Assumptions behind the ANOVA F-test

1. The samples are randomly selected in an independent manner from the k treatment populations. [Satisfied in a CRD]
2. All k populations have distributions that are approximately normal.
3. The k population variances are equal.

$$\sigma_1^2 = \sigma_2^2 = \cdots \sigma_k^2.$$

Example: Milk Quality Data.

The impact on milk protein level of three different diets is being studied.

Data: Measurements of milk protein levels for $n = 1337$ samples.

- **Response:** Milk Protein Level (%)
- **Factor:** DIET
- **Factor levels:** $k = 3$
  - 1: Barley
  - 2: Barley + Lupins
  - 3: Lupins

	TMT 1	TMT 2	TMT 3
n_i	425	459	453
$\overline{x}i$	3.532	3.430	2.312
s_i^2	0.102	0.091	0.114

$$\begin{aligned} SST &= 10.606 \\ SSE &= 136.432 \\ SS &= 147.038 \end{aligned}$$

$$\begin{aligned} k - 1 &= 2 \\ n - k &= 1334 \end{aligned}$$

Therefore

$$MST = \frac{SST}{k-1} = \frac{10.606}{2} = 5.303$$

$$MSE = \frac{SSE}{n-k} = \frac{136.432}{1334} = 0.102$$

and

$$F = \frac{MST}{MSE} = 51.851$$

If H_0 is true, that is,

$$\mu_1 = \mu_2 = \mu_3$$

then F should look like an observation from a

$$\text{Fisher-F}(k-1, n-k)$$

distribution.

Here we are dealing with the

$$\text{Fisher-F}(2, 1334)$$

distribution. From tables, we discover that if $\alpha = 0.05$, then

$$F_\alpha(2, 1334) = 3.002$$

and thus we

Reject H_0

and conclude that there is a significant impact on milk protein level due to diet.

Note: Tables in McClave and Sincich (p 901) only give

$$\begin{aligned} F_{0.05}(2, 120) &= 3.07 \\ F_{0.05}(2, \infty) &= 3.00 \end{aligned}$$

so we cannot look up $F_{0.05}(2, 1334)$. However, we know that

$$3.00 < F_{0.05}(2, 1334) < 3.07$$

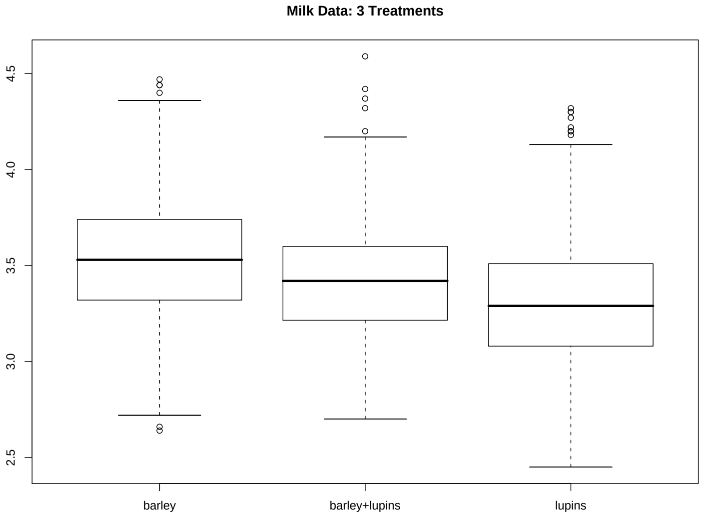
and here the test statistic is $F = 51.851$.

Are the assumptions met ?

1. **Independent samples** : Not possible to tell with current information. In fact, data comprise repeated measurements on 79 cows - potentially not independent, as observations on the same cow are likely to be more similar.
2. **Normal Distributions** : Visual inspection of boxplots indicates that this may be valid.
3. **Equal variances** :

$$s_1^2 = 0.102 \qquad s_2^2 = 0.091 \qquad s_3^2 = 0.114$$

so assumption appears to be valid
- can we test this formally ?


Milk Data: 3 Treatments
4.5
4.0
3.5
3.0
2.5
barley
barley+lupins
lupins

Example: Anxiety Response Treatment.

In a study of Alzheimer's disease and care of its sufferers, a medication designed to improve anxiety relief has been developed.

In a lab experiment, $n = 20$ rats were assigned to one of four ($k = 4$) treatment groups corresponding to dose-level of the medication.

A measure of response to a "flee stimulus" was recorded.

- **Response:** Pull response to stimulus (units of force)
- **Factor:** DOSE-LEVEL
- **Factor levels:** $k = 4$
  - Dose 0 (zero units)
  - Dose 1 (one unit)
  - Dose 2 (two units)
  - Dose 3 (three units)

0	1	2	2
27.0	22.8	21.9	23.5
26.2	23.1	23.4	19.6
28.8	27.7	20.1	23.7
33.5	27.6	27.8	20.8
28.8	24.0	19.3	23.9

We find that

$$SST = 140.094 \qquad SSE = 116.324 \qquad SS = 256.418$$

$$MST = 46.698 \qquad MSE = 7.270$$

and

$$F = 6.423$$

which we need to compare with the Fisher-F$(3, 16)$ distribution.

For $\alpha = 0.05$, from McClave and Sincich (p 901)

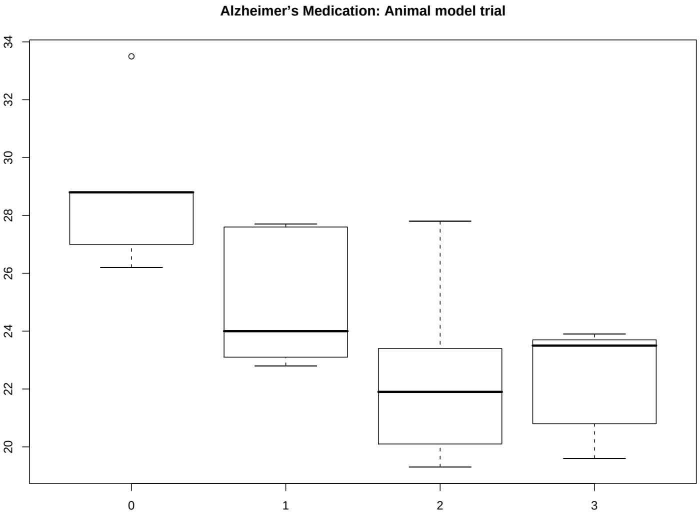
$$F_{0.05}(3, 16) = 3.24$$

and so we

$$\text{Reject } H_0$$

at $\alpha = 0.05$ and conclude that there is a significant difference between treatment groups.

p-value is 0.0046.


Alzheimer's Medication: Animal model trial
20
22
24
26
28
30
32
34
0
1
2
3