Lecture 03b – One-way ANOVA

ENVX2001 Applied Statistical Methods

Januar Harianto

Feb 2026

Why ANOVA?

The problem with multiple t-tests

Suppose we want to compare weight gain of chicks fed on **4 different diets**.

With t-tests, we'd need $\binom{4}{2} = 6$ pairwise comparisons:

1-2, 1-3, 1-4, 2-3, 2-4, 3-4

Even if there are no true differences, each test has a 5% chance of incorrectly finding significance. How does that add up?

```
prob_all_correct ← 0.95^6
prob_all_correct
```

```
[1] 0.7350919
```

We have a **26.5%** chance that at least one test result will be incorrect!

We need a better method

The problem of multiple comparisons requires a technique that considers **all** treatments at once.

ANOVA – Analysis of Variance

Before deciding *which* treatments differ, we first consider:

1. Differences **between** the treatment groups – **between-treatment effects**
2. Differences **within** each treatment group – **within-treatment effects**
   - due to random environmental fluctuations, genetics, experimental error

Chick weight data

The experiment

- 20 chicks
- 4 diets
- 5 replicates per diet
- Response: weight gain (g)

```
chicks_wide ← read.csv("data/chicks.csv")
chicks_wide
```

```
  Diet.1 Diet.2 Diet.3 Diet.4
1     99     61     42    169
2     88    112     97    137
3     76     30     81    169
4     38     89     95     85
5     94     63     92    154
```

Reshaping the data

```
chicks ← read.csv("data/chicks.csv") |>
  pivot_longer(
    cols = starts_with("Diet"),
    names_to = "diet",
    values_to = "weight"
  ) |>
  mutate(diet = as.factor(diet))

chicks
```

```
# A tibble: 20 × 2
   diet   weight
   <fct>   <int>
 1 Diet.1     99
 2 Diet.2     61
 3 Diet.3     42
 4 Diet.4    169
 5 Diet.1     88
```

```
 6 Diet.2    112
 7 Diet.3     97
 8 Diet.4    137
 9 Diet.1     76
10 Diet.2     30
11 Diet.3     81
12 Diet.4    169
13 Diet.1     38
14 Diet.2     89
15 Diet.3     95
16 Diet.4     85
17 Diet.1     94
18 Diet.2     63
19 Diet.3     92
20 Diet.4    154
```

ANOVA concepts

Terminology

- **Factor** (or treatment): the categorical variable of interest (here: diet)
- **Levels**: the categories within a factor (here: Diet 1, Diet 2, Diet 3, Diet 4)
- **Replicates**: the number of observations per level (here: $r = 5$)
- t = number of treatments; N = total observations = $r \times t = 20$

This is a **one-way** (or one-factor) ANOVA because there is only one factor (diet).

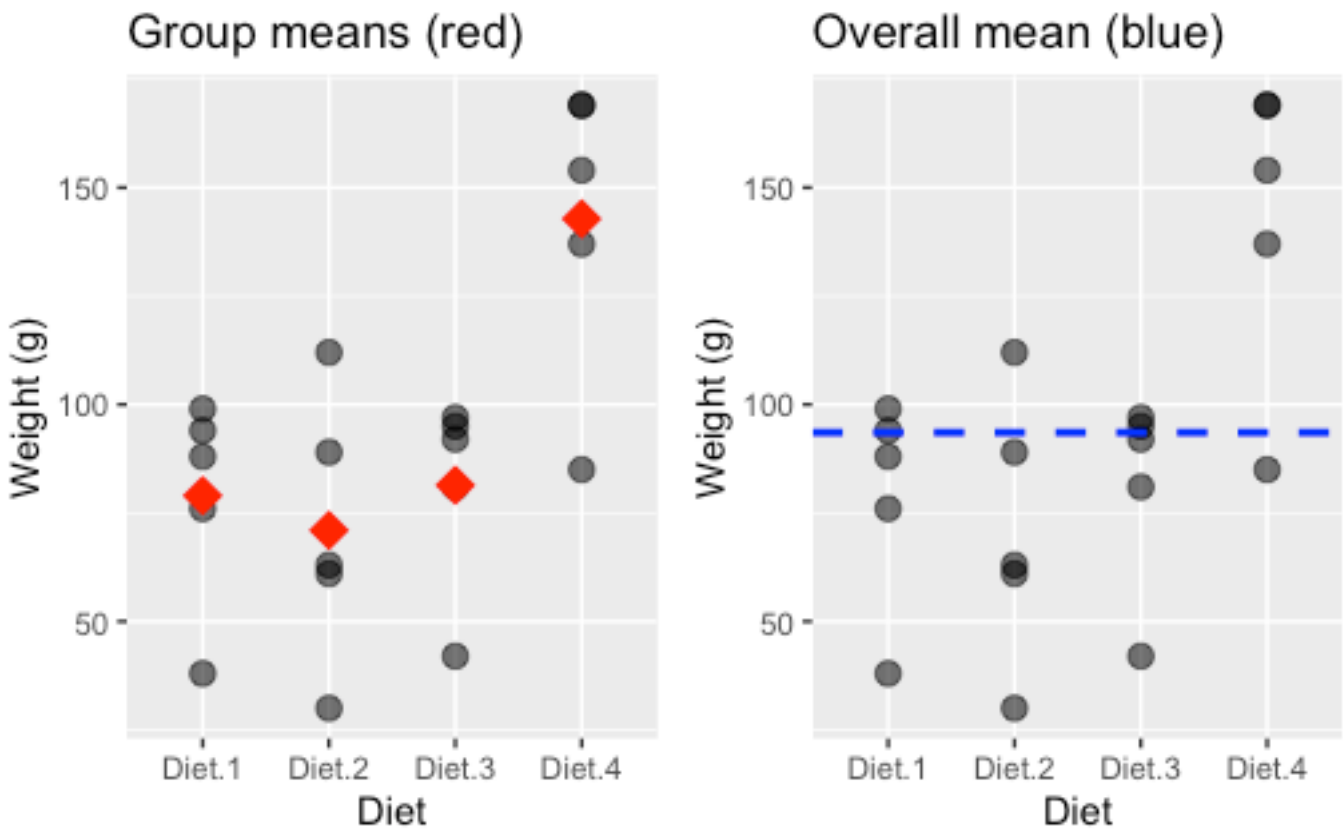
Which model best describes the data?

```
overall_mean ← mean(chicks$weight)
group_means ← chicks |>
  group_by(diet) |>
  summarise(mean_wt = mean(weight))

p1 ← ggplot(chicks, aes(diet, weight)) +
  geom_point(size = 3, alpha = 0.6) +
  geom_point(data = group_means, aes(diet, mean_wt),
    shape = 18, size = 5, colour = "red") +
  labs(title = "Group means (red)", x = "Diet", y = "Weight (g)")

p2 ← ggplot(chicks, aes(diet, weight)) +
  geom_point(size = 3, alpha = 0.6) +
  geom_hline(yintercept = overall_mean, colour = "blue",
    linewidth = 1, linetype = "dashed") +
  labs(title = "Overall mean (blue)", x = "Diet", y = "Weight (g)")

library(patchwork)
p1 + p2
```



Does the **group means model** (left) explain the data better than the **overall mean model** (right)?

Model equation

Same form as the t-test

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

- $i = 1, 2, ..., t$ (treatment); $j = 1, 2, ..., n_i$ (replicate)

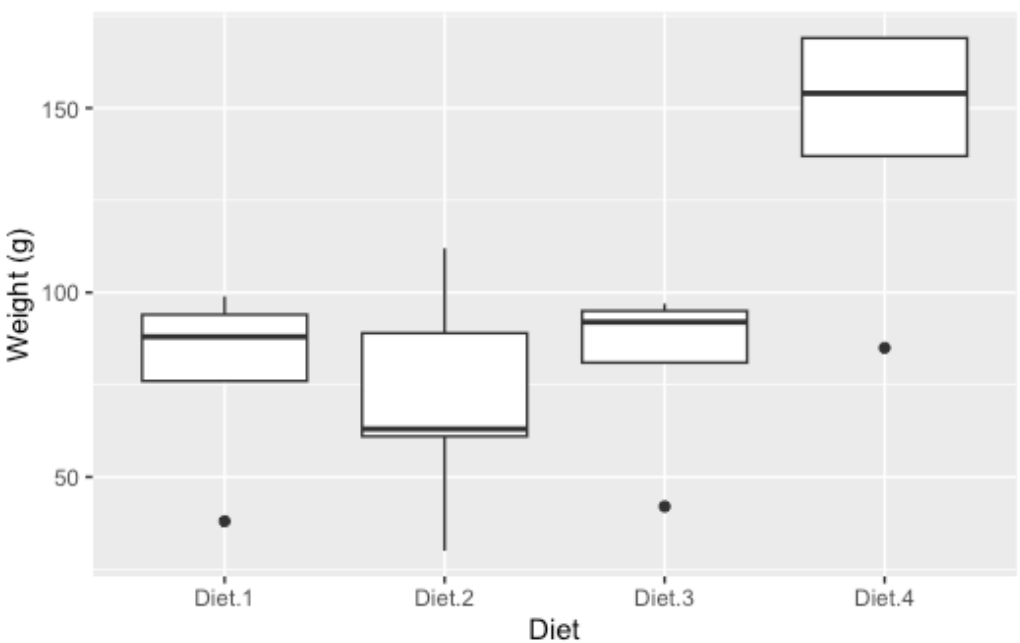
In the chick example:

- y_{ij} = observed weight gain for the jth chick on Diet i
- μ_i = mean weight gain for chicks on Diet i
- ε_{ij} = random error (residual)

Assumptions

Checking normality

```
ggplot(chicks, aes(diet, weight)) +
  geom_boxplot() +
  labs(x = "Diet", y = "Weight (g)")
```




```
shapiro.test(chicks$weight)
```

```
    Shapiro-Wilk normality test

data:  chicks$weight
W = 0.93272, p-value = 0.1742
```

$p > 0.05$: we can assume normality.

Checking equal variances

- $\sigma_1^2 = \sigma_2^2 = ... = \sigma_t^2$
- General guide: $\frac{\text{largest SD}}{\text{smallest SD}} < 2.0$
- Alternatively, use **Bartlett's test** of homogeneity of variance:

```
bartlett.test(weight ~ diet, data = chicks)
```

```
    Bartlett test of homogeneity of variances

data:  weight by diet
Bartlett's K-squared = 0.85164, df = 3, p-value = 0.8371
```

> ⚠ Warning
>
> Bartlett's test is **sensitive to non-normality**. Only use it when normality is reasonable.

Hypothesis test

Hypotheses

- **Null hypothesis**: $H_0 : \mu_1 = \mu_2 = ... = \mu_t$
- **Alternative hypothesis**: H_1: not all μ_i are equal

> ! Important
>
> ANOVA only tells us that **at least two** group means are different – not *which* ones.

The concept: partition variability

Partition the total variability into components:

$$SS_{\text{total}} = SS_{\text{treatment}} + SS_{\text{residual}}$$

- $SS_{\text{treatment}}$: variation **between** groups (due to diet)
- SS_{residual}: variation **within** groups (random error)

If the treatment has a real effect, $SS_{\text{treatment}}$ should be **large** relative to SS_{residual}.

Variance partitioning

Total sum of squares: SS_{total}

$$SS_{\text{total}} = \sum \left(y_{ij} - \bar{y}\right)^2$$

```
overall_mean ← mean(chicks$weight)
ss_total ← sum((chicks$weight - overall_mean)^2)
ss_total
```

```
[1] 29678.95
```

Treatment sum of squares: $SS_{\text{treatment}}$

$$SS_{\text{treatment}} = \sum n_i \times (\bar{y}_i - \bar{y})^2$$

```
group_means ← chicks |>
  group_by(diet) |>
  summarise(grp_mean = mean(weight))

ss_trt ← sum(5 * (group_means$grp_mean - overall_mean)^2)
ss_trt
```

```
[1] 16466.95
```

Residual sum of squares: SS_{residual}

$$SS_{\text{residual}} = \sum \left(y_{ij} - \bar{y}_i\right)^2$$

```
chicks_with_means ← chicks |>
  left_join(group_means, by = "diet")

ss_res ← sum((chicks_with_means$weight - chicks_with_means$grp_mean)^2)
ss_res
```

```
[1] 13212
```

Check: $SS_{\text{total}} = SS_{\text{treatment}} + SS_{\text{residual}}$

```
c(ss_total, ss_trt + ss_res)
```

```
[1] 29678.95 29678.95
```

ANOVA table

Structure

Source	df	SS	MS	F
Treatment	$t-1$	SS_{trt}	$SS_{\text{trt}}/(t-1)$	$MS_{\text{trt}}/MS_{\text{res}}$
Residual	$N-t$	SS_{res}	$SS_{\text{res}}/(N-t)$	
Total	$N-1$	SS_{total}		

- N = total observations, t = number of treatment levels
- **Mean Squares** (MS) standardise SS so they are comparable
- The larger the ratio $MS_{\text{trt}}/MS_{\text{res}}$, the stronger the treatment effect

Using `aov()` in R

```r
model ← aov(weight ~ diet, data = chicks)
summary(model)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
diet         3  16467    5489   6.647  0.004 **
Residuals   16  13212     826
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting the results

- $F = 6.65$ with $df = 3, 16$
- $p = 0.004$ which is < 0.05, so we **reject** H_0
- There are **significant differences** in mean weight gain amongst the 4 diets

How much variability is explained?

$$\frac{SS_{\text{treatment}}}{SS_{\text{total}}} = \frac{1.6467 \times 10^4}{2.9679 \times 10^4} = 55.5\%$$

The diets explain about 55% of the total variability in chick weight gain.

Post-hoc: which groups differ?

Confidence intervals for group means

We can examine 95% CIs for each treatment mean:

$$95\% \quad \text{CI} = \bar{y}_i \pm t_{N-t}^{0.025} \times SE(\bar{y}_i)$$

where $SE(\bar{y}_i) = \sqrt{MS_{\text{res}}/n_i}$

Using `emmeans`

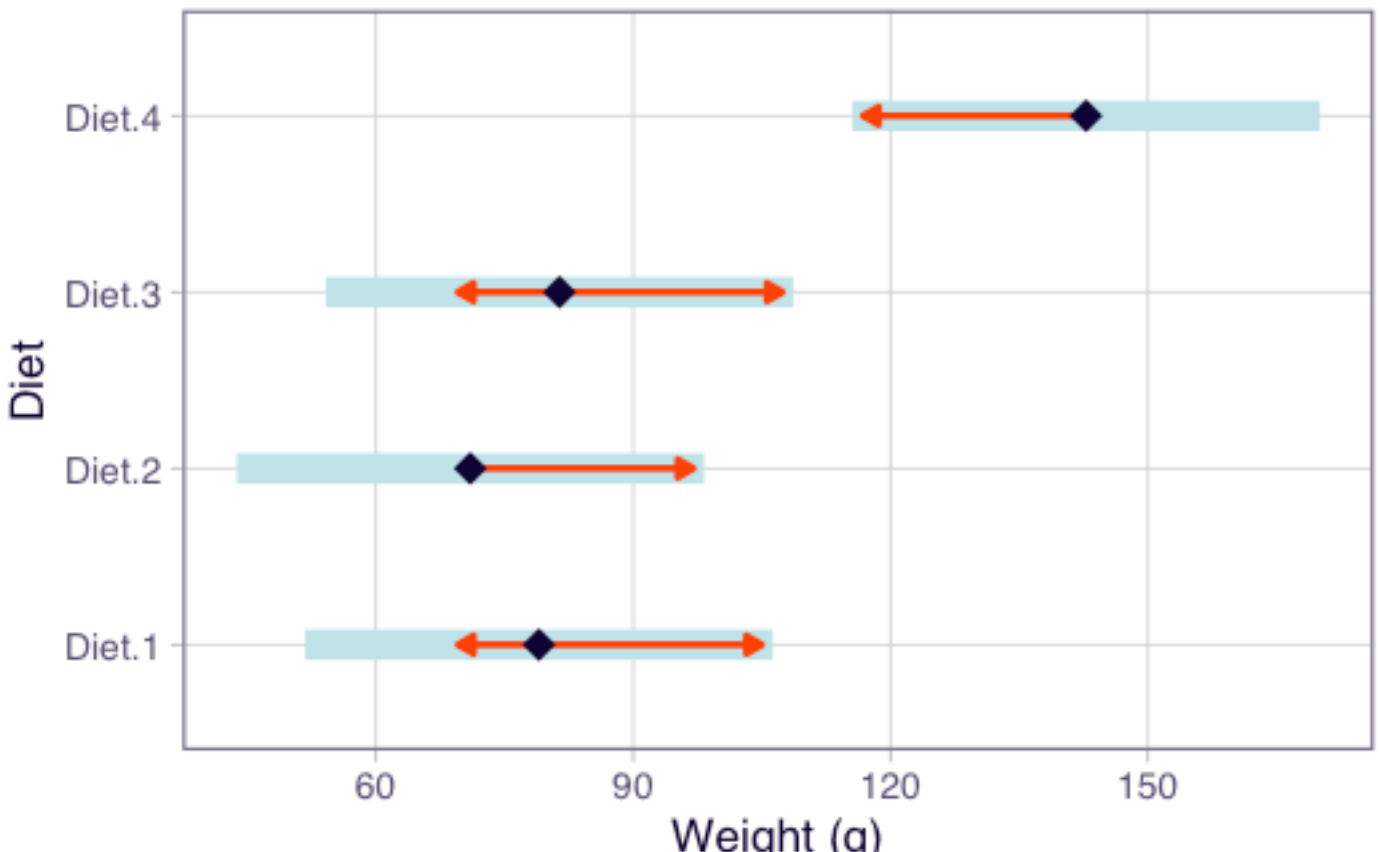
```
library(emmeans)
emm ← emmeans(model, "diet")
emm
```

```
 diet   emmean   SE df lower.CL upper.CL
 Diet.1   79.0 12.9 16     51.8    106.2
 Diet.2   71.0 12.9 16     43.8     98.2
 Diet.3   81.4 12.9 16     54.2    108.6
 Diet.4  142.8 12.9 16    115.6    170.0

Confidence level used: 0.95
```

Visualising group comparisons

```
plot(emm, comparisons = TRUE) +
  labs(x = "Weight (g)", y = "Diet")
```



Groups with **overlapping** arrows are **not** significantly different from each other.

Summary

t-test vs ANOVA

- **t-test**: compares means of **2** groups
- **ANOVA**: compares means of **2 or more** groups simultaneously
- ANOVA avoids the inflated Type I error rate from multiple t-tests
- The ANOVA F-test tells us *if* differences exist, but not *where*
- Post-hoc methods (e.g. `emmeans`) identify *which* pairs differ

Next week

- How to better identify which pairs are significantly different
- How to test and interpret model assumptions using **residual diagnostics**

References

- Quinn & Keough (2002), Chapter 7: Section 7.1
- Mead et al. (2002), Chapter 18: Sections 18.1–18.3

Thanks!

Questions?