MATH 4720 / MSSC 5720

Instructor: Mehdi Maadooliat

Chapter 9

Department of Mathematical and Statistical Sciences

MULTIPLE COMPARISON (POST HOC ANOVA)

MARQUETTE UNIVERSITY
Be The Difference.

- **In the ANOVA we test**

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_t$$

$$H_a: \mu_i \neq \mu_j \text{ for some pairs } (i, j)$$

- **If we reject H_0 in favor of H_a, then the question arises that for what pairs $\mu_i \neq \mu_j$.**

- **Now, we want to test**

$$H_0^{ij}: \mu_i = \mu_j$$

$$H_a^{ij}: \mu_i \neq \mu_j$$

- **for all pairs using, say, two-sample t-statistics.**

MULTIPLE COMPARISON CONT'D

MARQUETTE UNIVERSITY
Be The Difference.

- $t_{ij} = \frac{\bar{y}_i - \bar{y}_j}{s_p\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}$ **for all pairs** (i, j)

- **We will be testing many hypotheses. For example, if there are** 5 **treatments, there will be** 10 **hypotheses.**

- **If each hypothesis is tested at** $\alpha = 0.05$**, there is a good chance that one hypothesis will be falsely rejected.**

- **For illustration, suppose you are testing 100 hypothesis each at** $\alpha = 0.05$**, and suppose null** H_0 **is true for all hypotheses, then 5 (i.e. 5%) hypothesis will be falsely rejected.**

- **This problem always arises due to multiplicity of hypotheses.**

MULTIPLE COMPARISON CONT'D

MARQUETTE UNIVERSITY
Be The Difference.

- The problem of **multiplicity** is serious when you are testing many hypotheses, say thousands of hypotheses.

- **Example: Gene Expression Analysis**

- A data is collected on gene expressions on $10{,}000$ genes. This kind of research is done if you are trying to find genes responsible for cancer. If those genes are correctly detected, you can study their structure and come up with a cure.

- $$H_0^i: \mu_i = 0 \quad vs. \quad H_a^i: \mu_i \neq 0, \; i = 1, 2, \ldots, 10000$$

- If we test each of the hypotheses at $\alpha = 0.05$, then 5% of them (500 **hypotheses**) will be falsely discovered even if all nulls are true.

FAMILYWISE ERROR RATE (FEW)

MARQUETTE UNIVERSITY
Be The Difference.

- **Instead of using Type-I error rate, we should use a different type of error rate**

- α_F = **P(Falsely reject at least one hypotheses)**

- **If there are m hypotheses each tested at α_I, then approximately**

- $\alpha_F \leq 1 - (1 - \alpha_I)^m$
- **Or we can say:** $\alpha_F \leq m * \alpha_I$

- **This implies that**
- **if** $\alpha_I = \frac{0.05}{m}$, **then** $\alpha_F \leq 0.05$.

TABLE 9.5
A comparison of the familywise error rate α_F for m independent multiple tests

m, Number of multiple tests	α_I Probability of a Type I Error on an Individual Test		
	.10	.05	.01
1	.100	.050	.010
2	.190	.097	.020
3	.271	.143	.030
4	.344	.185	.039
5	.410	.226	.049
.	.	.	.
.	.	.	.
.	.	.	.
10	.651	.401	.096

BONFERRONI METHOD

MARQUETTE UNIVERSITY
Be The Difference.

- If there are m hypotheses, then we can test individual hypothesis at $\alpha = \frac{0.05}{m}$ using standard method such as t-test. This will guarantee that

- α_F = P(Falsely reject at least one null) ≤ 0.05.

- Problem with this approach: If m is large, $\alpha = \frac{0.05}{m}$ will be very small, and the chance of rejecting H_0 will be small. In other word, the power of **Bonferroni** is very poor.

BEYOND BONFERRONI

MARQUETTE UNIVERSITY
Be The Difference.

- In addition to Bonferroni, there are many methods of Multiple Comparisons. Some of them are

  1. Fisher's Least Significant Difference (LSD)
  2. Tukey's Method
  3. Dunnette's Method

- Recall that the purpose is to test

- $H_0^{ij}: \mu_i = \mu_j$
- $H_a^{ij}: \mu_i \neq \mu_j$

- In all of the methods we reject H_0^{ij} if
- $|\bar{y}_i - \bar{y}_j| > C$, a constant

BOOK EXAMPLE 9.3

MARQUETTE UNIVERSITY
Be The Difference.

- **A study was done to test 5 different agents used to control weeds. Each of these agents were applied to sample of 6 one-acre plots. The hay was harvested and the total yield was recorded.**

Agent	1	2	3	4	5
Type	**None**	**Bio1**	**Bio2**	**Chm1**	**Chm2**
$\bar{y}$	1.175	1.293	1.328	1.415	1.500
s	.1204	.1269	.1196	.1249	.1265
n	6	6	6	6	6

- **The question is that if there is a difference between the agents, which agent provides the best yield.**

EXAMPLE 9.3: ASSUMPTIONS

MARQUETTE UNIVERSITY
Be The Difference.

- **First, we check the validity of assumptions among the agents.**

- H_0: $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5$
  - **In R:** with(exmp9.3, levene.test(yield, agent))

Test for Equal Variances: yield versus Agent

95% Bonferroni confidence intervals for standard deviations

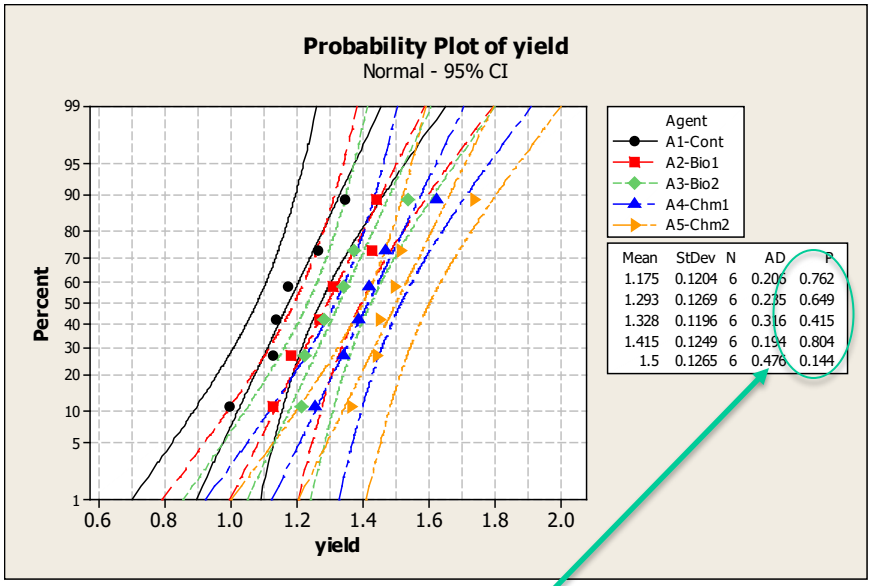
Agent	N	Lower	StDev	Upper
A1-Cont	6	0.0657789	0.120394	0.419543
A2-Bio1	6	0.0693292	0.126892	0.442187
A3-Bio2	6	0.0653430	0.119596	0.416762
A4-Chm1	6	0.0682488	0.124914	0.435296
A5-Chm2	6	0.0691068	0.126485	0.440768

Bartlett's Test (Normal Distribution)
Test statistic = 0.03, p-value = 1.000

Levene's Test (Any Continuous Distribution)
Test statistic = 0.04, p-value = 0.996

- **Levene's test $p - value$ is 0.996. Fail to reject equality of the variances**

- H_0: **Data is generated from normal distribution for each type of food.**



- **Large $p - values$. Fail to reject Normality assumption.**

EXAMPLE 9.3: ANOVA

- **Next, we perform ANOVA to test if there is a difference among the agents.**
- $H_0: \mu_1 = \mu_2 = \cdots = \mu_5$

  $H_a: \mu_i \neq \mu_j$ **for some pairs** (i, j)

➢ **In R:** summary(model <- aov(yield~agent, data= exmp9.3))

One-way ANOVA: yield versus Agent

Source	DF	SS	MS	F	P
Agent	4	0.3647	0.0912	5.96	0.002
Error	25	0.3825	0.0153		
Total	29	0.7472			

S = 0.1237 R-Sq = 48.81% R-Sq(adj) = 40.62%

- **TS.** $F = \dfrac{SS_B/df_B}{SS_E/df_E} = 5.96$
- $p - value = P(F > 5.96) = 0.0016$
- **Reject** H_0**, and conclude that there is a difference among the agents.**

POST HOC ANOVA

MARQUETTE UNIVERSITY
Be The Difference.

- **In the ANOVA, we only conclude that yields are different under different agents.**

- **How do we say which agent is the best? Are chemical agents better than the biological agents?**

- **All of these questions can be answered by all pairwise comparisons.**

- $$H_0^{ij}: \mu_i = \mu_j$$
  $$H_a^{ij}: \mu_i \neq \mu_j$$

- **There are 10 pairs.**

WHAT TO DO?

MARQUETTE UNIVERSITY
Be The Difference.

(a) We need to control

α_F = P(false rejections of even one null) = 0.05

(b) To achieve (a), we might lose power of true discovery.

- We will discuss four methods

  1. Bonferroni
  2. Fisher's LSD (Least Significant Difference)
  3. Tukey's
  4. Dunnette's

- Later we will look into the pros and cons for these methods. Keep two things in mind.

BONFERRONI'S METHOD

- **Use TS.** $t_{ij} = \frac{\bar{y}_i - \bar{y}_j}{s_p\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}$ **with Type-I error of** $\frac{\alpha}{m}$.

- **Here** s_p **is the pooled standard deviation that estimates the common standard deviation** σ**. Here, we can estimate** σ **by**

- $\hat{\sigma} = \sqrt{MSE}$, **where** $MSE = \frac{SS_E}{df_E}$

- **We say that pair** (μ_i, μ_j) **are significantly different if**
- **Formula**

$$|\bar{y}_i - \bar{y}_j| > t_{\alpha/2m}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$ (m **is # of hypotheses**)

EXAMPLE 9.3: BONFERRONI'S METHOD

MARQUETTE UNIVERSITY
Be The Difference.

Agent	1	2	3	4	5
Type	**None**	**Bio1**	**Bio2**	**Chm1**	**Chm2**
$\bar{y}$	1.175	1.293	1.328	1.415	1.500
s	.1204	.1269	.1196	.1249	.1265
n	6	6	6	6	6

ANOVA

One-way ANOVA: yield versus Agent

Source	DF	SS	MS	F	P
Agent	4	0.3647	0.0912	5.96	0.002
Error	25	0.3825	0.0153		
Total	29	0.7472			

S = 0.1237 R-Sq = 48.81% R-Sq(adj) = 40.62%

- $SS_E = \sum(n_i - 1)s_i^2 = 0.3825$, $df_E = 25$
- $MSE = \frac{SSE}{df_e} = \frac{0.3825}{25} = 0.0153$
- $\frac{\alpha}{2m} = \frac{0.05}{2*10} = 0.0025,\ t_{\alpha/2m} = 3.361$

Pairwise comparisons using t tests with pooled SD

data: yield and agent

	Bio1	Bio2	Chm1	Chm2
Bio2	1.0000	-	-	-
Chm1	0.9999	1.0000	-	-
Chm2	0.0770	0.2371	1.0000	-
None	1.0000	0.4208	0.0250	0.0012

P value adjustment method: bonferroni

- $\mu_i \neq \mu_j$ **if** $|\bar{y}_i - \bar{y}_j| > 3.361\sqrt{0.0153\left(\frac{1}{6} + \frac{1}{6}\right)} = 0.24002$

✓ **Agent 1 vs. 5: |1.500 – 1.175| = 0.325 > 0.24002**
✓ **Agent 1 vs. 4: |1.415 – 1.175| = 0.240 < 0.24002**
✓ **Agent 1 vs. 3: |1.328 – 1.175| = 0.153 < 0.24002**
✓ **Agent 1 vs. 2: |1.293 – 1.175| = 0.118 < 0.24002**

- **Thus, we only find that Chm2(Agent 5) is different from Control(Agent 1).**
  - **In R:** with(exmp9.3, pairwise.t.test(yield, agent, p.adj = "bonf"))

FISHER'S LSD

MARQUETTE UNIVERSITY
Be The Difference.

- **Formula:** $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$
- **Example 9.3**

Agent	1	2	3	4	5
$\bar{y}$	1.175	1.293	1.328	1.415	1.500

- $\frac{\alpha}{2} = 0.025,\ MSE = 0.0153,\ t_{\alpha/2} = 2.060,\ n_j = n_j = 6$
- $|\bar{y}_i - \bar{y}_j| > 2.06 * \sqrt{0.0153\left(\frac{1}{6} + \frac{1}{6}\right)} = 0.14711$

- ✓ **1 vs. 5: |1.500 – 1.175| = 0.325 > 0.14711**
- ✓ **1 vs. 4: |1.415 – 1.175| = 0.240 > 0.14711**
- ✓ **1 vs. 3: |1.328 – 1.175| = 0.153 > 0.14711**
- ✓ **1 vs. 2: |1.293 – 1.175| = 0.118 < 0.14711**

Agent 3,4 and 5 are different from control

- ✓ **2 vs. 5: |1.500 – 1.293| = 0.207 > 0.14711**
- ✓ **2 vs. 4: |1.415 – 1.293| = 0.122 < 0.14711**
- ✓ **2 vs. 3: |1.328 – 1.293| = 0.035 < 0.14711**

Agent 2 and 5 are different

➢ **In R:** with(exmp9.3, pairwise.t.test(yield, agent, p.adj = "none")) →

```
        Pairwise comparisons using t tests with pooled SD

data:  yield and agent

     Bio1    Bio2    Chm1    Chm2
Bio2 0.62849 -       -       -
Chm1 0.09999 0.23449 -       -
Chm2 0.00770 0.02371 0.24511 -
None 0.11091 0.04208 0.00250 0.00012

P value adjustment method: none
```

TUKEY'S METHOD

MARQUETTE UNIVERSITY
Be The Difference.

- **For the common sample sizes, $n_i = n$**

  Formula: $$|\bar{y}_i - \bar{y}_j| > q_\alpha(t, df_E)\sqrt{\frac{MSE}{n}}$$

- **where $q_\alpha(t, df_E)$ is the upper-tail value of the studentized range statistics (Given in Table 10)**
- **Example 9.3,**
  - $t = 5, df_E = 25, n = 6$,
  - **From Table 10,** $q_{0.05}(5, 25) = 4.158$
- $$q_\alpha(t, df_E)\sqrt{\frac{MSE}{n}} = 4.158 * \sqrt{\frac{0.0153}{6}} = 0.20997$$

- ✓ **1 vs. 5: |1.500 – 1.175| = 0.325 > 0.20997**
- ✓ **1 vs. 4: |1.415 – 1.175| = 0.240 > 0.20997**
- ✓ **1 vs. 3: |1.328 – 1.175| = 0.153 < 0.20997**
- ✓ **1 vs. 2: |1.293 – 1.175| = 0.118 < 0.20997**

Agents 4 and 5 are different from control

- ✓ **2 vs. 5: |1.500 – 1.293| = 0.207 < 0.20997**
- ✓ **2 vs. 4: |1.415 – 1.293| = 0.122 < 0.20997**
- ✓ **2 vs. 3: |1.328 – 1.293| = 0.035 < 0.20997**

No Significant difference between agents 2,3,4 and 5

➢ **In R:** TukeyHSD(model) →

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yield ~ agent, data = exmp9_3)

$agent
                 diff          lwr         upr     p adj
Bio2-Bio1  0.03498333 -0.174742851  0.24470952 0.9876001
Chm1-Bio1  0.12198333 -0.087742851  0.33170952 0.4472475
Chm2-Bio1  0.20698333 -0.002742851  0.41670952 0.0543204
None-Bio1 -0.11801667 -0.327742851  0.09170952 0.4796636
Chm1-Bio2  0.08700000 -0.122726185  0.29672618 0.7410383
Chm2-Bio2  0.17200000 -0.037726185  0.38172618 0.1460064
None-Bio2 -0.15300000 -0.362726185  0.05672618 0.2342349
Chm2-Chm1  0.08500000 -0.124726185  0.29472618 0.7567356
None-Chm1 -0.24000000 -0.449726185 -0.03027382 0.0192441
None-Chm2 -0.32500000 -0.534726185 -0.11527382 0.0010368
```

TUKEY'S METHOD CONT'D

- **Remark for unequal sample sizes:**

- **If sample sizes $n_i s$ are not same, then a modified Tukey's test is**

$$|\bar{y}_i - \bar{y}_j| > \frac{q_\alpha(t, df_E)}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- **This is also called Tukey-Kramer's Method.**

DUNNETT'S METHOD

MARQUETTE UNIVERSITY
Be The Difference.

- **Comparing with a control**
- **Assume that the sample sizes are same, i.e.,** $n_i = n$

$$|\bar{y}_i - \bar{y}_c| > d_\alpha\,(t-1, df_E)\sqrt{\frac{2MSE}{n}}$$

- **Note that this method is only used for comparing with a control.**

- $d_\alpha\,(t-1, df_E)$ **can be obtained from Table 11.**

- **For the book example 9.3, Agent 1 is the control.**
  - $t = 5, df_E = 25, n = 6,$
  - **From Table 11,** $d_{0.05}(5-1, 24) = 2.28$

- $d_\alpha\,(t-1, df_e)\sqrt{\frac{2MSE}{n}} = 2.28 * \sqrt{2 * \frac{0.0153}{6}} = 0.16282$

DUNNETT'S METHOD: EXAMPLE 9.3

- $H_0^i: \mu_i = \mu_c$
- $H_a^i: \mu_i \neq \mu_c$

- **Reject H_0^i if $|\bar{y}_i - \bar{y}_c| > 0.16282$**

Agent	1	2	3	4	5
$\bar{y}$	1.175	1.293	1.328	1.415	1.500

- ✓ **5 vs. 1: |1.500 – 1.175| = 0.325 > 0.16282**
- ✓ **4 vs. 1: |1.415 – 1.175| = 0.240 > 0.16282**
- ✓ **3 vs. 1: |1.328 – 1.175| = 0.153 < 0.16282**
- ✓ **2 vs. 1: |1.293 – 1.175| = 0.118 < 0.16282**

Only the chemical agents 4 and 5 are different from control. No biological agents are different from control.

```
  Dunnett's test for comparing several treatments with a control :
    95% family-wise confidence level

$None
                diff      lwr.ci    upr.ci    pval
Bio1-None 0.1180167 -0.06822422 0.3042576 0.30714
Bio2-None 0.1530000 -0.03324089 0.3392409 0.12932
Chm1-None 0.2400000  0.05375911 0.4262409 0.00887 **
Chm2-None 0.3250000  0.13875911 0.5112409 0.00044 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **In R:**
  - library("DescTools")
  - with(exmp9.3, DunnettTest(yield, agent, control = "None"))

COMPARISON OF BONFERRONI, LSD, TUKEY'S AND DUNNETT'S METHODS.

MARQUETTE UNIVERSITY
Be The Difference.

- For the Book example 9.3, we had $\mu_i \neq \mu_j$ if

- Bonferroni: $\left|\bar{y}_i - \bar{y}_j\right| > 0.24002$
- Fisher's LSD: $\left|\bar{y}_i - \bar{y}_j\right| > 0.14711$
- Tukey's: $\left|\bar{y}_i - \bar{y}_j\right| > 0.20997$
- Dunnett's: $\left|\bar{y}_i - \bar{y}_j\right| > 0.16282$

- Not all method control familywise error rate. Although the power of discovery for Fisher's LSD is better than other methods, you cannot say that

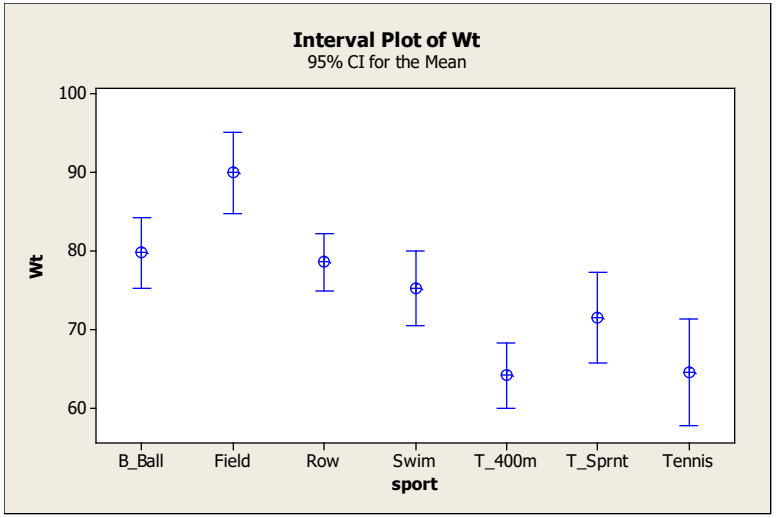
P(False discovery of at least one null) = 0.05

AUSTRALIAN INSTITUTE OF SPORT

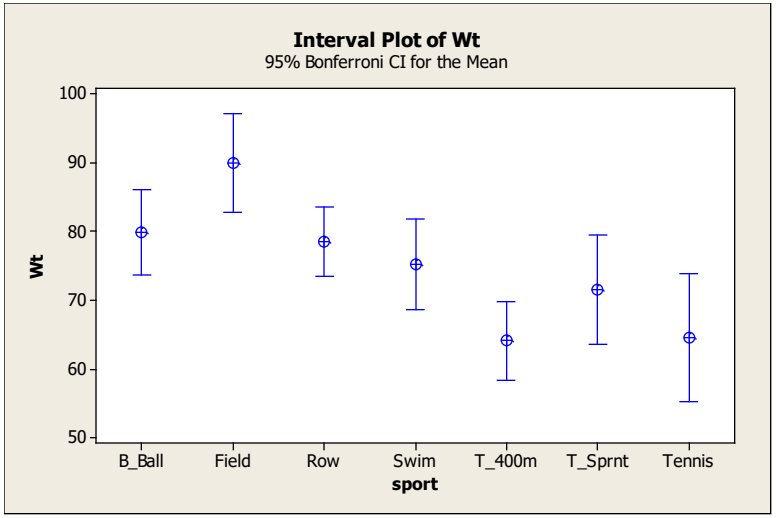
MARQUETTE UNIVERSITY
Be The Difference.

- **Confidence Interval of Wt for different Sports**

- **Individual Confidence Interval**
  - Does NOT consider the familywise error
  - $\bar{y}_i \pm t_{\alpha/2} * s_p \sqrt{\frac{1}{n_i}}$
  - where $s_p = \frac{\sum_{i=1}^{t}(ni-1)s_i^2}{(\sum_{i=1}^{t} n_i)-t}$



- **Bonferroni's Method**
  - Conservative
  - $\bar{y}_i \pm t_{\alpha/2m} * s_p \sqrt{\frac{1}{n_i}}$
  - where $m = 7$



- **Minitab**

AUSTRALIAN INSTITUTE OF SPORT

MARQUETTE UNIVERSITY
Be The Difference.

- **Pairwise comparison of Wt for different Sports**

- **Fisher's Method**
  - Does NOT consider the familywise error

Grouping Information Using Fisher Method

sport	N	Mean	Grouping
Field	19	89.97	A
B_Ball	25	79.78	B
Row	37	78.54	B
Swim	22	75.15	B C
T_Sprnt	15	71.51	C D
Tennis	11	64.47	D E
T_400m	29	64.05	E

Means that do not share a letter are significantly different.

- **Bonferroni's Method**
  - Conservative

Grouping Information Using Bonferroni Method

sport	N	Mean	Grouping
Field	19	89.97	A
B_Ball	25	79.78	A B
Row	37	78.54	B
Swim	22	75.15	B C
T_Sprnt	15	71.51	B C D
Tennis	11	64.47	C D
T_400m	29	64.05	D

Means that do not share a letter are significantly different.

- **Tukey's Method**
  - Consider the familywise error

Grouping Information Using Tukey Method

sport	N	Mean	Grouping
Field	19	89.97	A
B_Ball	25	79.78	A B
Row	37	78.54	B
Swim	22	75.15	B C
T_Sprnt	15	71.51	B C D
Tennis	11	64.47	C D
T_400m	29	64.05	D

Means that do not share a letter are significantly different.

- **Dunnett's Method**
  - Comparing with a Control group

Grouping Information Using Dunnett Method

Level	N	Mean	Grouping
B_Ball (control)	25	79.78	A
Field	19	89.97	
Row	37	78.54	A
Swim	22	75.15	A
T_Sprnt	15	71.51	A
Tennis	11	64.47	
T_400m	29	64.05	

Means not labeled with letter A are significantly different from control level mean.

CONCLUSION

MARQUETTE UNIVERSITY
Be The Difference.

- Excluding Fisher LSD, the other methods control familywise error rate. However, Bonferroni has the poorest discovery rate. Discovery rate for the Tukey's is better than Bonferroni, but not as good as Dunnett's. However, Dunnett's used only to compare with the control.

- In other words, Bonferroni, Tukey and Dunnett's all have familywise error rate of 0.05. If the objective is to compare only with a control, then Dunnett's is more powerful among three. Otherwise, Tukey's is more powerful than Bonferroni.

- Although Bonferroni is not very powerful, it does have advantage that it can be used in any situation (whether it is one factor or multifactor analyses) whenever there are multiple hypotheses.