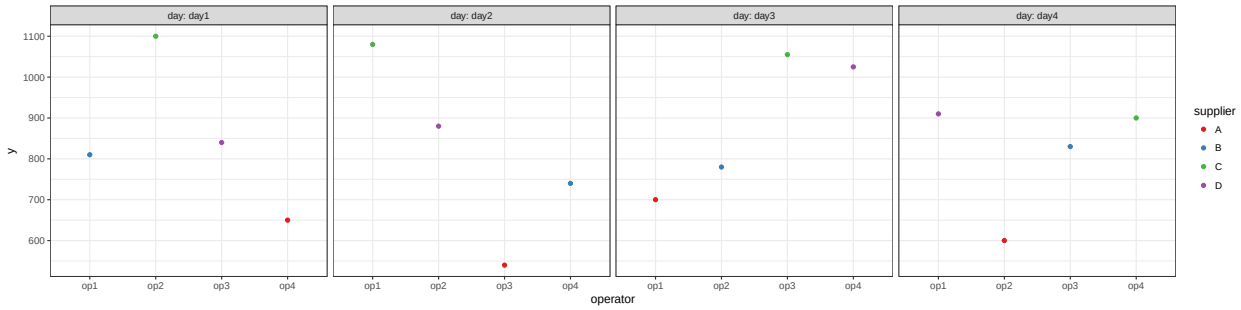
```
ggplot(breaking.df,
       aes(y = y, x = operator, colour = day, shape = supplier)) +
  geom_point() +
  scale_colour_brewer(palette = 'Set1')
```




```
ggplot(breaking.df,
       aes(y = y, x = operator, colour = day)) +
  geom_point() +
  scale_colour_brewer(palette = 'Set1') +
  facet_wrap(~ supplier, labeller = 'label_both', nrow = 1)
```




```
ggplot(breaking.df,
       aes(y = y, x = operator, colour = supplier)) +
  geom_point() +
  scale_colour_brewer(palette = 'Set1') +
  facet_wrap(~ day, labeller = 'label_both', nrow = 1)
```



There appears to be a suggestion of differences among suppliers but not among operators or days.

Part b

$X \in \mathbb{R}^{16\times 4}$ (16 rows $\times$ 4 factor levels)

$\beta \in \mathbb{R}^4$ (4 factor levels)

$y, \epsilon \in \mathbb{R}^{16}$ (16 rows)

$\gamma \in \mathbb{R}^7$ (1 intercept + $(3-1)$ from factor 1 + $(3-1)$ from factor 2)

$Z \in \mathbb{R}^{16\times 7}$, or the number of rows $\times$ dim(γ)

Part c

```
fixed.mod <- lm(y ~ day + operator + supplier, data = breaking.df)
summary(fixed.mod)
```

```
Call:
lm(formula = y ~ day + operator + supplier, data = breaking.df)

Residuals:
   Min     1Q Median     3Q    Max
-92.50 -25.94  -6.25  31.88  93.75

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   667.50      62.29  10.716  3.9e-05 ***
dayday2       -40.00      55.72  -0.718 0.499782
dayday3        40.00      55.72   0.718 0.499782
dayday4       -40.00      55.72  -0.718 0.499782
operatorop2   -35.00      55.72  -0.628 0.553020
operatorop3   -58.75      55.72  -1.054 0.332266
operatorop4   -46.25      55.72  -0.830 0.438247
supplierB     167.50      55.72   3.006 0.023812 *
supplierC     411.25      55.72   7.381 0.000317 ***
supplierD     291.25      55.72   5.227 0.001962 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 78.79 on 6 degrees of freedom
```

```
Multiple R-squared:  0.9141,    Adjusted R-squared:  0.7853
F-statistic: 7.094 on 9 and 6 DF,  p-value: 0.01348
```

```
drop1(fixed.mod, test = 'F')
```

```
Single term deletions

Model:
y ~ day + operator + supplier
         Df Sum of Sq    RSS    AIC F value   Pr(>F)
<none>                 37250 144.04
day       3     17600  54850 144.24  0.9450 0.475896
operator  3      7662  44912 141.04  0.4114 0.750967
supplier  3    371137 408387 176.36 19.9268 0.001602 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fixed.mod, lm(y ~ supplier, data = breaking.df))
```

```
Analysis of Variance Table

Model 1: y ~ day + operator + supplier
Model 2: y ~ supplier
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1      6 37250
2     12 62513 -6    -25262 0.6782 0.6754
```

```
fixed.mod <- lm(y ~ supplier, data = breaking.df)
summary(fixed.mod)
```

```
Call:
lm(formula = y ~ supplier, data = breaking.df)

Residuals:
     Min       1Q   Median       3Q      Max
-133.750  -37.813    8.125   41.563  111.250

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   622.50      36.09  17.250 7.79e-10 ***
supplierB     167.50      51.04   3.282  0.00656 **
supplierC     411.25      51.04   8.058 3.49e-06 ***
supplierD     291.25      51.04   5.707 9.81e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 72.18 on 12 degrees of freedom
Multiple R-squared:  0.8558,    Adjusted R-squared:  0.8198
F-statistic: 23.75 on 3 and 12 DF,  p-value: 2.464e-05
```

It appears that `supplier` is the only significant covariate.

Part d

```
mixed.mod <- lmer(y ~ supplier + (1 | operator) + (1 | day), data = breaking.df)
summary(mixed.mod)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ supplier + (1 | operator) + (1 | day)
   Data: breaking.df

REML criterion at convergence: 142.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8299 -0.4718  0.1027  0.6518  1.4691

Random effects:
 Groups   Name        Variance  Std.Dev.
 operator (Intercept) 7.275e-13 8.529e-07
 day      (Intercept) 2.191e+02 1.480e+01
 Residual             4.990e+03 7.064e+01
Number of obs: 16, groups:  operator, 4; day, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)   622.50      36.09  17.250
supplierB     167.50      49.95   3.353
supplierC     411.25      49.95   8.233
supplierD     291.25      49.95   5.831

Correlation of Fixed Effects:
          (Intr) spplrB spplrC
supplierB -0.692
supplierC -0.692  0.500
supplierD -0.692  0.500  0.500
```

The experiment was conducted to select a supplier, so it is the effect of interest.

Supplier C has the highest breaking point, as we saw in the fixed effect model and the plots.

Part e

```
mixed.op.mod <- lmer(y ~ supplier + (1 | operator), data = breaking.df)
mixed.day.mod <- lmer(y ~ supplier + (1 | day), data = breaking.df)

exactRLRT(mixed.op.mod, mixed.mod, mixed.day.mod)
```

```
    simulated finite sample distribution of RLRT.

    (p-value based on 10000 simulated values)

data:
RLRT = 0, p-value = 1
```

```
exactRLRT(mixed.day.mod, mixed.mod, mixed.op.mod)
```

```
    simulated finite sample distribution of RLRT.

    (p-value based on 10000 simulated values)

data:
RLRT = 0.030235, p-value = 0.3716
```

Neither are significant.