Lecture 19: ANOVA Part I

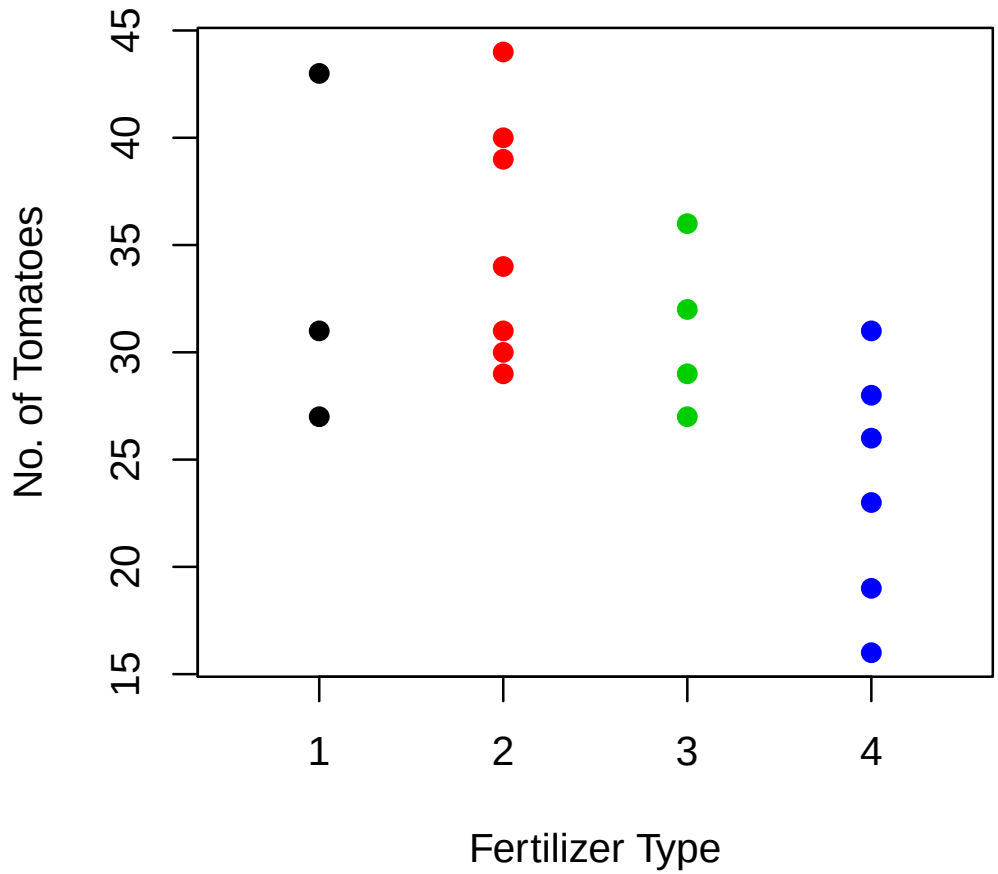
Chapter 5.5

Analysis of Variance (ANOVA)

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

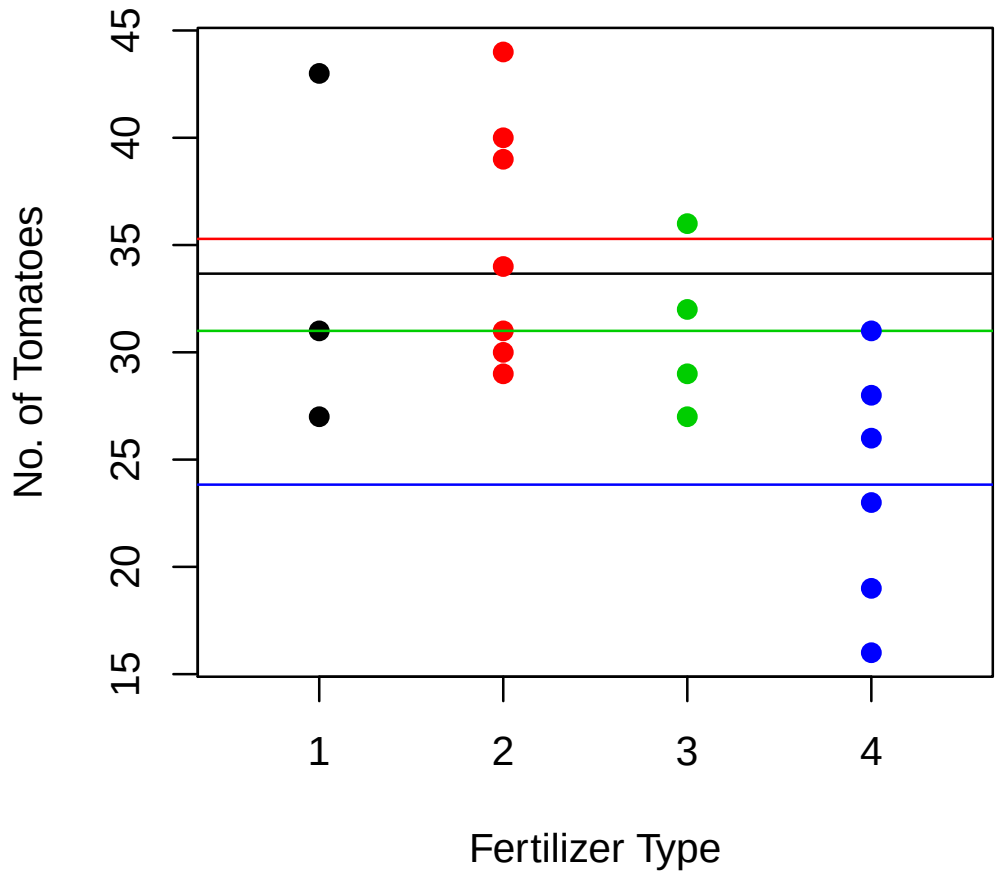
Analysis of Variance (ANOVA)

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



Analysis of Variance (ANOVA)

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



Analysis of Variance (ANOVA)

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

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

Analysis of Variance (ANOVA)

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

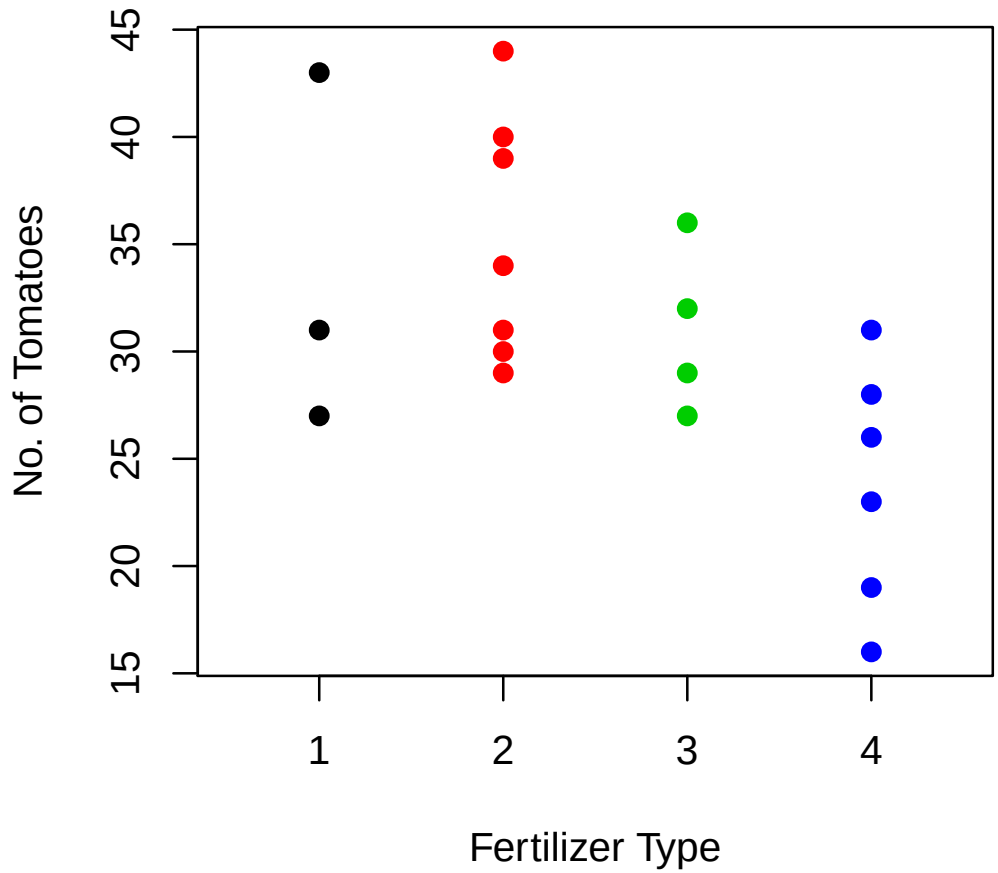
- n_i plants assigned to each of the $k = 4$ fertilizers:

n_1	n_2	n_3	n_4	total n
3	7	4	6	20

- Count the number of tomatoes on each plant

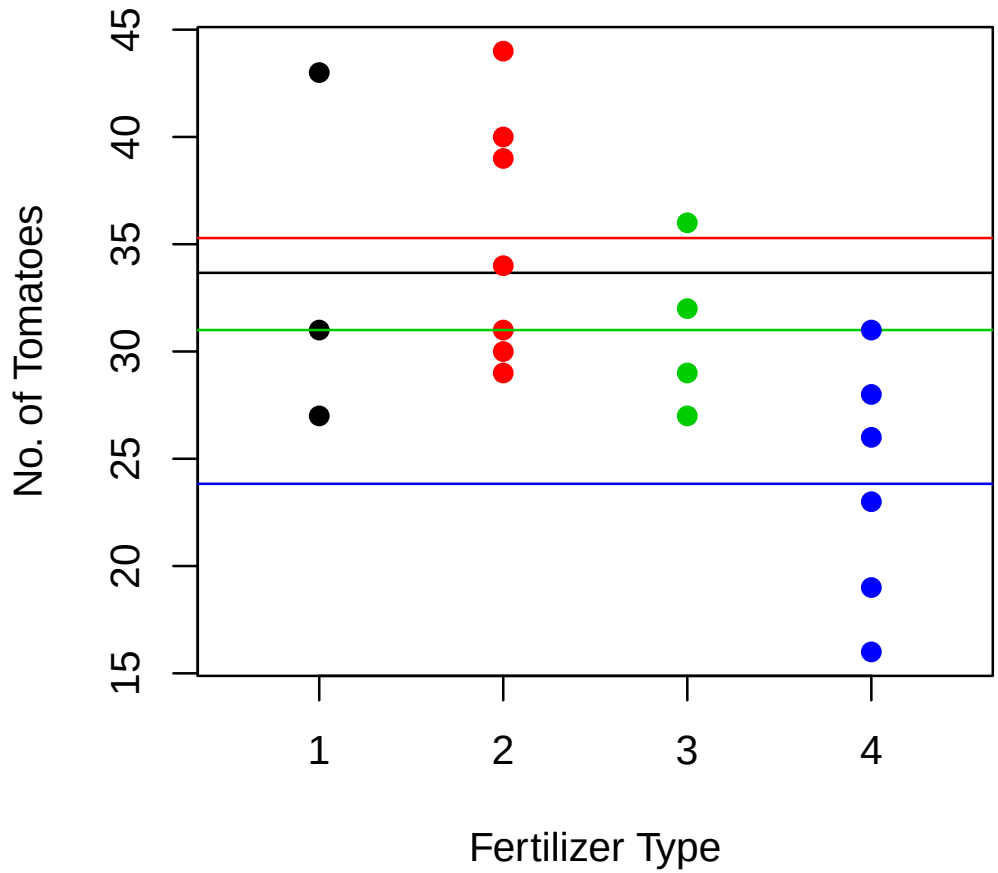
Tomato Fertilizer

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



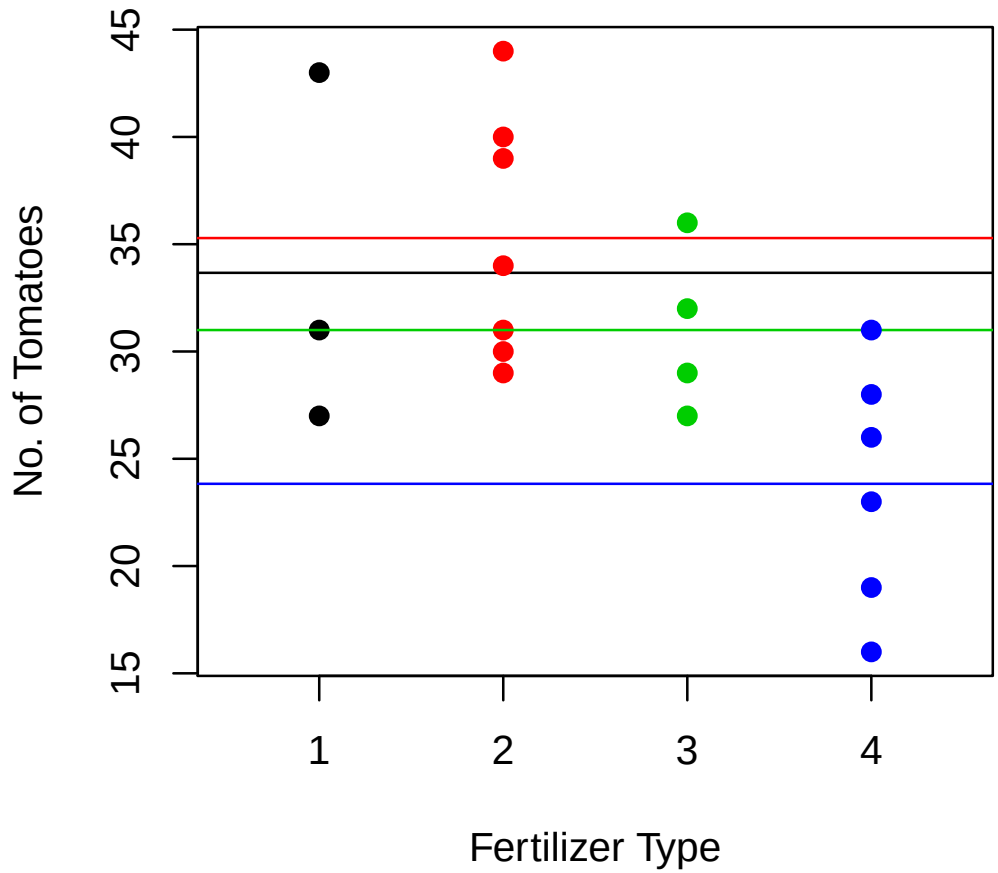
Tomato Fertilizer

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



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?



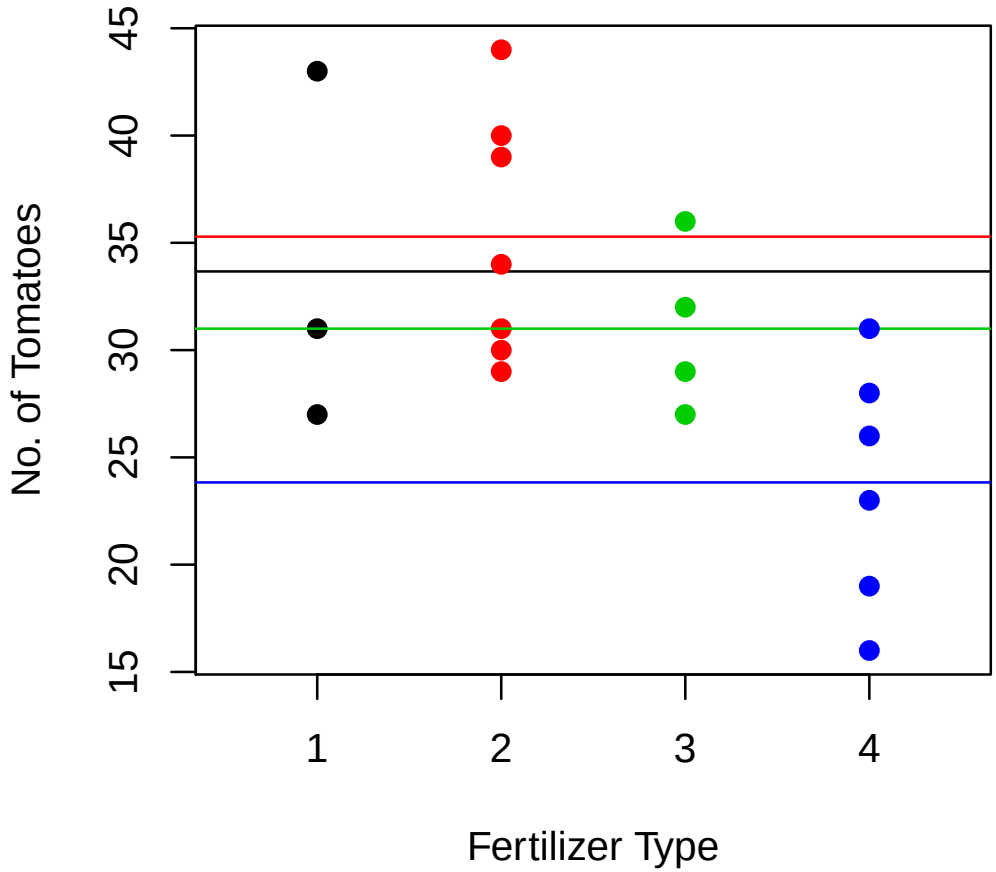
Analysis of Variance

Analysis of Variance

Analysis of Variance

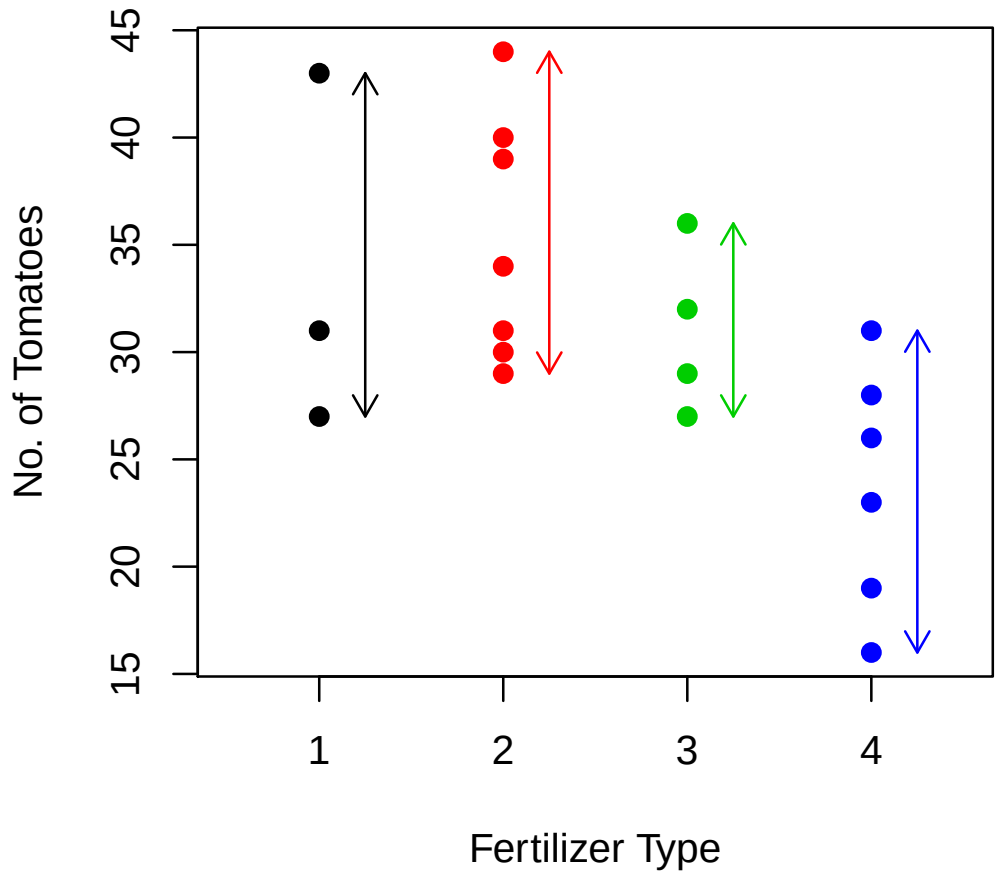
Tomato Fertilizer Example

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



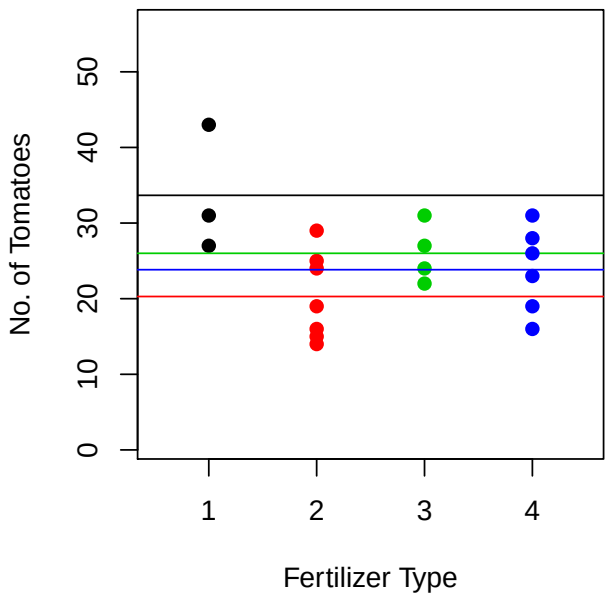
Tomato Fertilizer Example

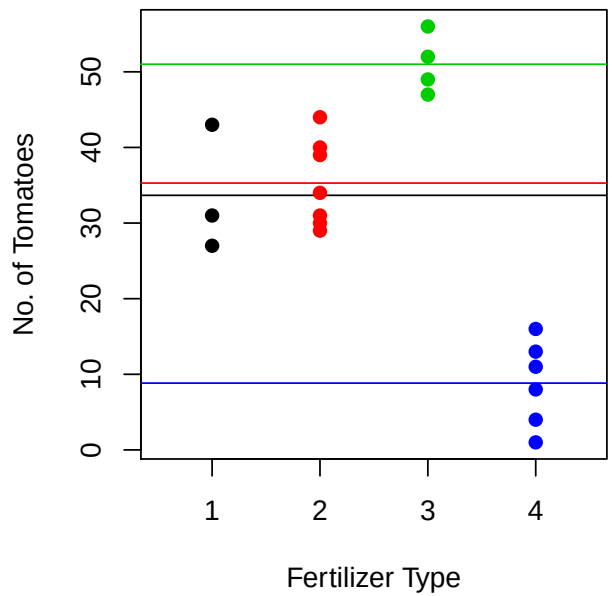
Denominator: the within-group variation refers to the variability within each level (the 4 vertical arrows):



Tomato Fertilizer Example

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



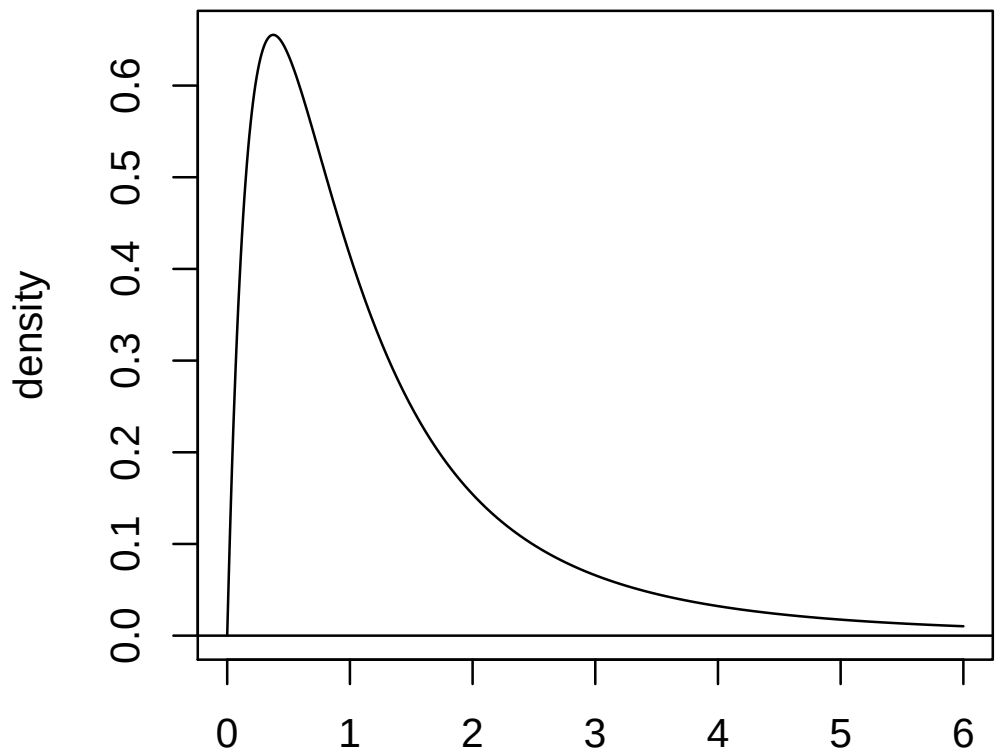


Tomato Fertilizer Example

F Distributions

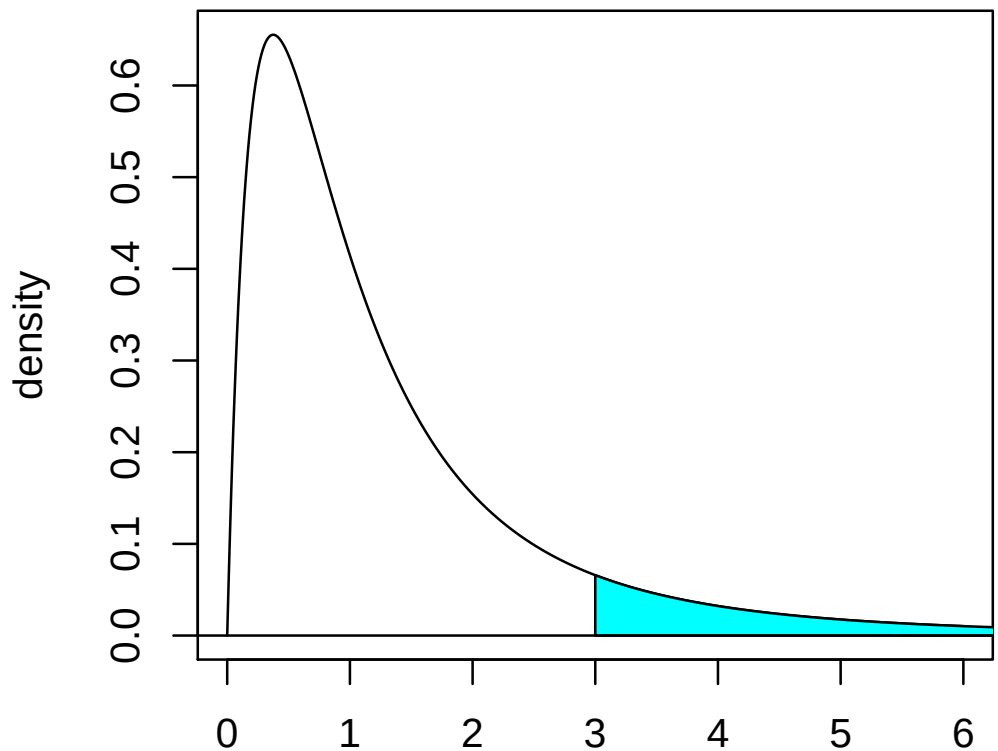
F Distributions

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



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)`



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			

Conditions