

Lecture 19: ANOVA Part I

Chapter 5.5

1 / 25

Previously: Conditions for Using t Distribution

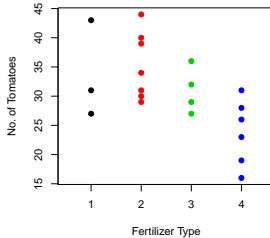
We use the t distribution when you have

- ▶ n is small.
- ▶ Independence: $n \leq 10\%$ rule
- ▶ Observations come from a nearly normal distribution:
 - ▶ Look at a histogram of the data (difficult when n is small)
 - ▶ Consider whether any previous experiences alert us that the data may be normal

2 / 25

Analysis of Variance (ANOVA)

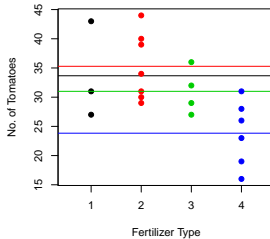
A farmer has the choice of four tomato fertilizers and wants to compare their performance in terms of crop yield.



3 / 25

Analysis of Variance (ANOVA)

A farmer has the choice of four tomato fertilizers and wants to compare their performance in terms of crop yield.



4 / 25

Analysis of Variance (ANOVA)

We have $k = 4$ groups AKA **levels of a factor**: the 4 types of fertilizer.

- ▶ n_i plants assigned to each of the $k = 4$ fertilizers:

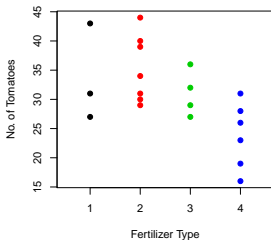
n_1	n_2	n_3	n_4	total n
3	7	4	6	20

- ▶ Count the number of tomatoes on each plant

5 / 25

Tomato Fertilizer

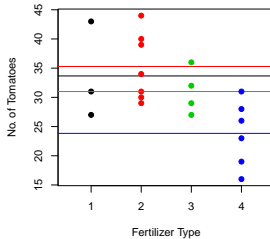
We observe the following, where each point is one tomato plant.



6 / 25

Tomato Fertilizer

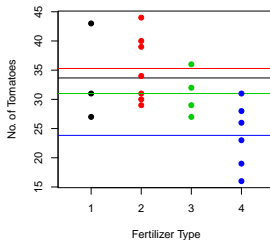
We observe the following, where each point is one tomato plant.
Plot the sample mean of each level.



6 / 25

Tomato Fertilizer

We observe the following, where each point is one tomato plant.
Plot the sample mean of each level. Question: are the mean tomato yields different?



6 / 25

Analysis of Variance

Say we have k groups and want to compare the k means:

$$\mu_1, \mu_2, \dots, \mu_k$$

We could do $\binom{k}{2}$ individual two-sample tests.

Ex. for groups 1 & 2:

$$\begin{array}{ll} H_0 : & \mu_1 = \mu_2 \\ \text{vs. } H_a : & \mu_1 \neq \mu_2 \end{array}$$

Analysis of Variance

Or we do a single overall test via Analysis of Variance ANOVA:

The hypothesis test is:

$$\begin{array}{ll} H_0 : & \mu_1 = \mu_2 = \dots = \mu_k \\ \text{vs. } H_a : & \text{at least one of the } \mu_i\text{'s are different} \end{array}$$

Analysis of Variance

ANOVA asks: where is the overall variability of the observations originate from?

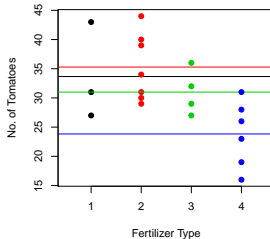
The **test statistic** used to compute a p -value is now the **F-statistic**:

$$F = \frac{\text{measure of between-group variability}}{\text{measure of within-group variability}}$$

9 / 25

Tomato Fertilizer Example

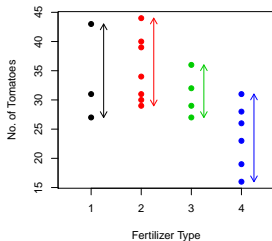
Numerator: the **between-group variation** refers to the variability **between** the levels (the 4 horizontal lines):



10 / 25

Tomato Fertilizer Example

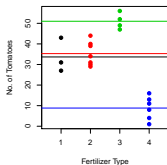
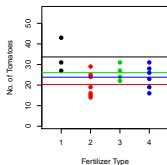
Denominator: the **within-group variation** refers to the variability **within** each level (the 4 vertical arrows):



11 / 25

Tomato Fertilizer Example

Now compare the following two plots. Which has “more different” means?



12 / 25

Tomato Fertilizer Example

- ▶ They have the **same within-group variability**. Call this value W
- ▶ The right plot has **higher between group variability** b/c the 4 means are more different. Call these values B_{left} and B_{right} with $B_{left} < B_{right}$
- ▶ Recall $F = \frac{\text{measure of between-group variability}}{\text{measure of within-group variability}}$
- ▶ Since $\frac{B_{left}}{W} < \frac{B_{right}}{W}$, thus $F_{left} < F_{right}$ The right plot as a larger F -statistic

13 / 25

F Distributions

Assuming H_0 is true (that $\mu_1 = \mu_2 = \dots = \mu_k$), the F -statistic

$$F = \frac{\text{measure of between-group variability}}{\text{measure of within-group variability}}$$

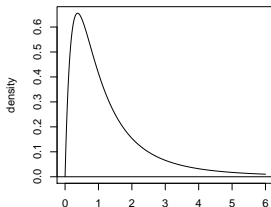
follows the F distribution with $df_1 = k - 1$ and $df_2 = n - k$ degrees of freedom where

- ▶ n = total number of observations
- ▶ k = number of groups

14 / 25

F Distributions

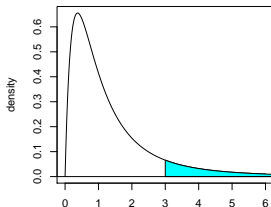
For $df_1 = 4$ and $df_2 = 6$, the F distribution looks like:



15 / 25

F Distributions

p -values are computed where "more extreme" means **larger**. Say the $F = 3$, the p -value is the **area to the right of 3** and is computed in R: `pf(3,df1=4,df2=6,lower.tail=FALSE)`



16 / 25

Conducting An F -Test

The results are typically summarized in an ANOVA table:

Source of Variation	df	SS	MS	F	p -value
Between groups	$k - 1$	$SSTr$	$MSTr = \frac{SSTr}{k-1}$	$\frac{MSTr}{MSE}$	p
Within groups	$n - k$	SSE	$MSE = \frac{SSE}{n-k}$		
Total	$n - 1$	SST			

17 / 25

Conditions

1. The observations have to be independent. 10% rule.
2. Trade off of n and normality of observations within each group.
3. Each of the groups has constant variance $\sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$.
Check via:
 - ▶ boxplots
 - ▶ comparing the sample standard deviations s_1, \dots, s_k

18 / 25

Discussion of Quiz

Question 1: Why did $\frac{1}{20}$ studies yield a positive/significant result
i.e. that there is a link between jelly beans and acne?

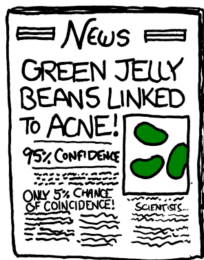
Not that the p-value is 0.05, rather that $\alpha = 0.05$:

- ▶ significance level AKA
- ▶ type I error rate AKA
- ▶ false positive rate

i.e. we expect 1 out of 20 results to be significant even if there is no effect.

19 / 25

Publication Bias



20 / 25

Publication Bias

Publication bias: people only highlight significant/positive results.
From Wikipedia: "Publication bias occurs when the publication of research results **depends on their nature and direction.**"

To counter this, some prominent medical journals including

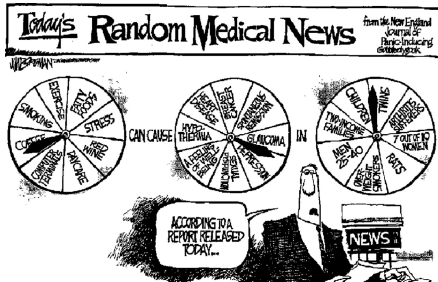
- ▶ New England Journal of Medicine
- ▶ The Lancet
- ▶ Journal of the American Medical Association

require registration of a trial **before** it starts so that unfavorable results are not withheld from publication.

Journal of **Negative Results:** <http://www.jnrnm.com/>

21 / 25

Publication Bias



From: Sterne JA, Davey Smith G (2001) Sifting the evidence - What's wrong with significance tests. BMJ 322: 226231.

22 / 25

What α to Use?

Should I use $\alpha = 0.05$ as my significance level? Before using it, put some thought into the balance between:

- ▶ **Type I errors.** Setting a smaller α yields a more **conservative** procedure: all things being equal, you will reject H_0 less often.
- ▶ **Type II errors.** Setting a bigger α yields a more **liberal** procedure: all things being equal, you will reject H_0 more often.

23 / 25

Multiple Testing

A related issue is the statistical concept of **multiple testing**.

Say we are conducting many experiments, and H_0 is true for all of them.

If you repeat experiments many times, you're bound to get a significant result eventually just by **chance alone**.

24 / 25

Multiple Testing

What do people do? Make the α stricter! i.e.

- ▶ make the α smaller
- ▶ i.e. less chance the p-value is smaller than α
- ▶ i.e. less chance of incorrectly rejecting H_0 when it is true

Use the **Bonferroni correction** to α : If you are conducting n tests, use $\alpha^* = \frac{\alpha}{n}$. You'll study its properties in HW8.