Sections 11.9 and 11.10

11.9: Introduction: ANOVA for Means

Analysis of Variance

ANOVA for Means

- In the **Light at Night** dataset, in a previous analysis we found that there was significant evidence for a difference in BodyMass gain for rats who were exposed to dim light at night and those who were not.
- That data was for 8 weeks and only compared two treatments.

ANOVA for Means: Light at Night

- In a similar study, rats were exposed to three levels of light at night over 4 weeks, and BMG (body mass gain) was measured. (Dataset: LightAtNight4)
- We want to test the claim that the THREE means are the same versus the claim that at least pair of the means differs.

Hypotheses

$$H_0 : \mu_{LD} = \mu_{DM} = \mu_{LL}$$

$$H_A : \text{At least one pair of means are different}$$

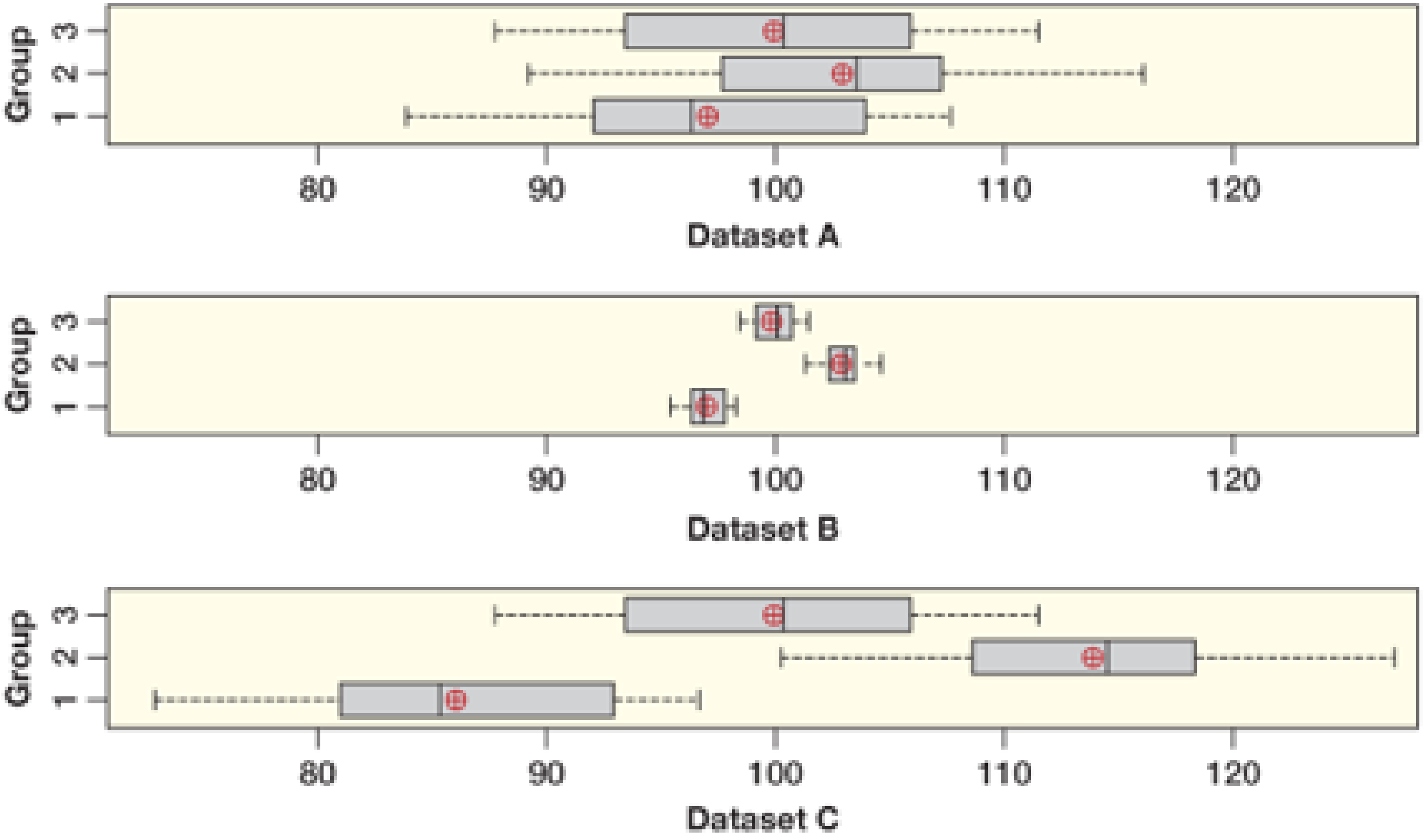
Descriptive Statistics

Let's think about how we might usefully visualize data for such a study.

- On the next slide are possible comparative boxplots of hypothetical data on three quantitative variables.
- Note that the center line in each here is at the median, not the mean. While these aren't exactly what we might want to look at, they are useful to see the main idea.
- Which of these (A, B, C) gives stronger evidence that there is a difference in the means among the three quantitative variables? Discuss.

Reading comparative boxplots

Image: From Lock, et. al., *Statistics: Unlocking the Power of Data*, Wiley, 2nd ed. p.



Descriptive Statistics

- Which of those gave the strongest evidence that the means are different?
  Answer: Dataset B (Middle)
- Which is next?
  Answer: Dataset C (Bottom)
- Which is the least strong evidence that the means are different?
  Answer: Dataset A (Top)

Why?

- Dataset B: (Middle graphs)
  The entire datasets barely overlap.
- Which is next? Dataset C
  The datasets clearly overlap, but
  the middle 50% doesn’t overlap.
- Which is least strong evidence that the means are different? Dataset A.
  The datasets substantially overlap.

Data description for our problem

`LightatNight4Weeks` *Light at Night for Mice - After 4 Weeks*

Description

Data from an experiment with mice having different nighttime light conditions

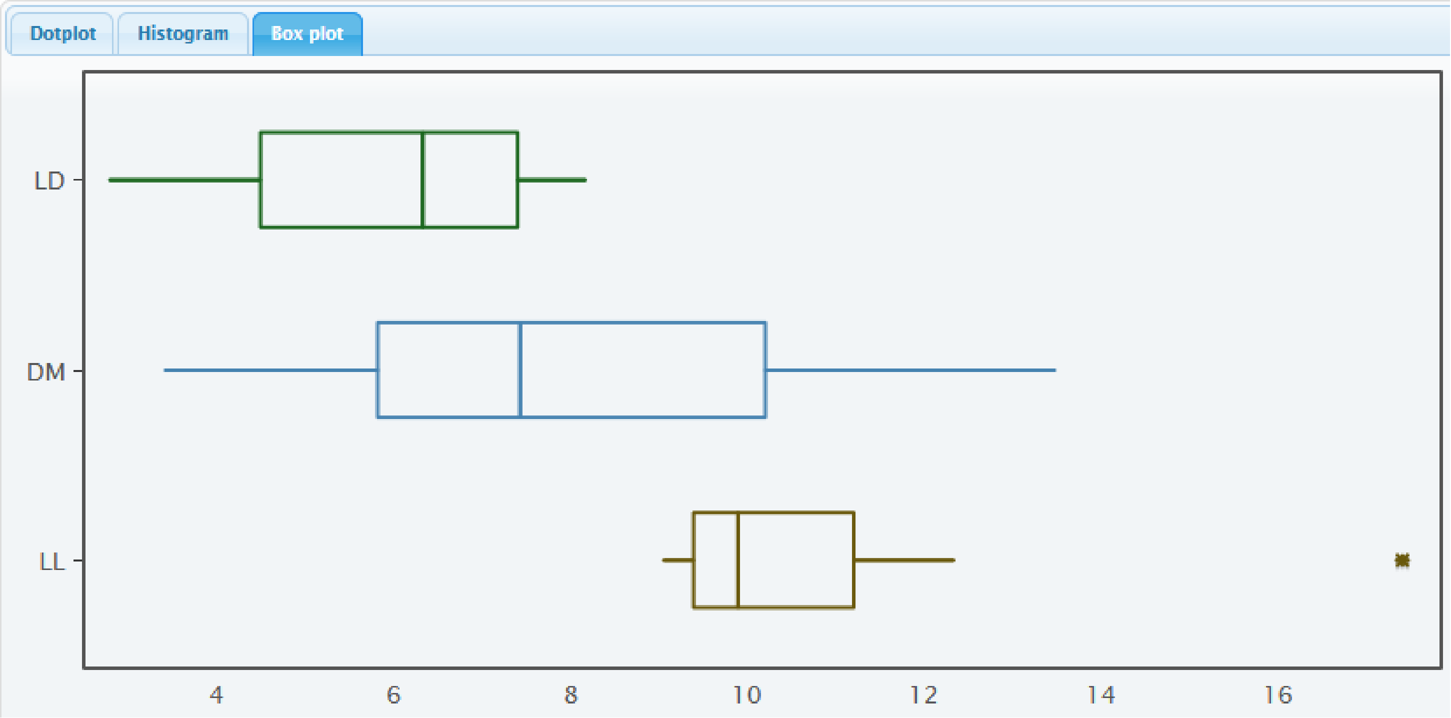
Format

A dataset with 27 observations on the following 9 variables.

`Light`	`DM`=dim light at night, `LD`=dark at night, or `LL`=bright light at night
`BMGain`	Body mass gain (in grams over a four week period)
`Corticosterone`	Blood corticosterone level (a measure of stress)

Graphs for our problem

Here are the comparative boxplots for our LightAtNight groups.



Summary of the Graphs

- It appears that there is a significant difference in the average BMG among the groups.

BMG (body mass gain)

- What is an appropriate test statistic for this?
- (Answer: an F statistic. It takes quite a bit of explanation!)

ANOVA for Means: F test

Developing the F statistic

- The main idea of Analysis of Variance (ANOVA) is to split the variability in the data into different parts so that they can be compared.
- Each part represents something important about the question of interest.

ANOVA for Means

- We want to think about how far apart the means are.
- We compare the variability between the groups

variance value

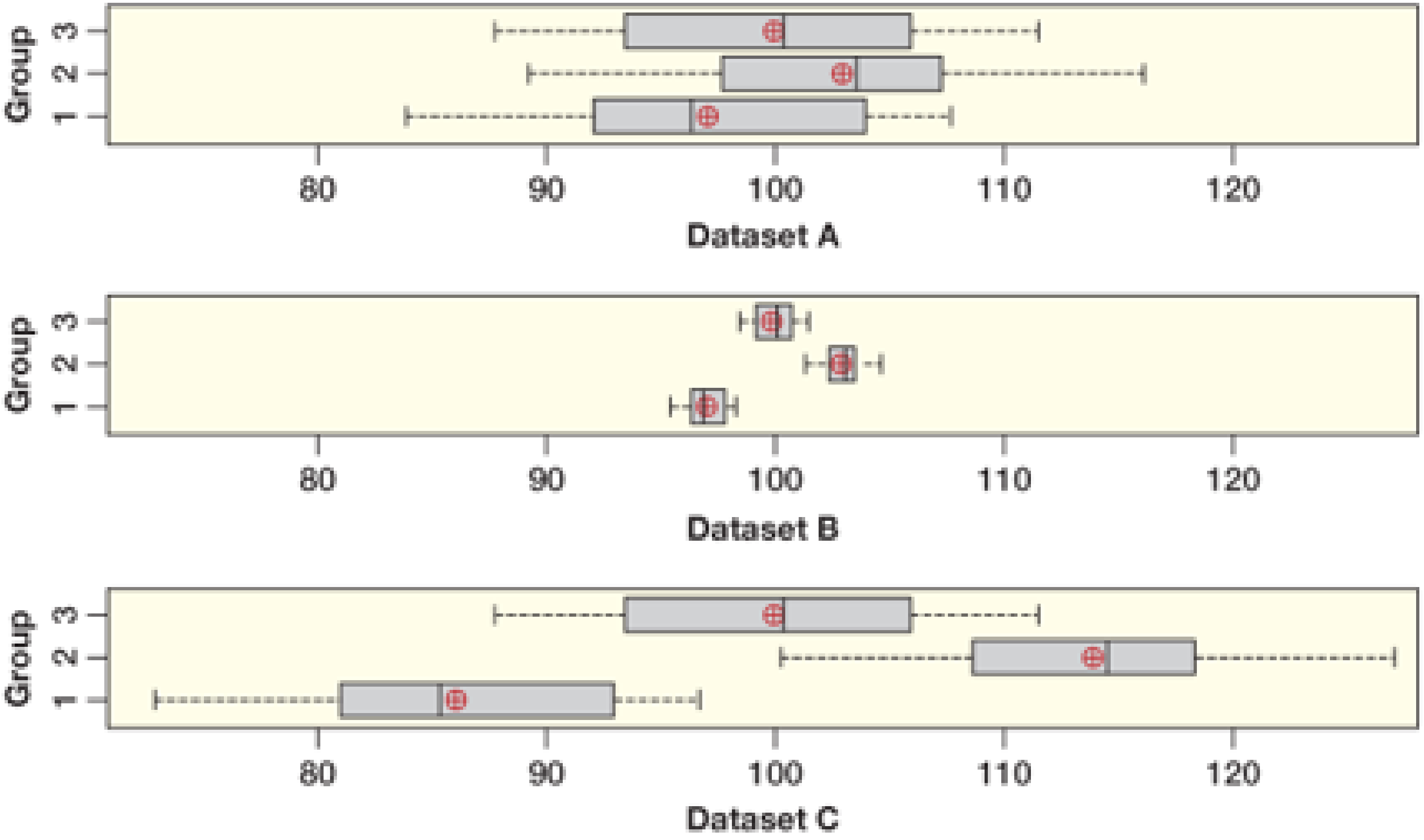
with

- the variability within the groups.

(Use this language to describe what was illustrated by the graphs in the Introduction to ANOVA.

Reading comparative boxplots

Image: Lock, et. al., *Statistics: Unlocking the Power of Data*, Wiley, 2nd ed. p. 542.



Discussing the graphs

Dataset B: Relatively speaking, the variability between the means is much larger than the variability within the groups.

Dataset A: The variability within the groups is quite a lot larger than the variability between the means.

Measuring Variability

We typically measure variability using the sum of squared deviations.

SSTotal (Sum of Squares of Total)

- TOTAL Variability: Variability of all data values from overall mean. SSTotal
- Variability BETWEEN groups: How far apart the means of the groups are. SSG SSG (Sum of Squares of Groups)
- Variability WITHIN the groups: How far apart the values within each group are. SSE SSE (Sum of Squares of error)

$$SSTotal = SSG + SSE$$

SS(Total) = SS(Between) + SS(Error)

Details

It is unrealistic to compute the sums of the squares "by hand" instead of with software, but it is instructive to see the actual formulas.

$\bar{x}$ = mean of all observations

$\bar{x}_i$ = mean of observations in the ith group

$$\text{SSTotal} = \sum (x - \bar{x})^2$$

$$\text{SSGroups} = \sum n_i (\bar{x}_i - \bar{x})^2$$

$$\text{SSError} = \sum (x - \bar{x}_i)^2$$

From SS to MS to F

To compare these deviations, we divide by the degrees of freedom to obtain "Mean Squared Error" of each.

MSG = SSG/(k-1)

MSE = SSE/(n-k)

Our F statistic is

$$F = \frac{MSG}{MSE} = \frac{113.1/2}{161.8/24} = \frac{56.541}{6.743} = 8.385$$

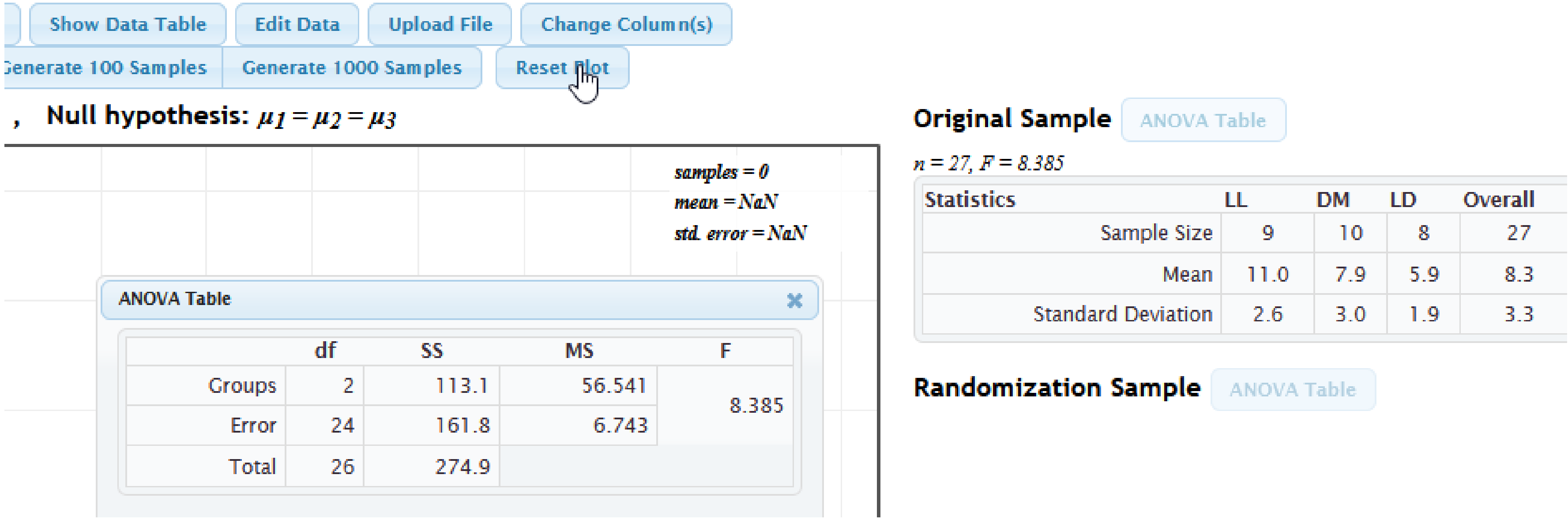
The ANOVA Table

ANOVA Table

Source	df	Sum of Sq	Mean Square	F-statistic	p-value
Groups	$k-1$	SSG	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSE}$	Upper tail of $F_{k-1,n-k}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$		
Total	$n-1$	SSTotal			

ANOVA for Means Output Table

- We have three light levels (groups) so the degrees of freedom is 3 – 1 = 2
- The overall degrees of freedom is 27 – 1 = 26



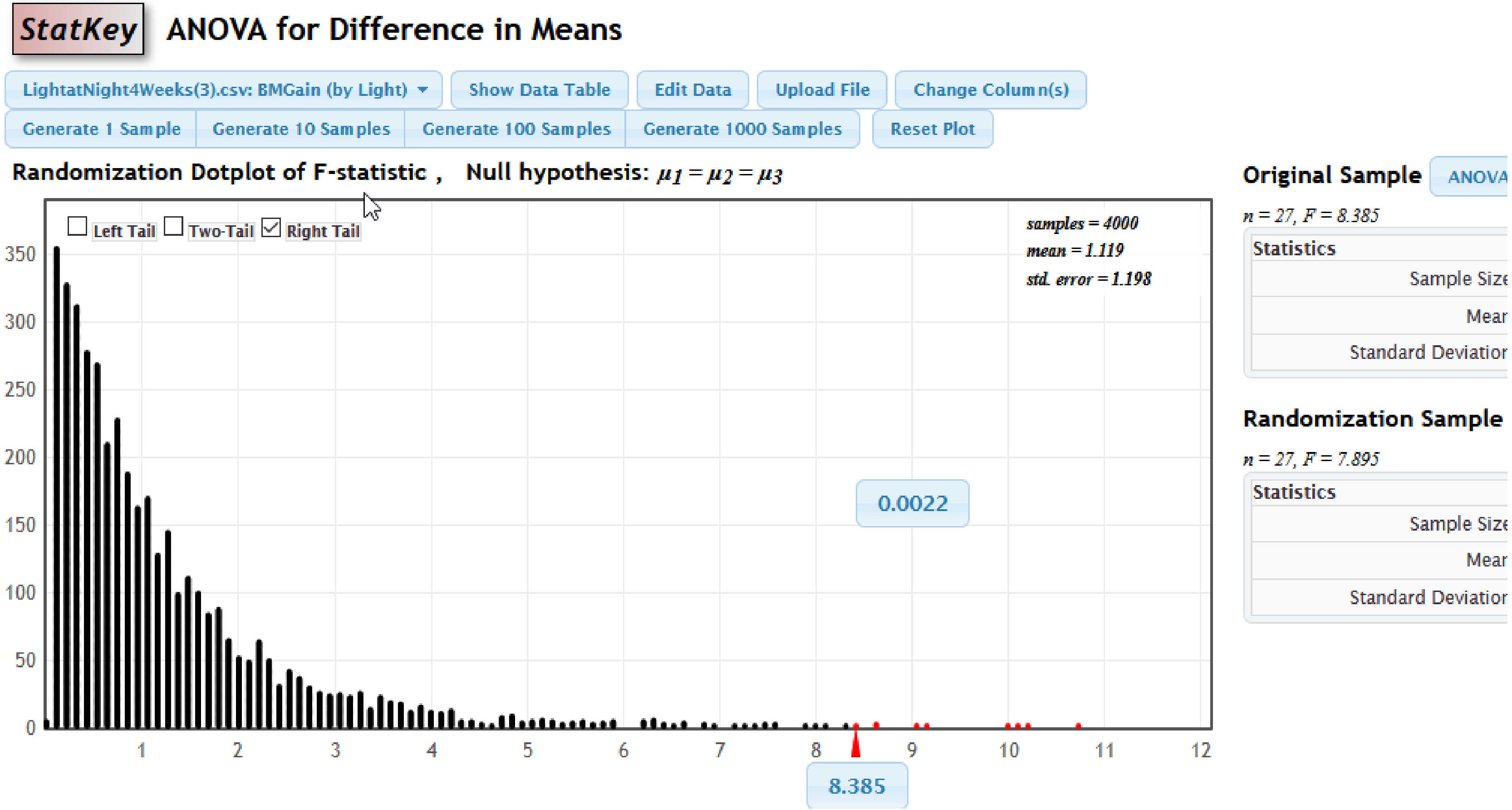
ANOVA Table

	df	SS	MS	F
Groups	2	113.1	56.541	8.385
Error	24	161.8	6.743	
Total	26	274.9		

Statistics	LL	DM	LD	Overall
Sample Size	9	10	8	27
Mean	11.0	7.9	5.9	8.3
Standard Deviation	2.6	3.0	1.9	3.3

Finding the p-value from a randomization test

p = 0.0022. Very strong evidence for Ha.



Conditions for using a theoretical dist’n

- To use the theoretical dist’n, we need to have sample sizes large enough for the Central Limit Theorem to give us the normality we need

  OR

- the individual populations should be normally distributed.
- Also, we need the sample standard deviations to be close to equal (the largest is no more than twice the smallest.)

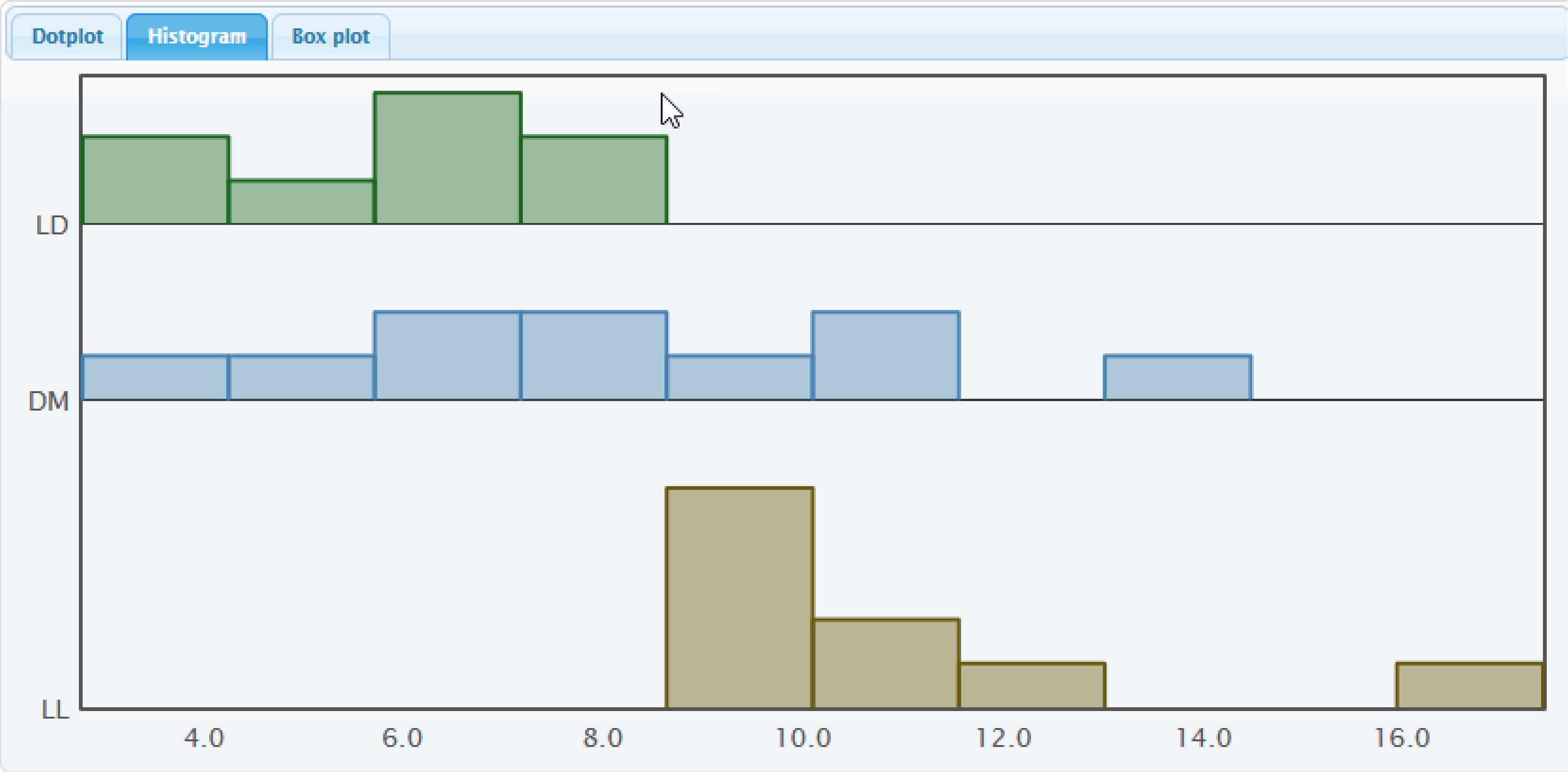
  From our output, that is 3.0/1.9 < 2.   So that’s good.

Conditions for using a theoretical dist’n

- The boxplots we saw aren’t really useful for thinking about the possible normality of the shape.

- Of course, for small sample sizes the shape of the graph of the data can vary widely from the shape of the population data.

Conditions for using a theoretical dist'n



Conditions for using a theoretical dist’n

The standard deviations of the samples are close enough together.

The sample sizes are so small it is not clear whether the data meet the condition about normality.

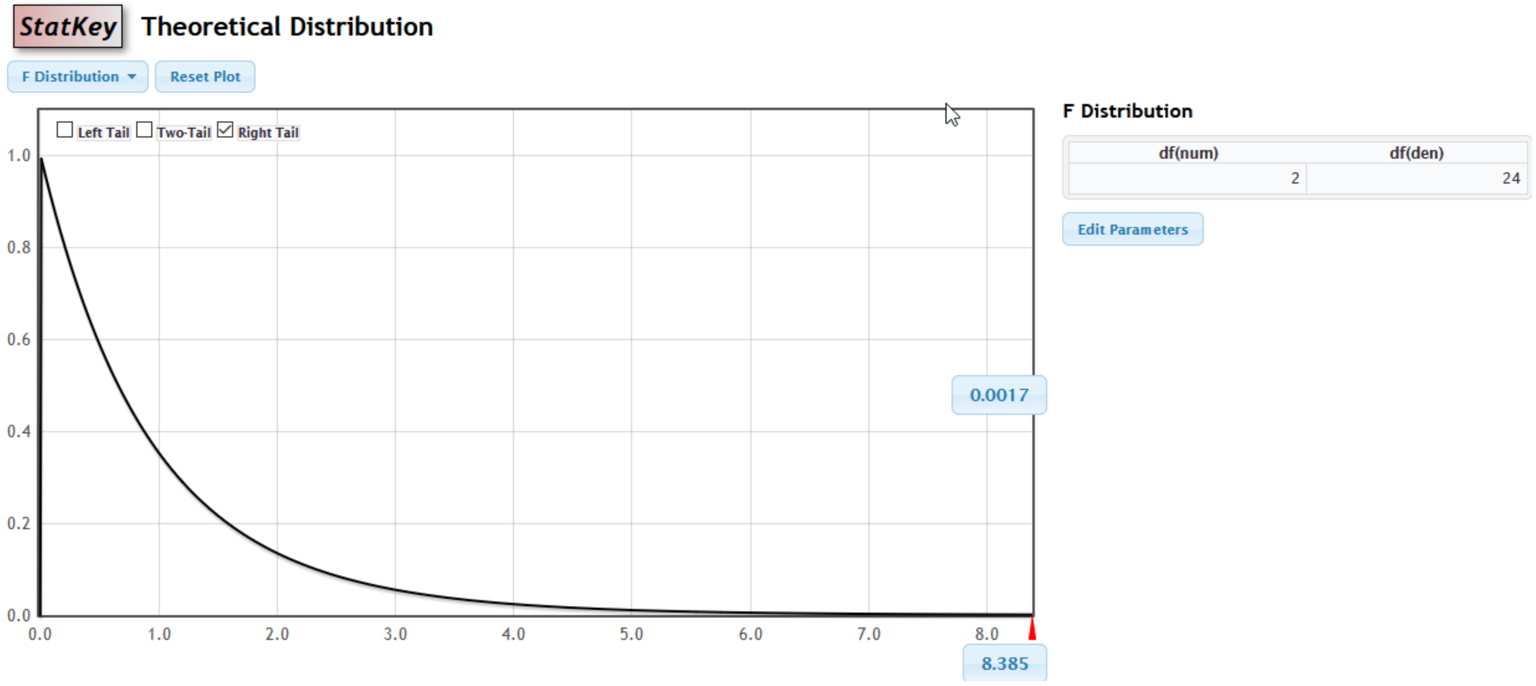
I will rely on the Randomization test result.

However .....

It is interesting to see what the p-value from the theoretical dist’n is.

Finding the p-value in the F dist’n

Here, the numerator df = 2 and the denominator df = 24, so, using the F dist’n, we find the p-value = 0.0017.



Conclusion

These data give quite strong evidence that not all the mean BMG values are equal for the three different light conditions. ($p = 0.002$)

Theoretical F dist’n

- The mathematical statistics derivation of the F distribution (and how this fits it) will be mentioned in a later chapter of this course.
- It relies on several results based on bivariate transformations of random variables (starting with data that is normally distributed.)