Introduction to Probability and Statistics Fourteenth Edition

Chapter 11
The Analysis of Variance

11.1 The design of an experiment

EXPERIMENTAL DESIGN

- The **sampling plan** or **experimental design** determines the way that a sample is selected.
- In an **observational study,** the experimenter observes data that already exist. The **sampling plan** is a plan for collecting this data.
- In a **designed experiment,** the experimenter imposes one or more experimental conditions on the experimental units and records the response.

DEFINITIONS

- An **experimental unit** is the object on which a measurement or measurements) is taken.
- A **factor** is an independent variable whose values are controlled and varied by the experimenter.
- A **level** is the intensity setting of a factor.
- A **treatment** is a specific combination of factor levels.
- The **response** is the variable being measured by the experimenter.

EXAMPLE

- A group of people is randomly divided into an experimental and a control group. The control group is given an aptitude test after having eaten a full breakfast. The experimental group is given the same test without having eaten any breakfast.

Experimental unit = person

Factor = meal

Response = Score on test

Levels = Breakfast or no breakfast

Treatments: Breakfast or no breakfast

EXAMPLE

- The experimenter in the previous example also records the person's gender. Describe the factors, levels and treatments.

Experimental unit = person

Response = score

Factor #1 = meal

Levels = breakfast or no breakfast

Factor #2 = gender

Levels = male or female

Treatments:

male and breakfast, female and breakfast, male and no breakfast, female and no breakfast

The Analysis of Variance (ANOVA)

- All measurements exhibit **variability**.
- The total variation in the response measurements is broken into portions that can be attributed to various **factors**.
- These portions are used to judge the effect of the various factors on the experimental response.

The Analysis of Variance

- If an experiment has been properly designed,

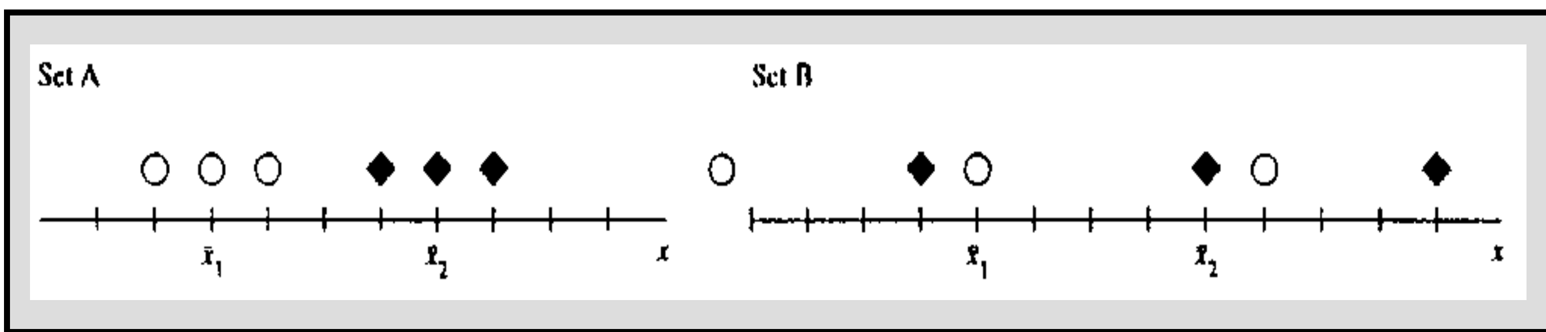
Total variation → Factor 1

Total variation → Factor 2

Total variation → Random variation

- We compare the variation due to any one factor to the typical random variation in the experiment.

The variation between the sample means is larger than the typical variation within the samples.	The variation between the sample means is about the same as the typical variation within the samples.



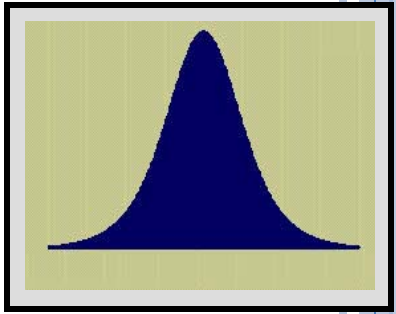
Assumptions

- Similar to the assumptions required in Chapter 10.

1. The observations within each population are normally distributed with a common variance σ^2.
2. Assumptions regarding the sampling procedures are specified for each design.

•Analysis of variance procedures are fairly robust when sample sizes are equal and when the data are fairly mound-shaped.

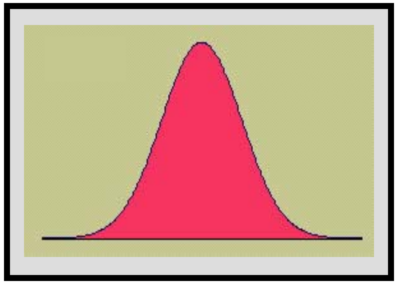
Three Designs



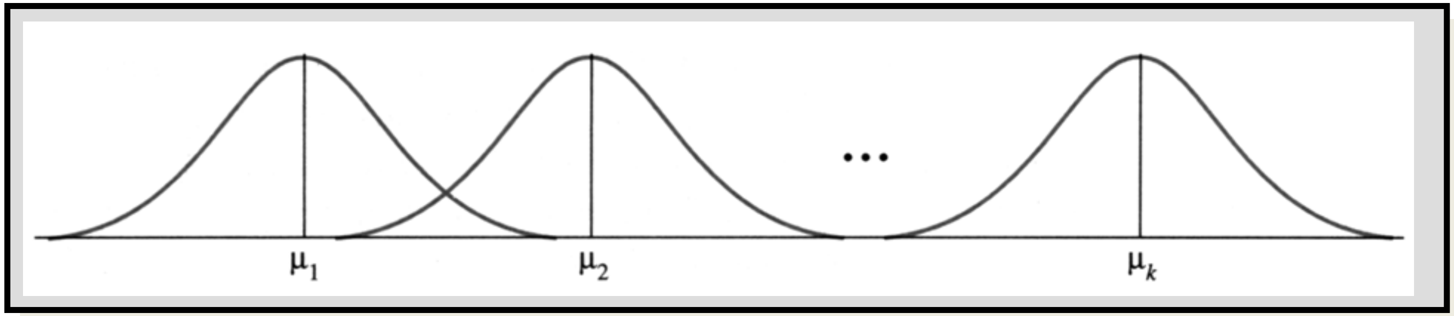
- **Completely randomized design:** an extension of the two independent sample t-test.
- **Randomized block design:** an extension of the paired difference test.
- **a × b Factorial experiment:** we study two experimental factors and their effect on the response.

11.2 The completely randomized design: a one-way classification

The Completely Randomized Design



- A **one-way classification** in which one factor is set at k different levels.
- The k levels correspond to k different normal populations, which are the **treatments**.
- Are the k population means the same, or is at least one mean different from the others?



Example

Is the attention span of children affected by whether or not they had a good breakfast? Twelve children were randomly divided into three groups and assigned to a different meal plan. The response was attention span in minutes during the

No Breakfast	Light Breakfast	Full Breakfast
8	14	10
7	16	12
9	12	16
13	17	15

k = 3 treatments. Are the average attention spans different?

The Completely Randomized Design

- Random samples of size n_1, n_2, …,n_k are drawn from k populations with means μ_1, μ_2,…, μ_k and with common variance σ^2.
- Let x_{ij} be the j-th measurement in the i-th sample.
- The total variation in the experiment is measured by the **total sum of squares**:

$$\text{Total SS} = \sum (x_{ij} - \bar{x})^2$$

The Completely Randomized Design

Random samples of size $n_1, n_2, \ldots, n_k$ are drawn from k populations with means $mu_1, \ldots, mu_k$ and with common variance σ^2.

Let x_{ij} be the j-th measurement in the i-th sample.

Model

$$x_{ij} \sim N(\mu_i, \sigma^2),$$

for $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, $n_1 + \cdots + n_k = n$.

Hypothesis test

1. $H_0 : \mu_1 = \cdots = \mu_k$ versus H_a : at least one mean is differ
2. Set $\alpha = 0.05$
3. Test statistic and its distribution:

$$F_{STAT} = \frac{MST}{MSE} \sim F_{(k-1),(n-k)}.$$

4. Calculate the realized F^*.
5. This is a right-tailed test.
    - The rejection region is $\{F : F > F_{(k-1),(n-k),\alpha}\}$.
    - The p-value is $P(F_{(k-1),(n-k)} > F^*)$.
6. Conclude.

The total sum of squares, total SS

The grand mean $\bar{G} = \frac{1}{n}\sum_{ij} x_{ij}$.

$$\text{Total SS} = \sum_{i,j}(x_{ij} - \bar{G})^2.$$

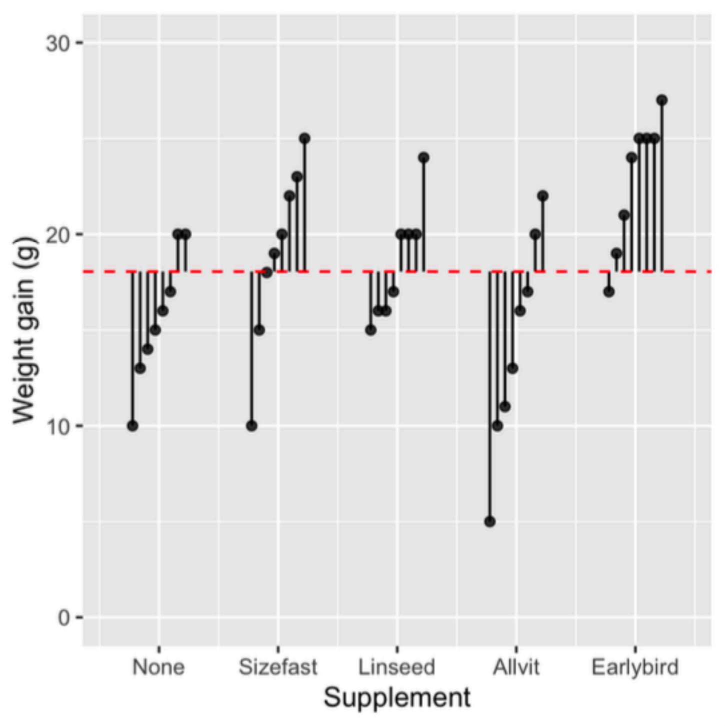
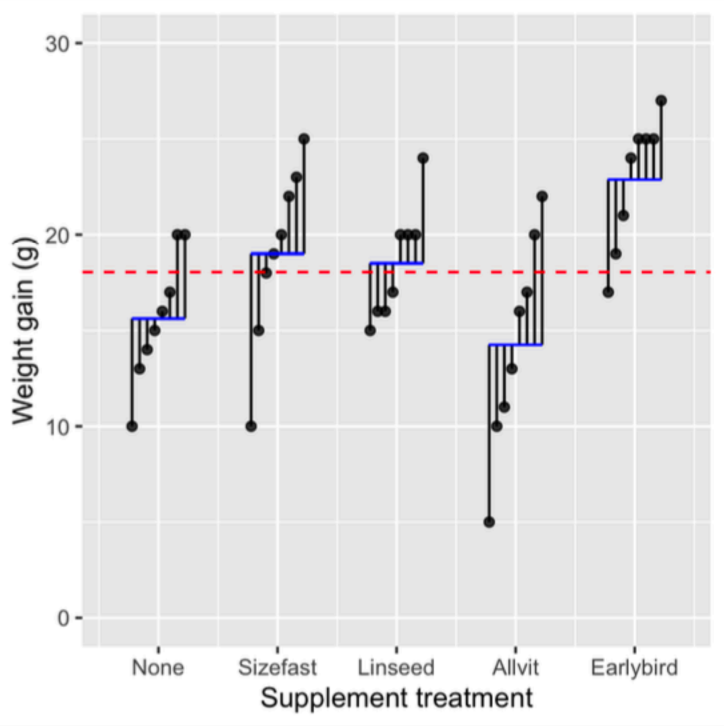


Photo credit: https://dzchilds.github.io/stats-for-bio

Sum squares of treatment, SST

The mean for treatment i: $\bar{T}_i = \frac{1}{n_i} \sum_{j=1,\ldots,n_i} x_{ij}$.

$$\text{SST} = \sum_i n_i(T_i - \bar{G})^2$$



ANOVA table

Sum squares of errors, SSE,

$$\text{SSE} = \sum_{ij}(x_{ij} - \bar{T}_i) = \text{Total SS} - \text{sST}.$$

Soure	SS	df	MS	F
Treatment	SST	$(k-1)$	MST = (SST/($k-1$))	$F = \frac{MST}{MSE}$
Error	SSE	$(n-k)$	MSE = (SSE/($n-k$))	
Total	Total SS	$(n-1)$		

Degrees of freedom and mean squares

These sums of squares behave like the numerator of a sample variance.

When divided by the appropriate degrees of freedom, each provides a mean square, an estimate of variation in the experiment.

Degrees of freedom are additive, just like the sums of squares.

The Analysis of Variance

The **Total SS** is divided into two parts:

- ✓ **SST** (sum of squares for treatments): measures the variation among the *k* sample means.
- ✓ **SSE** (sum of squares for error): measures the variation within the *k* samples.

in such a way that:

$$\text{Total SS} = \text{SST} + \text{SSE}$$

Computing Formulas

$$\text{Total SS} = \Sigma x_{ij}^2 - \text{CM}$$

$$= (\text{Sum of squares of a}$$

with

$$\text{CM} = \frac{(\Sigma x_{ij})^2}{n} = \frac{G^2}{n}$$

$$\text{SST} = \Sigma \frac{T_i^2}{n_i} - \text{CM}$$

$$\text{SSE} = \text{Total SS} - \text{SST}$$

and

G = Grand total of all n observations

T_i = Total of all observations in sample i

n_i = Number of observations in sample i

$n = n_1 + n_2 + \cdots + n_k$

1. $G = \sum_{ij} x_{ij}$
2. CM = G^2/n
3. $T_i = \sum_{j=1,\ldots,n_i} x_{ij}$
4. $\text{Total SS} = \sum_{ij} x_{ij}^2 - CM$
5. $\text{SST} = \sum \frac{T_i^2}{n_i} - CM$
6. $\text{SSE} = \text{Total SS} - \text{SST}.$

Why?

Simplified computing formulas for the ANOVA table I

Define

$$G = \sum x_{ij} = \text{grand total of all } n \text{ observations},$$

$$T_i = \sum_{j}^{n_i} x_{ij} = \text{Total of all observations in sample } i,$$

$$n_i = \text{Number of observations in sample } i,$$

$$n = n_1 + n_2 + \cdots + n_k.$$

Second, define

$$CM = \frac{1}{n} G^2.$$

Simplified computing formulas for the ANOVA table II

First, we have

$$CM = \frac{G^2}{n} = \frac{1}{n}\left(\frac{G}{n}\right)^2 = \frac{\bar{G}^2}{n}$$

Second, recall that

$$\bar{T}_i = \frac{1}{n_i} T_i.$$

Taking squares of both sides, we have

$$\bar{T}_i^2 = \frac{1}{n_i^2} T_i^2,$$

and hence

$$n_i \bar{T}_i^2 = \frac{1}{n_i} T_i^2.$$

Simplified computing formulas for the ANOVA table III

With $\sum x_{ij} = n\bar{G}$, we can calculate

$$\begin{aligned}
\text{Total SS} &= \sum (x_{ij} - \bar{G})^2 \\
&= \sum (x_{ij}^2 - 2\bar{G}x_{ij} + \bar{G}^2) \\
&= \sum x_{ij}^2 - 2\bar{G}(n\bar{G}) + n\bar{G}^2 \\
&= \sum x_{ij}^2 - n\bar{G}^2 \\
&= \sum x_{ij}^2 - CM.
\end{aligned}$$

Simplified computing formulas for the ANOVA table IV

Similarly, we have

$$
\begin{aligned}
\mathrm{SST} &= \sum_i n_i(\bar{T}_i - \bar{G})^2 \\
&= \sum_i n_i(\bar{T}_i - 2\bar{T}_i\bar{G} + \bar{G}^2) \\
&= \sum_i (n_i\bar{T}_i - 2n_i\bar{T}_i\bar{G} + n_i\bar{G}^2) \\
&= \sum_i \frac{T_i^2}{n_i} - 2(n\bar{G})\bar{G} + n\bar{G}^2 \\
&= \sum_i \frac{T_i^2}{n_i} - n\bar{G}^2 \\
&= \sum_i \frac{T_i^2}{n_i} - CM.
\end{aligned}
$$

The Breakfast Problem

No Breakfast	Light Breakfast	Full Breakfast
8	14	10
7	16	12
9	12	16
13	17	15
$T_1 = 37$	$T_2 = 59$	$T_3 = 53$

$G = 149$

$$\mathrm{CM} = \frac{149^2}{12} = 1850.0833$$

$$\text{Total SS} = 8^2 + 7^2 + \ldots + 15^2 - \mathrm{CM} = 1973 - 1850.0833 = 122.9167$$

$$\mathrm{SST} = \frac{37^2}{4} + \frac{53^2}{4} + \frac{59^2}{4} - \mathrm{CM} = 1914.75 - \mathrm{CM} = 64.6667$$

$$\mathrm{SSE} = \text{Total SS} - \mathrm{SST} = 58.25$$

Degrees of Freedom and Mean Squares

- These **sums of squares** behave like the numerator of a sample variance. When divided by the appropriate **degrees of freedom**, each provides a **mean square**, an estimate of variation in the experiment.
- **Degrees of freedom** are additive, just like the sums of squares.

$$\text{Total } df = \text{Trt } df + \text{Error } df$$

THE ANOVA TABLE

Total df = $n_1+n_2+\ldots+n_k -1 = n - 1$

Treatment df = $k - 1$

Error df = $n - 1 - (k - 1) = n - k$

Mean Squares

MST = SST/(k-1)

MSE = SSE/(n-k)

Source	**df**	**SS**	**MS**	**F**
Treatments	k -1	SST	SST/(k-1)	MST/MSE
Error	n - k	SSE	SSE/(n-k)	
Total	n -1	Total SS		

The Breakfast Problem

$$CM = \frac{149^2}{12} = 1850.0833$$

$$\text{Total SS} = 8^2 + 7^2 + \ldots + 15^2 - CM = 1973 - 1850.0833 = 122.9167$$

$$SST = \frac{37^2}{4} + \frac{53^2}{4} + \frac{59^2}{4} - CM = 1914.75 - CM = 64.6667$$

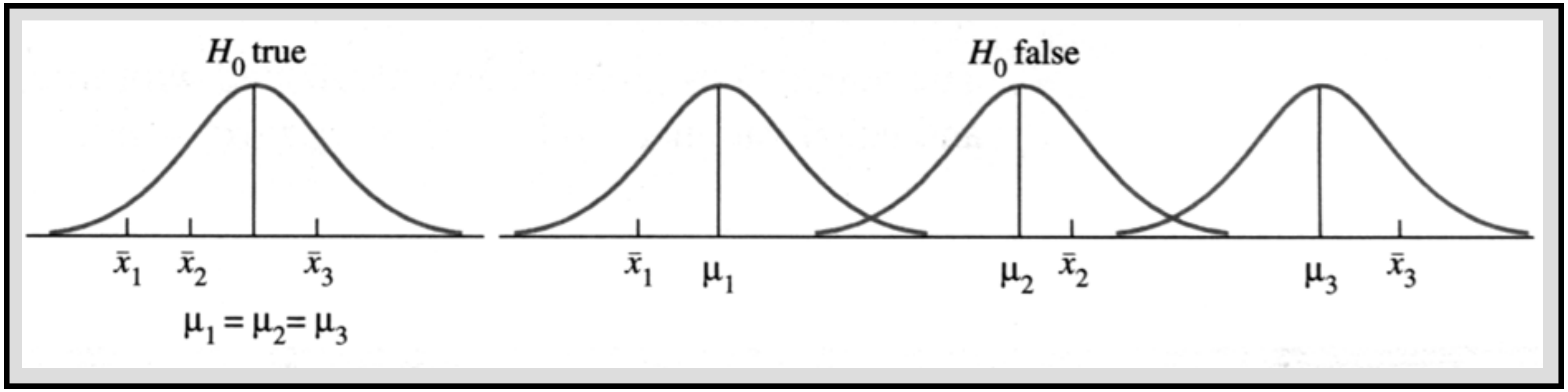
$$SSE = \text{Total SS} - SST = 58.25$$

Source	df	SS	MS	F
Treatments	2	64.6667	32.3333	5.00
Error	9	58.25	6.4722	
Total	11	122.9167		

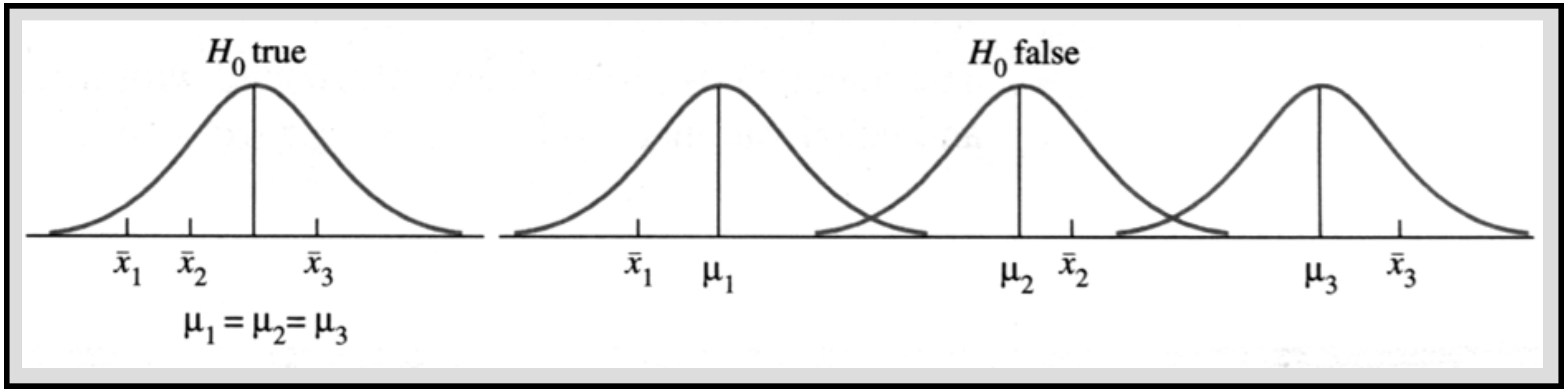
Testing the Treatment Means

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ versus

H_a : at least one mean is different

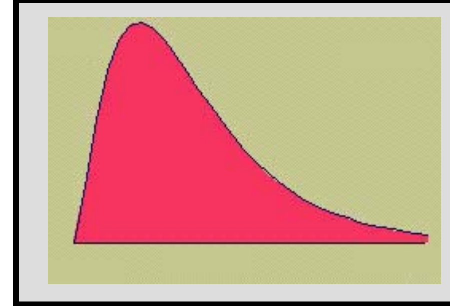


Remember that σ^2 is the common variance for all k populations. The quantity MSE = SSE/$(n - k)$ is a pooled estimate of σ^2, a weighted average of all k sample variances, whether or not H_0 is true.



- If H_0 is true, then the variation in the sample means, measured by MST = [SST/ $(k - 1)$], also provides an unbiased estimate of σ^2.
- However, if H_0 is false and the population means are different, then MST— which measures the variance in the sample means — is unusually **large.** The test statistic **F = MST/ MSE** tends to be larger that usual.

MY APPLET



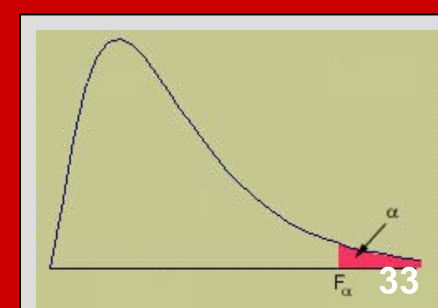
THE F TEST

- Hence, you can reject H 0 for large values of F, using a right-tailed statistical test.
- When H_0 is true, this test statistic has an F distribution with $df_1 = (k - 1)$ and $df_2 = (n - k)$ degrees of freedom and **right-tailed** critical values of the F distribution can be used.

To test $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

Test Statistic: $F = \dfrac{\text{MST}}{\text{MSE}}$

Reject H_0 if $F > F_\alpha$ with $k - 1$ and $n\text{-}k$ df.



The Breakfast Problem

Source	df	SS	MS	F
Treatments	2	64.6667	32.3333	5.00
Error	9	58.25	6.4722	
Total	11	122.9167		

$H_0 : \mu_1 = \mu_2 = \mu_3$ versus

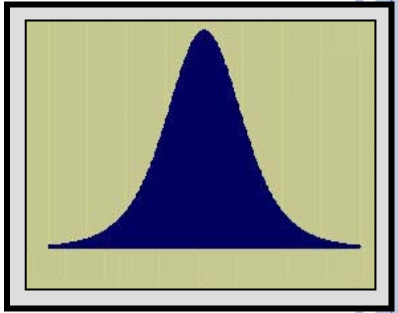
H_a : at least one mean is different

$$F = \frac{MST}{MSE} = \frac{32.3333}{6.4722} = 5.00$$

Rejection region : $F > F_{.05} = 4.26$.

We reject H_0 and conclude that there is a difference in average attention spans.

Confidence Intervals



- If a difference exists between the treatment means, we can explore it with confidence intervals.

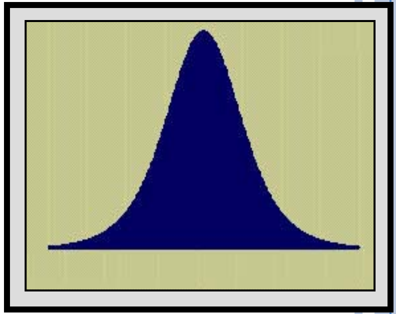
A single mean, $\mu_i : \bar{x}_i \pm t_{\alpha/2} \dfrac{s}{\sqrt{n_i}}$

Difference $\mu_i - \mu_j : (\bar{x}_i - \bar{x}_j) \pm t_{\alpha/2} \sqrt{s^2\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}$

where $s = \sqrt{\text{MSE}}$ and t is based on error df.

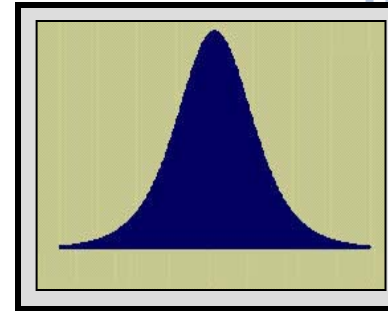
11.3 Ranking population means

Tukey's Method for Paired Comparisons



- Designed to test all pairs of population means simultaneously, with an **overall error rate of** α.
- Based on the **studentized range**, the difference between the largest and smallest of the *k* sample means.
- Assume that the **sample sizes are equal** and calculate a "ruler" that measures the distance required between any pair of means to declare a significant difference.

TUKEY'S METHOD



$$\text{Calculate:} \quad \omega = q_\alpha(k, df)\frac{s}{\sqrt{n_i}}$$

where k = number of treatment means

$s = \sqrt{\text{MSE}}$ $\qquad df$ = error df

n_i = common sample size

$q_\alpha(k, df)$ = value from Table 11.

If any pair of means differ by more than ω, they are declared different.

The Breakfast Problem

Use Tukey's method to determine which of the three population means differ from the others.

	No Breakfast	Light Breakfast	Full Breakfast
	$T_1 = 37$	$T_2 = 59$	$T_3 = 53$
Means	37/4 = 9.25	59/4 = 14.75	53/4 = 13.25

Percentage Points of the Studentized Range, $q(k, df)$; Upper 5% Points

	k				
df	2	3	4	5	6
1	17.97	26.98	32.82	37.08	40.41
2	6.08	8.33	9.80	10.88	11.74
3	4.50	5.91	6.82	7.50	8.04
4	3.93	5.04	5.76	6.29	6.71
5	3.64	4.60	5.22	5.67	6.03
6	3.46	4.34	4.90	5.30	5.63
7	3.34	4.16	4.68	5.06	5.36
8	3.26	4.04	4.53	4.89	5.17
9	3.20	3.95	4.41	4.76	5.02

$$\omega = q_{.05}(3,9)\frac{s}{\sqrt{4}} = 3.95\frac{\sqrt{6.4722}}{\sqrt{4}} = 5.02$$

The Breakfast Problem

List the sample means from smallest to largest.

$\bar{x}_1$	$\bar{x}_3$	$\bar{x}_2$
9.25	13.25	14.75

$\omega = 5.02$

We can declare a significant difference in average attention spans between "no breakfast" and "light breakfast", but not between the other pairs.

Since the difference between 9.25 and 13.25 is less than ω = 5.02, there is no significant difference. There is a difference between population means 1 and 2 however.

There is no difference between 13.25 and 14.75.