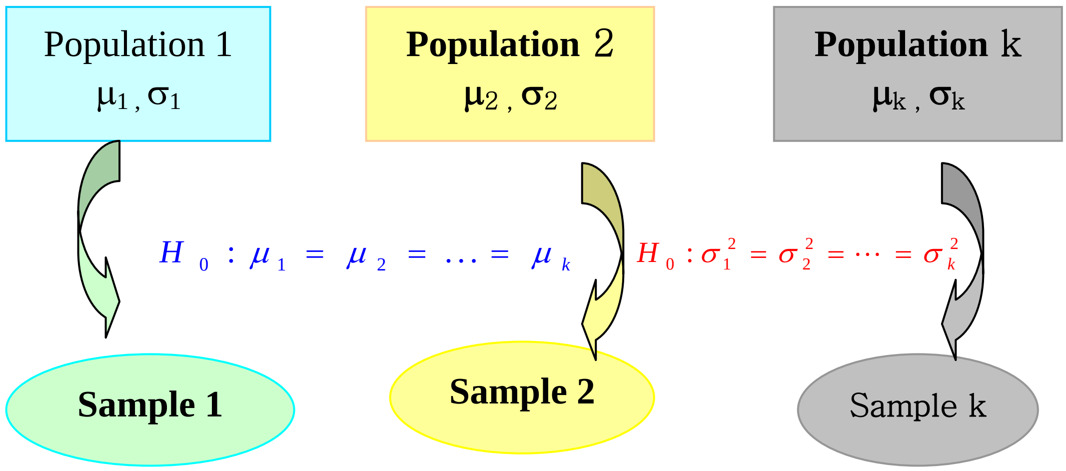
CH3: Single Factor Experiments with No Restriction On Randomization

3.1 Introduction

☐ **Completely Randomized Single Factor Experiment (One-Way Factorial)**



- **Model:** $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$, $i = 1, \cdots, n_j$ and $j = 1, \cdots, k$

  Y_{ij} : Response of i-th observation on the j-th treatment

  μ : A common effect for the whole experiment

  τ_j : The j-th treatment effect

  → $\tau_j = \mu_j - \mu$, where μ_j is true mean of j-th population

  ε_{ij} : Random error in the i-th observation on the j-th treatment,

  following $NID(0, \sigma^2)$ with common variance σ^2

- **Assumptions:**

  i) **Fixed Effect Model:** $H_0: \tau_1 = \cdots = \tau_k = 0$

  - $\tau_j,\ j = 1, \cdots, k$ are fixed parameters
  - $\sum_{j=1}^{k} \tau_j = 0$ → $\mu = \sum_{j=1}^{k} \mu_j / k$, where $\mu_j = \mu + \tau_j$

  ii) **Random Effect Model:** $H_0: \sigma_\tau^2 = 0$

  - $\tau_j,\ j = 1, \cdots, k$ are random variables
  - $\tau_j \sim NID(0,\ \sigma_\tau^2)$

3.2 Analysis of Variance Rationale

- Model: $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$

- Data Structure:

	Treatment						
	1	**2**	**…**	**j**	**…**	**k**	
	Y_{11}	Y_{12}	$\cdots$	Y_{1j}	$\cdots$	Y_{1k}	
	Y_{21}	Y_{22}	$\cdots$	Y_{2j}	$\cdots$	Y_{2k}	
	$\cdots$		$\cdots$		$\cdots$		
	Y_{i1}	Y_{i2}	$\cdots$	Y_{ij}	$\cdots$	Y_{ik}	
	$\cdots$		$\cdots$		$\cdots$		
	$Y_{n_1 1}$	$Y_{n_2 2}$	$\cdots$	$Y_{n_j j}$	$\cdots$	$Y_{n_k k}$	
Total	$T_{\bullet 1}$	$T_{\bullet 2}$	$\cdots$	$T_{\bullet j}$	$\cdots$	$T_{\bullet k}$	$T_{\bullet\bullet}$
Number	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k	N
Means	$\overline{Y}_{\bullet 1}$	$\overline{Y}_{\bullet 2}$	$\cdots$	$\overline{Y}_{\bullet j}$	$\cdots$	$\overline{Y}_{\bullet k}$	$\overline{Y}_{\bullet\bullet}$

Where $T_{\bullet j} = \sum_{i=1}^{n_j} Y_{ij}$, $T_{\bullet\bullet} = \sum_{j=1}^{k} (\sum_{i=1}^{n_j} Y_{ij}) = \sum_{j=1}^{k} T_{\bullet j}$,

$N = \sum_{j=1}^{k} n_j$, and $\overline{Y}_{\bullet\bullet} = \sum_{j=1}^{k} n_j \overline{Y}_{\bullet j} / N = T_{\bullet\bullet} / N$

- Fundamental Equation of Analysis of Variance

$$Y_{ij} - \mu \;=\; (\mu_j - \mu) \;-\; (Y_{ij} - \mu_j)$$

$$\rightarrow \quad Y_{ij} - \overline{Y}_{\bullet\bullet} \;=\; (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet}) \;-\; (Y_{ij} - \overline{Y}_{\bullet j})$$

$$\rightarrow \quad \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_{\bullet j})^2$$

$$SS_{Total} \quad = \quad SS_{Between} \quad + \quad SS_{Within}$$

$$SS_{Total} \quad = \quad SS_{Treatment} \quad + \quad SS_{Error}$$

(Note: $\sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})(Y_{ij} - \overline{Y}_{\bullet j}) = 0$) ← **Why?** **(HW)**

- Mean Squares:

- **Error Mean Square:**

$$MS_{error} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\bullet j})^2 / (N-k) = SS_{error} / (N-k)$$

← An Unbiased Estimate of σ^2

- **Treatment Mean Square:**

$$MS_{treatment} = \sum_{j=1}^{k} n_j (\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2 / (k-1) = SS_{tret} / (k-1)$$

← An Unbiased Estimate of σ^2, under H_0

(**Why? (HW): assume that** $n_j = n$, for all j)

- **Test Statistic:** $F \sim \dfrac{MS_{tret}}{MS_{error}} \sim F_{k-1,\, N-k}$ (under H_0)

- The One-Way ANOVA

- **Hypothesis:** $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$ versus $H_1 : \tau_j \neq 0$ for some j

Source	df	SS	MS	F	$p\text{-}value$
Treatment	$\nu_1 = k-1$	SS_{Tret}	MS_{Tret}	$f = MS_{Tret} / MS_{Error}$	$P(F_{(\nu_1,\nu_2)} \geq f)$
Error	$\nu_2 = N-k$	SS_{error}	MS_{error}		
Totals	$N-1$	SS_{Total}			

← For Computational Purpose:

i) $SS_{Total} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij}^2 - T_{\bullet\bullet}^2 / N$

ii) $SS_{Tret} = \sum_{j=1}^{k} n_j (\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{k} T_{\bullet j}^2 / n_j - T_{\bullet\bullet}^2 / N$

iii) $SS_{error} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\bullet j})^2 = SS_{Total} - SS_{Tret}$

☐ Example: Fabric Wear Resistance (Weight Loss in grams)

- Y: weight loss (gram)
- X: for each **Qualitative** Levels (fabric,: A, B, C, D), 4 replications ($n = 4$)
- Problem: Any difference in the average weight loss among 4 fabrics

- Data

	Fabric A	B	C	D	
	1.93	2.55	2.40	2.33	
	2.38	2.72	2.68	2.40	
	2.20	2.75	2.31	2.28	
	2.25	2.70	2.28	2.25	
Sum ($T_{.j}$)	8.76	10.72	9.67	9.26	$T_{..} = 38.41$
Average($\bar{Y}_{.j}$)	2.190	2.680	2.418	2.315	$N = 16$
Variance	.0358	.0079	.0332	.0043	$\sum\sum Y_{ij}^2 = 92.972$

$$\rightarrow \; SS_{Total} = \sum_{j=1}^{k}\sum_{i=1}^{n_j} Y_{ij}^2 - T_{..}^2 / N = 97.972 - 38.41^2 / 16 = 0.7639$$

$$SS_{Tret} = \sum_{j=1}^{k} T_{.j}^2 / n_j - T_{..}^2 / N = 8.76^2 / 4 + \cdots + 9.26^2 / 4 - 38.41^2 / 16 = 0.5201$$

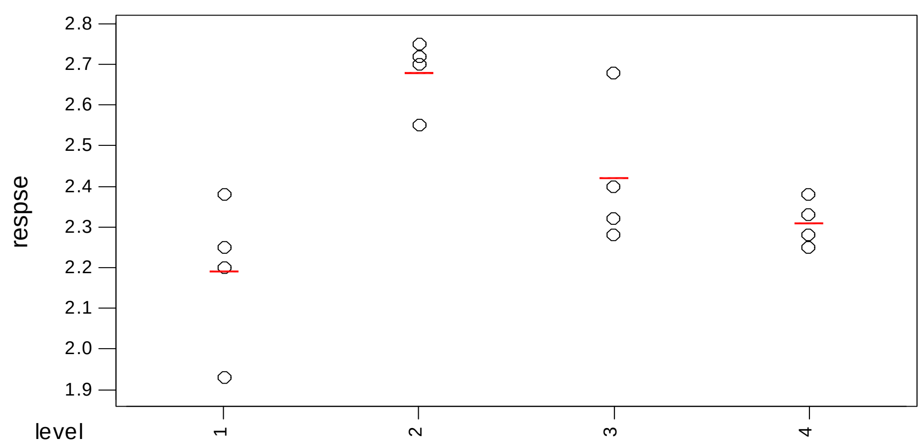
$$SS_{error} = SS_{Total} - SS_{Tret} = 0.7639 - 0.5201 = 0.2438$$

- ANOVA

- **Hypothesis:** $H_0 : \tau_1 = \tau_2 = \cdots = \tau_4 = 0$ versus $H_1 : \tau_j \neq 0$ for some j

Source	*df*	*SS*	*MS*	*F*	*p-value*
Treatment	$\nu_1 = 3$	0.5201	0.1734	f = 8.53	0.0026
Error	$\nu_2 = 12$	0.2438	0.0203		
Totals	15	0.7639			

→ **Reject** H_0 : There is some difference in the means ?



➔ **Further Questions (When Null Hypothesis is rejected): 3.3 After ANOVA**

i) **Which treatment is best?**

ii) **Mean wear of Fabric A differ from that of C? ($\mu_A = \mu_C$)**

iii) **Mean of A and B together differ from that of C and D?**

→ Need to consider:

- What design of experiment was employed?
- How the experiment was conducted?
- Are the model assumptions reasonable?

3.4 Tests on Means

☐ Orthogonal Contrasts

← Can be applied when set up prior to the running of experiment

$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$, $i = 1, \cdots, n_j$ and $j = 1, \cdots, k$

Denote $T_{\bullet j} = \sum_{i=1}^{n_j} Y_{ij}$

- **Contrast C_m in the Treatment Totals**: (when $n_j = n,\ j = 1, \cdots, k$)

  $C_m = c_{1m}T_{\bullet 1} + c_{2m}T_{\bullet 2} + \cdots + c_{km}T_{\bullet k}$, where $c_{1m} + c_{2m} + \cdots + c_{km} = 0$

  → $C_m = \sum_{j=1}^{k} c_{jm} T_{\bullet j}$ with $\sum_{j=1}^{k} c_{jm} = 0$

- **Orthogonal Contrasts:** (when $n_j = n,\ j = 1, \cdots, k$)

  C_m and C_q: Two Contrasts are Orthogonal if

  $$c_{1m}c_{1q} + c_{2m}c_{2q} + \cdots + c_{km}c_{kq} = 0$$

- **Sum of Squares for Contrasts:** (when $n_j = n,\ j = 1, \cdots, k$)

  $$SS_{C_m} = \frac{C_m^2}{n\sum_{j=1}^{k} c_{jm}^2}, \text{ where } C_m = c_{1m}T_{\bullet 1} + c_{2m}T_{\bullet 2} + \cdots + c_{km}T_{\bullet k}$$

Example of Fabric Wear: (Equal Sample Size)

$C_1 = T_{\bullet 1} \qquad\qquad\qquad - T_{\bullet 4}$ → 1(8.76) -1(9.26) = -0.50

$C_2 = \qquad T_{\bullet 2} - T_{\bullet 3}$ → 1(10.72) -1(9.67) = 1.05

$C_1 = T_{\bullet 1} - T_{\bullet 2} - T_{\bullet 3} + T_{\bullet 4}$ → 1(8.76) -1(10.72) -1(9.67) +1(9.26) = -2.37

	$T_{\bullet 1}$	$T_{\bullet 2}$	$T_{\bullet 3}$	$T_{\bullet 4}$	$n\sum_{j=1}^{k} c_{jm}^2$	SS_{C_i}
C_1	$+1$	0	0	-1	8	**0.0312**
C_2	0	$+1$	-1	0	8	**0.1378**
C_3	$+1$	-1	-1	$+1$	16	**0.3511**

← Note that: $SS_{Tret} = SS_{C_1} + SS_{C_2} + SS_{C_3} = 0.5201$

- Hypothesis Test Using Contrasts:

H_0 : $c_{1m}\mu_1 + \cdots + c_{km}\mu_k = 0$, for each $m = 1, 2, 3$

← $C_m = c_{1m}T_{\bullet 1} + c_{2m}T_{\bullet 2} + \cdots + c_{km}T_{\bullet k}$

$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} = \mu_j + \varepsilon_{ij}$ and $T_{\bullet j} = \sum_{i=1}^{n} Y_{ij} = n\mu_j + \sum_{i=1}^{n} \varepsilon_{ij}$

$E[C_m] = n(c_{1m}\mu_1 + c_{2m}\mu_2 + \cdots + c_{km}\mu_k)$ ← $E[T_{\bullet j}] = n\mu_j$

$Var[C_m] = (c_{1m}^2 + c_{2m}^2 + \cdots + c_{km}^2)(n\sigma^2)$ ← $Var[T_{\bullet j}] = n\sigma^2$

$$\Rightarrow \; T = \frac{C_m - E(C_m)}{\sqrt{Var(C_m)}} = \frac{C_m - n(c_{1m}\mu_1 + c_{2m}\mu_2 + \cdots + c_{km}\mu_k)}{\sqrt{n(c_{1m}^2 + \cdots + c_{km}^2)(\text{MSE})}} \sim t_{N-k}\text{, under } H_0$$

- Contrast Expressed in Terms of Sample Means:

$C_q = c_{1q}\bar{Y}_{\bullet 1} + c_{2q}\bar{Y}_{\bullet 2} + \cdots + c_{kq}\bar{Y}_{\bullet k}$ with $c_{1q} + c_{2q} + \cdots + c_{kq} = 0$

→ SS_{C_q} is the same as that of Contrast with Treatment Totals **(Why? HW)**

- Case of Different Sample Sizes:

- Contrast in the Treatment Totals: $C_m = c_{1m}T_{\bullet 1} + c_{2m}T_{\bullet 2} + \cdots + c_{km}T_{\bullet k}$,

  where $n_1c_{1m} + n_2c_{2m} + \cdots + n_kc_{km} = 0$

- C_m and C_q are Orthogonal: $n_1c_{1m}c_{1q} + n_2c_{2m}c_{2q} + \cdots + n_kc_{km}c_{kq} = 0$

- **Sum of Squares:** $SS_{C_m} = \dfrac{C_m^2}{n_1c_{1m}^2 + \cdots + n_kc_{km}^2}$

● Example of Fabric Wear: (continue)

$C_1 = T_{.1} \quad - T_{.4}$ → $H_0: \mu_1 - \mu_4 = 0 \quad (H_1: \mu_1 - \mu_4 \neq 0)$

$C_2 = \quad T_{.2} \quad - T_{.3}$ → $H_0: \mu_2 - \mu_3 = 0$

$C_3 = T_{.1} - T_{.2} - T_{.3} + T_{.4}$ → $H_0: (\mu_1 + \mu_4) - (\mu_2 + \mu_3) = 0$

→ **ANOVA Summary:**

Source	*df*	*SS*	*MS*	*F*	*P-value*
Treatment	**3**	**0.5201**	**0.1734**	**8.53**	**0.003**
Contrast 1	1	0.0312	0.0312	1.54	0.238
Contrast 2	1	0.1378	0.1378	6.76	**0.023**
Contrast 3	1	0.3511	0.3511	17.30	**0.001**
Error	**12**	**0.2438**	**0.0203**		
Totals	**15**	**0.7639**			

→ **i) Fabric 2 (B) differs from Fabric 3 (C) , in mean wear resistance!**

ii) Fabric 1 and 4 (A & D) differs from that of Fabric 2 and 3 (B & C)

□ Multiple Comparison Procedures

← **Can be applied even after the data have been examined**

● Student-Newman-Keuls Range Test: Comparing two Means

Step 1: Arrange the k sample means in order from low to high

$k = 4$ means:	2.19	2.32	2.42	2.68
Treatment	A	D	C	B

Step 2: Take MSE and error df from ANOVA Table

$MSe = 0.0203$ and $df_{error} = 12$

Step 3: Obtain the $s.e.(\bar{Y}_{.j})$ for each treatment by $s_{\bar{Y}_{.j}} = \sqrt{MSE / n_j}$

$$s_{\bar{Y}_{.j}} = \sqrt{0.0203/4} = 0.0712$$

Step 4: Find significant ranges (from Table E.1 or E.2) at α desired

→ **Using** $n_2 = df$ of error **and** $p = 2, 3, \cdots, k$ **, list** $k-1$ **ranges**

From Table E.1 with $\alpha = 0.05$ and $n_2 = 12$:

p:	***2***	***3***	***4*** $(= k-1)$
Ranges:	3.08	3.77	4.20

Step 5: Obtain $k-1$ Least Significant Ranges (LSR) by $s_{\bar{Y}_{\cdot j}} \times$ ranges

p:	*2*	*3*	*4* $(=k-1)$
LSR:	**0.22**	**0.27**	**0.30 (←0.0712* Ranges)**
(Ranges:	3.08	3.77	4.20)

Step 6: Test the observed ranges between means, beginning with the largest versus smallest, comparing with the LSR for $p = k$; then test largest versus 2nd smallest with the LSR for $p = k-1$; and so on …

→ When $\alpha = 0.05$ is applied:

$\bar{y}_B - \bar{y}_A = 2.68 - 2.19 = 0.49 > 0.30$ → conclude $\mu_B > \mu_A$

$\bar{y}_B - \bar{y}_D = 2.68 - 2.32 = 0.36 > 0.27$ → conclude $\mu_B > \mu_D$

$\bar{y}_B - \bar{y}_C = 2.68 - 2.42 = 0.26 > 0.22$ → conclude $\mu_B > \mu_C$

$\bar{y}_C - \bar{y}_A = 2.42 - 2.19 = 0.23 < 0.27$ → may be $\mu_c \approx \mu_A$

$\bar{y}_C - \bar{y}_D = 2.42 - 2.32 = 0.10 < 0.22$ → may be $\mu_c \approx \mu_D$

$\bar{y}_D - \bar{y}_A = 2.32 - 2.19 = 0.13 < 0.22$ → may be $\mu_A \approx \mu_D$

→ In Summary,

```
Treatment      A       D       C                     B
          ----------+--------+--------+-----------------------+----
Means:        2.19    2.32    2.42                  2.68
              ______________________                ____
```

→ The effect of B differs significantly from those of A, D, C,
But the effects of A, D, C do not differ significantly from each other

- **Question: Interpretation of SNK Test**

→ Seek a minimum response as the best treatment (in Red)

i. Case I

```
B_____A       D       C       E
      ----------------------
                      ---------------
```

ii. Case II

```
B_____A_______D       C       E
              -------------------------
```

- **Scheffe's Test: Examine Contrast (Combination of Treatments)**
  ← Can be applied to Contrasts which are Non-Orthogonal

Step 1: Set up all contrast of interest

- **Example:** **Fabric Wear Data, Non-Orthogonal Contrasts**

Contrast	Observed Values
$C_1 = T_{\bullet 1} - T_{\bullet 2}$	$c_1 = 8.76 - 10.72 = -1.96$
$C_2 = 3T_{\bullet 1} - T_{\bullet 2} - T_{\bullet 3} - T_{\bullet 4}$	$c_2 = -3.37$

Step 2: Determine the f **(**$= \alpha -$**quantile) of** $F_{k-1,N-k}$

If $\alpha = 0.05$, $f = F_{k-1,N-k}(\alpha) = F_{3,\,12}(0.05) = 3.49$

Step 3: Compute $A = \sqrt{(k-1)f}$

$$A = \sqrt{(k-1)f} = \sqrt{3 \times 3.49} = 3.24$$

Step 4: Compute the $s.e. = s_{C_m} = \sqrt{MSe(n_1 c_{1m}^2 + \cdots + n_k c_{km}^2)}$ **of Contrast**

← Note that $C_m = c_{1m} T_{\bullet 1} + c_{2m} T_{\bullet 2} + \cdots + c_{km} T_{\bullet k}$

And $T_{\bullet j} = \sum_{i=1}^{n_j} Y_{ij} = n\mu_j + \sum_{i=1}^{n_j} \varepsilon_{ij} \quad (Var[T_{\bullet j}] = n_j \sigma^2)$

$$s_{C_1} = \sqrt{0.0203(4(1)^2 + 4(-1)^2 + 4(0)^2 + 4(0)^2)} = 0.40$$

$$s_{C_2} = \sqrt{0.0203(4(3)^2 + 4(-1)^2 + 4(-1)^2 + 4(-1)^2)} = 0.99$$

Step 5: Reject H_0 **(true contrast among means is 0), if** $|c_m| > A \cdot s_{C_m}$

$|c_1| = 1.96 > 1.30 = A \cdot s_{C_1} (= 3.24 \times 0.40)$ → **Reject** H_0

$|c_2| = 3.37 > 3.21 = A \cdot s_{C_2} (= 3.24 \times 0.99)$ → **Reject** H_0

3.5 Confidence Limits on Means

- Under Normal Assumption, $\bar{y}_{\bullet j} \sim N(\mu + \tau_j = \mu_j, \sigma^2 / n_j)$
- $100(1-\alpha)\%$ Confidence Limits on μ_j : $\bar{y}_{\bullet j} \pm t_{N-k,1-\alpha/2}\sqrt{MSe / n_j}$

  where MSe came from ANOVA
- See Example 3.8

3.6 Components of Variance

- **Model: Single-Factor Experiment with No Restriction on Randomization**

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

	Fixed Effect Model	**Random Effect Model**
Levels	Intentionally Selected	Randomly Selected
Assumption	$\sum_{j=1}^{k} \tau_j = 0$	$\tau_j \sim NID(0,\ \sigma_\tau^2)$
Hypothesis	$H_0 : \tau_1 = \cdots = \tau_k = 0$	$H_0 : \sigma_\tau^2 = 0$
Decision	Restricted to selected levels	All levels of population
Interest	**Mean Difference**	**Component of Variance**

i) **Fixed Model:** Testing hypothesis/ Confidence Intervals/ Contrast in Means

ii) **Random Model:** Estimating Components of Variance.

→ Variance attributed from the true differences in treatment means, and
Variance due to random error about these means

□ Expected Mean Square (EMS: Cases for $n_j = n$)

- **Model:** $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$ where $i = 1, \cdots, n_j$ and $j = 1, \cdots, k$

- **Facts:** $\bar{Y}_{\cdot j} = \sum_{i=1}^{n} Y_{ij} / n = \mu + \tau_j + \bar{\varepsilon}_{\cdot j}$, where $\bar{\varepsilon}_{\cdot j} = \sum_{i=1}^{n} \varepsilon_{ij} / n \sim N(0, \sigma^2 / n)$

$$\bar{Y}_{\cdot\cdot} = \sum_{j=1}^{k} \sum_{i=1}^{n} Y_{ij} / (nk) = \mu + \sum_{j=1}^{k} \tau_j / k + \bar{\varepsilon}_{\cdot\cdot} \quad \text{where} \quad \bar{\varepsilon}_{\cdot\cdot} \sim N(0, \sigma^2 / nk)$$

- **EMS of Treatment:** Denoting $\bar{\tau} = \sum_{j=1}^{k} \tau_j / k$,

$$SS_{Tret} = n \sum_{j=1}^{k} (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$$
$$= n \sum_{j=1}^{k} (\tau_j - \bar{\tau})^2 + n \sum_{j=1}^{k} (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2 + 2n \sum_{j=1}^{k} (\tau_j - \bar{\tau})(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})$$

i) **Fixed:** $E[SS_{Tret}] = n \sum_{j=1}^{k} \tau_j^2 + (k-1)\sigma^2$

ii) **Random:** $E[SS_{Tret}] = n(k-1)\sigma_\tau^2 + (k-1)\sigma^2$

- **Theorem 3.1 (EMS of Treatment)**

i) **Fixed Effect:**
$$E[MS_{Tret}] = E[SS_{Tret} / (k-1)]$$
$$= n \sum_{j=1}^{k} \tau_j^2 / (k-1) + \sigma^2 = \sigma^2 + n\phi_\tau \leftarrow \phi_\tau = \frac{\sum_{j=1}^{k} \tau_j^2}{(k-1)}$$

ii) **Random Effect:** $E[MS_{Tret}] = \sigma^2 + n\sigma_\tau^2$

- **Theorem 3.2 (EMS of Error)**

$$SS_{error} = \sum_{j=1}^{k} \sum_{i=1}^{n} (Y_{ij} - \bar{Y}_{\cdot j})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{\cdot j})^2$$

$\rightarrow$ $E[SS_{error}] = \sum_{j=1}^{k} E[\sum_{i=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{\cdot j})^2] = k(n-1)\sigma^2$

➔ $E[MS_{error}] = \sigma^2$

3.7 Checking the Model

- **Model:** $Y_{ij} = \mu + \tau_j + \varepsilon_{ij} = \mu_j + \varepsilon_{ij}$, where $\tau_j = \mu_j - \mu$

  → $\hat{Y}_{ij} = \overline{Y}_{\bullet j}$ since $\hat{\mu}_j = \overline{Y}_{\bullet j}$

  → $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{\bullet j}$: **Residual from** Y_{ij}

- **Diagnostics:** Checking $e_{ij} \approx \varepsilon_{ij} \sim NID(0, \sigma^2)$

  → **1) Independence 2) Equal Variance 3) Normality**

□ Example (Fabric Wear)

Level	Y	Fitted	Residual	Level	Y	Fitted	Residual
A	1.93	2.1900	-0.2600	C	2.40	2.4175	-0.0175
A	2.38	2.1900	0.1900	C	2.68	2.4175	0.2625
A	2.20	2.1900	0.0100	C	2.31	2.4175	-0.1075
A	2.25	2.1900	0.0600	C	2.28	2.4175	-0.1375
B	2.55	2.6800	-0.1300	D	2.33	2.3150	0.0150
B	2.72	2.6800	0.0400	D	2.40	2.3150	0.0850
B	2.75	2.6800	0.0700	D	2.28	2.3150	-0.0350
B	2.70	2.6800	0.0200	D	2.25	2.3150	-0.0650

- **Estimation**

One-way ANOVA: response versus level

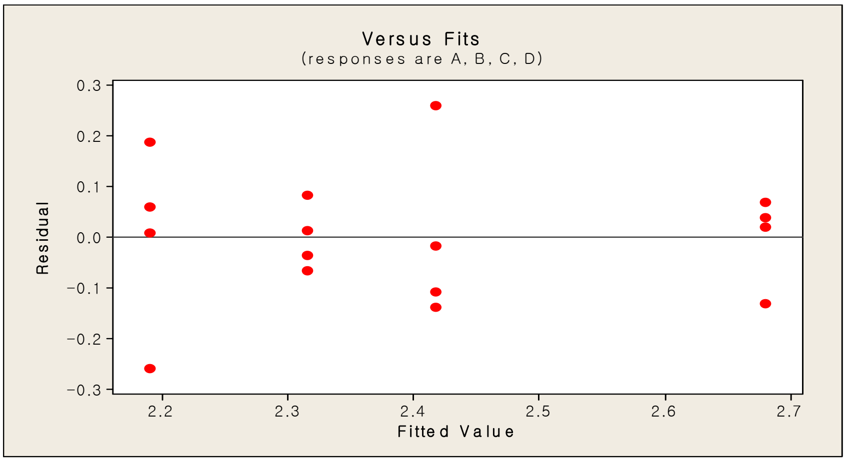
Source	DF	SS	MS	F	P
level	3	0.5201	0.1734	8.53	**0.003**
Error	12	0.2438	0.0203		
Total	15	0.7639			

S = 0.1425 R-Sq = 68.09% R-Sq(adj) = 60.11%

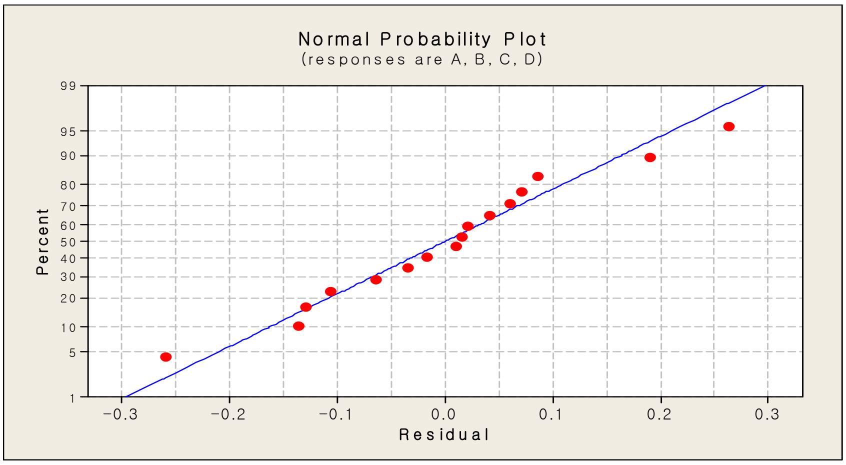
- Individual 95% CIs For Mean Based on **Pooled StDev (= 0.1425)**

```
Level  N    Mean    StDev   ---------+---------+---------+---------+
A      4  2.1900  0.1892  (------*-----)
B      4  2.6800  0.0891                             (-----*-----)
C      4  2.4175  0.1823               (------*-----)
D      4  2.3150  0.0656          (------*-----)
                          ---------+---------+---------+---------+
                                 2.25      2.50      2.75      3.00
```

● **Residual Plot (Equal Variance & Independence)**



● **Normal Probability Plot (Normality)**



→ Plot of i) Residuals versus experimental order

ii) Residuals versus Fitted values

- Test for Independence:

i) D-W Statistics

ii) Plot of e_j versus e_{j-1}

- **Transforming Y to achieve Normality/ Equal Variance**

→ Equal Variance:

Relationship (μ and σ^2)	**Transformation**
$s_j^2 \propto \bar{y}_{\cdot j}$	$\sqrt{y_{ij}}$ (Poisson)
$s_j^2 \propto \bar{y}_{\cdot j}(1-\bar{y}_{\cdot j})$	$\arcsin\sqrt{y_{ij}}$ (Binomial)
$s_j \propto \bar{y}_{\cdot j}$	$\ln(y_{ij})$ or $\ln(y_{ij}+1)$
$s_j \propto \bar{y}_{\cdot j}^{2}$	$1/y_{ij}$