EMSE 4765: DATA ANALYSIS
For Engineers and Scientists

Session 13: One-Way Analysis of Variance (ANOVA)

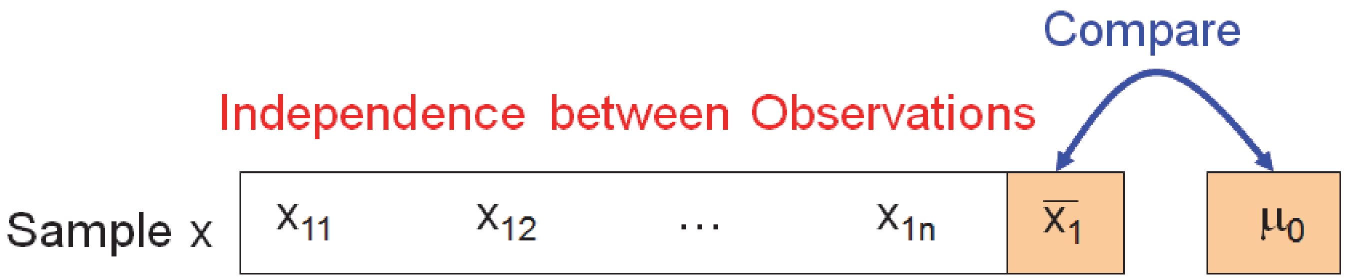
Version: 4/12/2021

Lecture Notes by: J. René van Dorp[1]

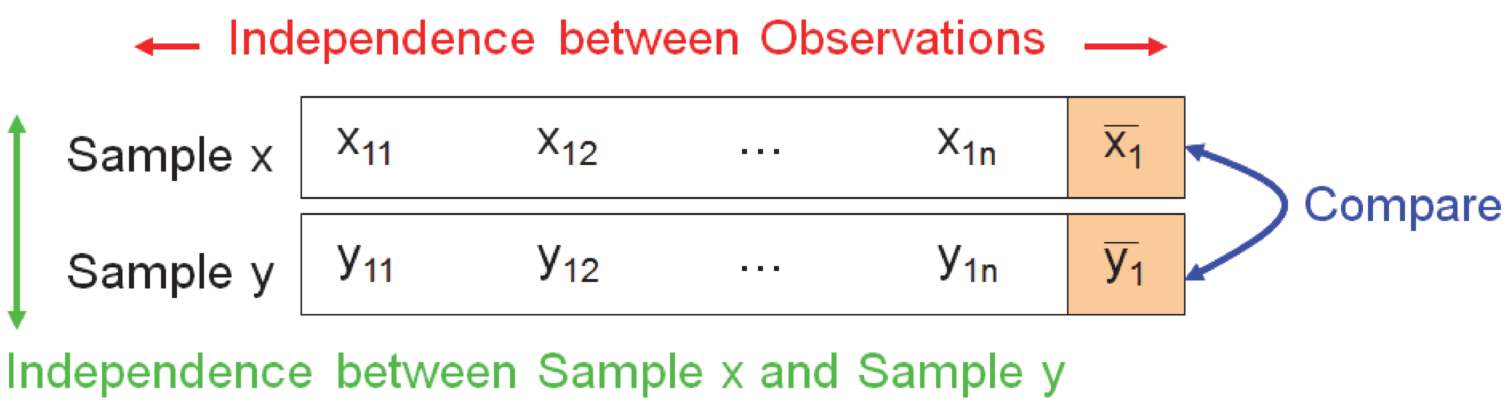
www.seas.gwu.edu/~dorpjr

1 Department of Engineering Management and Systems Egineering, School of Engineering and Applied Science, The George Washington University, 800 22nd Street, N.W., Suite 800, Washington D.C. 20052. E-mail: dorpjr@gwu.edu.

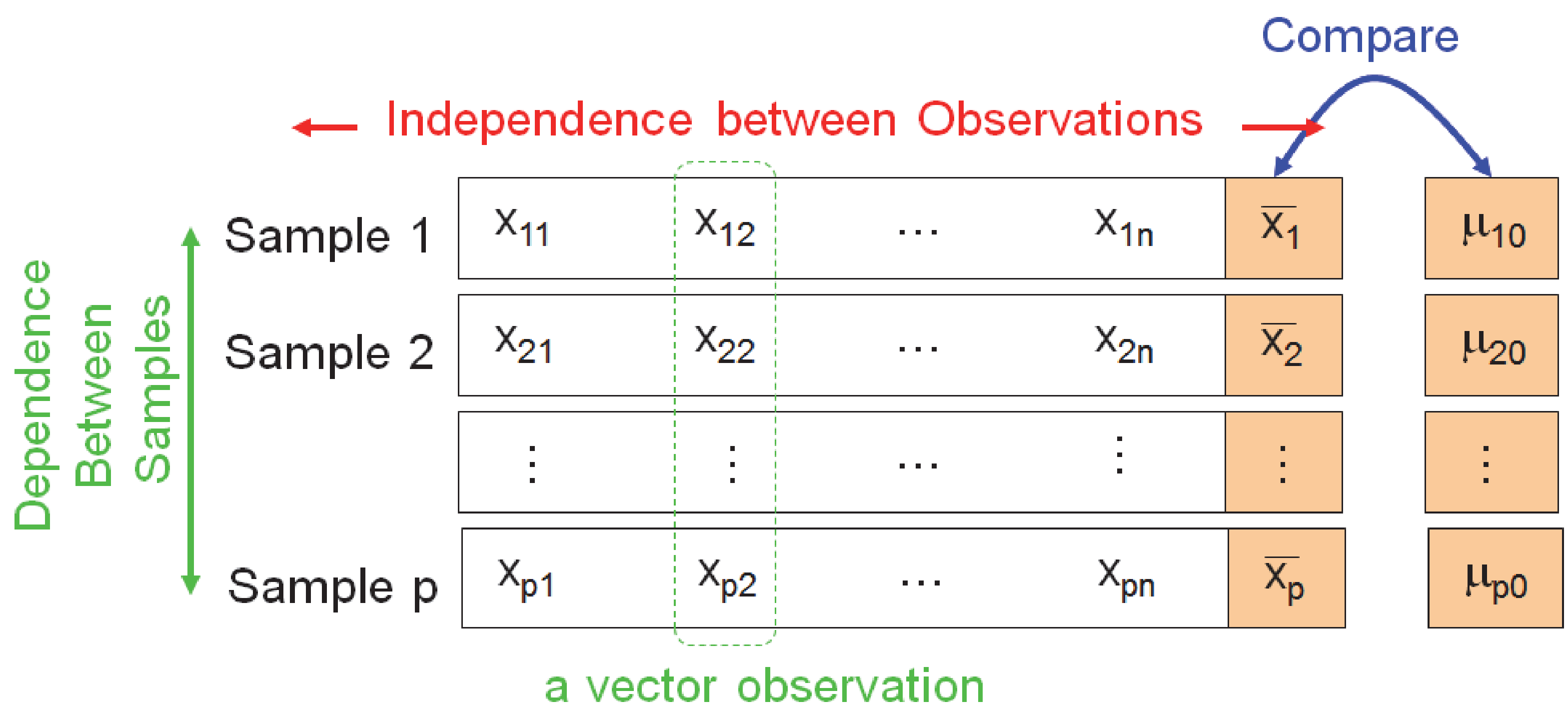
- Univariate T-test: $H_0 : \mu = \mu_0,\ H_1 : \mu \neq \mu_0$. (Scalar μ_0 is specified)



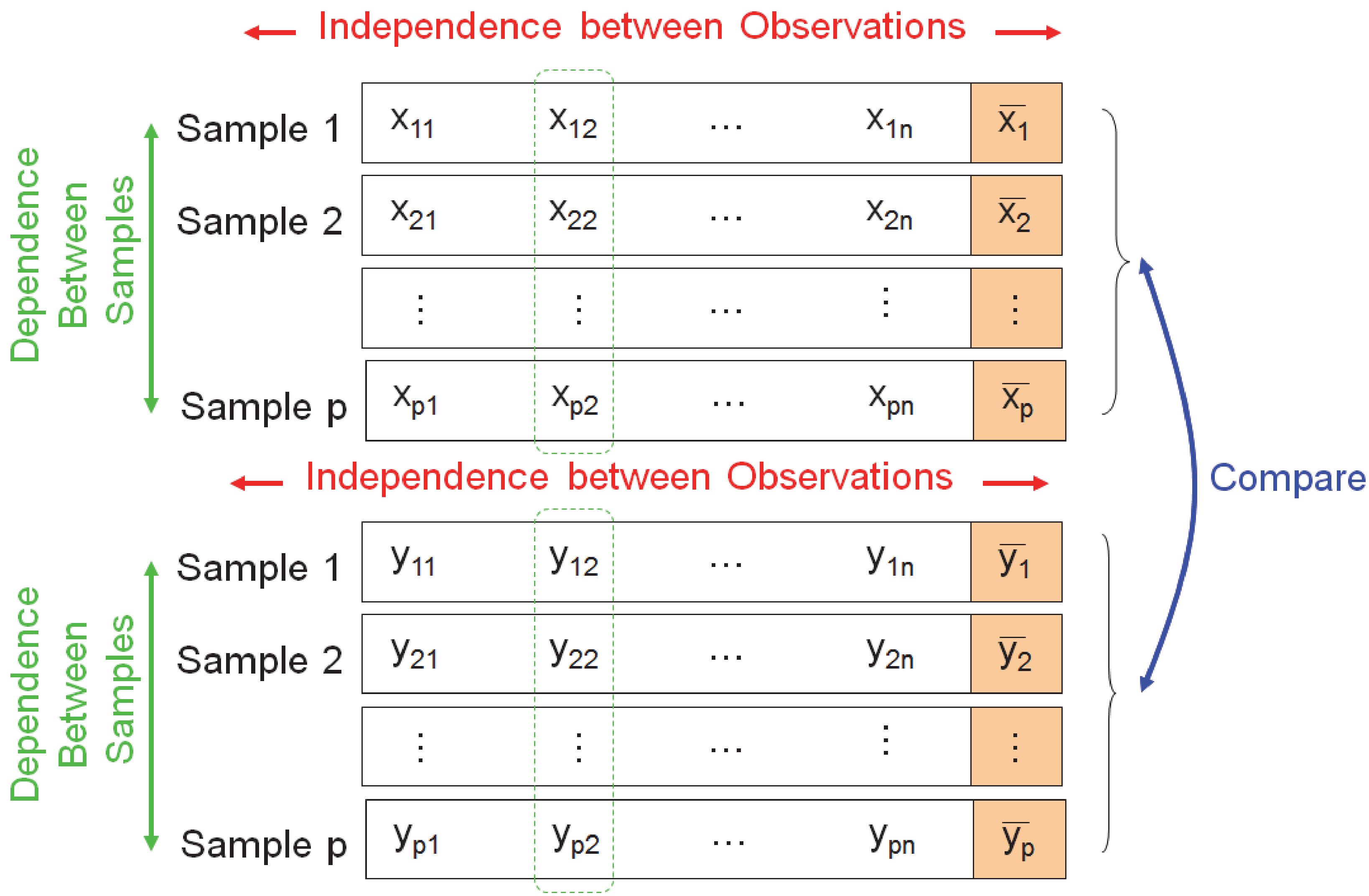
- Two Sample Univariate T-test: $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 \neq \mu_2$.



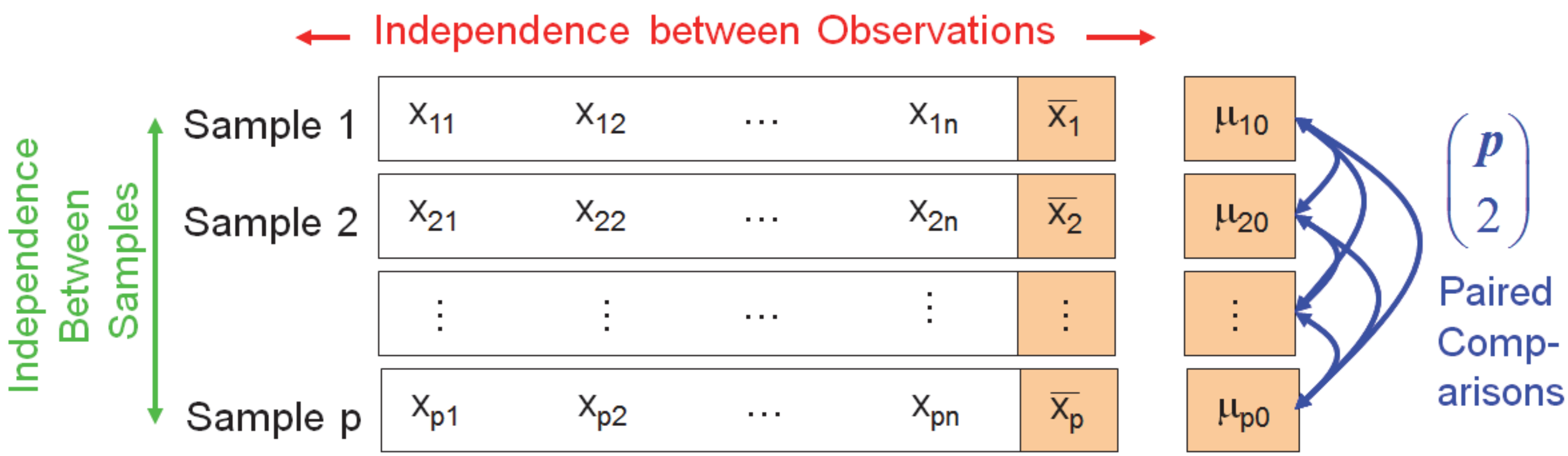
- Hotelling T^2-test: $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. (Vector $\boldsymbol{\mu}_0$ is specified)



- Two-Sample Hotelling T^2-test: $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2},\ H_1 : \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$.



- Objective of Analysis of Variance (ANOVA):



Tensile Strength Example: The tensile strength of synthetic fiber used to make cloth for men's shirts is of interest to a manufacturer. It is suspected that **the strength** is affected by **the percentage of cotton in the fiber**. Five levels of cotton percentages are of interest, 15%, 20%, 25%, 30%, and 35%. Five observations are to be taken at each level of cotton percentage, and the **25 total observations are to be run in random order**:

$$\text{Total number of paired comparisons: } \binom{5}{2} = 10$$

- It seems that this problem can be solved by performing 10 two-sample t tests on **all possible pairs**. However, this solution could lead to **a considerable distortion in the type I error**.

Tensile Strength Example:
We have 10 possible pairs. If the probability of **failing to reject the null hypothesis** (i.e. there is no difference between a pair) for all 10 tests is $1-\alpha = 0.95$, then the probability of correctly failing to reject the null hypothesis for all 10 tests (**i.e. there is no difference between the 10 samples**) equals:

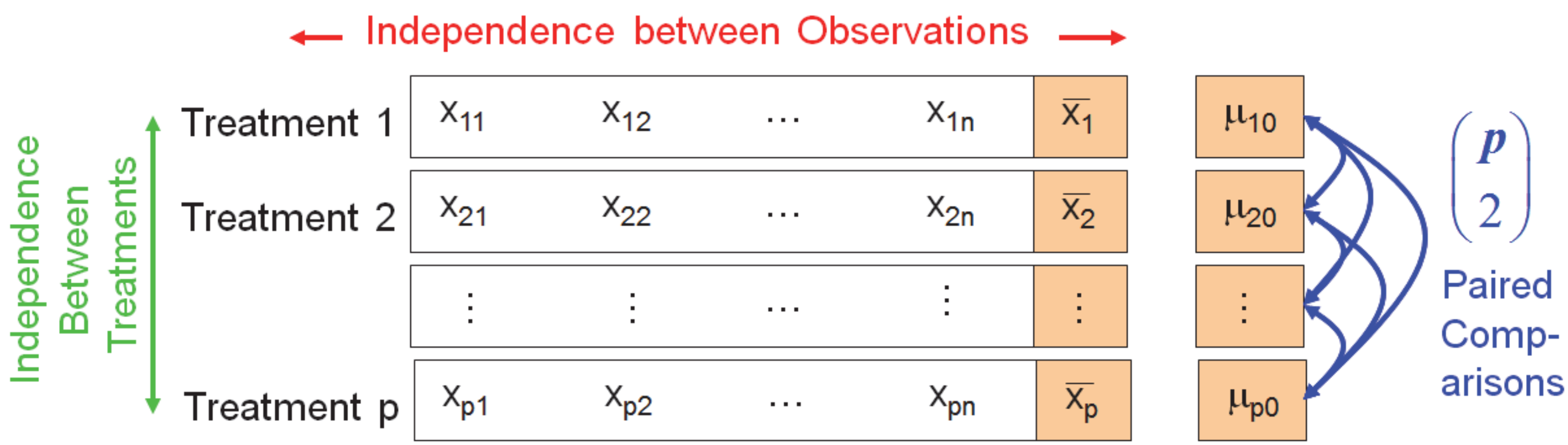
$$Pr(\text{"no differences between pairs"} \mid \text{no difference between pairs}) = 0.95 \ \Rightarrow$$
$$Pr(\text{"no differences between 10 pairs"} \mid \text{no difference between 10 pairs}) = (0.95)^{10} \approx 0.60$$

If the tests are independent (which is questionable).

$$Pr(\text{"A difference in at least one of the 10 pairs"} \mid \text{no difference between 10 pairs}) \\ = 1-(0.95)^{10} \approx 0.40$$

Thus **we observe a substantial increase of Type I error** from 5% to 40%.

A more appropriate procedure for **testing equality of several means** in the setting above is through an **ANALYSIS OF VARIANCE,** by assuming that **the variance within each sample is the same in the ANOVA model**. In ANOVA samples are referred to as **"treatments"**. Hence, we have:



$$X_{ij} = \mu + \tau_i + \epsilon_{ij}, \begin{cases} i = 1, \ldots, p \\ j = 1, \ldots, n \end{cases}$$

μ: a parameter common to all treatments called *the overall mean*

τ_i : a parameter unique to the i-th treatment called *the treatment effect*,

ϵ_{ij} : **a random error component**, $\epsilon_{ij} \sim N(0, \boldsymbol{\sigma})$ for all i, j and $i.i.d.$

- **Thus the mean of treatment i** equals the sum of **the overall mean** + **the i-th treatment effect:**

$$E[X_{ij}] = \mu + \tau_i,\ i = 1, \ldots, p,\ \Rightarrow \text{ one can choose } \mu \text{ such that: } \sum_{i=1}^{p} \tau_i \equiv 0.$$

- We are interested in testing the equality of the p treatment means as follows:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_p,\ H_1 : \mu_i \neq \mu_j, \text{ for a least one } i, j$$

- **If H_0 is true, all treatments have common mean μ.** An equivalent way to write the hypothesis test is in terms of the treatment effects τ_i is as follows :

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_p = 0,\ H_1 : \tau_i \neq 0, \text{ for a least one } i$$

Notation:

$$x_{i\bullet} = \sum_{j=1}^{n} x_{ij},\ \overline{x}_{i\bullet} = \frac{1}{n} x_{i\bullet}, \text{ assuming } n = \text{equal \# observations in each treatment}$$

$$x_{\bullet\bullet} = \sum_{j=1}^{p} \sum_{j=1}^{n} x_{ij}, \quad \overline{x}_{\bullet\bullet} = \frac{1}{N} x_{\bullet\bullet},\ N = np\ (\equiv \text{ total number of observations})$$

- **ANALYSIS OF VARIANCE (ANOVA) TABLE:**

Source	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_{Treatments}$	$p-1$	$MS_{Treatment}$	$\frac{MS_{Treatments}}{MS_E}$
Error (within treatments)	SS_E	$N-p$	MS_E	
Total	SS_T	$N-1$		

$$SS_T = \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij} - \overline{x}_{\bullet\bullet})^2, \; SS_E = \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij} - \overline{x}_{i\bullet})^2$$

$$SS_{Treatments} = \sum_{i=1}^{p} \sum_{j=1}^{n} (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet})^2 = n \times \sum_{i=1}^{p} (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet})^2,$$

$$SS_T = SS_E + SS_{Treatments}$$

Tensile Strength Example: The tensile strength of synthetic fiber used to make cloth for men's shirts is of interest to a manufacturer. It is suspected that the strength is affected by the percentage of cotton in the fiber. Five levels of cotton percentage are of interest, 15%, 20%, 25%, 30%, and 35%. Five observations are to be taken at each level of cotton percentage, and the 25 total observations are to be run in random order.

Table: Tensile Strength of Synthetic Fiber (lb/in.^{2})

Percentage of	Observations					
Cotton	1	2	3	4	5	$x_{i\bullet}$
15%	7	7	15	11	9	49
20%	12	17	12	18	18	77
25%	14	18	18	19	19	88
30%	19	25	22	19	23	108
35%	7	10	11	15	11	54
					$x_{\bullet\bullet}$	376

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
$SS_{Treatments}$	475.76	4	118.94	14.76	9.13E-06
SS_E	161.2	20	8.06		
SS_T	636.96	24			

- p-value $< \alpha$ for $\alpha \in \{1\%, 5\%, 10\%\} \Rightarrow$ Reject H_0 for all these α's

Conclusion: At least one of the treatment means differs!

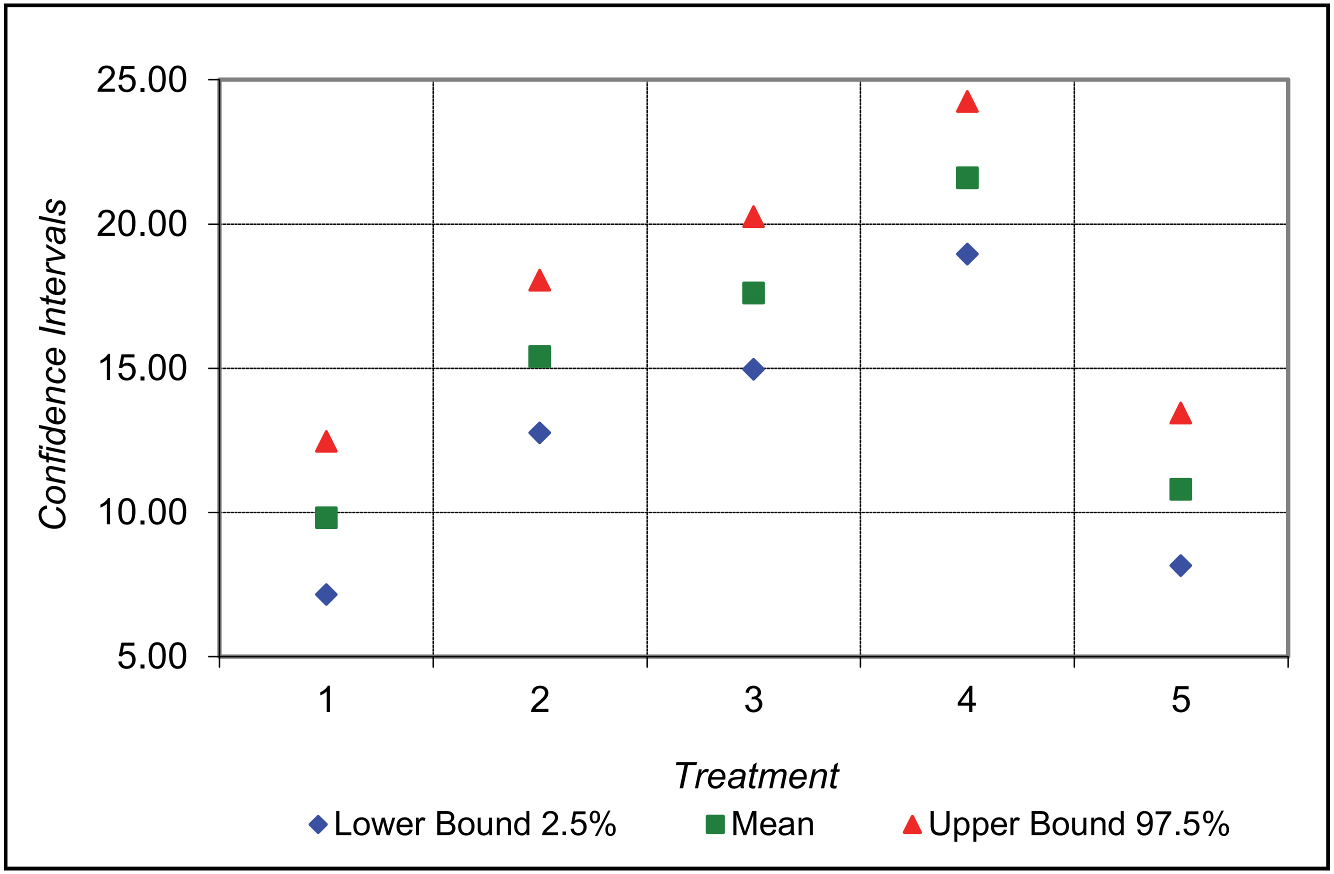
- **Estimation of parameters** can be done using the least squares approach (similar to **linear regression analysis**). Recall: $X_{ij} = \mu + \tau_i + \epsilon_{ij}$

$$\widehat{\mu} = \overline{X}_{\bullet\bullet}, \; \widehat{\tau}_i = \overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}, i = 1, \ldots, p,$$

$$\widehat{\mu_i} = \widehat{\mu} + \widehat{\tau}_i = \overline{X}_{i\bullet}, \; \widehat{\sigma}^2 = MS_E = SS_E/(N-p)$$

- **$100(1-\alpha)\%$ confidence intervals treatment means μ_i:**

$$\overline{X}_{i\bullet} \pm t_{\alpha/2,N-p}\sqrt{MS_E/n}$$



Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Treatment	4	475.8	118.940	14.76	0.000
Error	20	161.2	8.060		
Total	24	637.0			

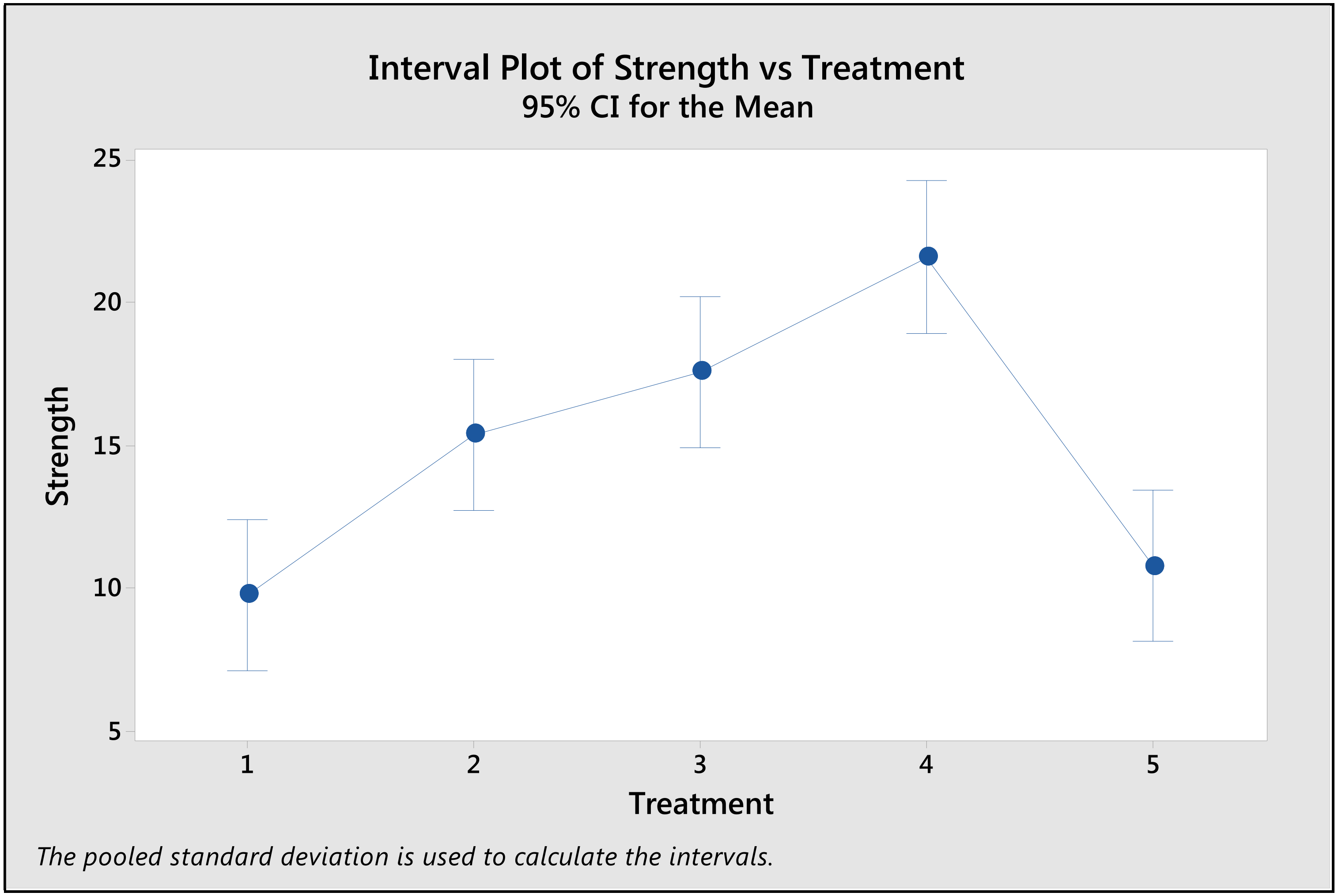
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.83901	74.69%	69.63%	60.46%

Means

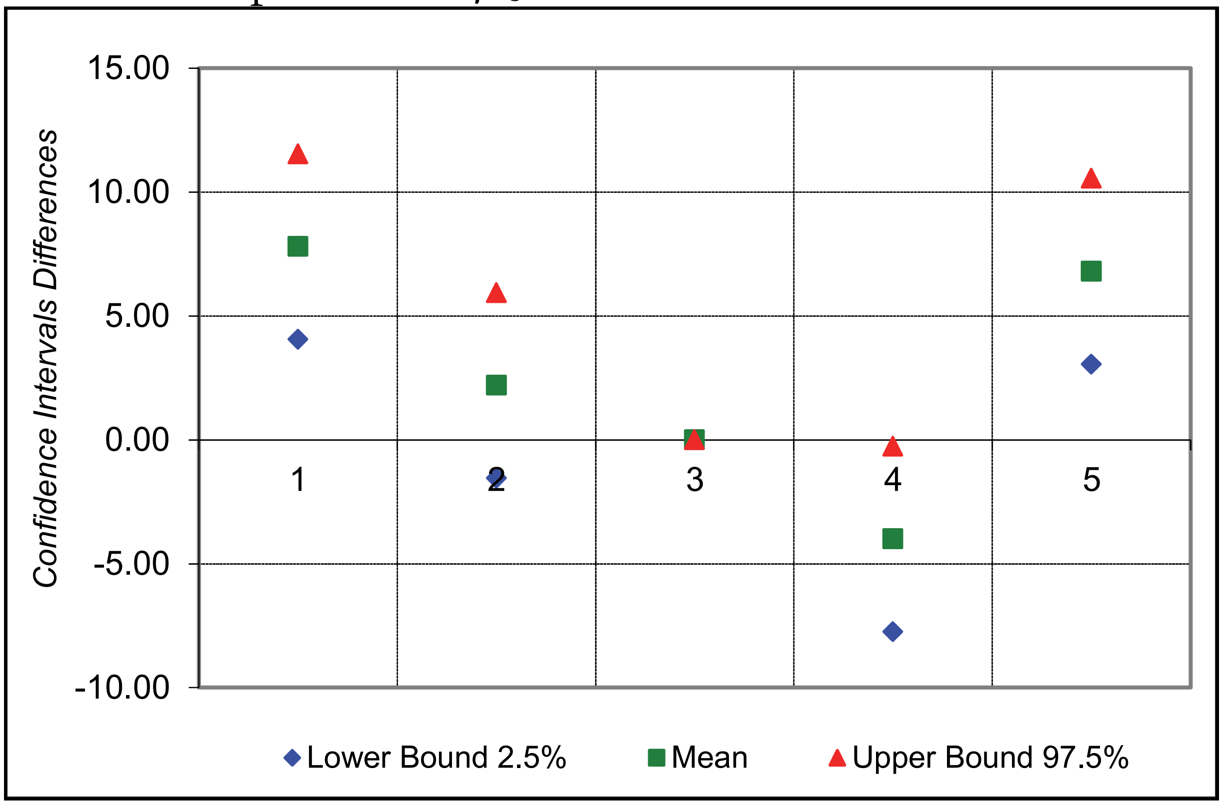
Treatment	N	Mean	StDev	95% CI
1	5	9.80	3.35	(7.15, 12.45)
2	5	15.40	3.13	(12.75, 18.05)
3	5	17.600	2.074	(14.952, 20.248)
4	5	21.60	2.61	(18.95, 24.25)
5	5	10.80	2.86	(8.15, 13.45)

MINITAB Plot of 95% Confidence Intervals



- **$100(1-\alpha)\%$ confidence intervals difference treatment means $\mu_i - \mu_k$:** $\overline{X}_{i\bullet} - \overline{X}_{k\bullet} \pm t_{\alpha/2,N-p}\sqrt{2MS_E/n}$

Comparison of μ_3 to other treatment means.



Conclusion: **We only fail to reject the null-Hypothesis that $\mu_3 = \mu_2$!!!**

- It was assumed in the model that the error terms **ϵ_{ij} are normal distributed with a mean 0 and a variance σ^2.**

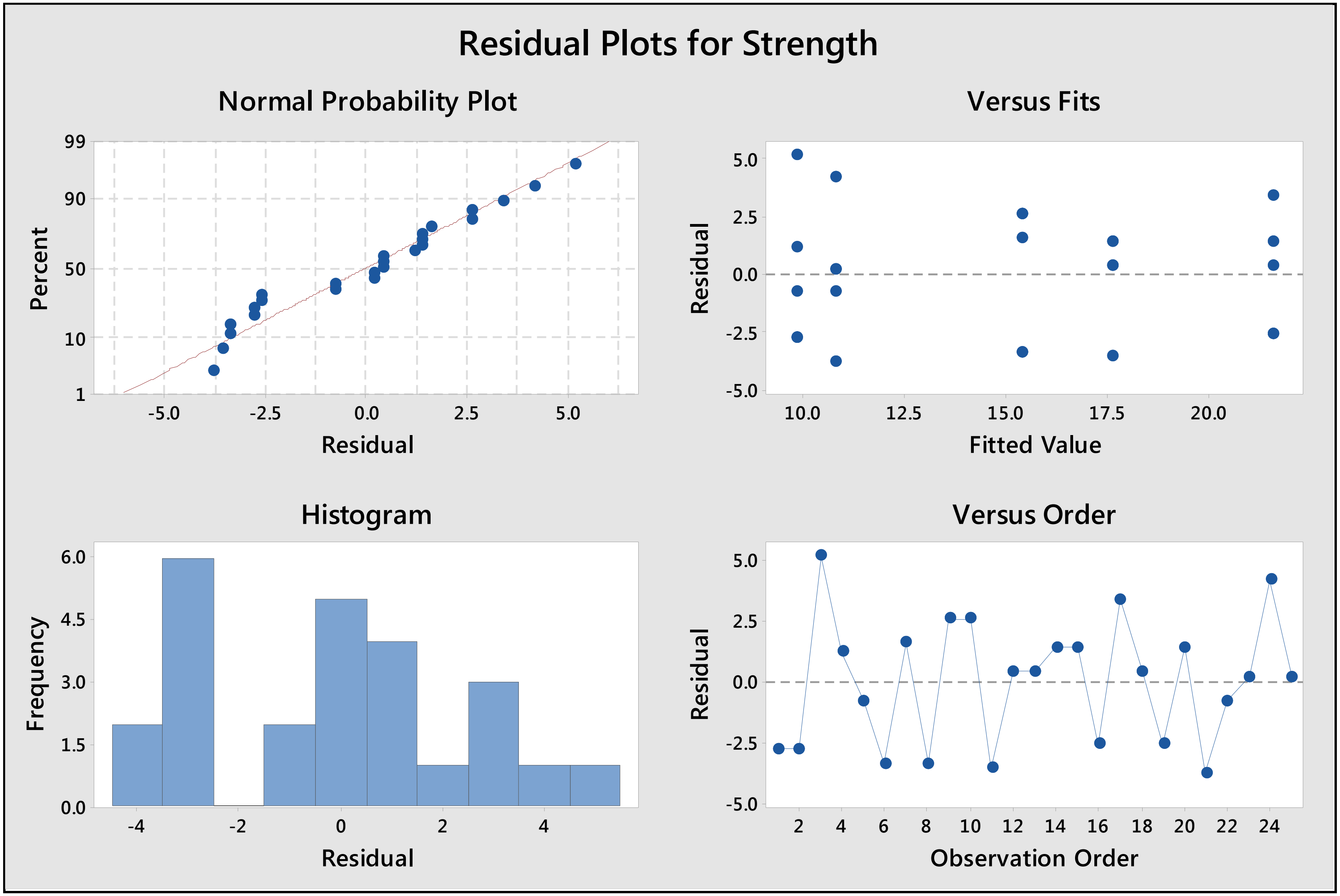
- The normality assumptions of the residuals ϵ_{ij} can be checked via **a normal probability plot.**

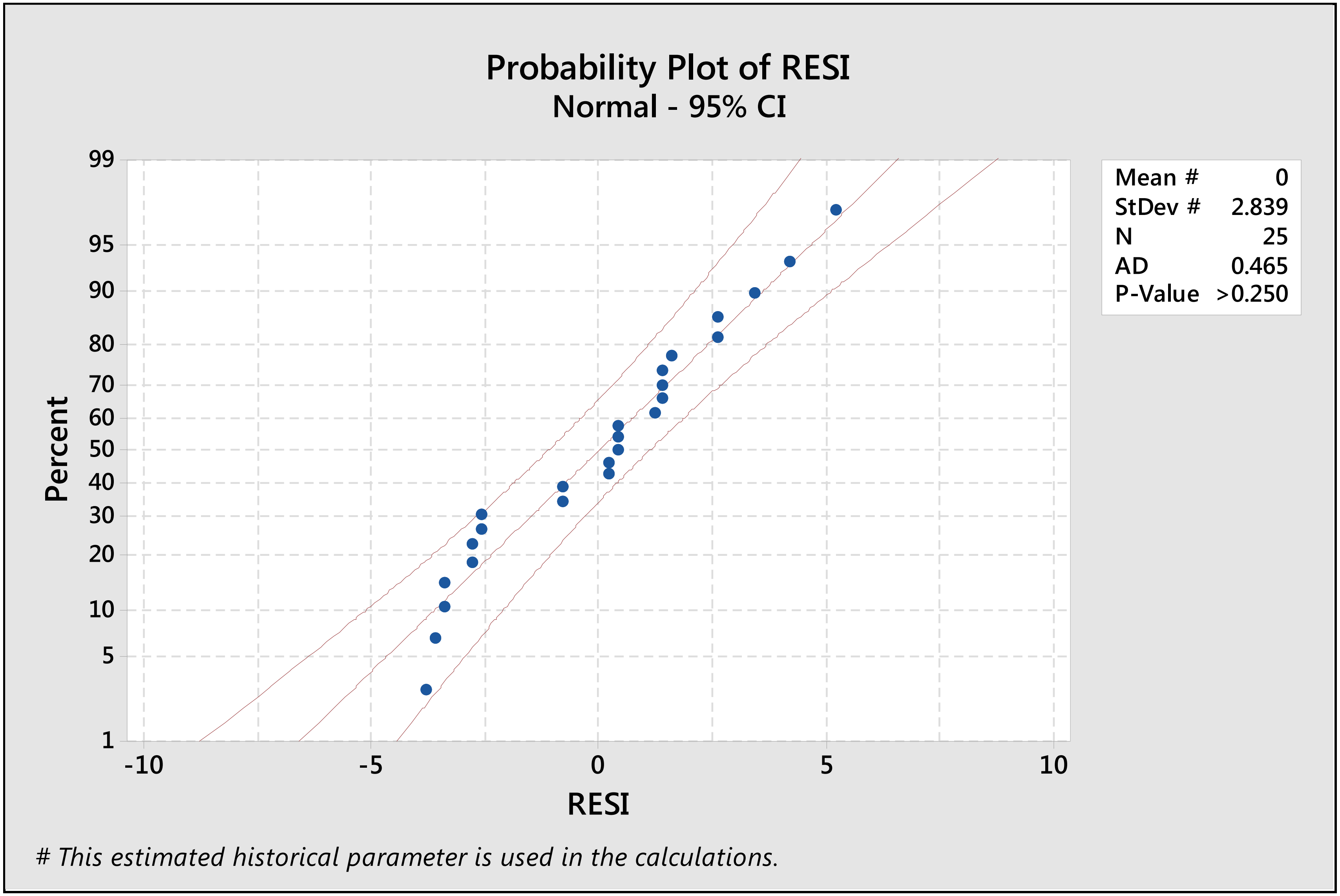
- It is important to recognize that we are testing **the equality of treatment means by testing for the equality of variances.**

- The required assumption that allows us to do this is that **the variance of the error terms ϵ_{ij} is constant across treatments $i = 1, \ldots, p$.**

- The assumption of equality of variance may be **visually verified by plotting the residuals of each treatment against one another.**

- Alternatively, we may also use Bartlett's test, to test for equality of variance across treatment.


Residual Plots for Strength
Normal Probability Plot
Percent
99
90
50
10
1
-5.0
-2.5
0.0
2.5
5.0
Residual
Versus Fits
Residual
5.0
2.5
0.0
-2.5
-5.0
10.0
12.5
15.0
17.5
20.0
Fitted Value
Histogram
Frequency
6.0
4.5
3.0
1.5
0.0
-4
-2
0
2
4
Residual
Versus Order
Residual
5.0
2.5
0.0
-2.5
-5.0
2
4
6
8
10
12
14
16
18
20
22
24
Observation Order


Probability Plot of RESI
Normal - 95% CI
Mean # 0
StDev # 2.839
N 25
AD 0.465
P-Value >0.250
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-10
-5
0
5
10
RESI
This estimated historical parameter is used in the calculations.

Bartlett's Test for Equality of Variance across treatments

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_p^2,\ H_1 : \text{Not true for at least one } \sigma_i^2$$

Test Statistic:

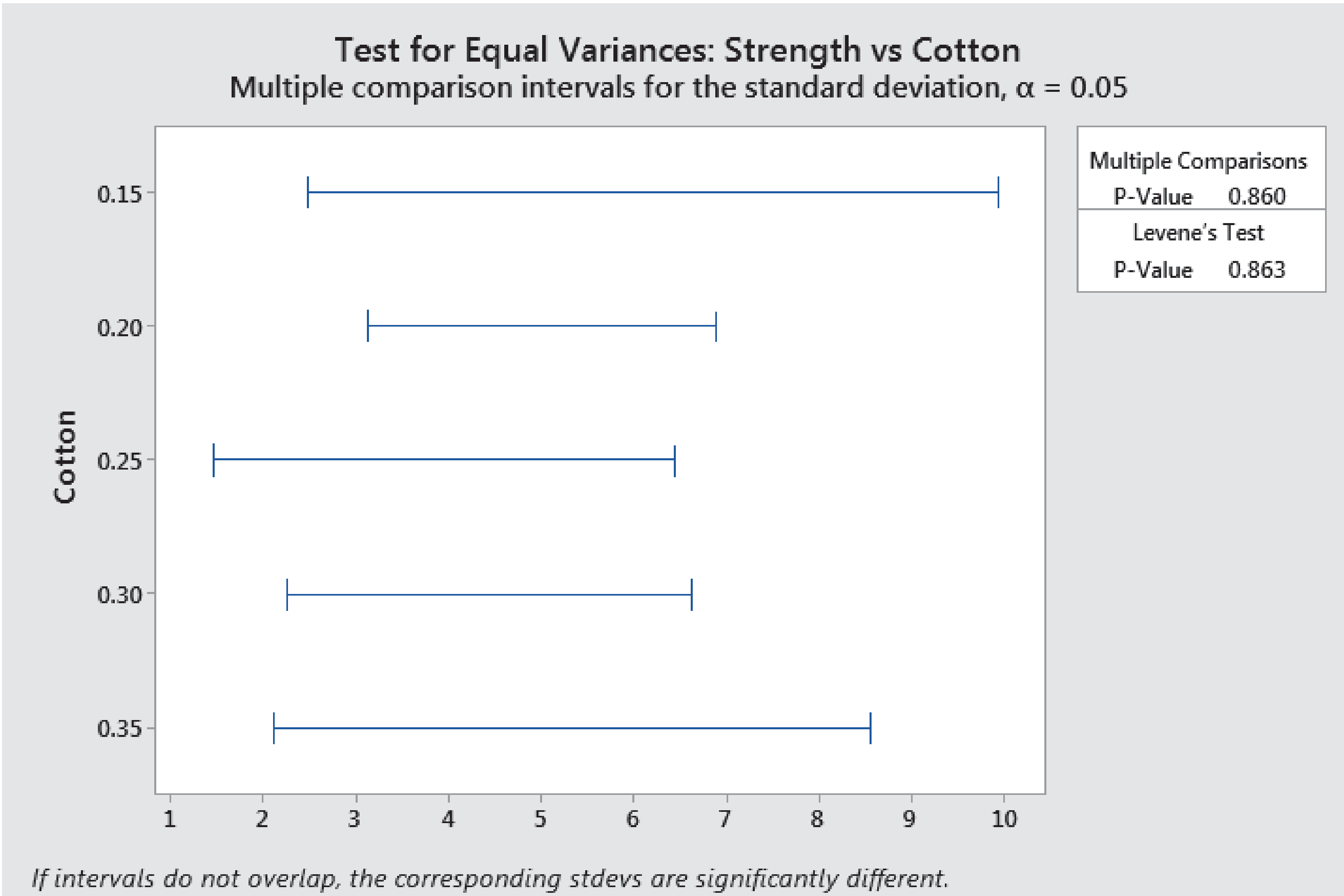
$$\chi_0^2 = \frac{q}{c} \sim \chi_{p-1}^2$$

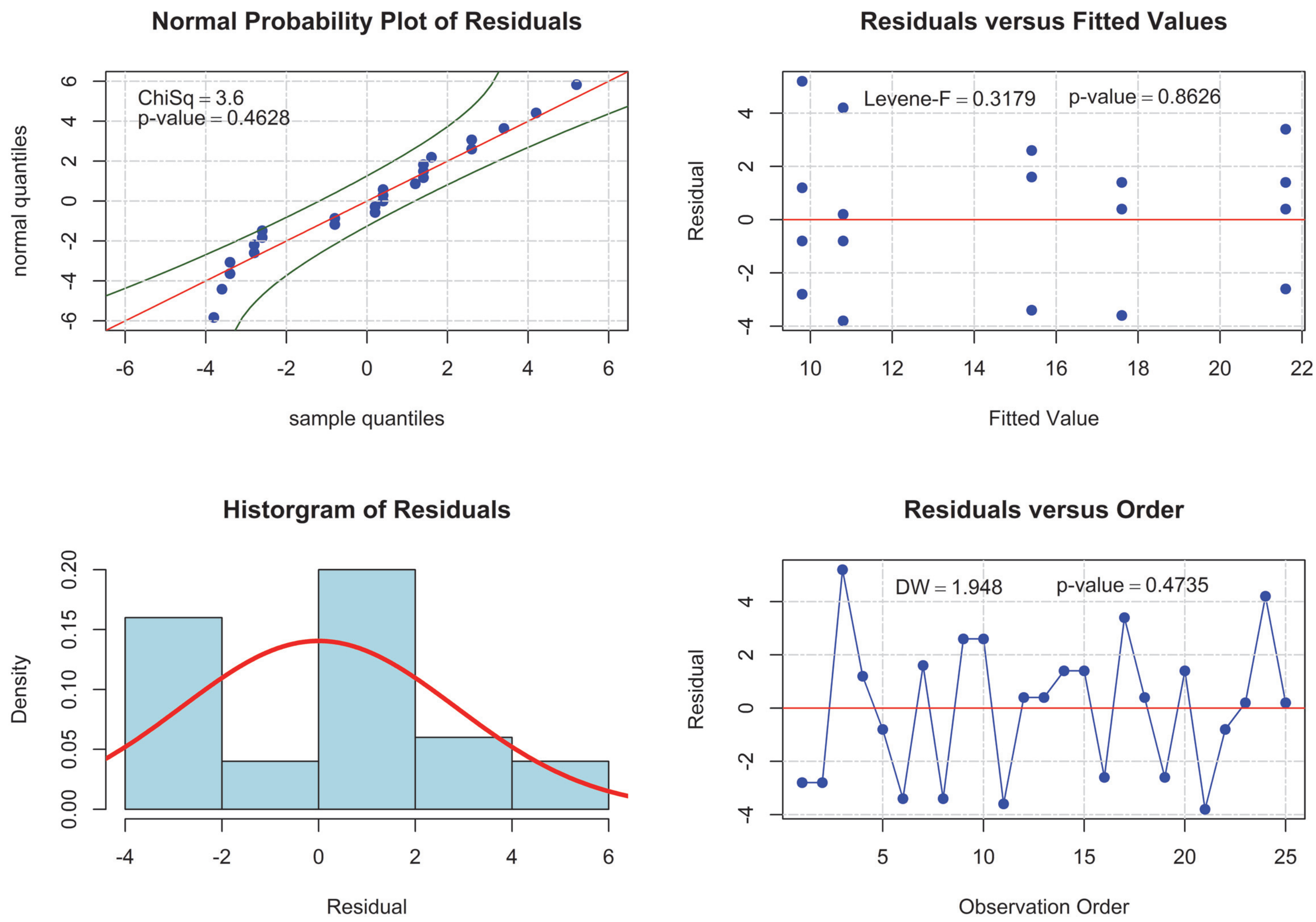
$$q = (N-p) \times Ln\left(S_{pooled}^2\right) - (n-1)\sum_{i=1}^{p} Ln\ S_i^2,\ N = n \times p$$

$$S_{pooled}^2 = \frac{1}{p}\sum_{i=1}^{p} S_i^2,\ c = 1 + \frac{1}{3(p-1)}\left[\frac{p}{(n-1)} - \frac{1}{(N-p)}\right]$$

Tensile Strength Example:

$N = 25,\ p = 5, S_{pooled}^2 \approx 8.06, q \approx 1.03, c \approx 1.10, \chi_0^2 \approx 0.93,$
$p - value \approx 0.92$. Conclusion: Fail to Reject the null-Hypothesis


Test for Equal Variances: Strength vs Cotton
Multiple comparison intervals for the standard deviation, α = 0.05
Multiple Comparisons
P-Value 0.860
Levene's Test
P-Value 0.863
Cotton
0.15
0.20
0.25
0.30
0.35
1
2
3
4
5
6
7
8
9
10
If intervals do not overlap, the corresponding stdevs are significantly different.

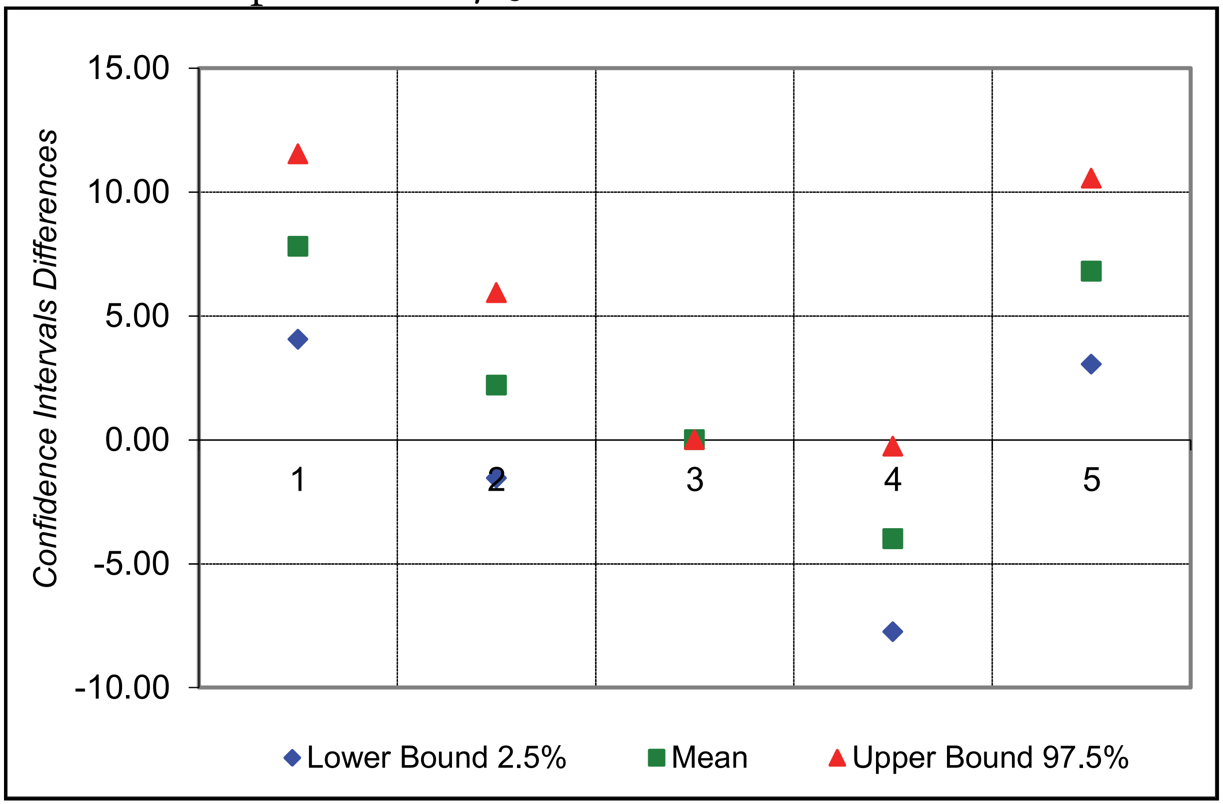


Same Analysis in "Tensile_Strength_Analysis.R"

- **$100(1-\alpha)\%$ confidence intervals difference treatment means $\mu_i - \mu_k$:**

$$\overline{X}_{i\bullet} - \overline{X}_{k\bullet} \pm t_{\alpha/2,N-p}\sqrt{2MS_E/n}$$

Comparison of μ_3 to other treatment means.



Conclusion: We only fail to reject the null-Hypothesis that $\mu_3 = \mu_2$!!!

- Using the $100(1-\alpha)\%$ confidence intervals for differences of treatment means of $\mu_3 - \mu_k,\ k = 1, 2, 4$ and 5 we tested the hypotheses:

$$H_0 : \mu_3 = \mu_k,\ H_1 : \mu_3 \neq \mu_k, k = 1, 2, 4 \text{ and } 5$$

- Hypothesis test could be tested by investigating **an appropriate linear combination of treatment totals**, for example:

Is $X_{3\bullet} - X_{k\bullet} = 0$? **(since n is same for each treatment).**

If we suspect that **the combined average of cotton percentages 1 and 3** did not differ from **the combined average of cotton percentages 4 and 5**, then the hypothesis to be tested is:

$$H_0 : \mu_1 + \mu_3 = \mu_4 + \mu_5; H_1 : \mu_1 + \mu_3 \neq \mu_4 + \mu_5$$

which implies **the following linear combination of treatment totals**:

$$X_{1\bullet} + X_{3\bullet} = X_{4\bullet} + X_{5\bullet}? \Leftrightarrow X_{1\bullet} + X_{3\bullet} - X_{4\bullet} - X_{5\bullet} = 0?$$

- A linear combination of treatments totals $C = \sum_{i=1}^{p} c_i X_{i\bullet}$ such that $\sum_{i=1}^{p} c_i = 0$ is called **a contrast or a contrast sum.**

- **The sum of squares for any contrast sum** equals:

$$SS_C = \Big(\sum_{i=1}^{p} c_i X_{i\bullet}\Big)^2 \Big/ \Big(n \times \sum_{i=1}^{p} c_i^2\Big),$$

where n is the number of observations in Treatment i and the contrast SS_C has **a single degree of freedom** and **hence $MS_C = SS_C/1 = SS_C$.**

- Therefore,

$$\frac{MS_C}{MS_E} = \frac{SS_C}{SS_E/N-p} \sim F_{1,N-p}$$

- **Conclusion:** Many important comparisons regarding treatment means, or their combinations, can be conducted using these contrast sums.

- **In addition,** two contrasts $\{c_i\}$ and $\{d_i\}$ are orthogonal when $\sum_{i=1}^{p} c_i d_i = 0$

- For p treatments, **a set of $(p-1)$ orthogonal contrasts partitions the sum of squares due to treatments** into $(p-1)$ independent single degree-of-freedom **contrast sum of square components.**

- **Due to orthogonalility of contrasts,** the contrast tests are independent. Hence, if the Type 1 error of each individual contrast test is $(1-\alpha)$, the Type 1 error of the $(p-1)$ orthogonal contrast tests equals $(1-\alpha)^{p-1}$.

- **Without orthogonality of these contrasts**, we cannot say anything about the combined Type 1 error probability.

- **There are many ways to choose orthogonal contrast coefficient for a given set of treatments**. For example, if there are $p=3$ treatments, with Treatment 1 being a control and Treatments 2 and 3 change levels of the factor of interest, then appropriate orthogonal contrast might be as follows

	Treatment 1(Control)	Treatment 2(Level 1)	Treatment 3(Level2)
c_i	-2	1	1
d_i	0	-1	1

- **Contrast coefficients must be chosen prior to running the experiment and prior to examining the data.**

Tensile Strength Example:

$H_0 : \mu_4 = \mu_5$ $\qquad C_1 = -X_{4\bullet} + X_{5\bullet}$

(Compares the average of Treatment 4 and with that of Treatment 5)

$H_0 : \mu_1 + \mu_3 = \mu_4 + \mu_5$ $\qquad C_2 = X_{1\bullet} + X_{3\bullet} - X_{4\bullet} - X_{5\bullet}$

(Compares the averages of Treatments 1 and 3 with that of Treatments 4 and 5)

$H_0 : \mu_1 = \mu_3$ $\qquad C_3 = X_{1\bullet} - X_{3\bullet}$

(Compares the average of Treatment 1 and with that of Treatment 3)

$H_0 : 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$ $\qquad C_4 = -X_{1\bullet} + 4X_{2\bullet} - X_{3\bullet} - X_{4\bullet} - X_{5\bullet}$

(Compares the average of Treatments 2 with that of Treatments 1, 3, 4 and 5)

Notice that the contrast coefficients are orthogonal!

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
C1	291.6	1	291.6	36.18	7.01E-06
C2	31.25	1	31.25	3.88	6.30%
C3	152.1	1	152.1	18.87	3.15E-04
C4	0.81	1	0.81	0.10	75.5%
$SS_{Treatments}$	475.76	4	118.94	14.76	9.13E-06
SS_E	161.2	20	8.06		
SS_T	636.96	24			

Conclusion at a 5% Significance Level:

- There are differences between the treatment means.
- Furthermore, differences are observed between Treatment 4 and Treatment 5 (C1), and differences of Treatment 1 and Treatment 3 (C3).
- No difference is observed between the average sum of Treatmens 1 and 3 and the average sun of Treatments 4 and 5 combined (C2).
- No difference is observed between the average of Treatment 2 and the average sum of Treatments 1, 3, 4 and 5 (C4)

Same Analysis in "Tensile_Strength_Analysis.R"

SOURCE	SS	df	MS	F	p-value
C1	291.60	1	291.60	36.18	0.00 %
C2	31.25	1	31.25	3.88	6.30 %
C3	152.10	1	152.10	18.87	0.03 %
C4	0.81	1	0.81	0.10	75.45 %
Treatments	475.76	4	118.94	14.76	0.00 %
Error	161.20	20	8.06		
Total	636.96	24			

- **Total sum of squares:**

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{n}(x_{ij}-\overline{x}_{\bullet\bullet})^2 = \sum_{i=1}^{p}\sum_{j=1}^{n}\left[(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})+(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})\right]^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{n}\left[(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})^2+2(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})+(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})^2\right]$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{n}(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})^2+2\sum_{i=1}^{p}\sum_{j=1}^{n}(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})+\sum_{i=1}^{p}\sum_{j=1}^{n}(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{n}(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})^2+2\sum_{i=1}^{p}(\boldsymbol{\overline{x}_{i\bullet}}-\boldsymbol{\overline{x}_{\bullet\bullet}})\sum_{j=1}^{n}(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}})+n\sum_{i=1}^{p}(\overline{x}_{i\bullet}-\overline{x}_{\bullet\bullet})^2$$

- **Cross product term equals zero, because:**

$$\sum_{j=1}^{n}(\boldsymbol{x_{ij}}-\boldsymbol{\overline{x}_{i\bullet}}) = \sum_{j=1}^{n}x_{ij}-n\,\overline{x}_{i\bullet} = n\,\overline{x}_{i\bullet}-n\,\overline{x}_{i\bullet}=0$$

- **Total sum of squares:**

$$SS_T = \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij} - \overline{x}_{\bullet\bullet})^2 = \sum_{i=1}^{p} \sum_{j=1}^{n} (\boldsymbol{x_{ij}} - \boldsymbol{\overline{x}_{i\bullet}})^2 + n \sum_{i=1}^{p} (\boldsymbol{\overline{x}_{i\bullet}} - \boldsymbol{\overline{x}_{\bullet\bullet}})^2$$

or

$$SS_T = SS_E + SS_{Treatments}$$

where;

$$SS_E = \sum_{i=1}^{p} \sum_{j=1}^{n} (\boldsymbol{x_{ij}} - \boldsymbol{\overline{x}_{i\bullet}})^2$$

The sum of squares within an treatment i, summed over all treatments

$$SS_{Treatments} = n \sum_{i=1}^{p} (\boldsymbol{\overline{x}_{i\bullet}} - \boldsymbol{\overline{x}_{\bullet\bullet}})^2$$

The sum of squares of treatment means $\overline{x}_{i\bullet}$ against the overall mean $\overline{x}_{\bullet\bullet}$

- The sample variance in the i-th treatment equals:

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{x_{ij}} - \boldsymbol{\overline{x}_{i\bullet}})^2 \Leftrightarrow (n-1)S_i^2 = \sum_{j=1}^{n}(\boldsymbol{x_{ij}} - \boldsymbol{\overline{x}_{i\bullet}})^2$$

- These S_i^2's can be combined to get an estimate of pooled variance as follows

$$\frac{1}{p}\sum_{i=1}^{p} S_i^2 = \frac{(n-1)\sum_{i=1}^{p} S_i^2}{(n-1)p} = \frac{(n-1)S_1^2 + \ldots + (n-1)S_p^2}{np-p} =$$

$$= \frac{\sum_{i=1}^{p}\left[\sum_{j=1}^{n}(\boldsymbol{x_{ij}} - \boldsymbol{\overline{x}_{i\bullet}})^2\right]}{N-p} = \frac{SS_E}{N-p}$$

- Recalling $\epsilon_{ij} \sim N(0,\sigma)$ and denoting :

$$MS_E = \frac{SS_E}{N-p} \Rightarrow E[MS_E] = E\left[\frac{1}{p}\sum_{j=1}^{p} S_i^2\right] = \frac{1}{p}\sum_{j=1}^{p}\sigma^2 = \sigma^2$$

- Recalling $\epsilon_{ij} \sim N(0, \sigma)$, observe that the estimators of the i-th treatment means

$$\overline{X}_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

are all random variables $i = 1, \ldots, p$ with common variance $V[\overline{X}_{i\bullet}] = \sigma^2/n$.

- **If the treatments means are all equal**, then $E[\overline{X}_{i\bullet}] = \mu$, we have that

$$\overline{X}_{\bullet\bullet} = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{n} X_{ij} = \frac{1}{p}\sum_{i=1}^{p}\left[\frac{1}{n}\sum_{j=1}^{n} X_{ij}\right] = \frac{1}{p}\sum_{i=1}^{p}\overline{X}_{i\bullet} \Rightarrow E[\overline{X}_{\bullet\bullet}] = \mu$$

and thus $\overline{X}_{\bullet\bullet}$ is an unbiased estimate of the **common** treatment mean μ.

- Hence, **if the treatments means are all equal,**

$$E\left[\frac{1}{p-1}\sum_{i=1}^{p}(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2\right] = \frac{\sigma^2}{n} \Leftrightarrow E\left[\frac{n}{p-1}\sum_{i=1}^{p}(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2\right] = \sigma^2$$

- Denoting:

$$MS_{Treatments} = \frac{n}{p-1}\sum_{i=1}^{p}(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2 = \frac{n\sum_{i=1}^{p}(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet})^2}{p-1} = \frac{SS_{Treatment}}{p-1}$$

we have, **when all the treatments means are equal,**

$$E[MS_{Treatments}] = \sigma^2.$$

- It can be shown that **if the treatment means are not necessarily equal,**

$$E[MS_{Treatments}] = \sigma^2 + \frac{n}{p-1}\sum_{i=1}^{p}\tau_i^2.$$

- We have shown that (**regardless of the value of the treatment means**)**:**

$$E[MS_E] = E[\frac{SS_E}{N-p}] = \sigma^2.$$

- Conclusion: If the value of $MS_{Treatments}$ **is close to that of** MS_E this can be seen as an **indication that the treatment means are equal.** Moreover, if the treatment means are different it follows that $MS_{Treatments}$ is larger than MS_E.

- But how large does $MS_{Treatments}$ have to be, before we decide that the treatment means are different?

$$\frac{MS_{Treatments}}{MS_E} = \frac{SS_{Treatments}/(p-1)}{SS_E/(N-p)} \sim F_{p-1,N-p}$$

- Hence, when

$$F_0 = \frac{MS_{Treatments}}{MS_E} > F_{p-1,N-p}, 1-\alpha$$

we reject the null-hypothesis of no differences between the treatment means.

- p-value of this hypothesis test equals: $Pr(F_{p-1,N-p} > MS_{Treatments}/MS_E)$