# Analysis of Variance Two Factors Part 1

**Dr. Supaporn Erjongmanee**

Department of Computer Engineering
Kasetsart University
fengspe@ku.ac.th



## Outline

- Two-Factor ANOVA
  - Additive Factors
  - Interaction Factors

## Two-Factor ANOVA : Additive Factors

- Compare two or more populations on <u>two</u> factors
  - Factor A has I level of treatments
  - Factor B has J level of treatments
  - Example: test washing detergents on pens
    - A = Brand of pens, I = 3
    - B = Washing detergent, J = 4

| | | Washing detergent | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 |



## Additive Two-Factor ANOVA : Effects Model

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$ $\sum \alpha_i = 0$ $\sum \beta_j = 0$

- $X_{ij}$ = random sample j of treatment i
- μ = overall mean of treatment
- $\alpha_i$ = effect due to factor A at level i
- $\beta_j$ = effect due to factor B at level j
- $\varepsilon_{ij}$ = random error from sample j of treatment i
  - Assumed to be independent and normally distributed with mean = 0, variance = $\sigma^2$
- Factor A is independent of factor B

## Additive Two-Factor ANOVA : Effects Model (cont.)

$\boxed{E(X_{ij}) = \mu + \alpha_i + \beta_j}$ $\boxed{\sum \alpha_i = 0}$ $\boxed{\sum \beta_j = 0}$

- If $\alpha_i = 0$ and $\beta_j = 0$, then all treatments have the same response

$\boxed{E(X_{ij}) = \mu}$

- Thus, null hypotheses for two additive factor ANOVA
  - Hypothesis on A : factor A at any level i has no effect on overall mean.
    - $H_{0A}: \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$
  - Hypothesis on B : factor B at any level j has no effect on overall mean.
    - $H_{0B}: \beta_1 = \beta_2 = \ldots = \beta_J = 0$



## Two-Factor ANOVA : Additive Factors (cont.)

- Let
  - $\mu_{ij}$ = mean of treatment i of factor A and treatment j of factor B
  - $I$ = number of treatments from factor A
  - $J$ = number of treatments from factor B
- Hypothesis on A : factor A at any level i has no effect on true mean.
  - $H_{0A}: \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$
  - $H_{aA}$: Not all $\alpha_i$'s are equal (Factor A has effect.)
- Hypothesis on B : factor B at any level j has no effect on true mean.
  - $H_{0B}: \beta_1 = \beta_2 = \ldots = \beta_J = 0$
  - $H_{aB}$: Not all $\beta_j$'s are equal (Factor B has effect.)

# Two-Factor ANOVA : Additive Factors (cont.)

- Test statistic (cont.):
  - I = Number of treatments from factor A
  - J = Number of treatments from factor B

SST = $\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X})^2$ df = $IJ - 1$

SSA = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_i - \bar{X})^2$ df = $I - 1$

SSB = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_j - \bar{X})^2$ df = $J - 1$

SSE = $\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2$ df = $(\mathrm{I} - 1)(J - 1)$

$$\boxed{SST = SSA + SSB + SSE}$$



# Two-Factor ANOVA : Additive Factors (cont.)

- Test statistic (cont.):
  - I = Number of treatments from factor A
  - J = Number of treatments from factor B

Another option: Sample-based computation

SST = $\sum_{i=1}^{I}\sum_{j=1}^{J}X_{ij}^2 - \frac{1}{IJ}(\sum_{i=1}^{I}\sum_{j=1}^{J}X_{ij})^2$ df = $IJ - 1$

SSA = $\frac{1}{J}\sum_{i=1}^{I}(\sum_{j=1}^{J}X_{ij})^2 - \frac{1}{IJ}(\sum_{i=1}^{I}\sum_{j=1}^{J}X_{ij})^2$ df = $I - 1$

SSB = $\frac{1}{I}\sum_{j=1}^{J}(\sum_{i=1}^{I}X_{ij})^2 - \frac{1}{IJ}(\sum_{i=1}^{I}\sum_{j=1}^{J}X_{ij})^2$ df = $J - 1$

SSE = SST – SSA - SSB df = $(\mathrm{I} - 1)(J - 1)$

## Two-Factor ANOVA : Additive Factors (cont.)

- Test statistic (cont.):

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | SSA/(I-1) | $f_A$ = MSA / MSE | $f_A > F_{\alpha, I-1, (I-1)(J-1)}$ |
| $H_{0B}$ vs. $H_{aB}$ | SSB/(J - 1) | $f_B$ = MSB / MSE | $f_B > F_{\alpha, J-1, (I-1)(J-1)}$ |



## Example

- Test 4 washing detergents on 3 brands of pens at significance level = 0.05

| | | Washing detergent | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 | 0.598 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 | 0.345 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 | 0.455 |
| | | 0.803 | 0.337 | 0.423 | 0.300 | 0.466 |

- A = Brand of pens, I = 3
- B = Washing detergents, J = 4
- SST* = 0.6947, df = 11
- SSA* = 0.1282, df = 2
- SSB* = 0.4797, df = 3
- SSE = 0.6947 – 0.1282 – 0.4797 = 0.0868, df = 11 – 2 – 3 = 6

# Example

- Test 4 washing detergents on 3 brands of pens at significance level = 0.05

| | | Washing detergent | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 | 0.598 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 | 0.345 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 | 0.455 |
| | | 0.803 | 0.337 | 0.423 | 0.300 | 0.466 |

SST = $\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X})^2$

SSA = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_i - \bar{X})^2$

SSB = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_j - \bar{X})^2$

SSE = $\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2$

$SST = SSA + SSB + SSE$

- A = Brand of pens, I = 3
- B = Washing detergents, J = 4
- SST* = 0.6947, df = 11
- SSA* = 0.1282, df = 2
- SSB* = 0.4797, df = 3
- SSE = 0.6947 – 0.1282 – 0.4797 = 0.0868, df = 11 – 2 – 3 = 6



# Example (cont.)

- Test 4 washing detergents on 3 brands of pens (cont.):

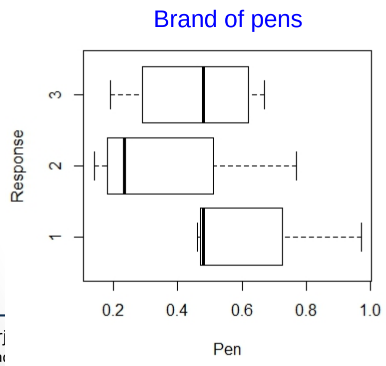
| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 0.1282/2 = 0.0641 | $f_A$ = 4.43 | $F_{0.05, 2, 6}$ = 5.14 |
| $H_{0B}$ vs. $H_{aB}$ | 0.4797 / 3 = 0.1599 | $f_B$ = 11.05 | $F_{0.05, 3, 6}$ = 4.76 |
| Error | 0.0868/6 =0.0144 | | |

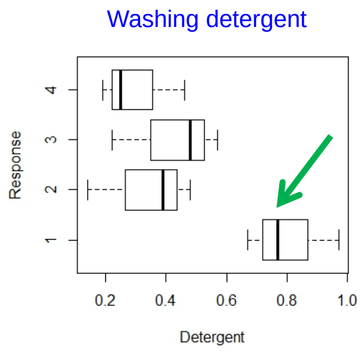
- $H_{0A}$ is not rejected. Factor A (brand of pens) has no effect on mark removal
- $H_{0B}$ is rejected. Factor B (washing detergents) has effect on mark removal

# Example

- Factor A (brand of pens) has no effect on mark removal
- Factor B (washing detergents) has effect on mark removal

| | | Washing detergent | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 | 0.598 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 | 0.345 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 | 0.455 |
| | | 0.803 | 0.337 | 0.423 | 0.300 | 0.466 |






# Multiple Comparisons

- To specify which mean is different from others
  - Fix on factor A or B
  - Find Tukey's Honestly Significant Difference (HSD) using the following formula

Factor A: $w_A = q_{\alpha,I,(I-1)(J-1)}\sqrt{\frac{MSE}{J}}$

Factor B: $w_B = q_{\alpha,\, J,(I-1)(J-1)}\sqrt{\frac{MSE}{I}}$

  - $q_{\alpha,\, m,\, n}$ = q-value from studentized range distribution with 2 degrees of freedom *m, n*
- Follow the same steps as finding other HSD

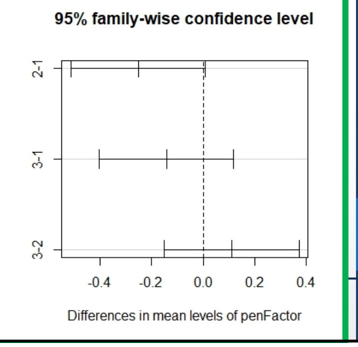
# Example

Factor A: $w_A = q_{\alpha,I,(I-1)(J-1)}\sqrt{\frac{MSE}{J}}$

- Test 4 washing detergents on 3 brands of pens at significance level = 0.05

| | | Washing detergents | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 | 0.598 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 | 0.345 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 | 0.455 |
| | | 0.803 | 0.337 | 0.423 | 0.300 | 0.466 |

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 0.1282/2 = 0.0641 | $f_A$ = 4.43 | $F_{0.05,2,6}$ = 5.14 |
| $H_{0B}$ vs. $H_{aB}$ | 0.4797 / 3 = 0.1599 | $f_B$ = 11.05 | $F_{0.05,3,6}$ = 4.76 |
| Error | 0.0868/6 =0.0144 | | |




- Fixed at brands of pens:

$$w_A = q_{0.05,3,6}\sqrt{\frac{MSE}{J}} = 4.34\sqrt{\frac{0.0144}{4}} = 0.261$$

- Sort factor-A sample means: $\overset{\bar{x}_2}{0.345},\ \overset{\bar{x}_3}{0.455},\ \overset{\bar{x}_1}{0.598}$
- 1 group of means: $\{\bar{x}_1, \bar{x}_2, \bar{x}_3\}$



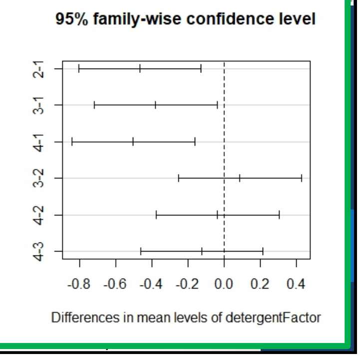
# Example (cont.)

Factor B: $w_B = q_{\alpha,\,J,(I-1)(J-1)}\sqrt{\frac{MSE}{I}}$

- Test 4 washing detergents on 3 brands of pens at significance level = 0.05

| | | Washing detergents | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Brand of pens | 1 | 0.97 | 0.48 | 0.48 | 0.46 | 0.598 |
| | 2 | 0.77 | 0.14 | 0.22 | 0.25 | 0.345 |
| | 3 | 0.67 | 0.39 | 0.57 | 0.19 | 0.455 |
| | | 0.803 | 0.337 | 0.423 | 0.300 | 0.466 |

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 0.1282/2 = 0.0641 | $f_A$ = 4.43 | $F_{0.05,2,6}$ = 5.14 |
| $H_{0B}$ vs. $H_{aB}$ | 0.4797 / 3 = 0.1599 | $f_B$ = 11.05 | $F_{0.05,3,6}$ = 4.76 |
| Error | 0.0868/6 =0.0144 | | |




- Fixed at detergent factor:

$$w_B = q_{0.05,4,6}\sqrt{\frac{MSE}{I}} = 4.90\sqrt{\frac{0.0144}{3}} = 0.340$$

- Sort factor-B sample means: $\overset{\bar{x}_4}{0.300},\ \overset{\bar{x}_2}{0.337},\ \overset{\bar{x}_3}{0.423},\ \overset{\bar{x}_1}{0.803}$
- 2 groups of means: $\{\bar{x}_4, \bar{x}_2, \bar{x}_3\}$ and $\bar{x}_1$

# Example 2

- Test 4 coatings on 3 soil type for corrosion at significance level = 0.05

| | | Soil Type (B) | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | $\bar{x}_j$ |
| Coating (A) | 1 | 64 | 49 | 50 | 54.33 |
| | 2 | 53 | 51 | 48 | 50.67 |
| | 3 | 47 | 45 | 50 | 47.33 |
| | 4 | 51 | 43 | 52 | 48.67 |
| | $\bar{x}_i$ | 53.75 | 47.00 | 50.00 | 50.25 |

  - A = Coatings, I = 4, B = Soil Types, J = 3
  - SST = 242.063, df = 11  SST = $\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X})^2$
  - SSA = 83.583, df = 3  SSA = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_i - \bar{X})^2$
  - SSB = 91.500, df = 2  SSB = $\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_j - \bar{X})^2$
  - SSE = SST – SSA – SSB = 66.979, df = 11 – 3 – 2 = 6



# Example 2 (cont.)

- Test 4 coatings on 3 soil types for corrosion (cont.):

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | **83.583/3 = 27.861** | $f_A$ = **2.495** | $F_{0.05, 3, 6}$ = |
| $H_{0B}$ vs. $H_{aB}$ | **91.500 / 2 = 45.750** | $f_B$ = **4.098** | $F_{0.05, 2, 6}$ = |
| Error | **66.97/6 =11.163** | | |

- $H_{0A}$ is not rejected. Factor A (coatings) has no effect on corrosion
- $H_{0B}$ is not rejected. Factor B (soil types) has no effect on corrosion

# Outline

- Two-Factor ANOVA
  - Additive Factors
  - Interaction Factors



# Two-Factor ANOVA : Interaction Factors

- Compare two or more populations on <u>two interaction</u> factors
  - Factor A has I level of treatments
  - Factor B has J level of treatments
  - K Samples from treatment of factors and B are collected
  - Example: 3 varieties of tomatoes on 4 planting density

| | Planting Density | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety | 10,000 | | | 20,000 | | | 30,000 | | | 40,000 | | |
| H | 10.5 | 9.2 | 7.9 | 12.8 | 11.2 | 13.3 | 12.1 | 12.6 | 14.0 | 10.8 | 9.1 | 12.5 |
| Ife | 8.1 | 8.6 | 10.1 | 12.7 | 13.7 | 11.5 | 14.4 | 15.4 | 13.7 | 11.3 | 12.5 | 14.5 |
| P | 16.1 | 15.3 | 17.5 | 16.6 | 19.2 | 18.5 | 20.8 | 18.0 | 21.0 | 18.4 | 18.9 | 17.2 |

## Interaction Two-Factor ANOVA : Effects Model

$$\boxed{X_{ijk} = \mu \quad + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}} \quad \boxed{\sum \alpha_i = 0} \quad \boxed{\sum \beta_j = 0}$$

- $X_{ij}$ = random sample j of treatment i
- $\mu$ = overall mean of treatment i on factor j
- $\alpha_i$ = effect due to factor A at level i
- $\beta_j$ = effect due to factor B at level j
- $\Upsilon_{ij}$ = interaction parameter between factors A and B
- $\varepsilon_{ij}$ = random error from sample j of treatment i
  - Assumed to be independent and normally distributed with mean = 0, variance = $\sigma^2$
- Factor A is not independent of factor B
  - If $\Upsilon_{ij}$ 's are all zeros, factors A and B are independent.



## Interaction Two-Factor ANOVA : Effects Model (cont.)

$$\boxed{X_{ijk} = \mu \quad + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}} \quad \boxed{\sum \alpha_i = 0} \quad \boxed{\sum \beta_j = 0}$$

- If $\alpha_i = 0$ , $\beta_j = 0$, and $\gamma_{ij} = 0$, then all treatments have the same response

  $\boxed{E(X_{ijk}) = \mu}$

- Thus, null hypotheses for interaction factor ANOVA
  - Hypothesis on A and B : factors A and B at any level i has no effect on overall mean.
    - $H_{0AB}$: $\Upsilon_{ij} = 0$ for all i,j ⬅ Test first   If reject, no need to test $H_{0A}$, $H_{0B}$
  - Hypothesis on A : factor A at any level i has no effect on overall mean.
    - $H_{0A}$: $\alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$
  - Hypothesis on B : factor B at any level j has no effect on overall mean.
    - $H_{0B}$: $\beta_1 = \beta_2 = \ldots = \beta_J = 0$

## Two-Factor ANOVA : Interaction Factors (cont.)

- Let
  - $\mu_{ijk}$ = mean of treatment i of factor A and treatment j of factor B
  - I = number of treatments from factor A
  - J = number of treatments from factor B
  - K = number of samples per treatments from factors A and B
- Hypothesis :
  - $H_{0AB}$: $\Upsilon_{ij} = 0$ for all i,j
  - $H_{aAB}$: At least $\Upsilon_{ij} \neq 0$
  - $H_{0A}$: $\alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$
  - $H_{aA}$: At least $\alpha_i \neq 0$
  - $H_{0B}$: $\beta_1 = \beta_2 = \ldots = \beta_I = 0$
  - $H_{aB}$: At least $\beta_j \neq 0$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

## Two-Factor ANOVA : Interaction Factors

- Test statistic (cont.):

SST = $\sum_i \sum_j \sum_k (X_{ijk} - \bar{X})^2$ df = $IJK - 1$

SSA = $\sum_i \sum_j \sum_k (\bar{X}_i - \bar{X})^2$ df = $I - 1$

SSB = $\sum_i \sum_j \sum_k (\bar{X}_j - \bar{X})^2$ df = $J - 1$

SSAB = $\sum_i \sum_j \sum_k (\bar{X}_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2$ df = $(I-1)(J-1)$

SSE = $\sum_i \sum_j \sum_k (X_{ijk} - \bar{X}_{ij})^2$ df = $IJ(K-1)$

$$\boxed{SST = SSA + SSB + SSAB + SSE}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

## Two-Factor ANOVA : Additive Factors (cont.)

- Test statistic (cont.):

Another option:
Sample-based computation

SST = $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - \frac{1}{IJK}(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk})^2$ df = $IJK-1$

SSA = $\frac{1}{JK}\sum_{i=1}^{I}(\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk})^2 - \frac{1}{IJK}(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk})^2$ df = $I-1$

SSB = $\frac{1}{IK}\sum_{j=1}^{J}(\sum_{i=1}^{I}\sum_{k=1}^{K}X_{ijk})^2 - \frac{1}{IJK}(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk})^2$ df = $J-1$

SSE = $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - \frac{1}{K}\sum_{i=1}^{I}\sum_{j=1}^{J}[(\sum_{k=1}^{K}X_{ijk})^2]$ df = $(I-1)(J-1)$

SSAB =SST – SSA – SSB - SSE



## Two-Factor ANOVA : Interaction Factors (cont.)

- Test statistic (cont.):
  - I = Number of treatments from factor A
  - J = Number of treatments from factor B
  - K = number of samples per treatments from factors A and B

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | SSA/(**I**-1) | $f_A$ = MSA / MSE | $f_A > F_{\alpha, I-1, IJ(K-1)}$ |
| $H_{0B}$ vs. $H_{aB}$ | SSB/(**J** - 1) | $f_B$ = MSB / MSE | $f_B > F_{\alpha, J-1, IJ(K-1)}$ |
| $H_{0AB}$ vs. $H_{aAB}$ | $\frac{SSAB}{(\mathbf{I}-1)(\mathbf{J}-1)}$ | $f_{AB}$ = MSAB / MSE | $f_{AB} > F_{\alpha, (I-1)(J-1), IJ(K-1)}$ |

# Example

- Test 3 varieties of tomatoes on 4 planting density at significance level = 0.01

| | Planting Density | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety | 10,000 | | | 20,000 | | | 30,000 | | | 40,000 | | |
| H | 10.5 | 9.2 | 7.9 | 12.8 | 11.2 | 13.3 | 12.1 | 12.6 | 14.0 | 10.8 | 9.1 | 12.5 |
| Ife | 8.1 | 8.6 | 10.1 | 12.7 | 13.7 | 11.5 | 14.4 | 15.4 | 13.7 | 11.3 | 12.5 | 14.5 |
| P | 16.1 | 15.3 | 17.5 | 16.6 | 19.2 | 18.5 | 20.8 | 18.0 | 21.0 | 18.4 | 18.9 | 17.2 |

- A = Varieties of tomatoes, I = 3
- B = Planting densities, J = 4
- K = Number of samples per factors A and B = 3
- IJK = 36



# Example (cont.)

- Test 3 varieties of tomatoes on 4 planting density at significance level = 0.01

| | Planting Density | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety | 10,000 | | | 20,000 | | | 30,000 | | | 40,000 | | |
| H | 10.5 | 9.2 | 7.9 | 12.8 | 11.2 | 13.3 | 12.1 | 12.6 | 14.0 | 10.8 | 9.1 | 12.5 |
| Ife | 8.1 | 8.6 | 10.1 | 12.7 | 13.7 | 11.5 | 14.4 | 15.4 | 13.7 | 11.3 | 12.5 | 14.5 |
| P | 16.1 | 15.3 | 17.5 | 16.6 | 19.2 | 18.5 | 20.8 | 18.0 | 21.0 | 18.4 | 18.9 | 17.2 |

- SST = 460.36, df = 35 (IJK -1)
- SSA = 327.60, df = 2 (I-1)
- SSB = 86.69, df = 3 (J-1)
- SSE = 38.04, df = 24 (IJ(K-1)) -> MSE = 38.04 / 24 = 1.59
- SSAB = 460.36 – 327.60 – 86.69 – 38.04 = 8.03, df = 35 – 2 – 3 - 24 = 6

## Example (cont.)

- Test statistic :

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 163.8 | $f_A$ = 103.02 | $F_{0.01, 2, 24} = 5.61$ |
| $H_{0B}$ vs. $H_{aB}$ | 28.9 | $f_B$ = 18.18 | $F_{0.01, 3, 24} = 4.72$ |
| $H_{0AB}$ vs. $H_{aAB}$ | 1.34 | $f_{AB}$ = 0.84 | $F_{0.01, 6, 24} = 3.67$ |

- $H_{0AB}$ is not rejected. Interaction has no effect.
- $H_{0A}$ is rejected. Factor A (varieties of tomatoes) has effect on average product
- $H_{0B}$ is rejected. Factor B (planting densities) has effect on average product



## Multiple Comparisons

- When interaction is rejected and one or both factors has the effect, we can perform multiple comparisons.
- To specify which mean is different from others
  - Fix on factor A or B
  - Find Tukey's Honestly Significant Difference (HSD) using the following formula

    Factor A: $w_A = q_{\alpha,I,IJ(K-1)}\sqrt{\frac{MSE}{JK}}$

    Factor B: $w_B = q_{\alpha,J,IJ(K-1)}\sqrt{\frac{MSE}{IK}}$
  - $q_{\alpha, m, n}$ = q-value from studentized range distribution with 2 degrees of freedom *m, n*
- Follow the same steps as finding other HSD

## Example (cont.)

I = 3, J = 4, K = 3    IJ(K-1) = 24

- Test 3 varieties of tomatoes on 4 planting density at α = 0.01

| | Planting Density | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety | 10,000 | | | 20,000 | | | 30,000 | | | 40,000 | | | $\bar{x}_i$ |
| H | 10.5 | 9.2 | 7.9 | 12.8 | 11.2 | 13.3 | 12.1 | 12.6 | 14.0 | 10.8 | 9.1 | 12.5 | 11.33 |
| Ife | 8.1 | 8.6 | 10.1 | 12.7 | 13.7 | 11.5 | 14.4 | 15.4 | 13.7 | 11.3 | 12.5 | 14.5 | 12.21 |
| P | 16.1 | 15.3 | 17.5 | 16.6 | 19.2 | 18.5 | 20.8 | 18.0 | 21.0 | 18.4 | 18.9 | 17.2 | 18.13 |

Factor A: $w_A = q_{\alpha,I,IJ(K-1)}\sqrt{\frac{MSE}{JK}}$

$$w_A = q_{0.01,3,24}\sqrt{\frac{MSE}{JK}} = 4.55\sqrt{\frac{1.59}{12}} = 1.66$$

- Sort sample means ($\bar{x}_i$) : $\underline{\overset{\bar{x}_1}{11.33},\ \overset{\bar{x}_2}{12.21}}$, $\overset{\bar{x}_3}{18.13}$
- 2 groups of means: $\{\bar{x}_1, \bar{x}_2\}$ and $\bar{x}_3$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

## Example (cont.)

I = 3, J = 4, K = 3    IJ(K-1) = 24

- Test 3 varieties of tomatoes on 4 planting density at α = 0.01

| | Planting Density | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variety | 10,000 | | | 20,000 | | | 30,000 | | | 40,000 | | |
| H | 10.5 | 9.2 | 7.9 | 12.8 | 11.2 | 13.3 | 12.1 | 12.6 | 14.0 | 10.8 | 9.1 | 12.5 |
| Ife | 8.1 | 8.6 | 10.1 | 12.7 | 13.7 | 11.5 | 14.4 | 15.4 | 13.7 | 11.3 | 12.5 | 14.5 |
| P | 16.1 | 15.3 | 17.5 | 16.6 | 19.2 | 18.5 | 20.8 | 18.0 | 21.0 | 18.4 | 18.9 | 17.2 |
| $\bar{x}_j$ | 11.48 | | | 14.39 | | | 15.78 | | | 13.91 | | |

Factor B: $w_B = q_{\alpha,J,IJ(K-1)}\sqrt{\frac{MSE}{IK}}$

$$w_B = q_{0.01,4,24}\sqrt{\frac{MSE}{IK}} = 4.91\sqrt{\frac{1.59}{9}} = 2.06$$

- Sort sample means ($\bar{x}_i$) : $\overset{\bar{x}_1}{11.48}$, $\underline{\overset{\bar{x}_4}{13.91},\ \overset{\bar{x}_2}{14.39},\ \overset{\bar{x}_3}{15.78}}$
- 2 groups of mean: $\bar{x}_1$ and $\{\bar{x}_4, \bar{x}_2, \bar{x}_3\}$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

# Example 2

- Test 2 varieties of Iron (Fe) on 3 concentration doses
- 18 samples per category

| $Fe^{2+}$ | | | $Fe^{3+}$ | | |
|---|---|---|---|---|---|
| **10.2** | **1.2** | **0.3** | **10.2** | **1.2** | **0.3** |
| 0.71 | 2.20 | 2.25 | 2.20 | 4.04 | 2.71 |
| 1.66 | 2.93 | 3.93 | 2.69 | 4.16 | 5.43 |
| 2.01 | 3.08 | 5.08 | 3.54 | 4.42 | 6.38 |
| 2.16 | 3.49 | 5.82 | 3.75 | 4.93 | 6.38 |
| 2.42 | 4.11 | 5.84 | 3.83 | 5.49 | 8.32 |
| 2.42 | 4.95 | 6.89 | 4.08 | 5.77 | 9.04 |
| 2.56 | 5.16 | 8.50 | 4.27 | 5.86 | 9.56 |
| 2.60 | 5.54 | 8.56 | 4.53 | 6.28 | 10.01 |
| 3.31 | 5.68 | 9.44 | 5.32 | 6.97 | 10.08 |
| 3.64 | 6.25 | 10.52 | 6.18 | 7.06 | 10.62 |
| 3.74 | 7.25 | 13.46 | 6.22 | 7.78 | 13.80 |
| 3.74 | 7.90 | 13.57 | 6.33 | 9.23 | 15.99 |
| 4.39 | 8.85 | 14.76 | 6.97 | 9.34 | 17.90 |
| 4.50 | 11.96 | 16.41 | 6.97 | 9.91 | 18.25 |
| 5.07 | 15.54 | 16.96 | 7.52 | 13.46 | 19.32 |
| 5.26 | 15.89 | 17.56 | 8.36 | 18.40 | 19.87 |
| 8.15 | 18.30 | 22.82 | 11.65 | 23.89 | 21.60 |
| 8.24 | 18.59 | 29.13 | 12.45 | 26.39 | 22.25 |

- A = Forms of Irons, I = 2, B = Concentration of doses, J = 3
- K = Number of samples per factors A and B = 18, IJK = 108



# Example 2 (cont.)

- Test 2 varieties of Iron (Fe) on 3 concentration doses

$\bar{x} = 8.64$

$\bar{x}_{Fe2+} = 7.88$
$\bar{x}_{Fe3+} = 9.40$

$\bar{x}_{10.2} = 4.82$
$\bar{x}_{1.2} = 8.92$
$\bar{x}_{0.3} = 12.19$

| $Fe^{2+}$ | | | $Fe^{3+}$ | | |
|---|---|---|---|---|---|
| **10.2** | **1.2** | **0.3** | **10.2** | **1.2** | **0.3** |
| 0.71 | 2.20 | 2.25 | 2.20 | 4.04 | 2.71 |
| 1.66 | 2.93 | 3.93 | 2.69 | 4.16 | 5.43 |
| 2.01 | 3.08 | 5.08 | 3.54 | 4.42 | 6.38 |
| 2.16 | 3.49 | 5.82 | 3.75 | 4.93 | 6.38 |
| 2.42 | 4.11 | 5.84 | 3.83 | 5.49 | 8.32 |
| 2.42 | 4.95 | 6.89 | 4.08 | 5.77 | 9.04 |
| 2.56 | 5.16 | 8.50 | 4.27 | 5.86 | 9.56 |
| 2.60 | 5.54 | 8.56 | 4.53 | 6.28 | 10.01 |
| 3.31 | 5.68 | 9.44 | 5.32 | 6.97 | 10.08 |
| 3.64 | 6.25 | 10.52 | 6.18 | 7.06 | 10.62 |
| 3.74 | 7.25 | 13.46 | 6.22 | 7.78 | 13.80 |
| 3.74 | 7.90 | 13.57 | 6.33 | 9.23 | 15.99 |
| 4.39 | 8.85 | 14.76 | 6.97 | 9.34 | 17.90 |
| 4.50 | 11.96 | 16.41 | 6.97 | 9.91 | 18.25 |
| 5.07 | 15.54 | 16.96 | 7.52 | 13.46 | 19.32 |
| 5.26 | 15.89 | 17.56 | 8.36 | 18.40 | 19.87 |
| 8.15 | 18.30 | 22.82 | 11.65 | 23.89 | 21.60 |
| 8.24 | 18.59 | 29.13 | 12.45 | 26.39 | 22.25 |

- SST = 3992.37, df = 107 (IJK -1)
- SSA = 62.26, df = 1 (*I*-1)
- SSB = 983.62, df = 2 (J-1)
- SSE = 2938.20, df = 102 (*I*J(K-1))
- SSAB = SST – SSA – SSB – SSE = 8.29, df = 2

## Example 2(cont.)

- SST = 3992.37, df = 107
- SSA = 62.26, df = 1
- SSB = 983.62, df = 2
- SSE = 2938.20, df = 102
- SSAB = 8.29, df = 2

**MSE = 2938.20 / 102 = 28.81**

- Test statistic :

| Hypothesis | Mean Square (MS) | Test statistic (f) | Rejection region | |
|---|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 62.26/1= 62.26 | $f_A$ = 62.26/28.81 = 2.16 | $F_{0.01, 1, 102}$= 6.89<br>$F_{0.05, 1, 102}$= 3.93<br>$F_{0.1, 1, 102}$= 2.76 | Not reject at 0.1 |
| $H_{0B}$ vs. $H_{aB}$ | 983.62/2 = 491.81 | $f_B$ = 491.81/28.81 = 17.07 | $F_{0.01, 2, 102}$= 4.82<br>$F_{0.05, 2, 102}$= 3.09<br>$F_{0.1, 2, 102}$= 2.36 | Reject at 0.01 |
| $H_{0AB}$ vs. $H_{aAB}$ | 8.29/2 = 4.15 | $f_{AB}$ = 4.15/28.81 = 0.14 | $F_{0.01, 2, 102}$= 4.82<br>$F_{0.05, 2, 102}$= 3.09<br>$F_{0.1, 2, 102}$= 2.36 | Not reject at 0.1 |

- $H_{0AB}$ is not rejected. Interaction has no effect.
- $H_{0A}$ is not rejected. Factor A (forms of Iron) has no effect on % of retained iron
- $H_{0B}$ is rejected. Factor B (Concentration) has effect on % of retained iron

## Example 2(cont.)

- SST = 3992.37, df = 107
- SSA = 62.26, df = 1
- SSB = 983.62, df = 2
- SSE = 2938.20, df = 102
- SSAB = 8.29, df = 2

**MSE = 2938.20 / 102 = 28.81**

- Test statistic :

| Hypothesis | Mean Square (MS) | Test statistic (f) | P-value |
|---|---|---|---|
| $H_{0A}$ vs. $H_{aA}$ | 62.26/1= 62.26 | $f_A$ = 62.26/28.81 = 2.16 | 0.14<br>Not reject at 0.1 |
| $H_{0B}$ vs. $H_{aB}$ | 983.62/2 = 491.81 | $f_B$ = 491.81/28.81 = 17.07 | $4.02 \times 10^{-7}$<br>Reject at 0.01 |
| $H_{0AB}$ vs. $H_{aAB}$ | 8.29/2 = 4.15 | $f_{AB}$ = 4.15/28.81 = 0.14 | 0.87<br>Not reject at 0.1 |

- $H_{0AB}$ is not rejected. Interaction has no effect.
- $H_{0A}$ is not rejected. Factor A (forms of Iron) has no effect on % of retained iron
- $H_{0B}$ is rejected. Factor B (Concentration) has effect on % of retained iron

# References


1. J.L. Devore and K.N. Berk, Modern Mathematical Statistics with Applications, Springer, 2012.
2. J.A. Rice, Mathematical Statistics and Data Analysis, Duxbury Press, 1995.