# Topic 3: 1-way Analysis of Variance

## ENVX2001 – Applied Statistical Methods

Januar Harianto

## Outline

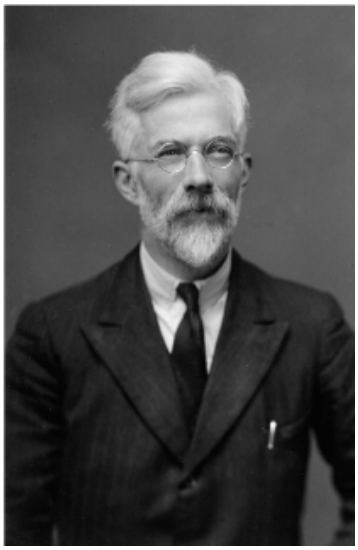
| Week | Lecturer | Lecture Topic | Assessment | Lab |
|---|---|---|---|---|
| **Part 1: Designed Studies** | | | | |
| 1 | Januar Harianto | Introduction: Surveys | | Lab 1 |
| 2 | Januar Harianto | Surveys: Sampling designs | | Lab 2 |
| **3** | **Januar Harianto** | **ANOVA I: One-way Analysis of Variance (ANOVA)** | | **Lab 3** |
| 4 | Aaron Greenville | ANOVA II: Introduction to experimental design | Report 1 | No Labs |
| 5 | Aaron Greenville | ANOVA III: ANOVA with blocking | | Lab 4 |
| 6 | Aaron Greenville | ANOVA IV: ANOVA with 2 or more factors | | Lab 5 |
| **Part 2: Finding Patterns in Data** | | | | |
| 7 | Liana Pozza | Regression I: Multiple linear regression | Report 2 | Lab 6 |
| 8 | Liana Pozza | Regression II: Variable selection | | Lab 7 |
| 9 | Liana Pozza | Regression III: Predictive modelling | | Lab 8 |
| 10 | Mathew Crowther | Multivariate analysis I: Principal component analysis (PCA) | | Lab 9 |
| 11 | Mathew Crowther | Multivariate analysis II: Clustering | | Lab 10 |
| 12 | Mathew Crowther | Multivariate analysis III: MDS and MANOVA | | Lab 11 |
| **Part 3: Revision** | | | | |
| 13 | TBA | Revision | Presentation | Lab 12 |

# Outline

## Two-sample t-test

William Gosset (1908)

## Analysis of Variance (ANOVA)



Ronald Fisher (1925)



# Learning outcomes

At the end of this topic students should be able to:

- Describe the concept of how the ANOVA is used to determine whether there is a statistically significant difference in the means of treatments;
- Demonstrate proficiency in the use of R (and interpretation of the output) for performing an analysis of variance (ANOVA) on experimental data with 1 treatment factor.

# Revisiting the t-test

## Example

- Weights of two breeds of cattle are to be compared
- Twelve cattle from Breed 1 were randomly sampled, and another 15 weights from Breed 2 were also recorded

| Breed1 | Breed2 |
|---|---|
| 187.6 | 148.1 |
| 180.3 | 146.2 |
| 198.6 | 152.8 |
| 190.7 | 135.3 |
| 196.3 | 151.2 |
| 203.8 | 146.3 |
| 190.2 | 163.5 |
| 201.0 | 146.6 |
| 194.7 | 162.4 |
| 221.1 | 140.2 |
| 186.7 | 159.4 |
| 203.1 | 181.8 |
| | 165.1 |
| | 165.0 |
| | 141.6 |



# Two-sample t-test

## Import data

```
cattle <- read_csv("assets/tables/cattle.csv") %>%
  pivot_longer(cols = starts_with("Breed"), names_to = "breed", values_to = "weight") %>%
  mutate(breed = as.factor(breed))
```

Tidying the data (FYI).

**Before**

```
## # A tibble: 15 x 2
##    Breed1 Breed2
##     <dbl>  <dbl>
##  1   188.   148.
##  2   180.   146.
##  3   199.   153.
##  4   191.   135.
##  5   196.   151.
##  6   204.   146.
```

**After**

```
## # A tibble: 30 x 2
##    breed  weight
##    <fct>   <dbl>
##  1 Breed1   188.
##  2 Breed2   148.
##  3 Breed1   180.
##  4 Breed2   146.
##  5 Breed1   199.
##  6 Breed2   153.
```

# Two-sample t-test

## Descriptive statistics (mean)

```
with(cattle, mean(weight[breed == "Breed1"],
  na.rm = TRUE))
```

```
## [1] 196.175
```

```
with(cattle, mean(weight[breed == "Breed2"],
  na.rm = TRUE))
```

```
## [1] 153.7
```

## Descriptive statistics (sd)

```
with(cattle, sd(weight[breed == "Breed1"],
  na.rm = TRUE))
```

```
## [1] 10.61604
```

```
with(cattle, sd(weight[breed == "Breed2"],
  na.rm = TRUE))
```

```
## [1] 12.30139
```



# Two-sample t-test

## Model assumptions: Equal variances

- $\sigma_1^2 \approx \sigma_2^2$
- General guide: $\frac{larger\ standard\ deviation}{smaller\ standard\ deviation} < 2.0$

```
12.30139/10.61604
```

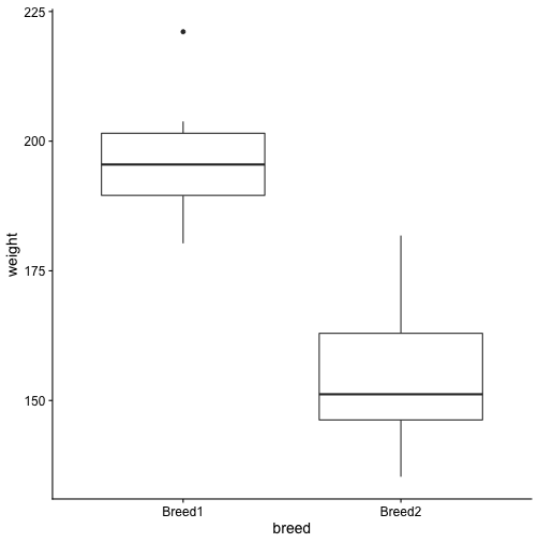
```
## [1] 1.158755
```

- Only difference between the two distributions is where the distribution located, otherwise the same

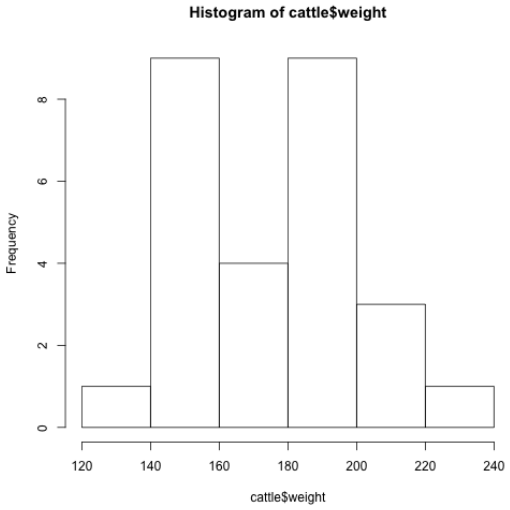
# Two-sample t-test

## Model assumptions: Normality

```
ggplot(cattle, aes(breed, weight)) +
  geom_boxplot() + cowplot::theme_cowplot()
```




```
hist(cattle$weight)
```






# Two-sample t-test

## Model assumptions: Normality

- $y_{i,j} \sim N(\mu_i, \sigma^2)$ or $\varepsilon_{i,j} \sim N(\mu_i, \sigma^2)$

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

```
##
##     Shapiro-Wilk normality test
##
## data:  cattle$weight
## W = 0.93704, p-value = 0.103
```

- If p > 0.05, the distribution of the `cattle` data is not significantly different from a normal distribution, *i.e.* we can assume normality.

# Two-sample t-test

## Model equation

- Observed data = Group Mean + Random Error (residuals)

$$y_{i,j} = \mu_i + \varepsilon_{i,j}$$

- $i = 1, 2\ (group)$; $j = 1, 2, \ldots, n_i\ (replicate)$

In cattle example:

- $\mu_1$ = mean body weight (kg) for cattle in Breed 1
- $\mu_2$ = mean body weight (kg) for cattle in Breed 2



# Two-sample t-test

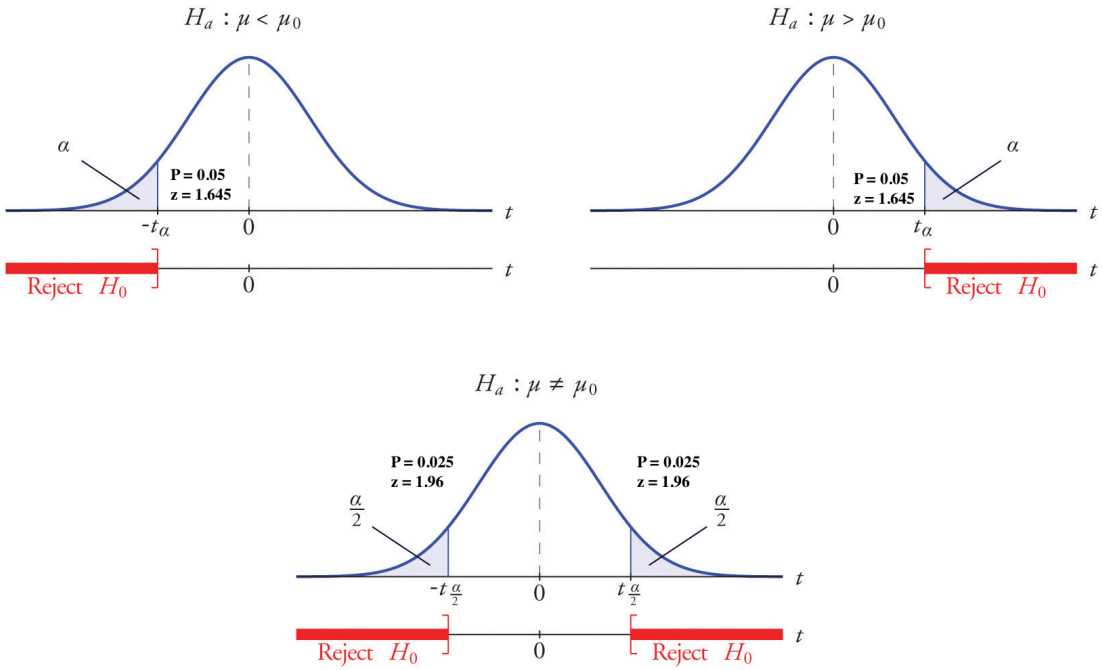
## T-test

```
with(cattle, t.test(weight[breed == "Breed1"], weight[breed == "Breed2"],
  var.equal = TRUE))
```

```
##
##     Two Sample t-test
##
## data:  weight[breed == "Breed1"] and weight[breed == "Breed2"]
## t = 9.4624, df = 25, p-value = 9.663e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  33.23011 51.71989
## sample estimates:
## mean of x mean of y
##   196.175   153.700
```

# Two-sample t-test

## Interpretation






# Two-sample t-test

## Hypothesis testing

- Null hypothesis:

$$H_0 : \mu_1 = \mu_2$$

- Alternate hypothesis:

$$H_1 : \mu_1 \neq \mu_2$$

- Test statistic:

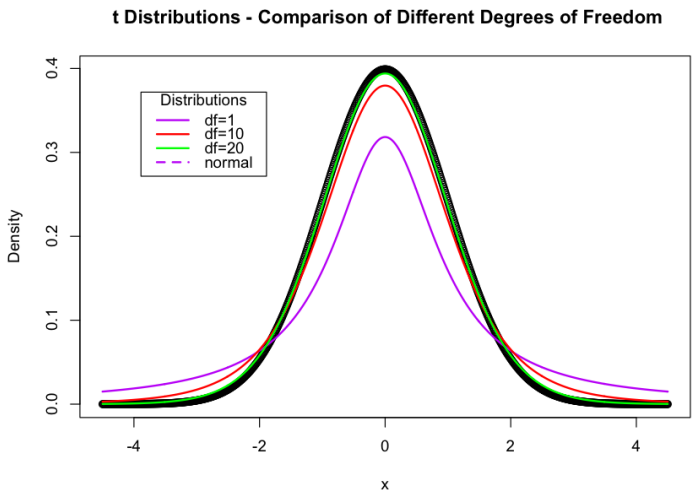
$$t = \frac{\bar{y}_2 - \bar{y}_1}{\sqrt{s^2(\frac{1}{n_2} + \frac{1}{n_1})}} = \frac{\bar{y}_2 - \bar{y}_1}{se(\bar{y}_2 - \bar{y}_1)} = \frac{\Delta\ in\ mean}{standard\ error\ of\ the\ \Delta\ in\ mean}$$

- Degrees of freedom:

$$n_1 + n_2 - 2$$

# T-distribution

- changes shape with datasets size – degrees of freedom (df)
- as n increases >> closer to normal distribution
- for standard normal distribution, 95% observations found in interval [-1.96,1.96]






# Analysis of Variance (ANOVA)

## Example

- A study was undertaken to compare the weight gains (g) of chicks on four different diets
- Twenty similar chicks were used in the study and were randomly allocated to one of the four groups
- The allocation was done in such a way as to have equal replication (five chicks) in each treatment group

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

# Should we use a t-test?

- We have 4 treatments
- We could do a series of t-tests for the 6 possible pairwise comparisons
  - 1 vs 2; 1 vs 3; 1 vs 4; 2 vs 3; 2 vs 4; 3 vs 4
- **Problem:** even if the true differences between treatment (population) means differ, each test has a 5% probability of incorrectly finding significant results
  - 6 tests, we have $0.95^6 = 0.735 = 73.5\%$ of getting all correct
  - 26.5% chance of getting at least 1 incorrect
- We need a method to test for the equality of treatments simultaneously
  - This avoids the problem of multiple comparisons



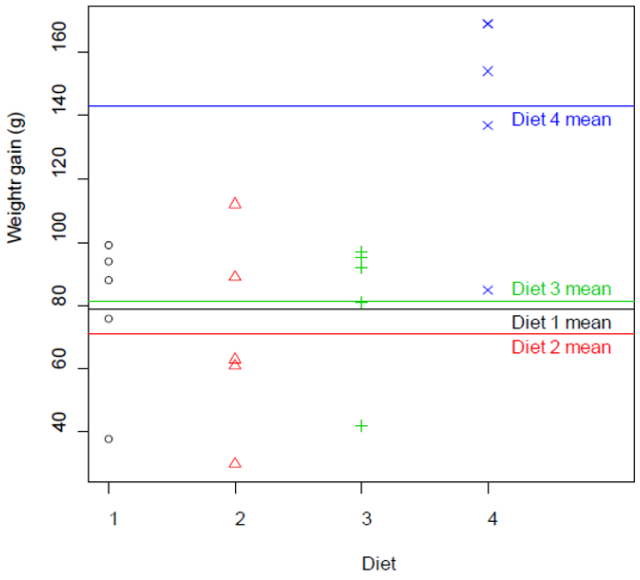
# ANOVA

- Differences between the 4 diets
  - treatment effect
- Differences within diets
  - due to background random environmental fluctuations, genetics, experimental error

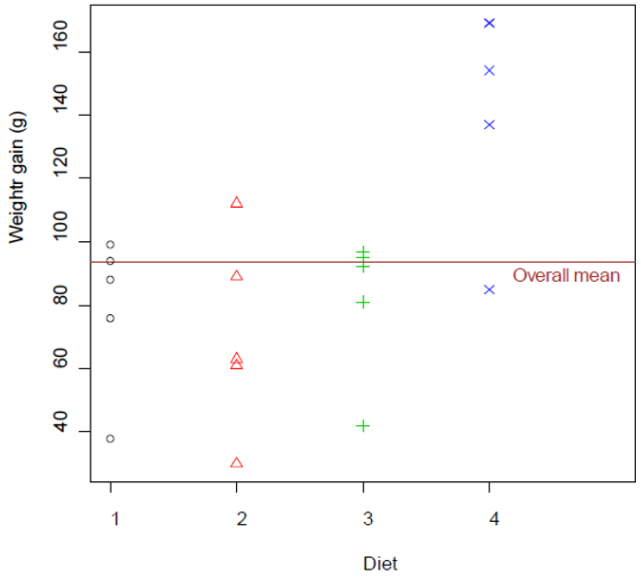
# ANOVA

Which model best describes data?

Group means shown separately




Overall mean






# ANOVA: Terminology

- Suppose in general that we have $t$ different treatments, and have drawn samples of size $n_1, n_2, \ldots, n_t$ from the 1st, 2nd, ..., $t^{\text{th}}$ population
- The total number of observations is $n_1, n_2, \ldots, n_t = N$. In the diets example, there are $t = 4$ treatments, with equal replication $(n_1 = n_2 = n_3 = n_4 = 5)$ with $N = 20$
- For equally replicated designs, we will use $r$ as the number of replicates per group (with $N = rt$)
- In the chick example, there is only **one factor** or treatment factor (diet)
- That factor has 4 levels (the 4 diet options).
- Hence the ANOVA conducted on these data is a **1-way (or 1-factor)** ANOVA

# ANOVA

## Model equation

- Observed data = Group Mean + Random Error (residuals)

$$y_{i,j} = \mu_i + \varepsilon_{i,j}$$

- $i = 1, 2\ (group)$; $j = 1, 2, \ldots, n_i\ (replicate)$

In cattle example:

- $y_{i,j}$ = observed weight gain for $j^{\text{th}}$ chicken on Diet $i$;
- $\mu_i$ = mean weight gain for chicks on Diet, $i$.



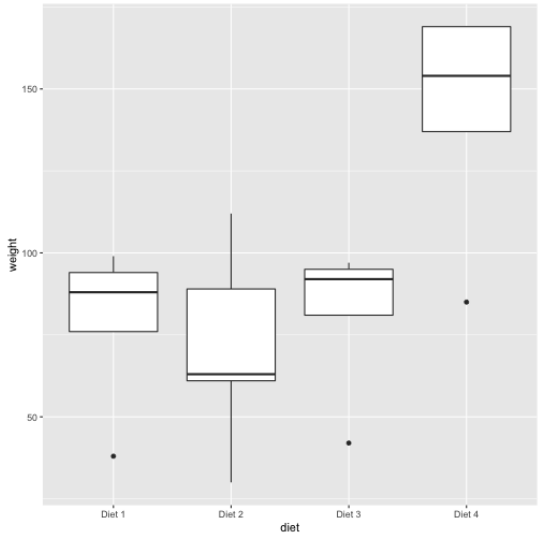
# ANOVA

## Model assumptions: Normality

- $y_{i,j} \sim N(\mu_i, \sigma^2)$ or $\varepsilon_{i,j} \sim N(\mu_i, \sigma^2)$
- Check this assumptions using a histogram, boxplot for each group
- Or examine residuals (Topic 4)

```
chicks <- read_csv("assets/tables/chicks.csv") %>%
  pivot_longer(cols = starts_with("Diet"),
    names_to = "diet",
    values_to = "weight") %>%
  mutate(diet = as.factor(diet))
```

```
ggplot(chicks, aes(diet, weight)) +
  geom_boxplot()
```



# ANOVA

## Model assumptions: Normality

```
hist(chicks$weight)
```

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

```
##
##     Shapiro-Wilk normality test
##
## data:  chicks$weight
## W = 0.93272, p-value = 0.1742
```

- If p > 0.05, the distribution of the `cattle` data is not significantly different from a normal distribution, *i.e.* we can assume normality.



# ANOVA

## Model assumptions: Equal variances

- $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$
- General guide: $\frac{largest\ standard\ deviation}{smallest\ standard\ deviation} < 2.0$
- Alternatively: perform a formal hypothesis test, e.g. 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
```

- **Bartlett's test is not reliable if data is not normal**

# ANOVA

## Hypothesis testing

- Null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_t$$

- Alternate hypothesis:

$$H_1 : not\ all\ \mu_i\ are\ equal$$

- **Important:** only tells us that at least 2 treatment (group) means are different



# ANOVA

## Concept

- Partition the variability of the data into components:
  - Differences due to treatments
  - Residual variation

Total Sums-of-Square (SS) = Treatment SS + Residual SS

# ANOVA

## Table

- Partition the variability of the data into components:
  - Differences due to treatments
  - Residual variation

| Source | df | Sums-of-square (SS) | Mean-square (MS) | F statistic |
|---|---|---|---|---|
| Treatment | $t-1$ | $SS_{trt}$ | $SS_{trt}/df_{trt}$ | $MS_{trt}/Ms_{res}$ |
| Residual | $N-t$ | $SS_{res}$ | $SS_{res}/df_{res}$ | |
| Total | $N-1$ | $SS_{tot}$ | $SS_{tot}/df_{tot}$ | |

$N$ = number of observations, $t$ = treatment levels



# ANOVA

## Calculations

Total sum-of-squares, $SS_{tot}$

$$SS_{tot} = \sum (data - overall\ mean)^2$$

```
library(dplyr) # load package
overall_mean <- mean(chicks$weight) # calculate overall mean
tot_ss <- mutate(chicks, sst = (weight - overall_mean)^2) # calculate (data - overall mean)^2
sum(tot_ss$sst) # sum for total ss
```

```
## [1] 29678.95
```

# ANOVA

## Calculations

### Treatment sum-of-squares, $SS_{trt}$

$$SS_{trt} = \sum n_i \times (group\ mean - overall\ mean)^2$$

```
# using dplyr again
chicks <- group_by(chicks, diet) # group by diet, so that we can summarise by group
grp <- summarise(chicks, grp_mean = mean(weight)) # summarise by group
trt_ss <- mutate(grp, sstr = (grp_mean - overall_mean)^2) # calculate (grp mean - overall mean)^2
5* sum(trt_ss$sstr) # sum for treatment ss
```

```
## [1] 16466.95
```



# ANOVA

## Calculations

### Residual sum-of-squares, $SS_{res}$

$$SS_{res} = \sum (data - group\ mean)^2$$

```
merged <- merge(chicks, grp)
res_ss <- mutate(merged, ssr = (weight - grp_mean)^2)
sum(res_ss$ssr)
```

```
## [1] 13212
```

# ANOVA

## Table

| Source | df | Sums-of-square (SS) | Mean-square (MS) | F statistic |
|---|---|---|---|---|
| Treatment | $t-1$ | $SS_{trt}$ | $SS_{trt}/df_{trt}$ | $MS_{trt}/Ms_{res}$ |
| Residual | $N-t$ | $SS_{res}$ | $SS_{res}/df_{res}$ | |
| Total | $N-1$ | $SS_{tot}$ | $SS_{tot}/df_{tot}$ | |

## Chick weight example

| Source | df | Sums-of-square (SS) | Mean-square (MS) | F statistic |
|---|---|---|---|---|
| Treatment | 3 | 16467 | 5489 | 6.65 |
| Residual | 16 | 13212 | 826 | |
| Total | 19 | 29679 | 1562 | |



# ANOVA

## 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
```

## Table

| Source | df | Sums-of-square (SS) | Mean-square (MS) | F statistic |
|---|---|---|---|---|
| Treatment | 3 | 16467 | 5489 | 6.65 |
| Residual | 16 | 13212 | 826 | |
| Total | 19 | 29679 | 1562 | |

# ANOVA

- Test statistic:

$$F = \frac{treatment\ MS}{residual\ MS}; df = t - 1,\ N - 1$$

- the **residual MS** is an estimate of $\sigma^2$, so $s = \sqrt{residual\ MS} = \sqrt{826} = 28.7$
- $s$ is the pooled standard deviation from pooling $t$ = 4 groups
- treatment MS is also an estimate of $\sigma^2$ (if the null hypothesis is true)



# ANOVA

- Test statistic:

$$F = \frac{treatment\ MS}{residual\ MS}; df = t - 1,\ N - t$$

- If the null hypothesis is true, the observed F statistic (variance ratio) will have a value around 1; large F values indicate the null hypothesis is false
- Hypothesis test: Compare observed F statistic with F distribution with $t - 1$ and $N - t$ degrees of freedom (d.f.), e.g. $Ft - 1,\ N - t$ or $F_{treat\ d.f., residual\ d.f.}$

# ANOVA

- Our example: $F = \frac{5489}{826} = 6.65$ with $d.f. = 3,\ 16$
- Probability of obtaining the observed test statistics or larger, $P = P(F_{3,16} > 6.65 = 0.04)$
- Since p-value is small ($<0.05$) we *reject* the null hypothesis
  - there are *significant differences* in mean weight gain amongst the 4 diets
- Proportion of variability explained by diets:

$$\frac{Treatment\ SS}{TotalSS} = 16647 \div 29679 = 0.55\ (55\%)$$



# ANOVA

Which pairs of groups means are (statistically) different?

- We could look at the 95% CI for each mean and see which overlap.

$$95\%CI = \bar{y} \pm t^{0.052}_{residual\ d.f.} \times se(\bar{y})$$

where

$$se(\bar{y}) = \sqrt{\frac{residual\ MS}{n_i}}$$

and $n_i$ is the number of replicates in treatment, $i$

- e.g. for `Diet 4`:

$$95\%CI = 142.8 \pm 2.12 \times \sqrt{\frac{826}{5}}$$

```
lower <- 142.8 - 2.12 * sqrt(826/5)
upper <- 142.8 + 2.12 * sqrt(826/5)
```

$95\%CI$ = [115.6, 170]

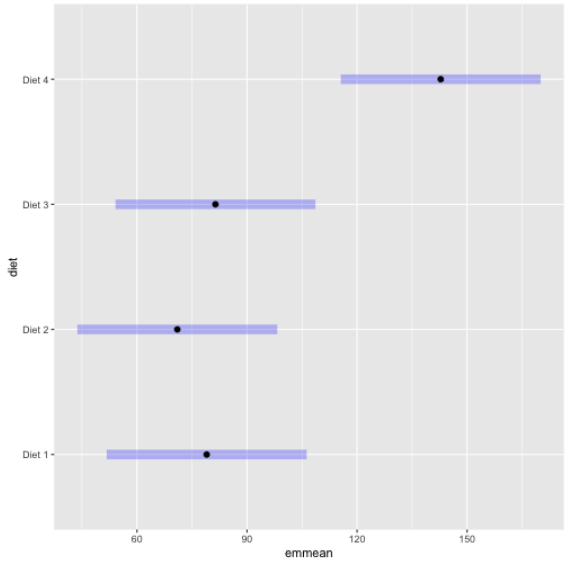
# ANOVA

Which pairs of groups means are (statistically) different?

```
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
```

```
plot(emm)
```






# Summary

- 2-sample t-tests are limited to situation when we have experiments with only 2 levels
- The ANOVA allows us to analyse experiments with 2 or more treatment levels
- It can be generalised to analyse any experiment, e.g. more than 1 treatment factors
- The ANOVA table helps us determine whether there is a significant different between at least one pair of treatment means

## Next week

- How to (better) identify which pair(s) are significant different
- How to test the model assumptions

# Thanks!

Slides created via the R package **xaringan**.



## Readings

- Quinn & Keough (2002)
  - Chapter 7: Section 7.1
- Mead et al. (2002)
  - Chapter 18: Sections 18.1-18.3 (most is for finite populations but useful for conceptual understanding)