第 2 期 医薬安全性研究会　基礎セミナー
◇ 日時：2012 年 6 月 2 日 （土） 10:00～12:00
◇ 場所：日本科学技術連盟、1 号館、3 階 講堂
協賛：サイエンティスト社

＊ じっくり勉強すれば身につく統計入門（10:00～12:00） **1**
基本に戻ろう: 多重比較の基礎
杉本 典子 3
基本に戻ろう: 多重比較法の比較
橘田 久美子
（シミックバイオリサーチセンター） 13

「じっくり勉強すれば身につく統計解析」を副題としたシリーズ全 3 巻がサイエンティスト社から刊行されました．タイトルは「医薬品開発のための統計解析，第 1 部 基礎，第 2 部 実験計画法，第 3 部 非線形モデル」です．また，「じっくり勉強すれば身につく統計～ Excel, JMP による基礎から応用統計解析実務者コース」（SAS（株）JMP ジャパン事業部主催，年 12 回）のテキストとして使用されています．この本を題材として，第 1 回目「基本統計量とデータの比較」，「第 2 回目：回帰モデル」，第 3 回目：「共分散分析」，と基礎セミナーを開催してきました．第 4 回目は，「多重比較の基礎」，「医薬品開発における多重比較」についてじっくりと勉強して身に付けたいと思われる方々の参加をお待ちしています．

基本に戻ろう：多重比較の基礎

杉本 典子

多群の実験データに，t 検定を何回も繰り返して適用すると，本当は差がないのに偶然によって有意差が出てしまうため，多重性を考慮することが必要，とよく言われています.「多重性」とはなんなのでしょうか．なにを「考慮」したらよいのでしょうか．多重性を無視したらどんなことが起こるのでしょうか．最近，多くの多重比較の方法が提案され臨床試験で実際に使われるようになってきましたが，まずは，多重性とはなにか多重比較法を学ぶ上で基本的な考え方を説明します.

基本に戻ろう：多重比較法の比較

橘田 久美子（シミックバイオリサーチセンター）

多群の実験データに多重性を考慮した解析をするのは当然であり，すでに我社のコンピュータに組み込まれていて，いまさら多重比較法の比較など必要性がないと思っていらっしゃいませんか．実験の目的にあった多重比較の方法になっていると自信がありますか．典型的な小動物を対象にした毒性試験において，コントロール群と幾つかの用量群の間で Dunnett の多重比較は，不適切といわれることもありますが，なぜなのでしょう．Williams の方法がいいとの論文もあったりします．どうしたら良いのでしょうか．自信を持って対処するためには，多重比較の統計的な性能を調べて，見えるようにすると良いのです．代表的な多重比較の方法の比較について，基本的な考え方を丁寧に解説します.

じっくり勉強すれば身につく統計解析入門

多重比較

杉本　典子

はじめに

グリーン本

「じっくり勉強すれば身につく統計入門」がテーマ。

医薬の分野で統計的方法を適用する際の基本的な考え方を説明しており、入門者向け。

→本日は第2部1.3章

参照するファイルのダウンロード先

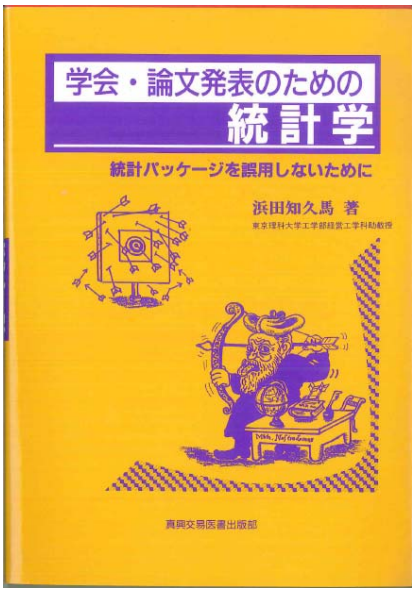
http://www.scientist-press.com/12_278.html

はじめに

浜田知久馬
「学会・論文発表のための統計学
ー統計パッケージを誤用しないためにー」,
真興交易医書出　版部(1999)

と

BioS22　予備コース10/19の
「多重比較」のスライドから
一部利用させて頂きます



今回の発表の前提として

多重比較を説明するうえで今回は2群の比較（t検定）
の基本知識があることを前提としています。

ご紹介したグリーン本のテキストの以下を参照
　　第1部　　基礎
　　3章　2組のデータ解析

発表内容

（1） はじめに
（2） 多重性
（3） 検定の多重性：例1
（4） 検定の多重性：例2
（5） 検定の多重性：例3

(1)はじめに

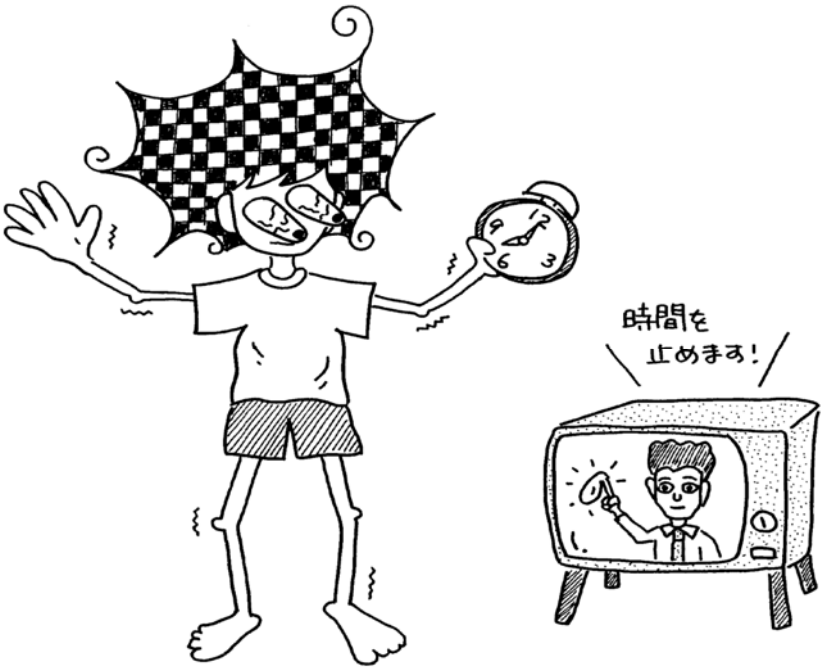
浜田先生：資料

時計を止める超能力

ユリ・ゲラーが「わたしがTVをご覧のみなさんのもとに念を送って，みなさんの手元にある時計を止めて見せます」というのです．たとえば，番組の15分間，彼が念を送りますと番組の視聴者から「TVを見ていたら時計が本当に止まってしまった！」という驚きの知らせが番組に何件も寄せられます

浜田先生：資料

時計を止める超能力



浜田先生：資料

超能力か？

1）1つの時計が15分間に止まる可能性は非常に低い．電池の寿命を1年として

1/（4×24×365）＝1/35040

2）当時視聴率は高かった．

500万世帯がみて，1世帯に4個時計があるとすると，2000万個の時計が対象

2000万/3.504万＝570個

が止まる個数の期待値

(1)　多重性

浜田先生：資料

ユリ・ゲラーの偽超能力で多くの時計が止まったという電話がテレビ番組に多数寄せられたように，全体としてどれかが有意となる確率はかなり高くなる

↓

本当は差がないのに偶然によって有意差が出る
→多重性

検定での多重性はどういうものがあるのでしょうか？

(3) 検定の多重性:例1

多重性による第1種の過誤の増大

(例) ラットを用いた薬剤の反復投与毒性試験

RBC(赤血球数)データの例

→通常，毒性の評価のため、化合物の用量を等比的に3ないし4段階に変化させ実験する

ラット(雄)の RBC（赤血球数） 単位（×10^4/mm^3）

	対照群(0mg)	低用量群(1mg)	中用量群(3mg)	高用量群(10mg)
平均値	926.0	911.9	891.5	893.0
標準偏差	25.7	20.1	39.8	35.4
N	10	10	10	10

実際の毒性試験データだそう

対照群と各用量群の比較を行います

ラット(雄)の RBC（赤血球数） 単位（×10^4/mm^3）

	対照群(0mg)	低用量群(1mg)	中用量群(3mg)	高用量群(10mg)
平均値	926.0	911.9	891.5	893.0
標準偏差	25.7	20.1	39.8	35.4
N	10	10	10	10

t検定を使用

t 検定（両側 5 %）の結果

	平均値の差	t 値	p 値
対照群－低用量群	－14.1	－1.3667	0.1895
対照群－中用量群	－34.5	－2.3052	0.0333*
対照群－高用量群	－33.0	－2.3874	0.0281*

棄却限界値(2.101) ＊：p＜0.05

Q：比較を1回だけではない
　1つの研究の中で複数の検定を同時に行ってもいいの？

t 検定（両側　5　%）の結果

	平均値の差	t 値	p 値
対照群－低用量群	－14.1	－1.3667	0.1895
対照群－中用量群	－34.5	－2.3052	0.0333*
対照群－高用量群	－33.0	－2.3874	0.0281*

例では3回の比較を
同時に行ってる

1回の比較あたりの有意水準を5%としても，
実際に3回の比較を行うと，偶然で有意差が出る確率は
5%より大きく（約15%）なる

検定の多重性とは
このように複数の検定を同時に行うことで
偶然により誤って有意になる確率(αエラー)が
大きくなる現象です

Q：検定の多重性がおこる？
　じゃあ、どうすればいいの？

1つの方法として多重比較法を利用することができます

多重比較法とは
　データに対して複数の比較を同時に行うとき，
　比較の組み合わせ全体を1つの解析とみなし
　比較全体で誤って有意とする確率を
　有意水準以下に抑える手法

(4) 検定の多重性：例2

異なる4種類の薬剤を比較
すべての組み合わせを検討したい場合

	薬剤A	薬剤B	薬剤C	薬剤D
平均	10	23	19	18

計6回

1回の比較あたりの有意水準を5%ならば
6回の比較すると偶然で有意差が出る確率は約30%
⇒ 5%よりかなり大きくなってしまう！

検定の多重性をおさえるには、どうしたらいい？

例：組み合わせが6回あるなら

	薬剤A	薬剤B	薬剤C	薬剤D
平均	10	23	19	18

計6回

＜アイデア＞
6回の比較があるならば、
5%にαを保つため、 0.05 ÷ 6 ≒ 0.008
と考えられる？
→ つまり6回やってもαを5%にすればいい！

(5) 検定の多重性：例3

効果が水準によって単調増減することがわかってる場合もあるでしょう

	対照（0mg）	薬剤A（低用量）	薬剤A（中用量）	薬剤A（高用量）
平均	38	23	19	18

前例のようなすべての組み合わせは必要なく、
高用量が有意でなければ、中用量や低用量は効果がないなら、比較する順序も仮定できる場合
→ こんなときはどうすればよいのでしょうか・・・

まとめ

組み合わせの数は実験の目的で異なります
その目的にあった多重比較法があります

すべての水準の組合せを比較したい場合や
ある基準となる水準との比較したい場合、
水準の順序に意味がある比較の場合など
各目的に対して、
どんな多重比較法を適用すればいいのか

これが今回のテーマです

謝辞

本発表原稿を作成するにあたり，
本発表に資料の提供と利用を許してくださいました

浜田 久知馬 先生（東京理科大 教授）

また，セミナー打合せを通して日々有益なご意見を
いただいておりますJMPセミナー講師の皆様，
ありがとうございました．

心より感謝申し上げます．

じっくり勉強すれば身につく
統計入門

多重比較

株式会社シミックバイオリサーチセンター
橘田久美子

今回の内容

(1) 取り扱うデータ
(2) 全組み合わせでの比較
- Bonferroniの方法
- Tukeyの方法

(3) コントロール群との比較
- Bonferroniの方法
- Dunnettの方法
- Williamsの方法

(4) 方法の比較
(5) 多重比較法の特徴
(6) 多重比較法を実施する時の注意

(1) 取り扱うデータ

今回は，母集団が

・等分散

・正規分布

・各水準の繰り返し数 (各群のデータ数) が等しい

データを使って説明します．

それ以外の場合は
永田・吉田先生の書かれた
『統計的多重比較法の基礎』
芳賀先生の
『医薬品開発のための統計解析
第2部　実験計画法』
を参考にしてください．

(2) 全組み合わせでの比較

	A0	A1	A2	A3	A4
平均	17.00	17.50	21.50	24.17	23.50
標準偏差	2.90	1.87	2.74	2.86	2.74

A0～A4の 5 水準で実験を行い上記のような結果を得た．
水準間に差があるかどうかを検定したい．

各水準を2群ずつ対にして t 検定を行うとすると
水準の組み合わせ数 = (水準数)×(水準数-1)÷2 通りの組合せができる．

5 水準の場合は 5×(5-1)÷2 = 10 通りの組合せができる．

1回あたりの危険率を α^*，検定の繰り返し数を m (この例ではm=10) とすると，全体の危険率 α は $\alpha = m \times \alpha^*$
となり全体の危険率は大きくなってしまう．

検定の多重性

全体の危険率を α にするために危険率に配慮した比較の方法が多重比較である．

(2-1) Bonferroniの方法

1回当たりの検定の危険率が α^* である検定を m回繰り返した場合，全体の危険率 α は $\alpha = m \times \alpha^*$ となる.

帰無仮説 (ここでは群間に差がない) が正しいとき，m 回行う検定でどれかが有意となる確率 (全体の危険率) α は $m \times \alpha^*$ 以下であるはずなので，最大限に見積もって

$\alpha^* = \alpha / m$

とすれば m回検定したときの全体の危険率は α 以下に抑えられる.

⇒ $\alpha^* = \alpha / m$ として検定を行なう方法がBonferroniの方法

検定の回数 m に注目して1回あたりの危険率 α^* を調整する方法

Bonferroniの方法は最大限の危険率を見積もるために，保守的になり，実際の第1種の誤りの危険率は期待する危険率よりも**小さくなる** (その分第2種の誤りの危険率が大きくなる).

第1種の誤り
帰無仮説 (群間に差がない) が正しいのに帰無仮説を棄却する (群間に差がある)と間違ってしまう危険率

第2種の誤り
帰無仮説 (群間に差がない) が間違っているのに帰無仮説を採用する(群間に差がみつからない)と間違ってしまう危険率

(2-1) Bonferroniの方法

表示2-1　全組み合わせでの比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.00	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	17.50	**15**	**16**	**17**	**18**	**19**	**20**
A2	6	21.50	**25**	**24**	**19**	**18**	**21**	**22**
A3	6	24.17	**20**	**26**	**28**	**22**	**24**	**25**
A4	6	23.50	**26**	**27**	**21**	**20**	**23**	**24**
全体	30							

水準番号		差	|t 値|	p値	Bonferroni
A0	A1	-0.500	0.327	0.7464	---
A0	A2	-4.500	2.943	*0.0069*	---
A0	A3	-7.167	4.687	*0.0001*	有意
A0	A4	-6.500	4.251	*0.0003*	有意
A1	A2	-4.000	2.616	*0.0149*	---
A1	A3	-6.667	4.360	*0.0002*	有意
A1	A4	-6.000	3.924	*0.0006*	有意
A2	A3	-2.667	1.744	0.0934	---
A2	A4	-2.000	1.308	0.2028	---
A3	A4	0.667	0.436	0.6666	---

5水準を全組み合わせで検定したい.

5水準なので検定の回数は10通り

全体の危険率 α =0.05とすると，10回検定する場合の1回あたりの危険率 α^*は

$\alpha^* = 0.05/10 =$ **0.0050**

(2-2) Tukeyの方法

Bonfferoniの方法は保守的になりやすい方法であった.

全組み合わせで同時に検定する場合に，全体としての危険率 α を調整する方法がTukey によって考案された.

全組み合わせで平均の差の検定を行う場合，第1種の誤りが最も起きやすいのは平均値の差が最大値と最小値の組み合わせと考えられる.

個々の値の最大値 (x_{max}) と最小値 (x_{min}) の差を“範囲”とする．水準の個数を a とすると，範囲が水準の個々の値 (x_i，i=1～a) の標準偏差 (s) の何倍になるか (q) を考える.

$$q = \frac{x_{max} - x_{min}}{s}$$

x が n 個の平均だとすると，q の分母は $\frac{s}{\sqrt{n}}$

この式をまとめると $q = \frac{x_i - x_{i'}}{s/\sqrt{1/n}}$

通常の t 検定のときの t 値の式は $t = \frac{x_i - x_{i'}}{s/\sqrt{2/n}} = \frac{q}{\sqrt{2}}$

この関係から，Tukeyの検定を行う場合は q の棄却限界値を √2 で割って適用する.

(2-2) Tukeyの方法

Tukey の方法で使う q 値の5%点 (棄却限界値) は統計数値表から求めるが『Tukeyの方法のための・・・』という名前の表はない.

Tukey の方法での棄却限界値は，統計数値表のなかの
『スチューデント化された範囲の分布の上側5%点』(q 値の 5%点に相当)
という表の該当する値を √2 で割った値を検定に使用する.

表示2-2　多重比較における t 値の5%点

	t	Tukey		
自由度	2	3	5	7
10	2.228	2.741	3.291	3.623
20	2.086	2.530	2.992	3.267
25	2.060	2.491	2.937	3.200
30	2.042	2.466	2.901	3.157
60	2.000	2.403	2.812	3.050
∞	1.960	2.343	2.728	2.949

←水準数

↑全体の自由度

水準数が 2 の場合は通常の平均の差の t 検定に対応する.

(2-2) Tukeyの方法

表示2-1　全組み合わせでの比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.00	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	17.50	**15**	**16**	**17**	**18**	**19**	**20**
A2	6	21.50	**25**	**24**	**19**	**18**	**21**	**22**
A3	6	24.17	**20**	**26**	**28**	**22**	**24**	**25**
A4	6	23.50	**26**	**27**	**21**	**20**	**23**	**24**

水準番号		差	|t 値|	p値	Bonferroni	Tukey
A0	A1	-0.500	0.327	0.7464	---	---
A0	A2	-4.500	2.943	*0.0069*	---	有意
A0	A3	-7.167	4.687	*0.0001*	有意	有意
A0	A4	-6.500	4.251	*0.0003*	有意	有意
A1	A2	-4.000	2.616	*0.0149*	---	---
A1	A3	-6.667	4.360	*0.0002*	有意	有意
A1	A4	-6.000	3.924	*0.0006*	有意	有意
A2	A3	-2.667	1.744	0.0934	---	---
A2	A4	-2.000	1.308	0.2028	---	---
A3	A4	0.667	0.436	0.6666	---	---

自由度 25
5水準
のときの
Tukeyの棄却
限界値は
t=2.937

(2-2) Tukeyの方法

Bonfferoniの方法は保守的になりやすい方法であった.
Tukey の方法と Bonfferoni の方法での t 値の5%点 (棄却限界値) を比較する.

表示2-2　多重比較における t 値の5%点

	t	Tukey					
自由度	2	3	Bonf.	5	Bonf.	7	Bonf.
10	2.228	2.741	2.870	3.291	3.581	3.623	4.035
20	2.086	2.530	2.613	2.992	3.153	3.267	3.477
25	2.060	2.491	2.566	2.937	3.078	3.200	3.380
30	2.042	2.466	2.536	2.901	3.030	3.157	3.319
60	2.000	2.403	2.463	2.812	2.915	3.050	3.173
∞	1.960	2.343	2.394	2.728	2.807	2.949	3.038
	m		3		10		21

Tukeyの t 値に比べ，Bonferroniの t 値の方が大きい.
⇒有意になりにくい (= 保守的である) ことがわかる.

* Bonf. の t 値はTINV関数を使い有意水準 (通常は0.05) を繰り返し数 m で割った値を確率とし，=TINV(0.05/m，自由度) という式で計算している.

(3) コントロール群との比較

毒性試験などでは**コントロール群と投与群との比較**だけが目的である場合も多い．このような場合に，Tukey の方法のような全組み合わせで同時に検定する方法を適用すると，検定の回数が多くなるために群間の差を見逃す可能性 (第2種の誤りの危険率) が非常に大きくなってしまう．

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.2	**17**	**23**	**25**	**19**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.03333						

コントロール群との比較の方法として

・Bonferroniの方法

・Dunnettの方法

・Williamsの方法

について紹介する．

(3-1) Bonferroniの方法

表示3-1　コントロール群との比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.3	**17**	**23**	**25**	**20**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.06667						

コントロール群 (A0) とA1，A2，A3，A4群とを比較するので検定の回数は m=4

全体の危険率 α =0.05とすると1回あたりの危険率は

$\alpha^* = 0.05/4 = 0.0125$

水準番号		差	|t 値|	p値	Bonferroni
A0	A1	-2.000	1.236	0.2279	---
A0	A2	-3.500	2.163	*0.0403*	---
A0	A3	-4.333	2.678	*0.0129*	---
A0	A4	-5.500	3.399	*0.0023*	有意

(3-2) Dunnettの方法

コントロール群とその他の群を同時に検定する場合に，全体としての危険率 α を調整する方法がとしてDunnett の方法がある．

Dunnettの方法では，A0 - A1群，A0 - A2群，・・・という比較を同時に行う．

A0 - A1 ，A0 - A2それぞれの t 値は

$$t_1 = \frac{\overline{x_1} - \overline{x_0}}{se[d]} \qquad t_2 = \frac{\overline{x_2} - \overline{x_0}}{se[d]}$$

どちらにも x0 が含まれているので独立といえない．

この相関関係を考慮して計算された棄却限界値の表が『Dunnett の方法のための両側 (or 上側) 5%点 (相関係数=〇の場合)』である．

今回は全ての水準で繰り返し数が等しい，ことを前提にしているので，両者の相関係数 $\rho = n_i / (n_0 + n_i)$ は

ρ=0.5

⇒「相関係数=0.5 の場合」の数値表の値を使う．

(3-2) Dunnettの方法

表示3-1　コントロール群との比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.3	**17**	**23**	**25**	**20**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.06667						

自由度 25，5水準のときの Dunnett の棄却限界値は t=2.607

水準番号		差	|t 値|	p値	Bonferroni	Dunnett
A0	A1	-2.000	1.236	0.2279	---	---
A0	A2	-3.500	2.163	*0.0403*	---	---
A0	A3	-4.333	2.678	*0.0129*	---	有意
A0	A4	-5.500	3.399	*0.0023*	有意	有意

(3-2) Dunnettの方法

Dunnett の方法と Bonfferoni の方法での t 値の5%点 (棄却限界値) を比較する.

表示3-2　多重比較における t 値の5%点 (Dunnettの両側5%点)

	t	Dunnett					
ve	2	3	Bonf.	5	Bonf.	7	Bonf.
10	2.228	2.568	2.634	2.891	3.038	3.070	3.277
20	2.086	2.379	2.423	2.651	2.744	2.802	2.927
25	2.060	2.344	2.385	2.607	2.692	2.752	2.865
30	2.042	2.321	2.360	2.578	2.657	2.720	2.825
60	2.000	2.265	2.299	2.508	2.575	2.642	2.729
∞	1.960	2.212	2.241	2.442	2.498	2.567	2.638
	m		2		4		6

Dunnett の t 値に比べ，Bonferroniの t 値の方が大きい.
⇒有意になりにくい (= 保守的である) ことがわかる.

(3-3) Williamsの方法

同じ薬剤を投与量を段階的に変えて投与する場合に，反応 y が単調に増加 / 減少することが予想されることがしばしばある.
Dunnett の方法は処理群の水準間の順序は考慮していない．水準間の順序を考慮する方法としてWilliams の方法がある.
Williams の方法では，対照群と処理群を順番に対比較する.

水準	A_0	A_1	A_2	・・・	A_{m-1}	A_m
投与量	0	d_1	d_2	・・・	d_{m-1}	d_m
反応	y_0	y_1	y_2	・・・	y_{m-1}	y_m

①
②

①の比較で水準間に有意差がなければこの薬剤は効果がない，と判断.
①の比較で水準間に有意差があれば，②の比較を実施
・・・
と投与量の多い水準から順に検定を行いA_0群と有意差のある水準の範囲を求める.

Williamsの方法では反応の方向がわかっているので，基本的には片側検定となる.

(3-3) Williamsの方法

表示3-1　コントロール群との比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.3	**17**	**23**	**25**	**20**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.06667						

水準番号		差	|t 値|	5%有意	2.5%有意
A0	A4	-5.500	3.399	有意	有意
A0	A3	-4.333	2.678	有意	有意
A0	A2	-3.500	2.163	有意	有意
A0	A1	-2.000	1.236	---	---

Williamsの5%点	自由度ν = 25			n0 = ni				
水準数	2	3	4	5	6	7	8	9
5%	1.708	1.788	1.814	1.827	1.835	1.839	1.843	1.845
2.5%	2.060	2.125	2.146	2.156	2.162	2.165	2.168	2.17

(3-4) 3つの方法の結果

表示3-1　コントロール群との比較

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.3	**17**	**23**	**25**	**20**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.06667						

帰無仮説の個数		棄却限界値		Student-t	Bonferroni	Dunnett	Williams
4				2.060	2.692	2.607	2.156
水準番号		差	|t 値|	p値	Bonferroni	Dunnett	Williams
A0	A1	-2.000	1.236	0.2279	---	---	---
A0	A2	-3.500	2.163	*0.0403*	---	---	有意
A0	A3	-4.333	2.678	*0.0129*	---	有意	有意
A0	A4	-5.500	3.399	*0.0023*	有意	有意	有意

棄却限界値を比較するとWilliamsの方法の検出力の良さがわかる.

(4) 方法の比較

Bonfferoni の方法では，検定を m 回繰り返した場合の全体の危険率 α は1回あたりの危険率 α*の m 倍と考え α= m×α* とした．

Tukeyの方法などでは，1回あたりの危険率 α* はいくらになるだろうか．
それが分れば， α= f ×α* という関係式を想定し，この係数 f を多重度と呼ぶことにすれば手法の比較に役立つと思われる．

Tukeyの方法で自由度=25，水準数=5の両側5%の t 値は 2.937 (表示2-2参照)．
t 値が2.937 (自由度25) となるときの p値は0.00703… とわかる．
この時の多重度 f = 0.05/0.00703 = 7.115 となった．

Bonferroniの方法では水準数=5 を全組み合わせで比較した場合 f = m = 10であった．
Tukey の方法では Bonferroni の方法に比べて多重性が少ない，つまり検定の回数が少ないことになる．このことからもTukeyの方法のほうが検出力が高くなることが判る．

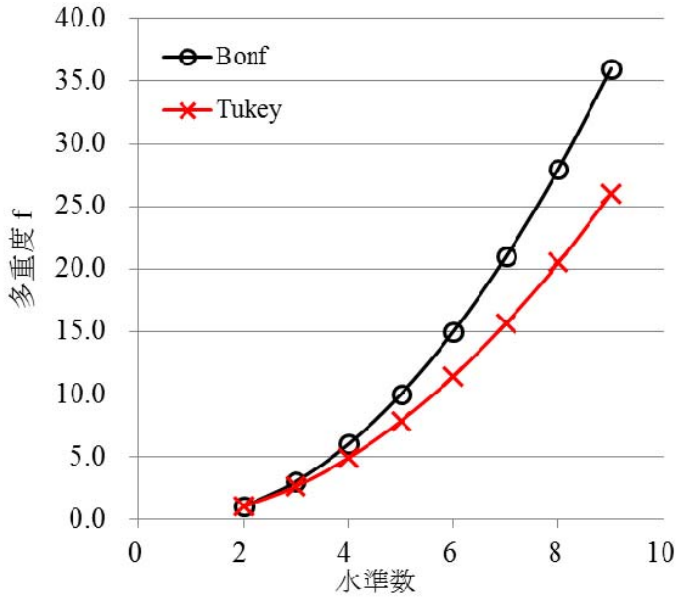
(4) 方法の比較

表示4-1　全組み合わせでの多重検定の比較

多重度 f =9×(9-1)÷2

水準数	2	3	4	5	6	7	8	9
Bonf	1	3	6	10	15	21	28	36
Tukey	1.000	2.616	4.901	7.847	11.424	15.665	20.490	26.019

両側 α =0.05，自由度 ν e=∞



今回説明した多重比較の方法のうち，全組み合わせでの方法 (Bonferroni, Tukey) について，各水準数での多重度 f (= 検定全体の危険率 α/個々の検定の α*) を比較した．

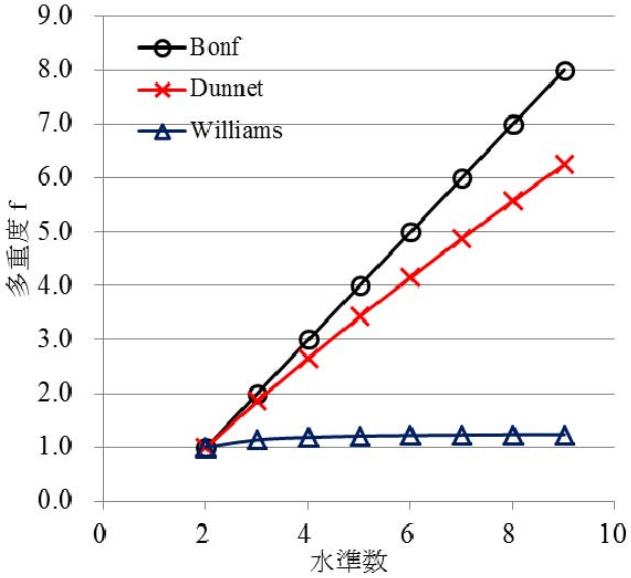
どちらの方法でも水準数に比例して，多重性が急激に増加していることがわかる．Tukey と比較して，Bonferroni の方法では水準数が多くなるほど多重性が増加 (=検出力が低下) することがわかる．

(4) 方法の比較

表示4-2　コントロール群と比較する場合の多重検定の比較

水準数	2	3	4	5	6	7	8	9
Bonf	1	2	3	4	5	6	7	8
Dunnet	1.00	1.85	2.66	3.42	4.15	4.87	5.57	6.25
Williams	1.00	1.14	1.19	1.21	1.22	1.23	1.24	1.24

両側 α =0.05，自由度 ν e=∞



同様にコントロール群と比較する場合の方法 (Bonferroni，Dunnett，Williams) についても各水準数での多重度 f を比較した．

Williamsの方法は水準数が増えても多重度 f はあまり変化していない．

このグラフを見れば (Williamsの方法の前提条件を満たす場合は) Williams の方法が優れていることが一目瞭然で良くわかる．

(5) 多重比較法の特徴

【Bonfferoniの方法】

・全体の危険率が α 以下になるように，検定の回数 m から1回あたりの危険率 α^* を $\alpha^* = \alpha/m$ と決める方法．

・検定の回数 m は自由に決められるが，非常に保守的になる．

【Tukeyの方法】

・全組み合わせで同時に検定する場合に，全体としての危険率 α を調整する方法．

【Dunnettの方法】

・対照群と比較群を比較する方法．

・比較群間の順序に意味はない．

【Williamsの方法】

・対照群と比較群を比較する場合で，結果に単調性が想定できる場合の検定方法．

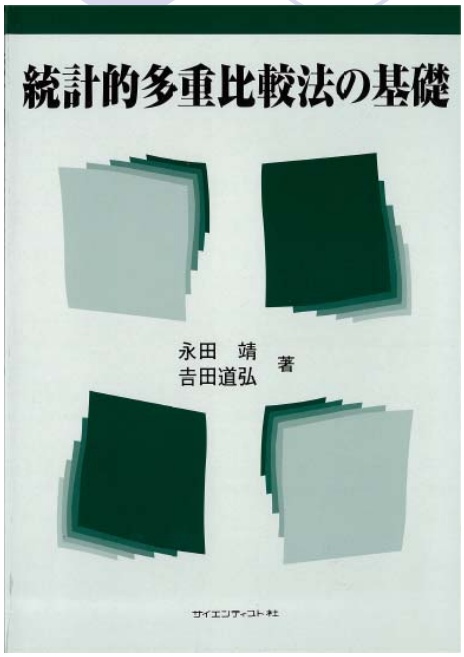
・非常に検出力が高く，検定回数が多くなっても検出力が落ちにくい．

(6) 多重比較法を実施する時の注意

- 多重比較を用いると第1種の誤りの確率 (帰無仮説(= 差がない) が正しいときに帰無仮説を棄却してしまう(=差があると結論する)確率) αを抑えることはできるが，検出力が下がるので第2種の誤りの確率 βが増える.
- 検定の回数に比例して検出力は下がる.
- 新薬を厚労省へ申請する場合， その有意性を保証することが必要であり，多重比較を用いることが常識である.
- 開発段階で有効性のある化合物を見逃してしまう危険は避けたいのでこの段階で多重比較を行なうことは不適切かもしれない (第2種の誤りは小さいほうが良い).
- 化合物の毒性や副作用を調べる試験では，毒性・副作用を見逃す危険を小さくしなければならない (第2種の誤りを小さくしたい) .
- 実験をデザインする場合には，各統計手法の特徴をよく考慮し，実験の目的を満たす統計手法を選択しなければならない.

【参考にした教科書】

サイエンティスト社
2009年12月15日 初版



サイエンティスト社
1997年11月30日 初版

【謝辞】

本発表に際し終始適切にご助言を戴き
丁寧に指導して下さいました
JMPセミナー講師の皆様

どうもありがとうございました.

心より感謝申し上げます.

表示2-1　p値を調整する方法

データ

水準	1	2	3	4	5	6	7	平均	標準偏差
A1	10.7	9.7	8.5	9.4	8.8	8.4	10.6	9.443	0.95
A2	8.1	8.3	8.7	6.9	5.7	9.5	6.7	7.700	1.32
A3	7.9	7.5	7.4	9.2	5.7	8.3	9.7	7.957	1.31
A4	6.2	7.1	5.5	4.7	6.3	6.9	7.5	6.314	0.97

水準数	繰返数
4	7

残差

水準	1	2	3	4	5	6	7
A1	1.257	0.257	-0.943	-0.043	-0.643	-1.043	1.157
A2	0.400	0.600	1.000	-0.800	-2.000	1.800	-1.000
A3	-0.057	-0.457	-0.557	1.243	-2.257	0.343	1.743
A4	-0.114	0.786	-0.814	-1.614	-0.014	0.586	1.186

残差平方和	自由度	平均平方	標準偏差	差の標準誤差
31.743	24	1.323	1.150	0.615

帰無仮説の個数	棄却限界値	Student-t	Bonferroni
6		2.064	2.875

水準番号		差	|t 値|	p 値	Bonferroni	p値(昇順)	Holm
A1	A2	1.743	2.835	**0.0091**	---	2	有意
A1	A3	1.486	2.417	*0.0236*	---	4	---
A1	A4	3.129	5.089	**0.0000**	有意	1	有意
A2	A3	-0.257	0.418	0.6794	---	6	---
A2	A4	1.386	2.254	*0.0336*	---	5	---
A3	A4	1.643	2.672	*0.0133*	---	3	---

危険率 α

Student	0.05						
Bonferroni	0.0083						
Horm	α	α1	α2	α3	α4	α5	α6
	0.05	0.0083	0.0100	0.0125	0.0167	0.0250	0.0500

表示2-1　全組み合わせでの比較

データ

水準	n	平均	1	2	3	4	5	6
A0	6	17.00	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	17.50	**15**	**16**	**17**	**18**	**19**	**20**
A2	6	21.50	**25**	**24**	**19**	**18**	**21**	**22**
A3	6	24.17	**20**	**26**	**28**	**22**	**24**	**25**
A4	6	23.50	**26**	**27**	**21**	**20**	**23**	**24**
全体	30							

水準数	繰返数
5	6

残差

水準	標準偏差	1	2	3	4	5	6
A0	2.90	-2.00	2.00	4.00	-4.00	1.00	-1.00
A1	1.87	-2.50	-1.50	-0.50	0.50	1.50	2.50
A2	2.74	3.50	2.50	-2.50	-3.50	-0.50	0.50
A3	2.86	-4.17	1.83	3.83	-2.17	-0.17	0.83
A4	2.74	2.50	3.50	-2.50	-3.50	-0.50	0.50

残差平方和	自由度	平均平方	標準偏差	差の標準誤差
175.3333	25	7.013	2.648	1.52898

帰無仮説の個数	棄却限界値	Student-t	Bonferroni	Tukey
10		2.060	3.078	2.937

水準番号		差	|t 値|	p値	Bonferroni	Tukey
A0	A1	-0.500	0.327	0.7464	---	---
A0	A2	-4.500	2.943	*0.0069*	---	有意
A0	A3	-7.167	4.687	*0.0001*	有意	有意
A0	A4	-6.500	4.251	*0.0003*	有意	有意
A1	A2	-4.000	2.616	*0.0149*	---	---
A1	A3	-6.667	4.360	*0.0002*	有意	有意
A1	A4	-6.000	3.924	*0.0006*	有意	有意
A2	A3	-2.667	1.744	0.0934	---	---
A2	A4	-2.000	1.308	0.2028	---	---
A3	A4	0.667	0.436	0.6666	---	---

表示 2-2　多重比較における t 値の5%点

	t	Tukey					
自由度	2	3	Bonf.	5	Bonf.	7	Bonf.
10	2.228	2.741	2.870	3.291	3.581	3.623	4.035
20	2.086	2.530	2.613	2.992	3.153	3.267	3.477
25	2.060	2.491	2.566	2.937	3.078	3.200	3.380
30	2.042	2.466	2.536	2.901	3.030	3.157	3.319
60	2.000	2.403	2.463	2.812	2.915	3.050	3.173
∞	1.960	2.343	2.394	2.728	2.807	2.949	3.038
	m		3		10		21

Bonf.=Bonferroni

α	0.05
α*	0.00703
f	7.1147

	3		5		7	
10	3.877	2.7415	4.654	3.2909	5.124	3.6232
20	3.578	2.53	4.232	2.9925	4.62	3.2668
25	3.523	2.4911	4.153	2.9366	4.526	3.2004
30	3.487	2.4657	4.102	2.9006	4.464	3.1565
60	3.399	2.4035	3.977	2.8122	4.314	3.0505
∞	3.314	2.3434	3.858	2.728	4.17	2.9486

表示3-1　コントロール群との比較

データ

水準	n	平均	1	2	3	4	5	6
A0	6	17.0	**15**	**19**	**21**	**13**	**18**	**16**
A1	6	19.0	**15**	**23**	**21**	**17**	**20**	**18**
A2	6	20.5	**24**	**23**	**18**	**17**	**20**	**21**
A3	6	21.3	**17**	**23**	**25**	**20**	**21**	**22**
A4	6	22.5	**25**	**26**	**20**	**19**	**22**	**23**
全体	30	20.0667						

水準数	繰返数
5	6

残差

水準	標準偏差	効果	1	2	3	4	5	6
A0	2.90	-3.07	-2.00	2.00	4.00	-4.00	1.00	-1.00
A1	2.90	-1.07	-4.00	4.00	2.00	-2.00	1.00	-1.00
A2	2.74	0.43	3.50	2.50	-2.50	-3.50	-0.50	0.50
A3	2.61	1.27	-4.33	1.67	3.67	-1.33	-0.33	0.67
A4	2.39	2.43	2.50	3.50	-2.50	-3.50	-0.50	0.50

分散分析表

要因	平方和	自由度	平均平方	F 比	p値
水準間	109.53	4	27.383	3.487	*0.0215*
残差	196.33	25	7.853	1.000	
全体	305.87	29	10.547		
（検算）	305.87	29			

差の標準誤差　1.61795

帰無仮説の個数			棄却限界値	Student-t	Bonferroni	Dunnett	Williams
4				2.060	2.692	2.607	2.156
水準番号		差	|t 値|	p値	Bonferroni	Dunnett	Williams
A0	A1	-2.000	1.236	0.2279	---	---	---
A0	A2	-3.500	2.163	*0.0403*	---	---	有意
A0	A3	-4.333	2.678	*0.0129*	---	有意	有意
A0	A4	-5.500	3.399	*0.0023*	有意	有意	有意

			Williams	5%有意	2.5%有意
				1.827	2.156
水準番号		差	|t 値|	5%有意	2.5%有意
A0	A4	-5.500	3.399	有意	有意
A0	A3	-4.333	2.678	有意	有意
A0	A2	-3.500	2.163	有意	有意
A0	A1	-2.000	1.236	---	---

表示 3-2　多重比較における t 値の5%点 (Dunnettの両側5%点)

	t	Dunnett					
ve	2	3	Bonf.	5	Bonf.	7	Bonf.
10	2.228	2.568	2.634	2.891	3.038	3.070	3.277
20	2.086	2.379	2.423	2.651	2.744	2.802	2.927
25	2.060	2.344	2.385	2.607	2.692	2.752	2.865
30	2.042	2.321	2.360	2.578	2.657	2.720	2.825
60	2.000	2.265	2.299	2.508	2.575	2.642	2.729
∞	1.960	2.212	2.241	2.442	2.498	2.567	2.638
	m		2		4		6

Bonf.=Bonferroni

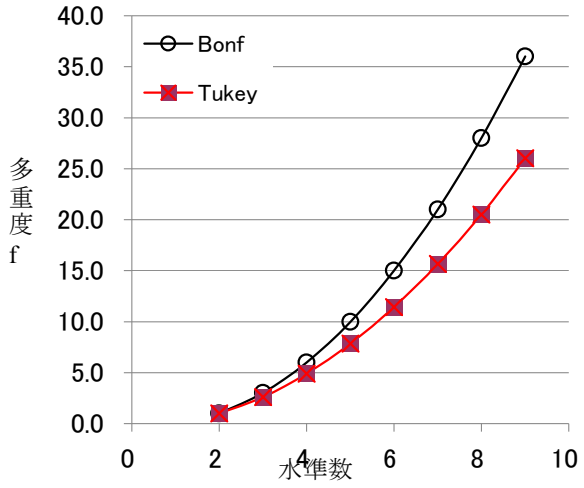
表示 3-3　自由度 25，n0=ni の場合の Williams の棄却限界値 (片側検定)

自由度ν =25　　n0 = ni

水準数	2	3	4	5	6	7	8	9
5%	1.708	1.788	1.814	1.827	1.835	1.839	1.843	1.845
2.5%	2.060	2.125	2.146	2.156	2.162	2.165	2.168	2.17

表示 4-1　全組み合わせでの多重検定の比較

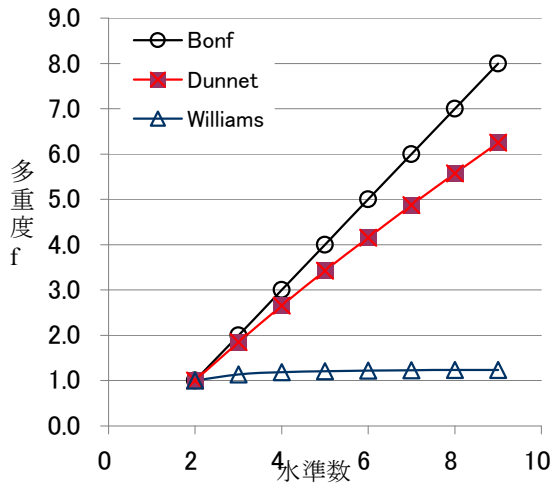
α	0.05							
水準数	2	3	4	5	6	7	8	9
Bonf	1.000	3.000	6.000	10.000	15.000	21.000	28.000	36.000
Tukey	1.000	2.616	4.901	7.847	11.424	15.665	20.490	26.019



	0.05							
水準数	2	3	4	5	6	7	8	9
Student	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387
Tukey	1.960	2.343	2.569	2.728	2.850	2.949	3.031	3.102
α*	0.025	0.010	0.005	0.003	0.002	0.002	0.001	0.001
f	1.000	2.616	4.901	7.847	11.424	15.665	20.490	26.019

表示 4-2　コントロール群と比較する場合の多重検定の比較

α	0.05							
水準数	2	3	4	5	6	7	8	9
Bonf	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000
Dunnet	1.000	1.854	2.656	3.423	4.153	4.874	5.571	6.249
Williams	1.000	1.139	1.186	1.209	1.221	1.230	1.236	1.239



	0.05							
水準数	2	3	4	5	6	7	8	9
Dunnett	1.960	2.212	2.349	2.442	2.511	2.567	2.613	2.652
α*	0.025	0.013	0.009	0.007	0.006	0.005	0.004	0.004
f	1.000	1.854	2.656	3.423	4.153	4.874	5.571	6.249
Williams	1.960	2.015	2.032	2.040	2.044	2.047	2.049	2.050
α*	0.025	0.022	0.021	0.021	0.020	0.020	0.020	0.020
f	1.000	1.139	1.186	1.209	1.221	1.230	1.236	1.239