One-Way ANOVA

One-Way ANOVA

One-way ANOVA model: Given a factor α occuring at $i = 1, \ldots, I$ levels, with $j = 1, \ldots, n_i$ observations per level. We assume

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, I; \quad j = 1, \ldots, n_i$$

where $\epsilon_{ij} \sim i.i.d. N(0, \sigma^2)$.

ANOVA – Analysis of Variance: partition the overall variance in the response to each of the factors and the error

Blood Coagulation Example

- Study of blood coagulation times
- Factor: 4 different diets
- $n = 24$ animals randomly assigned to four different diets.
- $I_1 = 4, I_2 = 6, I_3 = 6, I_4 = 8$

	A	B	C	D
mean	61.00	66	68.0	61.00
var	3.33	8	2.8	6.86
n_i	4.00	6	6.0	8.00

Identifiability

Let $\beta = (\mu, \alpha_1, ..., \alpha_I)$, β is not identifiable!

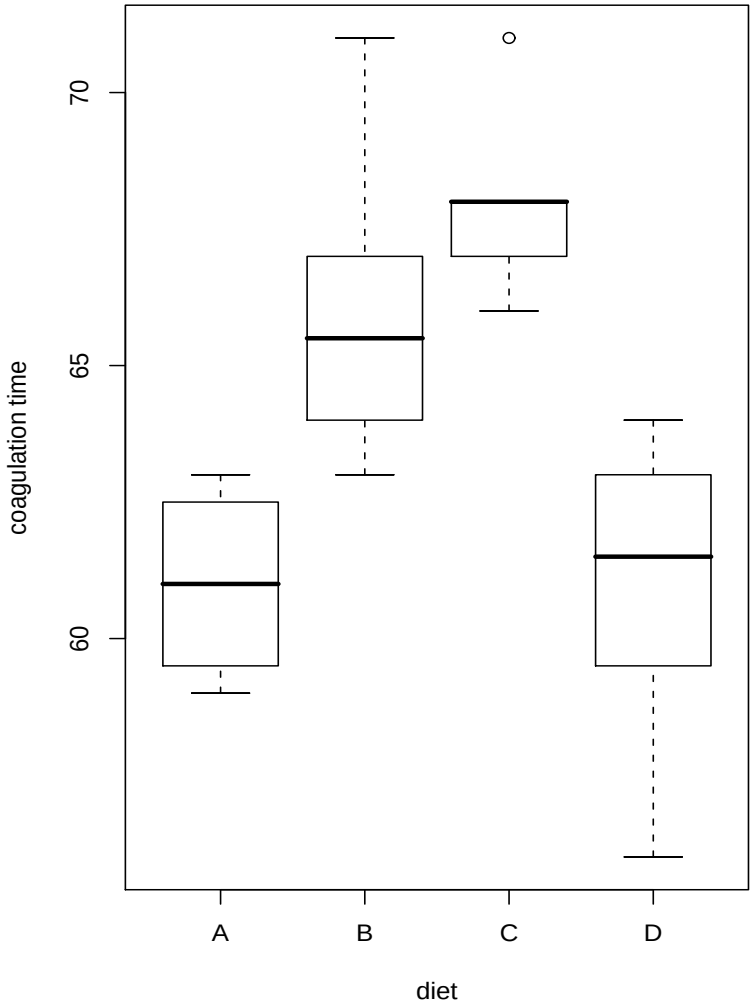
Solution:

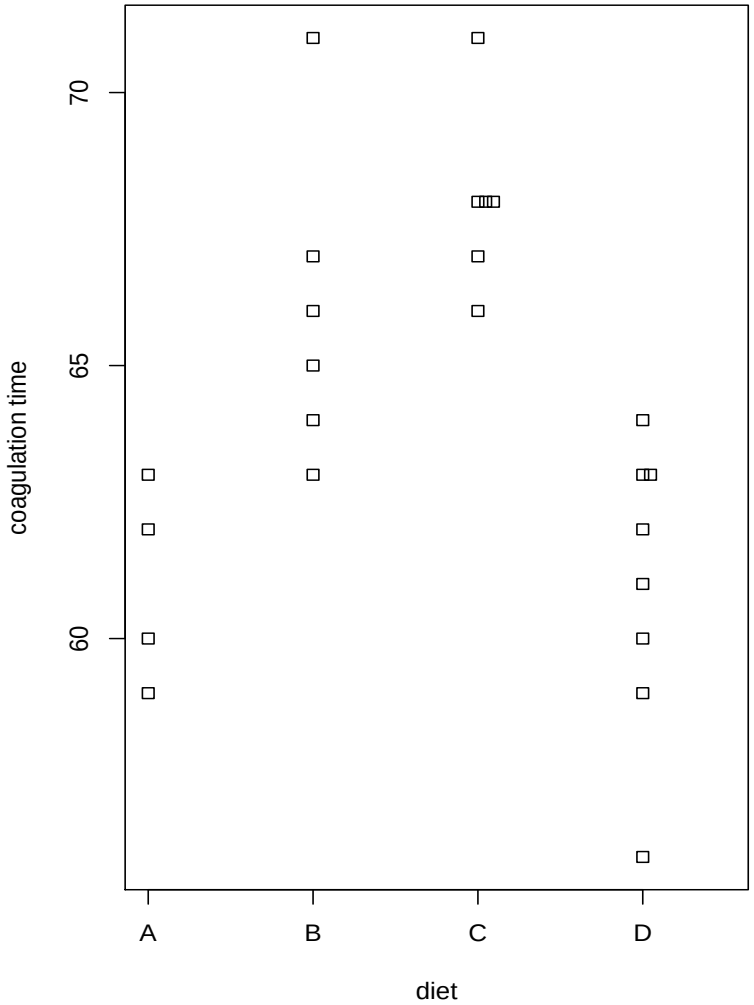
- Set $\mu = 0$
- Set $\alpha_1 = 0$.
- Set $\sum_i \alpha_i = 0$.

Blood Coagulation Example Continued

```
> library(faraway)
> data(coagulation)
> coagulation
   coag diet
1    62    A
2    60    A
... ...
23   63    D
24   59    D
```

```
> par(mfrow=c(1, 2))
> plot(coag ~ diet, coagulation)
> stripchart(coagulation$coag ~
    coagulation$diet, method="stack")
## Stripchart plot is preferred
## when there is little data
```


coagulation time
60
65
70
A
B
C
D
diet


coagulation time
60
65
70
A
B
C
D
diet

```
> g <- lm(coag ~ diet, coagulation)
> summary(g)
Coefficients:
            Estimate Std.Error t value Pr(>|t|)
Intercept 6.100e+01 1.183e+00  51.554  < 2e-16
dietB     5.000e+00 1.528e+00   3.273 0.003803
dietC     7.000e+00 1.528e+00   4.583 0.000181
dietD    -1.071e-14 1.449e+00  -7e-15 1.000000
Residual standard error:  2.366 on 20 DF
Multiple R-Squared: 0.6706
F-statistic: 13.57 on 3 and 20 DF
p-value: 4.658e-05
## Small p-value indicates some difference
## between the groups
```

Other constraints

```
> model.matrix(g)
> gi <- lm(coag ~ diet - 1, coagulation)
> summary(gi)
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
dietA  61.0000     1.1832   51.55   <2e-16
dietB  66.0000     0.9661   68.32   <2e-16
dietC  68.0000     0.9661   70.39   <2e-16
dietD  61.0000     0.8367   72.91   <2e-16
Residual standard error:
  2.366 on 20 degrees of freedom
Multiple R-Squared: 0.9989
```

```
Adjusted R-squared: 0.9986
F-statistic:  4399 on 4 and 20 DF
p-value: < 2.2e-16
## p-values not useful in this case
> gnull <- lm(coag ~ 1, coagulation)
> anova(gnull,gi)
Analysis of Variance Table
Model 1: coag ~ 1
Model 2: coag ~ diet - 1
  Res.Df RSS Df Sum of Sq      F    Pr(>F)
1     23 340
2     20 112  3       228 13.571 4.658e-05 ***

---
```

```
> options(contrasts=c("contr.sum","contr.poly"))
> gs <- lm(coag ~ diet , coagulation)
> summary(gs)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.0000     0.4979 128.537  < 2e-16 ***
diet1        -3.0000     0.9736  -3.081 0.005889 **
diet2         2.0000     0.8453   2.366 0.028195 *
diet3         4.0000     0.8453   4.732 0.000128 ***
Residual standard error: 2.366 on 20 DF
R-Squared: 0.6706,
F-statistic: 13.57 on 3 and 20 DF,
p-value: 4.658e-05
```

Diagnostics

```
### QQ plot to check normality assumption:
> qqnorm(residuals(g))
### Residual plots: no outlier, skewness or unequal varianc
> plot(jitter(fitted(g)), residuals(g),
    xlab="Fitted", ylab="Residuals")

## LRT for equal variance:
> bartlett.test(coag~diet, data=coagulation)
Bartlett test of homogeneity of variances
data:  coag by diet
Bartlett's K-squared = 1.668, df = 3, p-value = 0.6441
```

Blood Coagulation Example Continued