Analysis of Variance
Single Factor

Dr. Supaporn Erjongmanee

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

Outline

- Introduction
- Single-Factor ANOVA:
  - Equal sample size
    - F-test
    - Multiple comparison

Introduction : Analysis of Variance

- Analysis of quantitative responses
- In short, ANOVA
- Simplest ANOVA
  - Single-factor or One-way
  - Factorial or Multiple-way
- Examples
  - Study of five brands of gasoline on car efficiency
  - Study of four types of sugar on bacteria growth

Outline

- Introduction
- Single-Factor ANOVA:
  - Equal sample size
    - F-test
    - Multiple comparison

Single-Factor ANOVA : Equal Sample Size

- Compare two or more populations on one factor
- Let
  - μ_1 = mean of treatment (population) 1
  - μ_2 = mean of treatment (population) 2
  - …
  - μ_I = mean of treatment (population) I

I = number of compared treatments

- Hypothesis
  - H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$
  - H_a: Not all μ_i's are equal (at least two of the μ_i's are different)

Notation

- Let
  - X_{ij} = Random variable for measurement j of treatment i
  - x_{ij} = Sample value for measurement j of treatment i
  - **J** = Number of samples in one treatment
  - **I** = Number of treatments
- (Treatment) sample mean: $\bar{X}_i = \frac{\sum_{j=1}^{J} X_{ij}}{J}$

Divided by number of samples in one treatment

- Grand mean: $\bar{X} = \frac{\sum_{i=1}^{I} \bar{X}_i}{I} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}}{IJ}$

Divided by number of samples from treatments

- (Treatment) sample variance: $S_i^2 = \frac{\sum_{j=1}^{J}(X_{ij}-\bar{X}_i)^2}{J-1}$

Example

- Experiment testing strength of 4 shipping boxes

Type	Compression Strength (lb)						Sample Mean	Sample SD
1	655.5	788.3	734.3	721.4	679.1	699.4	713.00	46.55
2	789.2	772.5	786.9	686.1	732.1	774.8	756.93	40.34
3	737.1	639.0	696.3	671.7	717.2	727.1	698.07	37.20
4	535.1	628.7	542.4	559.0	586.9	520.0	562.02	39.87

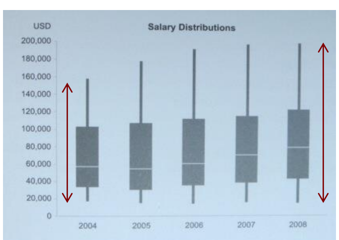
Grand Mean = (713.00+756.93+698.07+563.02) / 4
= 682.50

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Box Plots

- ✓ Analyze distribution within series
- ✓ Analyze change of distributions over time



- To analyze distribution change within series or over time
- Feature
  - <u>Length</u> represents distributed amount

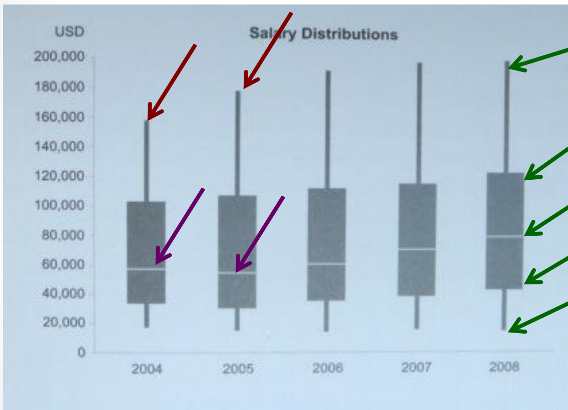
Image source: Figure 7.23, [2]

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Box Plots (cont.)

- Example



100%

75%

50%

25%

0%

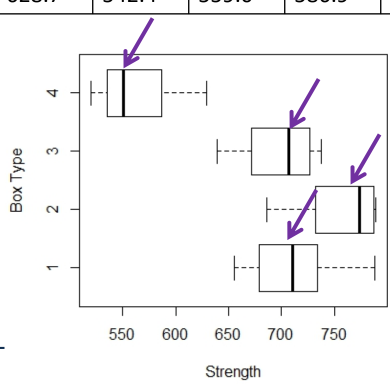
Median salary paid in 2004 is more than 2005.

Maximum salary in 2004 is less than in 2005.

Image source: Figure 7.23, [2]

Example (cont.)

Type	Compression Strength (lb)						Sample Mean	Sample SD
1	655.5	788.3	734.3	721.4	679.1	699.4	713.00	46.55
2	789.2	772.5	786.9	686.1	732.1	774.8	756.93	40.34
3	737.1	639.0	696.3	671.7	717.2	727.1	698.07	37.20
4	535.1	628.7	542.4	559.0	586.9	520.0	562.02	39.87



Basic Assumption

- Assume
  - Distribution of each population is normal with the same variance = σ^2
- Therefore, each sample X_{ij} comes from normal distribution with
  - $E(X_{ij}) = \mu_i$
  - $V(X_{ij}) = \sigma^2$
  - Hypothesis
    - $H_0: \mu_1 = \mu_2 = \ldots = \mu_I$
    - H_a: Not all μ_i's are equal (at least two of the μ_i's are different)

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Sum of Squares

- When H_0 is true, all sample means ($\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_I$) should be the same
- Therefore, test statistics will measured from differences of sample means
- Treatment sum of squares (SST_r) : Difference between different treatments
  - Sum of difference between each sample mean and grand mean
  - $SST_r = J(\bar{X}_1 - \bar{X})^2 + J(\bar{X}_2 - \bar{X})^2 + \cdots + J(\bar{X}_I - \bar{X})^2$
    $= J \sum_i (\bar{X}_i - \bar{X})^2$
- Error sum of squares (SSE): Difference within the same treatment
  - Sum of differences between samples and sample mean
  - $SSE = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Review: Sample Variance Distribution

- Let $X_1, X_2, \ldots, X_n$ be random sample from a normal distribution with mean value = μ and standard deviation = σ.
- Then, sample variance S^2 has distribution to be a chi-square distribution with degree of freedom = n-1

$$S^2 \approx \sigma^2 \frac{\chi^2_{n-1}}{(n-1)} \quad \Rightarrow \quad \frac{(n-1)S^2}{\sigma^2} \approx \chi^2_{(n-1)}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Sum of Squares (cont.)

- Treatment sum of squares (SST_r) :
  - $SST_r = J\ \sum_i (\bar{X}_i - \bar{X})^2$

$$= J \sum_i (Y_i - \bar{Y}\ )^2$$

$$= J\ (I-1) \frac{\sum_i (Y_i - \bar{Y}\ )^2}{(I-1)}$$

$$= (\mathrm{I}-1) J\ S_Y^2$$

$$\frac{SST_r}{\sigma^2} = \frac{(I-1)S_Y^2}{\sigma^2/J}$$

$$= \frac{(I-1)}{(\sigma^2/J)} \cdot \frac{\sigma_Y^2 \chi^2_{I-1}}{I-1}$$

$$= \frac{\cancel{(I-1)}\cancel{(\sigma^2/J)}\chi^2_{I-1}}{\cancel{(\sigma^2/J)}\cancel{(I-1)}} \approx \chi^2_{I-1}$$

$$\boxed{X_i \sim N(\mu, \sigma^2) \rightarrow Y_i = \bar{X}_i \sim N(\mu, \frac{\sigma^2}{J})}$$

$$\boxed{s^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n-1}}$$

$$\boxed{S_X{}^2 \approx \sigma_X^2 \frac{\chi^2_{n-1}}{(n-1)}}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Sum of Squares (cont.)

- Error sum of squares (SSE)
  - $SSE = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2$

$= \sum_j (X_{1j} - \bar{X}_1)^2 + \sum_j (X_{2j} - \bar{X}_2)^2 + \ldots + \sum_j (X_{Ij} - \bar{X}_I)^2$

$= (J-1)S_1^2 + (J-1)S_2^2 + \cdots + (J-1)S_I^2$

$= (J-1)[S_1^2 + S_2^2 + \ldots + S_I^2]$

Each X_i has the same variance

$=I(J-1)S^2$

$$\frac{SSE}{\sigma^2} = \frac{I(J-1)S^2}{\sigma^2} \approx \chi^2_{I(J-1)}$$

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

$$X_i \sim N(\mu, s^2)$$

$$\frac{(n-1)S^2}{\sigma^2} \approx \chi^2_{(n-1)}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Statistics in Computer Engineering Applications
Slide 15

Department of Computer Engineering
Kasetsart University

F distribution

- Let X_1 and X_2 be <u>independent</u> *chi-squared* random variables with v_1 and v_2 degrees of freedom

$$F_{v1, v2} = \frac{X_1/v_1}{X_2/v_2}$$

- Generally, sample variance has sampling distribution in term of chi-squared distribution with degree of freedom = n -1

$$S^2 \approx \sigma^2 \frac{\chi^2_{n-1}}{(n-1)}$$

- Let S_1^2 and S_2^2 be sample variances with chi-squared distribution

$$\frac{(m-1)S_1^2}{\sigma_1^2} \approx \chi^2_{(m-1)} \quad \text{and} \quad \frac{(n-1)S_2^2}{\sigma_2^2} \approx \chi^2_{(n-1)}$$

$$F_{m-1,n-1} = \frac{\frac{(m-1)S_1^2/\sigma_1^2}{m-1}}{\frac{(n-1)S_2^2/\sigma_2^2}{n-1}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Statistics in Computer Engineering Applications
Slide 16

Department of Computer Engineering
Kasetsart University

F distribution (cont.)

- Let X_1 and X_2 be independent chi-squared random variables with v_1 and v_2 degrees of freedom

$$F_{v1,\,v2} = \frac{X_1/v_1}{X_2/v_2}$$

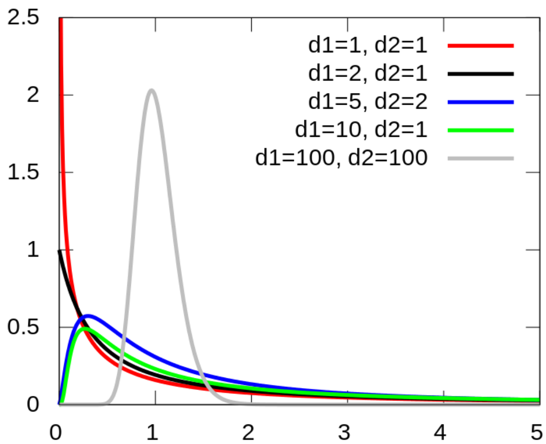


Image source: http://en.wikipedia.org/wiki/F-distribution

Sum of Squares (cont.)

- Treatment sum of squares (SST_r) :

$$\frac{SST_r}{\sigma^2} = \frac{(I-1)S_{\bar{Y}}^2}{\sigma^2/J} = \frac{(I-1)(\sigma^2/J)\chi_{I-1}^2}{(\sigma^2/J)(I-1)} \approx \chi_{I-1}^2$$

- Error sum of squares (SSE)

$$\frac{SSE}{\sigma^2} = \frac{I(J-1)S^2}{\sigma^2} \approx \chi_{I(J-1)}^2$$

Use as test statistic

$$\frac{\frac{SSTr}{\cancel{\sigma^2}}/(I-1)}{\frac{SSE}{\cancel{\sigma^2}}/I(J-1)} = \frac{SSTr/(I-1)}{SSE/I(J-1)} = \frac{\chi_{I-1}^2/(I-1)}{\chi_{I(J-1)}^2/I(J-1)} \approx F_{I-1,I(J-1)}$$

To Perform F-Test

Mean of chi-square distribution = degree of freedom

$$E\left(\frac{SST_r}{\sigma^2}\right) = I - 1 \qquad E\left(\frac{SSE}{\sigma^2}\right) = I(J-1)$$

- Test statistic:

$$f = F_{I-1,I(J-1)} = \frac{SSTr/(I-1)}{SSE/(I(J-1))} = \frac{MSTr}{MSE}$$

$$\boxed{E\left(\frac{SSTr}{I-1}\right) = \sigma^2} \qquad \boxed{E\left(\frac{SSE}{I(J-1)}\right) = \sigma^2}$$

$$\boxed{E(MSTr) = \sigma^2} \qquad \boxed{E(MSE) = \sigma^2}$$

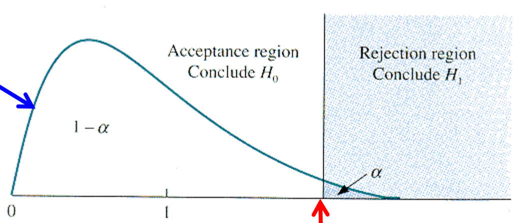
- If H_0 is true ($\bar{X}_i's$ are about the same), MSTr and MSE are unbiased estimates of σ^2 => test statistic f ~ 1 (f is small)
- If H_0 is false ($\bar{X}_i's$ are not the same), E(MSTr) > σ^2 => test statistic f is large.
- Hence, rejection region covers large test statistic f

F-Test

- Test statistic:

$$F_{I-1,I(J-1)} = \frac{\frac{SSTr}{\sigma^2}/(I-1)}{\frac{SSE}{\sigma^2}/(I(J-1))} = \frac{SSTr/(I-1)}{SSE/(I(J-1))} = \frac{MSTr}{MSE}$$

F distribution
for v_1 and v_2



$F_{\alpha, v1, v2}$

If test statistic > $F_{\alpha, v1, v2}$, reject H_0

Image source: http://www.unc.edu/~nielsen/soci708/m16/m2009.gif

Summary : ANOVA

- Test statistic:

$$F_{I-1,I(J-1)} = \frac{\frac{SSTr}{\sigma^2}/(I-1)}{\frac{SSE}{\sigma^2}/(I(J-1))} = \frac{SSTr/(I-1)}{SSE/(I(J-1))} = \frac{MSTr}{MSE}$$

- When f is large, we are about to reject H_0

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	I-1	SSTr	SSTr/(I-1)	MSTr / MSE
Error	I(J-1)	SSE	SSE/(I(J-1))	
Total	IJ-1	SST		

Summary : ANOVA (cont.)

- SST = SSTr + SSE

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i + \bar{x}_i - \bar{x})^2$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum_i \sum_j (\bar{x}_i - \bar{x})^2 - 2\sum_i \sum_j (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})$$

$$= SSE + SSTr \cancel{-2\sum_i \sum_j (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})}$$ 0

Summary : ANOVA (cont.)

- Test statistic:

$$F_{I-1,I(J-1)} = \frac{\frac{SSTr}{\sigma^2}/(I-1)}{\frac{SSE}{\sigma^2}/(I(J-1))} = \frac{SSTr/(I-1)}{SSE/(I(J-1))} = \frac{MSTr}{MSE}$$

Another option: Sample-based computation

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	I-1	$\frac{1}{J}\sum_{i=1}^{I}(\sum_{j=1}^{J} x_{ij})^2 - \frac{1}{IJ}(\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij})^2$	SSTr/(I-1)	MSTr / MSE
Error	I(J-1)	SST - SSTr	SSE/(I(J-1))	
Total	IJ-1	$\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}^2 - \frac{1}{IJ}(\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij})^2$		

Example 1

- Experiment degree of soiling on 3 mixtures of fabric and polymer
- Prove whether 3 mixture means are the same at α = 0.01

Mixture	Degree of soiling				
1:	0.56	1.12	0.90	1.07	0.94
2:	0.72	0.69	0.87	0.78	0.91
3:	0.62	1.08	1.07	0.99	0.93

- Let
  - μ_1 = mean of mixture 1
  - μ_2 = mean of mixture 2
  - μ_3 = mean of mixture 3
  - I = 3, J = 5
- Hypothesis
  - H_0: $\mu_1 = \mu_2 = \mu_3$
  - H_a: Not all μ_i's are equal (at least two of the μ_i's are different)

Example 1 (cont.)

- Experiment degree of soiling on 3 mixtures

Mixture	Degree of soiling					$\bar{x}_i$
1:	0.56	1.12	0.90	1.07	0.94	0.918
2:	0.72	0.69	0.87	0.78	0.91	0.794
3:	0.62	1.08	1.07	0.99	0.93	0.938

$\bar{x}$= 0.883

- Fill ANOVA table

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	**2**	**0.0608**	**0.0304**	**0.99**
Error	**12**	**0.3701 = 0.4309 – 0.0608**	**0.0308**	
Total	**14**	**0.4309**		

Example 1 (cont.)

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	**2**	**0.0608**	**0.0304**	**0.99**
Error	**12**	**0.3701 = 0.4309 – 0.0608**	**0.0308**	
Total	**14**	**0.4309**		

- Rejection region: given α = 0.01
  - $F_{0.01, 2, 12} = 6.93$
- H_0 is not rejected.
- All mixture means are equals.

Outline

- Introduction
- Single-Factor ANOVA:
  - Equal sample size
    - F-test
    - Multiple comparison

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Multiple Comparisons

- Question:
  - When H_0 for ANOVA is rejected, how many means are different from each other?
- Procedures:
  - Find confidence interval of pairwise difference $\mu_i - \mu_j$
  - If confidence interval for any pairwise difference $\mu_i - \mu_j$ does not include zero, we determine that μ_i, μ_j are significantly different from each other

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Studentized Range Distribution

- Let $Z_1, Z_2, \ldots, Z_m$ be *m* independent *standard normal* random variables
- Let W be a chi-squared random variable with degree of freedom = v, and independent of the Z_i's
- Then, Q distribution, called studentized range distribution is defined as

$$Q = \frac{\max|Z_i - Z_j|}{\sqrt{\frac{W}{v}}} \qquad \text{where } W = \frac{SSE}{\sigma^2} = \chi^2_{I(J-1)} = \frac{I(J-1)MSE}{\sigma^2}$$

- This Q distribution has 2 parameters: m and v
  - Hence, it is denoted by $Q_{\alpha,m,v}$

Studentized Range Distribution (cont.)

$$X_i \sim N(\mu, \sigma^2) \rightarrow \bar{X}_i \sim N(\mu, \frac{\sigma^2}{J})$$

$$Z_i = \frac{\bar{X}_i - \mu_i}{\sigma/\sqrt{J}} \sim N(0,1)$$

- From

$$Z_i = \frac{\bar{X}_i - \mu_i}{\sigma/\sqrt{J}}, \qquad W = \frac{SSE}{\sigma^2} = \chi^2_{I(J-1)} = \frac{I(J-1)MSE}{\sigma^2}, \qquad m = I, \qquad v = I(J-1)$$

$$Q = \frac{\max|Z_i - Z_j|}{\sqrt{\frac{W}{v}}} = \frac{\max\left|\frac{\bar{X}_i - \mu_i}{\sigma/\sqrt{J}} - \frac{\bar{X}_j - \mu_j}{\sigma/\sqrt{J}}\right|}{\sqrt{\frac{\frac{I(J-1)MSE}{\sigma^2}}{I(J-1)}}} = \frac{\max|\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)|}{\sqrt{MSE/J}}$$

$$1 - \alpha = P\left(\frac{\max|\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)|}{\sqrt{MSE/J}} \le Q_{\alpha,I,I(J-1)}\right)$$

$$= P\left(\frac{|\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)|}{\sqrt{MSE/J}} \le Q_{\alpha,I,I(J-1)} \; for\ all\ i,j\right)$$

Probability that maximum difference between sample mean difference $(\bar{X}_i - \bar{X}_j)$ true mean difference $(\mu_i - \mu_j)$ is less than $Q_{\alpha,I,I(J-1)}\sqrt{MSE/J}$ $= 1 - \alpha$

Studentized Range Distribution (cont.)

$$1-\alpha = P\left(\frac{|\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)|}{\sqrt{MSE/J}} \le Q_{\alpha,I,I(J-1)} \; for\ all\ i,j\right)$$

$$= P\left(-Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} \le \bar{X}_i - \bar{X}_j - (\mu_i - \mu_j) \le Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} \; for\ all\ i,j\right)$$

$$= P\left(\underbrace{\boxed{\bar{X}_i - \bar{X}_j - Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}}} \le \mu_i - \mu_j \le \boxed{\bar{X}_i - \bar{X}_j + Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}}}}_{\text{Confidence intervals between one pair of true mean difference } \mu_i - \mu_j} \; for\ all\ i,j\right)$$

Confidence intervals between one pair of true mean difference $\mu_i - \mu_j$

- There are $\binom{I}{2} = \frac{I(I-1)}{2}$ confidence intervals of $\mu_i - \mu_j$

Studentized Range Distribution (cont.)

$$1-\alpha = P\left(\bar{X}_i - \bar{X}_j - Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} \le \mu_i - \mu_j \le \bar{X}_i - \bar{X}_j + Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} \; for\ all\ i,j\right)$$

Expect difference between sample mean difference $(\bar{X}_i - \bar{X}_j)$ and true mean difference $(\mu_i - \mu_j)$ is not more than this value $Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}}$

- The value w = $Q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}}$ is called Tukey's honestly significantly difference (HSD)

Multiple Comparisons : Equal Sample Size

- When some means are not all equal, how to specify which mean is different from others
- Procedure
  1. Find Tukey's Honestly Significant Difference (HSD)

$$HSD_\alpha = q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}}$$

- $q_{\alpha,\, I,\, I(J-1)}$ = q-value from studentized range distribution with 2 degrees of freedom *I, I(J-1)*

  2. Sort sample means in increasing order
  3. Underline pairs that differ less than HSD_α
  4. Any pair without underline are considered as significantly different.

Example 1

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	2	0.0608	0.0304	0.99
Error	12	0.3701 = 0.4309 – 0.0608	0.0308	
Total	14	0.4309		

- Experiment degree of soiling on 3 mixtures
- Group mixture means at α = 0.01

Mixture	Degree of soiling					$\bar{x}_i$
1:	0.56	1.12	0.90	1.07	0.94	0.918
2:	0.72	0.69	0.87	0.78	0.91	0.794
3:	0.62	1.08	1.07	0.99	0.93	0.938

$$HSD_\alpha = q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} = q_{0.01,3,12}\sqrt{\frac{0.0308}{5}} = 5.05\sqrt{\frac{0.0308}{5}} = 0.396$$

- Sort sample means: 0.794, 0.918, 0.938 ($\bar{x}_2$, $\bar{x}_1$, $\bar{x}_3$)

  0.124 0.020

```
> qtukey(0.01, nmeans=anovaResult$rank,
+ df=anovaResult$df.residual, lower.tail=F)
[1] 5.045934
```

- One group of mixture means

Note that H_0 is not rejected.

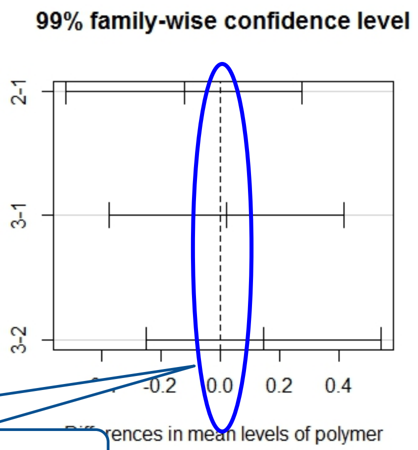
Example 1

```
> tk <- TukeyHSD(anovaResult, conf.level = 0.99)
> plot(tk)
```

- Experiment degree of soiling on 3 m
- Group mixture means at α = 0.01

Mixture	Degree of soiling				
1:	0.56	1.12	0.90	1.07	0.
2:	0.72	0.69	0.87	0.78	0.
3:	0.62	1.08	1.07	0.99	0.

$\bar{x}_i$



- Sort sample means:

  $\bar{x}_2$ $\bar{x}_1$ $\bar{x}_3$

  0.794, 0.918, 0.938

  0.124 0.020

- One group of mixture means

Confidence interval of difference $(\bar{X}_i - \bar{X}_j)$ includes zero. This means $\bar{X}_i \approx \bar{X}_j$

Example 2

- Test on 5 brands of automobile oil filters
  - Use 9 samples for each brands
- $\bar{x}_1$= 14.5, $\bar{x}_2$= 13.8, $\bar{x}_3$= 13.3, $\bar{x}_4$= 14.3, $\bar{x}_5$= 13.1

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	4	13.32	3.33	37.84
Error	40	3.53	0.088	
Total	44	16.85		

- Rejection region:
  - $F_{0.05, 4, 40}$ = 2.61
- H_0 is rejected
- Find Tukey's HSD to see mean differences

Example 2 (cont.)

- $\bar{x}_1$= 14.5, $\bar{x}_2$= 13.8, $\bar{x}_3$= 13.3, $\bar{x}_4$= 14.3, $\bar{x}_5$= 13.1

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	4	13.32	3.33	37.84
Error	40	3.53	0.088	
Total	44	16.85		

$$HSD_\alpha = q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} = q_{0.05,5,40}\sqrt{\frac{0.088}{9}} = 4.04\sqrt{\frac{0.088}{9}} = 0.399$$

$\bar{x}_5$ $\bar{x}_3$ $\bar{x}_2$ $\bar{x}_4$ $\bar{x}_1$

- Sort sample means: 13.1, 13.3, 13.8, 14.3, 14.5
- 3 groups of means:
  - $\bar{x}_5, \bar{x}_3$ are not significantly different from each other
  - $\bar{x}_4, \bar{x}_1$ are not significantly different from each other
  - $\bar{x}_2$ is significantly different from $\bar{x}_5, \bar{x}_3$ and $\bar{x}_4, \bar{x}_1$

Example 2 (cont.)

- If use another value for sample mean and same HSD_α:
  - $\bar{x}_1$= 14.5, $\bar{x}_2$= 14.15, $\bar{x}_3$= 13.3, $\bar{x}_4$= 14.3, $\bar{x}_5$= 13.1

$$HSD_\alpha = q_{0.05,5,40}\sqrt{\frac{0.088}{9}} = 4.04\sqrt{\frac{0.088}{9}} = 0.399$$

$\bar{x}_5$ $\bar{x}_3$ $\bar{x}_2$ $\bar{x}_4$ $\bar{x}_1$

- Sort sample means: 13.1, 13.3, 14.15, 14.3, 14.5
- 2 groups of means
  - $\bar{x}_5, \bar{x}_3$ are not significantly different from each other
  - $\bar{x}_2, \bar{x}_4, \bar{x}_1$ are not significantly different from each other

Example 3

- For another data set:
  - $\bar{x}_1$= 79.28, $\bar{x}_2$= 61.54, $\bar{x}_3$= 47.92, $\bar{x}_4$= 32.76

	df	Sum of Squares (SS)	Mean Square (MS)	f
Treatment	3	5882.3575	1960.7858	21.09
Error	16	1487.4000	92.9625	
Total	19	7369.7575		

- I = 4, J = 5

$$HSD_\alpha = q_{\alpha,I,I(J-1)}\sqrt{\frac{MSE}{J}} = q_{0.05,4,16}\sqrt{\frac{92.9625}{5}} = 4.05\sqrt{\frac{92.9625}{5}} = 17.47$$

- Sort sample means: 32.76, 47.92, 61.54, 79.28 ($\bar{x}_4$, $\bar{x}_3$, $\bar{x}_2$, $\bar{x}_1$)

Although $\bar{x}_4, \bar{x}_2$ are different from each other, they are not different from $\bar{x}_3$

- 2 groups of means
  - $\bar{x}_4, \bar{x}_3, \bar{x}_2$ are not significantly different from each other
  - $\bar{x}_1$ is significantly different from each other

Supaporn Erjongmanee
fengspe@ku.ac.th

Statistics in Computer Engineering Applications
Slide 39

Department of Computer Engineering
Kasetsart University

Example 3

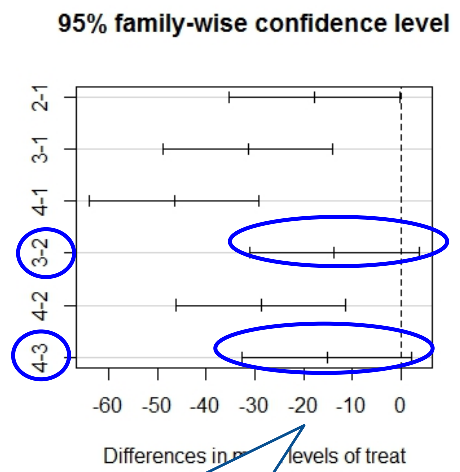
- For another data set:
  - $\bar{x}_1$= 79.28, $\bar{x}_2$= 61.54, $\bar{x}_3$= 47.92, $\bar{x}_4$= 32.76

$$HSD_\alpha = 4.05\sqrt{\frac{932.9625}{5}} = 17.47$$

- Sort sample means:

$\bar{x}_4$ $\bar{x}_3$ $\bar{x}_2$ $\bar{x}_1$
32.76, 47.92, 61.54, 79.28

- 2 groups of means
  - $\bar{x}_4, \bar{x}_3, \bar{x}_2$ are not significantly different from each other
  - $\bar{x}_1$ is significantly different from each other



Confidence intervals of $(\bar{X}_2 - \bar{X}_3)$ and $(\bar{X}_3 - \bar{X}_4)$ include zero. This means $\bar{X}_2 \approx \bar{X}_3 \approx \bar{X}_4$

Supap...
fengspe...
...neering

References

1. J.L. Devore and K.N.Berk, Modern Mathematical Statistics with Applications, Springer, 2012.
2. S. Few, Now You See It: Simple Visualization Techniques for Quantitative Analysis, Analytics Press, 2009

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University