Chapter 10
One-Way Analysis of Variance: Comparing Several Means

Objectives

- Comparing several means
- The analysis of variance *F* test
- The idea of analysis of variance
- Conditions for ANOVA
- Some details of ANOVA
- *F* distributions and degrees of freedom
- JMP example

Introduction

- The two-sample *t* procedures compared the means of two populations or the mean responses to two treatments in an experiment.

- ANOVA: Method of **testing the equality of three or more population means** by analyzing sample variances.

Note: We are comparing ***means,*** even though the procedure is called analysis of *variance.*

Initial Thought

Why can't we just test two samples at a time using the two-sample *t* test?

- With four populations, we need to perform six hypothesis tests:

$H_0: \mu_1 = \mu_2$ $H_0: \mu_1 = \mu_3$ $H_0: \mu_1 = \mu_4$ $H_0: \mu_2 = \mu_3$ $H_0: \mu_2 = \mu_4$ $H_0: \mu_3 = \mu_4$
$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 \neq \mu_3$ $H_1: \mu_1 \neq \mu_4$ $H_1: \mu_2 \neq \mu_3$ $H_1: \mu_2 \neq \mu_4$ $H_1: \mu_3 \neq \mu_4$

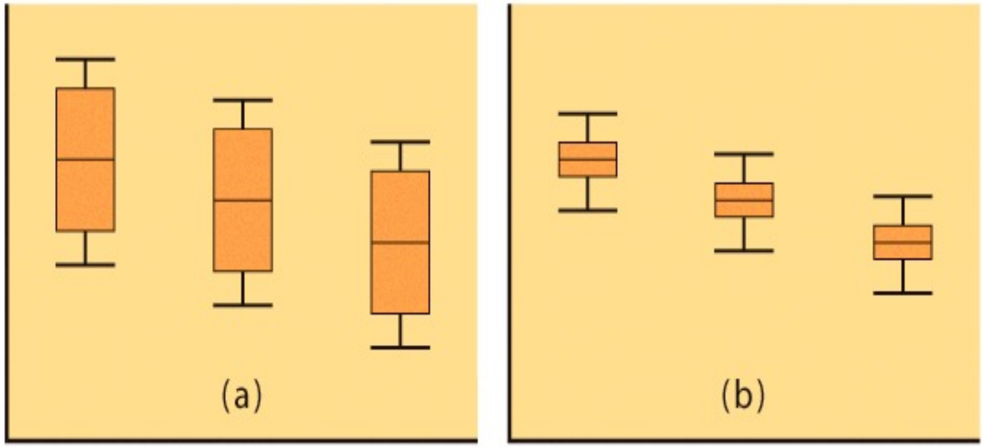
- If each individual test has a 95% confidence level ($1 - \alpha = 0.95$), our overall test has a much lower confidence level equal to $0.95^6 = 0.735$.
- If each individual test has a significance level of $\alpha = 0.05$, our overall test has a much higher significance level of $\alpha = 0.265$.
- As we increase the number of hypothesis tests, we increase the type I error (probability of rejecting a true H_0), i.e. Increase the likelihood of finding a difference in one of the pairs when no such difference actually exists!

The ANOVA question: Do all groups have the same population mean? The purpose of ANOVA is to assess whether the observed differences among sample means are *statistically significant*.

ANOVA tests whether several populations have the same means by comparing *how far apart the sample means are* with *how much variation there is within a sample*.

Comparing means

Difference in means small relative to overall variability



Difference in means large relative to overall variability

- ✓ The sample means for the three samples are the same for each set.
- ✓ The variation *among sample means* for (a) is identical to that for (b).
- ✓ The variation *among the individuals within* each of the three samples is much less for (b).

CONCLUSION: The samples in (b) contain a larger amount of variation among the sample means *relative to* the amount of variation within the samples, so ANOVA will find *more significant differences among the means in (b)*

- assuming equal sample sizes here for (a) and (b).
- **Note:** Larger samples will find more significant differences.

One-way ANOVA

Types of Variables:

1. Dependent/response (y) = Continuous
2. Independent/predictor (x) = Categorical with more than two categories (groups)

two categories simplifies to a two-sample *t*-test!!!

Two sample t-test with equal variances

Assume equal but unknown standard deviations,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

The square of this t-statistic is the same as ANOVA F statistic

$$t^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_p^2 (\frac{1}{n_1} + \frac{1}{n_2})}$$

Two-sample *t* statistic

A two-sample *t*-test assuming equal variance and an ANOVA comparing only two groups will give you the exact same *p*-value (for a two-sided hypothesis).

H_0: $\mu_1 = \mu_2$ H_a: $\mu_1 \neq \mu_2$	H_0: $\mu_1 = \mu_2$ H_a: $\mu_1 \neq \mu_2$
One-way ANOVA	*t*-test assuming equal variance
F-statistic	*t*-statistic

F = t^2 and both *p*-values are the same.

t-test is more flexible: (but can only compare two group means)

You may choose a one-sided alternative instead.

Or you may want to run a *t*-test assuming unequal variance if you are not sure that your two populations have the same standard deviation σ.

An Overview of ANOVA

- F**irst** examine the multiple populations to test for overall statistical significance as evidence of any difference among the population means → **ANOVA F-test**

- **If** that overall test showed statistical significance, then a detailed follow-up analysis is legitimate.

  - If we planned our experiment with specific alternative hypotheses in mind (before gathering the data), we can test them using **contrasts.**

  - If we do not have specific alternatives, we can examine all pair-wise parameter comparisons to define which parameters differ from which, using **multiple comparisons procedures.**

One way ANOVA

- Goal: comparison of *I* population or treatment means $\mu_1, \mu_2, \ldots \mu_I$

- Assumptions:
  - **Each of the *I* populations is normally distributed.**
  - $\sigma_1 = \sigma_2 = \ldots = \sigma_I$
  - **Independence** of observations both within and between the *I* groups.

- Test of Hypotheses: Let μ_i =the ith group mean, where i = 1, 2, … I.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_I$$

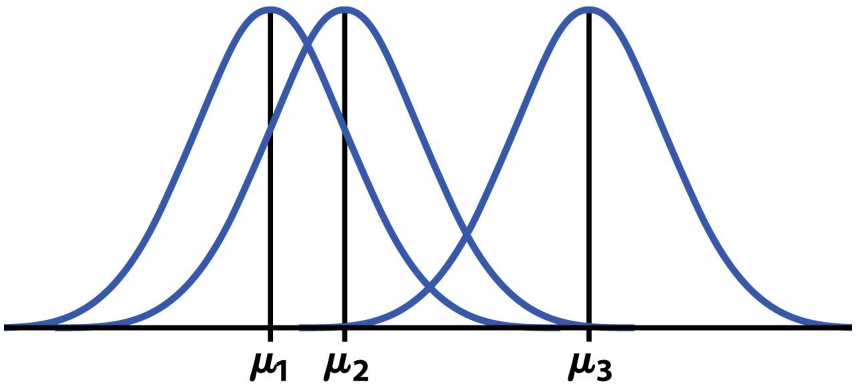
H_a: at least one μ_i is different from others

The One-Way ANOVA Model

Random sampling always produces chance variations. Any "factor effect" would thus show up in our data as the factor-driven differences plus chance variations ("error"):

Data = factor effect + error

or **Data = FIT + RESIDUAL**



The one-way ANOVA model analyzes data x_{ij} where chance variations (error terms) are Normally distributed $N(0,\sigma)$:

Cell means model $\qquad x_{ij} = \mu_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma)$

for $I = 1,\ldots,I$ and $j = 1,\ldots,n_i$. The **parameters of the model** are the population means $\mu_1,\mu_2,\ldots,\mu_I$ and the common standard deviation σ.

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

H_a: not all the μ_i are equal.

Use F test

Estimates of the Population Parameters

The unknown parameters in the model are the I population means μ_i and the common population standard deviation σ.

To estimate μ_i, we use the sample mean:

$$\bar{x}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

To estimate σ, we use the pooled standard deviation:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_I-1)s_I^2}{(n_1-1)+(n_2-1)+\cdots+(n_I-1)}}$$

Components of variation

Variation *between I groups: (variation due to treatment)*

$$SSG = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{x}_{i.} - \overline{x})^2 = \sum_{i=1}^{I} n_i(\overline{x}_{i.} - \overline{x})^2, DFG = I - 1$$

Variation *within the groups: (variation due to error)*

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2 = \sum_{i=1}^{I}(n_i - 1)s_i^2, DFE = N - I$$

Variation of data around overall mean:

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2, DFT = N - 1$$

The sum of squares represents variation in the data: SST = SSG + SSE.

The degrees of freedom in the ANOVA model: DFT = DFG + DFE

The ANOVA *F* Statistic

To determine statistical significance, we need a test statistic:

The ANOVA *F* Statistic

The **analysis of variance *F* statistic** for testing the equality of several means has this form:

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$

- *F* is always zero or positive
  - *F* is zero only when all sample means are the same
  - *F* gets larger as means move farther apart
- Large values of *F* are evidence against H_0: *equal means*

The ANOVA *F* Test

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}} \qquad F = \frac{\text{MSG}}{\text{MSE}} = \frac{\text{SSG}/(I-1)}{\text{SSE}/(N-I)}$$

The measures of variation in the numerator and denominator are ***mean squares:***

- Numerator: **Mean Square for Groups** (MSG)

$$MSG = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2}{I - 1}$$

- Denominator: **Mean Square for Error** (MSE)

$$MSE = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$

  - MSE is also called the **pooled sample variance**, written as s_p^2 (s_p is the ***pooled standard deviation***).
  - s_p^2 estimates the common variance σ^2.

The ANOVA table

Source of variation	Sum of squares SS	DF	Mean square MS	F	P value	F crit
Among or between "groups"	$\sum n_i(\bar{x}_{i.} - \bar{x})^2$	I -1	MSG=SSG/DFG	MSG/MSE	Tail area above F	Value of F for α
Within groups or "error"	$\sum (n_i - 1)s_i{}^2$	N - I	MSE=SSE/DFE			
Total	SST=SSG+SSE $\sum (x_{ij} - \bar{x})^2$	N – 1				

R^2 = SSG/SST
Coefficient of determination

√MSE = s_p
Pooled standard deviation

The sum of squares represents variation in the data: SST = SSG + SSE.

The degrees of freedom likewise reflect the ANOVA model: DFT = DFG + DFE.

Data ("Total") = fit ("Groups") + residual ("Error")

Testing hypotheses in one-way ANOVA

We have ***I* independent SRSs**, from *I* populations or treatments.

The i^{th} population has a **normal distribution** with unknown mean μ_i.

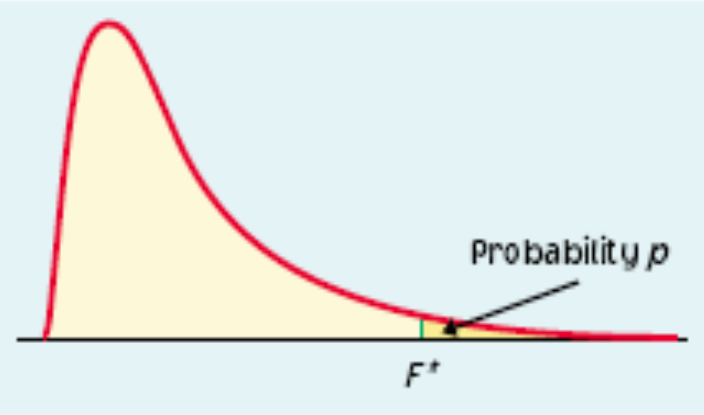
All *I* populations have the **same standard deviation *σ***, unknown.

The ANOVA *F* statistic tests:

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

H_a: not all the μ_i are equal.

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\text{SSG}/(I-1)}{\text{SSE}/(N-I)}$$



When H_0 is true, *F* has the **F distribution** with *I*– 1 (*numerator*) and *N*– *I* (*denominator*) degrees of freedom.

Using Table E

The F distribution is asymmetrical and has two distinct degrees of freedom. This was discovered by Fisher, hence the label "F."

Once again, what we do is calculate the value of F for our sample data and then look up the corresponding area under the curve in Table E.

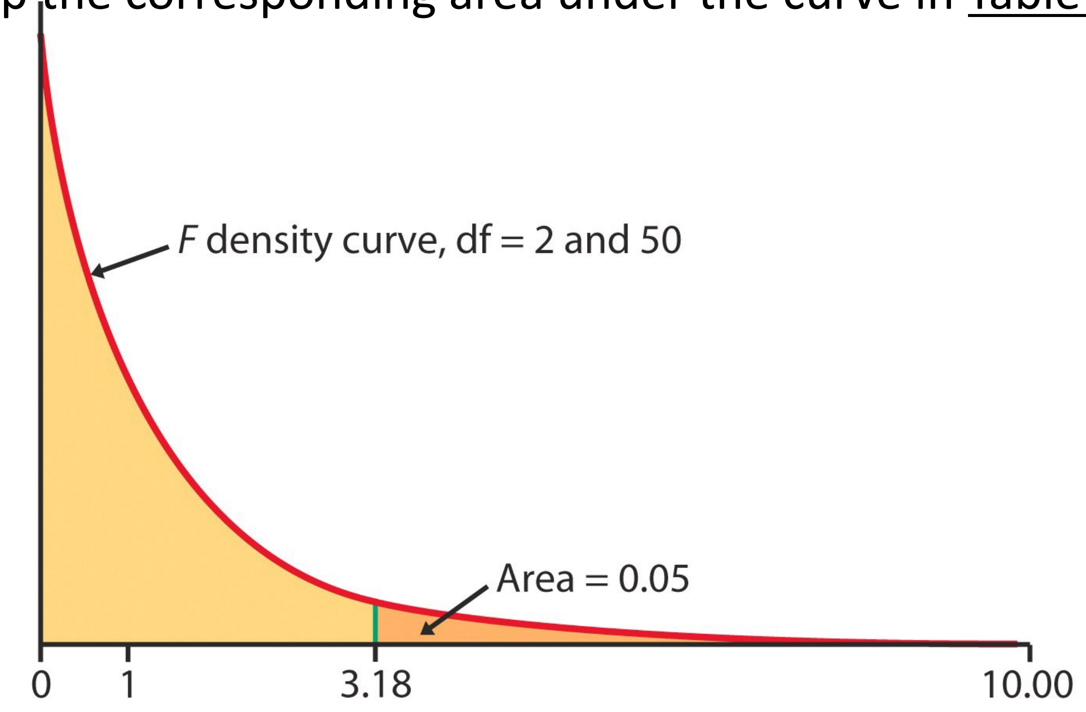


Table E F distribution critical values

$df_{num} = I - 1$

For df: 5,4

$df_{den} = N - I$

p F

Degrees of freedom in the denominator	p	Degrees of freedom in the numerator 1	2	3	4	5	6	7	8
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66
	0.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859	5928.4	5981.1
	0.001	405284	500000	540379	562500	576405	585937	592873	598144
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	0.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	0.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	0.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03

Comparing several means

EXAMPLE: Comparing tropical flowers

- **STATE:** On the island of Dominica, the relationship between varieties of the tropical flower *Heliconia* and the different species of hummingbirds that fertilize them was examined. Researchers wondered if the lengths of the flowers and the forms of the hummingbirds' beaks have evolved to match each other. The table below gives length measurements (in millimeters) for samples of three varieties of *Heliconia*, each fertilized by a different species of hummingbird. Do the three varieties display distinct distributions of length? In particular, are the mean lengths of their flowers different?

- **PLAN:** Use graphs and numerical descriptions to describe and compare the three distributions of flower length. Finally, ask whether the differences among the mean lengths of the three varieties are statistically significant.

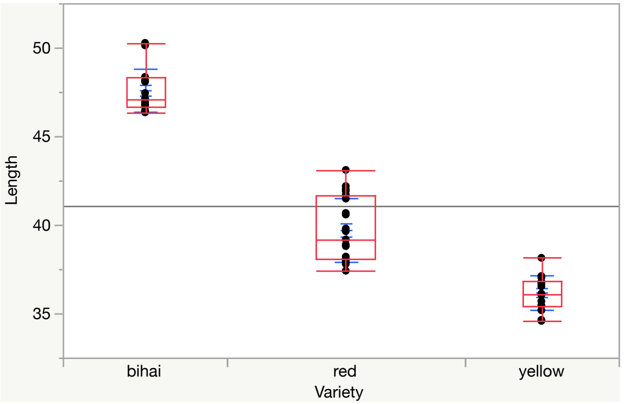
FLOWER LENGTHS (MILLIMETERS) FOR THREE *HELICONIA* VARIETIES							
H. BIHAI							
47.12	46.75	46.81	47.12	46.67	47.43	46.44	46.64
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36
***H. CARIBAEA* RED**							
41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07
38.10	37.97	38.79	38.23	38.87	37.78	38.01	
***H. CARIBAEA* YELLOW**							
36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10
35.17	36.82	36.66	35.68	36.03	34.57	34.63	

Comparing several means

EXAMPLE: Comparing tropical flowers

- Summary statistics we will use in further analysis:

Sample	Variety	Sample size	Mean length	Standard deviation
1	*bihai*	16	47.60	1.213
2	red	23	39.71	1.799
3	yellow	15	36.18	0.975



- The three varieties differ so much in flower length that there is little overlap among them. In particular, the flowers of *bihai* are longer than either red or yellow. The mean lengths are 47.6 mm for *H. bihai*, 39.7 mm for *H. caribaea* red, and 36.2 mm for *H. caribaea* yellow.
- Are these observed differences in sample means statistically significant? We must develop a test for comparing more than two population means.

Comparing several means

- (Sample) means:
  - *bihai*: 47.60
  - red: 39.71
  - yellow: 36.18

- Null hypothesis:
- $H_0: \mu_1 = \mu_2 = \mu_3$

The true means for length are the same for all groups (the three flower types).

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Variety	2	1082.8724	541.436	259.1193	<.0001*
Error	51	106.5658	2.090		
C. Total	53	1189.4381			

JMP tells us that, for the flower length data, the test statistic is F = 259.12 with P < 0.0001. There is very strong evidence that the three varieties of flowers do not have the same mean length. It appears from the boxplot and summary statistics that *bihai* flowers are distinctly longer than either red or yellow. Red and yellow are closer together, but the red flowers tend to be longer.

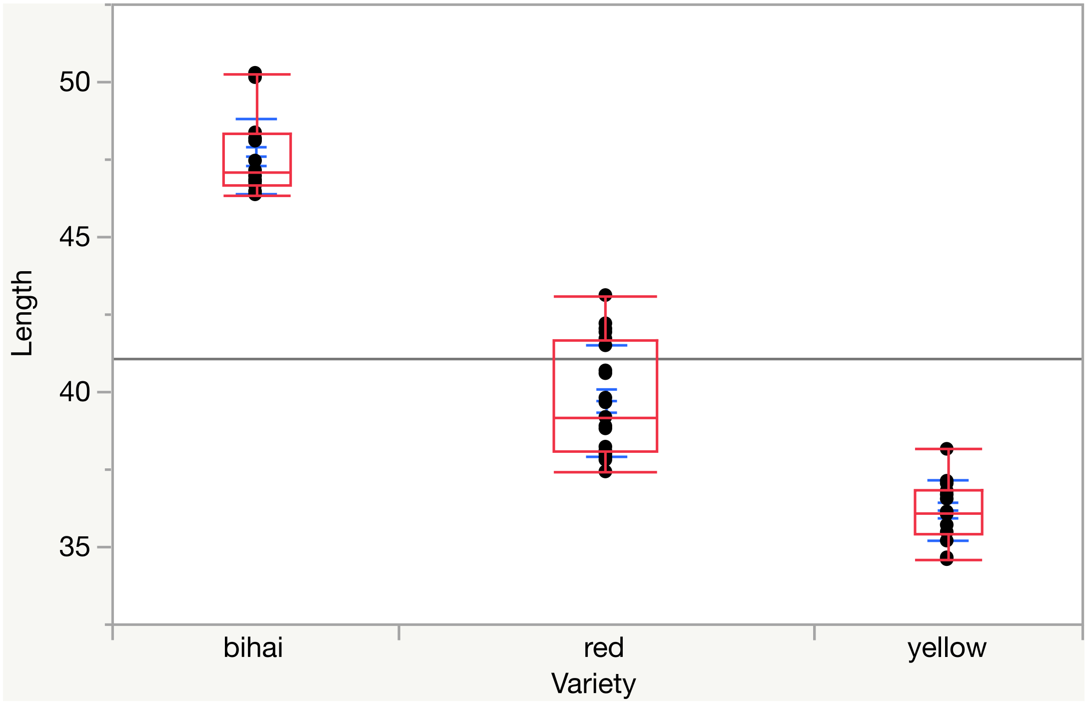
Follow-up (post-ANOVA) analysis

- If reject H0 in the overall ANOVA F test, we know at least one mean is different from other means.
- We need follow-up (post-ANOVA) analysis to figure out which group mean is different.
- This can be done through pre-planned contrast or multiple comparisons. (will be covered in Stat 442).

JMP example

- Open data (eg27-01FLOWER.jmp), make sure the grouping variable is a categorical variable, click Analyze, then Fit Y by X
- Click length into Y box, Variety into X box, then click ok.
- Click the red triangle near Oneway analysis of Length by Variety, click Quantiles to get side-by-side boxplot and 5-number summaries.
- Click Means and St.Dev to get sample group mean and s.d
- Click Means/Anova to perform an ANOVA from the popup menu

Oneway Analysis of Length By Variety



Quantiles

Level	Minimum	10%	25%	Median	75%	90%	Maximum
bihai	46.34	46.41	46.69	47.12	48.2925	50.162	50.26
red	37.4	37.816	38.07	39.16	41.69	42.112	43.09
yellow	34.57	34.606	35.45	36.11	36.82	37.512	38.13

Means for one-way ANOVA

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
bihai	16	47.5975	0.36138	46.872	48.323
red	23	39.7113	0.30141	39.106	40.316
yellow	15	36.1800	0.37323	35.431	36.929

Oneway Anova
Summary of Fit

Rsquare	0.910407
Adj Rsquare	0.906893
Root Mean Square Error	1.445519
Mean of Response	41.06704
Observations (or Sum Wgts)	54

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Variety	2	1082.8724	541.436	259.1193	<.0001*
Error	51	106.5658	2.090		
C. Total	53	1189.4381			

Std Error uses a pooled estimate of error variance