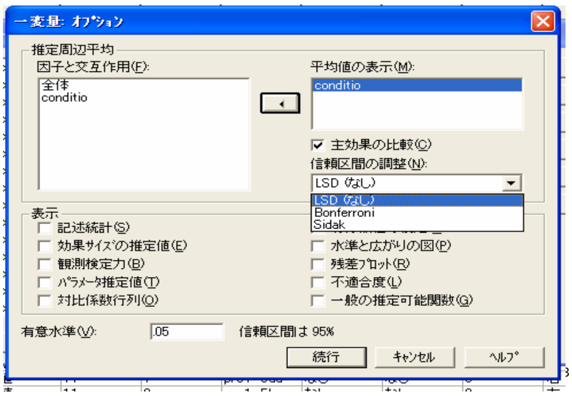
多重比較法

2006,**前期,データ解析演習**

M２　服部貴大

ＳＰＳＳの多重比較（1）

ＳＰＳＳの多重比較（2）



今回、主に発表するもの

適用状況	手法の特徴		検定方法
全ての群間の対比較	パラメトリック	シングルステップ	Tukey, Bonferroni
		ステップワイズ	Tukey-Welsch(R-E-G-WのQ, F), オリジナルRyan, Tukey-b
	ノンパラメトリック	シングルステップ	Bonferroni
		ステップワイズ	Tukey-Welsch, オリジナルRyan
対照群との対比較	パラメトリック	シングルステップ	Bonferroni, Dunnett
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Bonferroni
		ステップワイズ	
対比についての検定	パラメトリック	シングルステップ	Scheffe, Bonferroni
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Scheffe, Bonferroni
		ステップワイズ	

例（森・吉田，1990，p87の一部）

- 20**個の単語を覚える**
- **記憶方略**（3**条件；自由,反復,イメージ**）

自由(条件A)	反復 (条件B)	イメージ (条件C)
9	6	9
7	5	11
8	6	13
8	3	14
12	6	16
11	7	12
8	10	15
13	9	14
9.5	6.5	13

3つの記憶方略のうち、どれが良いかを知りたい

- A vs B,　B vs C,　C vs A**でそれぞれの条件間で**t**検定をし、それぞれが有意水準（通常α＝**0.05**または**0.01**）となるかどうかで、条件間の差を判定。**
  - **t検定の繰り返し（ＬＳＤ[最小有意差]）**
- **t検定の繰り返し（ＬＳＤ[最小有意差]）は良くないとされる**
  - **ちなみに、t検定の繰り返しとＬＳＤでは、検定で用いる誤差分散が異なる**
    - **t検定は2条件での誤差分散**
    - **ＬＳＤは全条件での誤差分散**

誤差の割合（有意水準）の概念的単位

- 第一種の過誤を犯す確率を、どのような単位において特定の値（α）に設定するかといったことに関する概念
  - 比較の集合
    - 比較したい組み合わせ全体の集合を単位とする
  - 個々の比較
    - 比較したい組み合わせ全体ではなく、その中の幾つかだけの集合を単位とする

どうしてＬＳＤはいけないか

- 先のＬＳＤでは、各条件ごとのt検定での特定の値（有意水準α）を0.05（または0.01）に設定
  - 個々の比較を誤差の割合の概念的単位としている
- しかし、検定の目的は「3つの記憶方略のうち、どれが良いかを調べる」こと。
- 比較の集合を、誤差の割合の概念的単位としている（3条件の比較の組合せ全体の集合を対象）
- だが、LSDでは、誤差の割合の概念的単位を、個々の比較に有意水準の基準を変更している。

では、検定目的と一貫した誤差の割合の概念的単位である比較の集合でＬＳＤを行うと

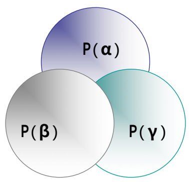
- 比較の集合全体で、有意水準を考えなくてはいけない
  - 比較の集合全体で、少なくとも1つは間違う（第Ｉ種の過誤を犯す）確率⇒有意水準
- ＬＳＤで2条件間の有意水準を５％で行ったとしたら
  - ⇒　１つの対比較で５％の確率で間違う
    - 誤って帰無仮説を棄却する確率５％
    - ５％の確率で、本当は有意差はないのに有意としてしまう（第Ｉ種の過誤）

3つの対比較（A vs B,　B vs C,　C vs A）のうち、少なくとも1つは間違う確率の計算①

- 1つの対比較で間違わない確率
  - 1ー0.05 = 0.95
- 3つとも間違わない確率
  - 0.95✕0.95✕0.95 = 0.857375
- 3つの対比較のうち少なくとも1つは間違う確率
  - 1ー0.857375 = 0.142625 (14.3%)

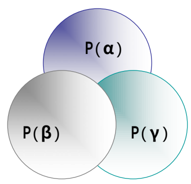
3つの対比較（A vs B,　B vs C,　C vs A）のうち、少なくとも1つは間違う確率の計算②

- 条件Ａ,Ｂ,Ｃを比較したい
  - A vs B,
    - ここで間違う確率：P(α)
  - B vs C,
    - ここで間違う確率：P(β)
  - C vs A,
    - ここで間違う確率：P(γ)



3つの対比較（A vs B,　B vs C,　C vs A）のうち、少なくとも1つは間違う確率の計算③

- 別解
- P(α∪β∪γ)
  = P(α)+ P(β)+ P(γ)
  ー P(α∩β)ー P(β∩γ)ー P(γ∩α)
  ＋ P(α∩β∩γ)
  = 0.05✕3
  ー0.05✕0.05✕3
  ＋ 0.05✕0.05✕0.05
  = 0.142625 (14.3%)



どうすれば いいか？（1）

- 全体としての有意水準を公称の値（有意水準5%のようにあらかじめ宣言しておく値）にコントロールできるように、1回1回の検定における棄却限界値を調整する
- この方法を⇒**多重比較法**
  (multiple comparison procedure)

どうすれば いいか？（2）

- 有意水準5%を保ったまま3条件を比べると
  - A vs B,　B vs C,　C vs A
  - 対比較を3回行うので、それぞれの有意水準を5%÷3 = 1.67%にして行う
    ⇒ボンフェローニ(Bonferroni)の方法
  - $P(\alpha)+P(\beta)+P(\gamma)$
    $= 0.0167 + 0.0167 + 0.167$
    $=0.0501$

ボンフェローニの不等式

- $P(\alpha)+P(\beta)+P(\gamma)$
  $= 0.0167 + 0.0167 + 0.167$
  $=0.0501$
- $P(\alpha\cup\beta\cup\gamma)$
  $= P(\alpha)+P(\beta)+P(\gamma)$
  $- P(\alpha\cap\beta) - P(\beta\cap\gamma) - P(\gamma\cap\alpha)$
  $+ P(\alpha\cap\beta\cap\gamma)$
- 青の部分を入れれば、必ず0.05より小さい数となる
- $P(\alpha\cup\beta\cup\gamma)$
  $= 0.0501$
  $- 0.000279\times 3$
  $+ 0.00000465746$
  $= 0.049825$

• ボンフェローニの不等式
  - $P(\alpha\cup\beta\cup\gamma) \leqq P(\alpha)+P(\beta)+P(\gamma)$

ボンフェローニの欠点

- 条件数が多いと、検出力が低くなる
  - 各対比較の有意水準が条件数4以上では0.01を切ってしまう

条件数	対比較の組合せ数	各対比較の有意水準	少なくとも1つ間違う確率
2	1	0.050000	0.050000
3	3	0.016667	0.049171
4	6	0.008333	0.048970
5	10	0.005000	0.048890
6	15	0.003333	0.048850
7	21	0.002381	0.048827
8	28	0.001786	0.048813
9	36	0.001389	0.048804
10	45	0.001111	0.048797

ボンフェローニの利点と改良型

- ボンフェローニの不等式に基づく事が、ボンフェローニ法の特徴
- ボンフェローニの不等式は各条件間に相関のある場合でも成り立つ
  ⇒各条件のデータが相関を持っている場合でもボンフェローニは用いることが可能（ボンフェローニの利点）
- この不等式を使ってボンフェローニの検出力の低さを改良したものも開発されている
  - ホルム(Holm)の方法、シェイファー(Shaffer)の方法
  - しかし検出力の面では、他のステップダウン法（ぺり法など）よりも低い

誤差の割合（有意水準）の概念的単位が**個々の比較**の場合

- **ア・プリオリな比較**（事前比較：a priori comparison)
  - 各処理水準の平均値の差について、あらかじめ特定の仮説が設けられている場合
- **ア・ポステリオリな比較**（事後比較：a posteriori comparison)
  - 当該の要因に何らかの効果があることだけを仮定し、個々の平均値の差については特に仮説を持たない場合

- ア・プリオリな比較の場合に、誤差の割合の概念的単位が**個々の比較**となりうる場合がある

個々の比較となる例

- 先の記憶方略を1条件増やして
- 4条件（自由A，反復B，イメージC，文章D）
- あらかじめ以下の3つの仮説が立てられていた場合
  - 仮説1：反復方略は再生成績に対し抑制効果を持つ（自由方略よりも再生数が少ない） A vs B
  - 仮説2：イメージ方略は文章方略よりも効果があるC vs D
  - 仮説3：深い処理の方略（イメージと文章）はそうでない方略よりも効果がある(A and B) vs (C and D)

↓

- 全ての組み合わせの多重比較を行う必要はなく、この仮説だけの比較でよい

ア・プリオリな比較

- ア・プリオリな比較の場合には、多重比較に先立っての主効果の有意性の検定は不要
- ア・プリオリな比較の検定法
  - 直交対比（比較）における多重t検定(標準誤差は全体のものを使用)
  - ダン法（ボンフェローニ法）
    - 対比（比較）が直交していない場合

（森・吉田，1990，pp164-166）

直交対比（直交比較）とは

- 直交対比（直交比較）
  - 互いに重複した情報を含まない比較
  - 記憶方略の例で言えば
  - 仮説1：A vs B
    - $\mu_A - \mu_B = (+1)\mu_A + (-1)\mu_B + (0)\mu_C + (0)\mu_D$
  - 仮説2：C vs D
    - $\mu_C - \mu_D = (0)\mu_A + (0)\mu_B + (+1)\mu_C + (-1)\mu_D$
  - 仮説3：(A and B) vs (C and D)
    - $(\mu_A + \mu_B)/2 - (\mu_C + \mu_D)/2$
      $= (1/2)\mu_A + (1/2)\mu_B + (-1/2)\mu_C + (-1/2)\mu_D$
  - それぞれの仮説の係数の積和が0となる⇒直交対比
    - 仮説1と2：$1 \times 0 + (-1) \times 0 + 0 \times 1 + 0 \times (-1) = 0$
    - 仮説1と3：$1 \times 1/2 + (-1) \times 1/2 + 0 \times (-1/2) + 0 \times (-1/2) = 0$
    - 仮説2と3：$0 \times 1/2 + 0 \times 1/2 + 1 \times (-1/2) + (-1) \times (-1/2) = 0$

検定の多重性の問題

- 森,吉田（1990）では、ア・プリオリの場合だけ主効果の有意差の検定の不要を主張
- 永田・吉田（1997）では、
  - 多重比較法を適用する際には、前もって1元配置分散分析により母平均の一様性の帰無仮説についての検定を行い、有意になった場合にのみ多重比較を行う、と述べられることがあるが、これは一般的には正しくない
  - 1元配置分散分析で用いる検定統計量（F統計量）が多重比較の手順の中に含まれていない場合には「1元配置分散分析」と「多重比較法」の2つの検定を行うことになり、ここにまた検定の多重性が生じるためである

永田・吉田（1997）では

- F統計量を用いない多重比較（例えば、テューキー、ボンフェローニなど）では
- 最初に、どの比較をするのか事前に決め、主効果の有意性の検定を行わず、多重比較を行う手順が書かれている

- 心理学では、情報の一つとして主効果の検定も載せる？
- 事前に分散分析を行うと検出力が少し落ちる」ことを心に留めておくべき
- 多要因の分散分析を行ってから（単純）主効果の多重比較を行うことが多く、一要因分散分析で分析が終了するのはまれ。この意味では、「分散分析→多重比較」が自然に用いられる。

テューキー(Tukey)の方法（1）

- すべての条件間の対比較
- 帰無仮説： $\mu_i = \mu_j$
- 対立仮説： $\mu_i \neq \mu_j$
- 検定統計量t_{ij}を計算

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{V_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

- 誤差分散V_Eが異なるだけで、t検定の検定統計量とほぼ同じ
  - 全条件の誤差分散か、2条件だけの誤差分散か

テューキー(Tukey)の方法（2）

- ステューデント化された範囲（q）の臨界値の表を用いて求める
  - t検定ではt分布の臨界値を使用
- 帰無仮説を棄却

$$|t_{ij}| \geq \frac{q}{\sqrt{2}}$$

- 全条件のサンプルサイズが等しい場合

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{V_E\left(\frac{2}{n}\right)}} \geq \frac{q}{\sqrt{2}} \quad \Longleftrightarrow \quad \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{V_E}{n}}} \geq q$$

25

ステューデント化された範囲の統計量Q

- 範囲（最大値ー最小値）という統計量の期待値は、処理水準数（条件数）が多いほど、大きくなる
- そこで、パラメータの中に処理水準数を含む次式で表される統計量
  - P：処理水準の集合

$$Q = \max_{i,j \in p} \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{V_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

- Qという統計量の臨界値は、有意水準（α）と誤差項の自由度（ϕ_e）とともに、処理水準数（a）によって規定（付表4を参照）

26

テューキー(Tukey)の方法の注意点

- 条件数が多く、条件間のサンプルサイズの比率が大きい場合には、有意水準よりも低い確率となる（保守性が高くなる・有意になりにくくなる）
  - 条件数5以上　かつ　比率3倍以上
  - これ以下であるならば0.045≦α＜0.05は保てる
- 母集団分布は正規分布、すべての条件の母分散は等しいことを仮定（**パラメトリック法**）
  - ⇒母集団分布を正規分布と仮定できない場合には使用できない
- ステップダウン法の方が検出力が高い

27

演習（1）

- 先の記憶方略の例を使って
- 4条件（自由A，反復B，イメージC，文章D）
- SPSSのデータ「演習（1）用」をダウンロード
- SPSSで1要因分散分析→その後の検定→BonferroniとTukeyにチェック
- それぞれの有意確率を比較

（SPSSでは、Bonferroniは対比較の数を掛けて0.05に合わせて算出される）

28

多重比較

従属変数: 再生数

	(I) 条件	(J) 条件	平均値の差 (I-J)	標準誤差	有意確率	95% 信頼区間 下限	95% 信頼区間 上限
Tukey HSD	自由 (A)	反復 (B)	3.0000	1.1495	.065	-.1386	6.1386
		イメージ (C)	-3.5000*	1.1495	.024	-6.6386	-.3614
		文章 (D)	-3.0000	1.1495	.065	-6.1386	.1386
	反復 (B)	自由 (A)	-3.0000	1.1495	.065	-6.1386	.1386
		イメージ (C)	-6.5000*	1.1495	.000	-9.6386	-3.3614
		文章 (D)	-6.0000*	1.1495	.000	-9.1386	-2.8614
	イメージ (C)	自由 (A)	3.5000*	1.1495	.024	.3614	6.6386
		反復 (B)	6.5000*	1.1495	.000	3.3614	9.6386
		文章 (D)	.5000	1.1495	.972	-2.6386	3.6386
	文章 (D)	自由 (A)	3.0000	1.1495	.065	-.1386	6.1386
		反復 (B)	6.0000*	1.1495	.000	2.8614	9.1386
		イメージ (C)	-.5000	1.1495	.972	-3.6386	2.6386
Bonferroni	自由 (A)	反復 (B)	3.0000	1.1495	.086	-.2634	6.2634
		イメージ (C)	-3.5000*	1.1495	.030	-6.7634	-.2366
		文章 (D)	-3.0000	1.1495	.086	-6.2634	.2634
	反復 (B)	自由 (A)	-3.0000	1.1495	.086	-6.2634	.2634
		イメージ (C)	-6.5000*	1.1495	.000	-9.7634	-3.2366
		文章 (D)	-6.0000*	1.1495	.000	-9.2634	-2.7366
	イメージ (C)	自由 (A)	3.5000*	1.1495	.030	.2366	6.7634
		反復 (B)	6.5000*	1.1495	.000	3.2366	9.7634
		文章 (D)	.5000	1.1495	1.000	-2.7634	3.7634
	文章 (D)	自由 (A)	3.0000	1.1495	.086	-.2634	6.2634
		反復 (B)	6.0000*	1.1495	.000	2.7366	9.2634
		イメージ (C)	-.5000	1.1495	1.000	-3.7634	2.7634

*. 平均の差は .05 で有意

29

ステップダウン法（1）

- ステップワイズ法
  - 1つの帰無仮説を検定して、その結果を見た後で次のステップに進むかどうかを決定していく方法
  - 今まで述べてきた方法は**シングルステップ法**
  - ステップワイズ法には**ステップダウン法**とステップアップ法がある
- コヒーレンス(coherence)
  - ある帰無仮説が保留されるならば、それが誘導するすべての帰無仮説も保留される
    - 例　帰無仮説：$\mu_A = \mu_B = \mu_C$が保留（成立）されれば、これが誘導する3つの帰無仮説：$\mu_A = \mu_B$，$\mu_C = \mu_A$，$\mu_B = \mu_C$も保留（成立）される

30

ステップダウン法（2）

- ステップダウン法は、コヒーレンスを保障するように手順を構成
- ステップダウン法の例を1つ紹介
  - テューキー・ウェルシュ(Tukey-Welsch)の方法
    - 別名、Ryan法、修正Ryan法、Ryan, Einot, Gabriel and Welsch法（略してR-E-G-WのQ, F）などとも呼ぶ
      - Ryan法は、Ryan(1959, 1960)が、ボンフェローニの不等式を使って、提案。
      - その後、Einot, Gabriel(1975)がボンフェローニの不等式よりも精度の高いシダックの不等式を使うことで、Ryan法を改善。
      - さらにWelsch(1977)が、有意水準の配分について改善を行い、現在の形が出来上がる。
    - 森・吉田(1990)p171にあるRyan法は、Ryanによるオリジナルな方法

テューキー・ウェルシュの方法

- ステップ数：p
  - 例えば、帰無仮説 $\mu_A = \mu_B = \mu_C$ を検定したいとき、この中に3つの処理水準の平均がある。このとき、このステップ数は3であるという。また、このとき{A,B,C}を用いて表す。
- 統計量としては、
  - ステューデント化された範囲の統計量Q（R-E-G-WのQ）
  - F統計量（永田・吉田, 1997, p108）（R-E-G-WのF）
  - その他、ノンパラメトリック法の統計量を用いることも可能
    - （ただどんな統計量でも可能かは不明。処理水準数を考慮したものだけかも）
  - 使用する統計量により、棄却限界値を判断する表が異なる程度で、その他の手順はどの統計量でも同じ
  - 平均値での多重比較の場合、統計量QとFを使えるが、検出力はFの方が若干良い

有意水準（1）

- $\alpha_{a-1} = \alpha_a = \alpha$　　(p= a, a－1のとき)
- $\alpha_p = 1-(1-\alpha)^{p/a}$　(2≦p≦a－2のとき)
  - α：全体の有意水準（自分で最初に設定したもの）
  - a：全体の処理水準数
  - p：ステップ数
- 例：α=0.05, a=4の場合
  - ステップ数4　$\alpha_4 = \alpha = 0.05$
  - ステップ数3　$\alpha_3 = \alpha = 0.05$
  - ステップ数2　$\alpha_2 = 1-(1-0.05)^{2/4} = 0.0253$

有意水準（2）

- 前ページの有意水準は、有意水準αを個々の帰無仮説の大きさpに応じ、シダックの不等式に基づいて配分したもの

  （テューキー・ウェルシュの有意水準配分）
- シダックの不等式

$$P(T_1 \le c_1, T_2 \le c_2, \cdots, T_k \le c_k) \ge \prod_{i=1}^{k} P(T_i \le c_i)$$

  - 少なくとも1つは $T_i \leqq c_i$ となる確率

    ≧　全てがともに $T_i \leqq c_i$ となる確率

テューキー・ウェルシュの方法 検定手順（1）

- 先の記憶方略の例を用いて
  - 4条件（自由A, 反復B, イメージC, 文章D）

自由(A)	反復(B)	イメージ(C)	文章(D)
9	6	9	10
7	5	11	13
8	6	13	8
8	3	14	13
12	6	16	12
11	7	12	14
8	10	15	14
13	9	14	16
9.5	6.5	13	12.5

  - 統計量はステューデント化された範囲の統計量Qを使用

テューキー・ウェルシュの方法 検定手順（2）

- ステップ数の最も大きいものから検定
  - 4条件（4水準）あるので、ステップ数p=4をまず検定
  - 統計量Qを算出
    - 右の式に4条件全ての対比較における値を算出し、その中で最も大きいものがQ

$$Q = \max_{i,j \in p} \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{V_E\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}}$$

テューキー・ウェルシュの方法 演習①

- **ステップ数p=4の統計量Qを算出**
  - p=4, {A,B,C,D}に含まれる対比較のt統計量を算出し、その中の最大のものがQ統計量となる
  - エクセルのデータ「Tukey-Welsch用」をダウンロード
    - エクセルを用いて、対比較のt統計量を算出
    - 式は、前ページのmaxの右にあるものを使用
      - x_iバー：各条件の平均
      - V_E：誤差分散（この例では5.286）
      - n_i：各条件のデータ数（この例では各条件ともn＝8）
- **ステップ数p=4の統計量Q（　　　　　）**

テューキー・ウェルシュの方法 演習①解答

- **ステップ数p=4の統計量Q（5.654）B vs C**

対比較			前者の平均	後者の平均	誤差分散	各条件のn	t統計量
A	vs	B	9.5	6.5	5.286	8	2.60968249
A	vs	C	9.5	13	5.286	8	-3.0446296
A	vs	D	9.5	12.5	5.286	8	-2.6096825
B	vs	C	6.5	13	5.286	8	-5.6543121
B	vs	D	6.5	12.5	5.286	8	-5.219365
C	vs	D	13	12.5	5.286	8	0.43494708

テューキー・ウェルシュの方法 検定手順(3)

- **棄却限界値c_pを付表を使って算出と比較**
  - 3≦pのとき

$$c_p = \max\left\{\frac{q(p,\phi_E;\alpha_p)}{\sqrt{2}}, c_{p-1}\right\}$$

  - p=2のとき

$$c_p = \frac{q(p,\phi_E;\alpha_p)}{\sqrt{2}}$$

テューキー・ウェルシュの方法 演習②

- **ステップ数p=4の棄却限界値c_4を付表を使って算出**
  - $q(p,\phi_E;\ \alpha_4)/\sqrt{2}$ =（　　　　　）$/\sqrt{2}$ =（　　　　　）
    付表4を使用
    - p：ステップ数（ここで比較対照したい処理水準数）=（　　）
    - ϕ_E：誤差項の自由度
      - ϕ_E = n – a = 全データ数－条件数 =（　　）
    - 有意水準：α_4（有意水準(1)のスライドから）=（　　）
  - $c_{p-1} = c_{4-1} = c_3$ =（　　　　）
    - $q(p,\phi_E;\ \alpha_3)/\sqrt{2}$ =（　　　　　）$/\sqrt{2}$ =（　　　　　）
    - $c_2 = q(2,\phi_E;\ \alpha_2)/\sqrt{2}$ =（　　　　　）$/\sqrt{2}$ =（　　　　　）
      付表7を使用
  - $q(p,\phi_E;\ \alpha_4)$ と c_3 の大きい方がp=4の棄却限界値c_4
    - c_4 =（　　　　　）

テューキー・ウェルシュの方法 演習②解答

- **ステップ数p=4の棄却限界値c_4を付表を使って算出**
  - $q(p,\phi_E;\ \alpha_4)/\sqrt{2}$ = (3.861) $/\sqrt{2}$ = (2.730)
    付表4を使用
    - p：ステップ数（ここで比較対照したい処理水準数）= (4)
    - ϕ_E：誤差項の自由度
      - ϕ_E = n – a = 全データ数－条件数 = 32-4 = (28)
    - 有意水準：α_4（有意水準(1)のスライドから）= (0.05)
  - $c_{p-1} = c_{4-1} = c_3$ = (2.474)
    - $q(p,\phi_E;\ \alpha_3)/\sqrt{2}$ = (3.499) $/\sqrt{2}$ = (2.474)
      (p=3, ϕ_E=28, α_3=0.05)
    - $c_2 = q(2,\phi_E;\ \alpha_2)/\sqrt{2}$ = (3.341) $/\sqrt{2}$ = (2.362)
      (p=2, ϕ_E=28, α_3=0.0253)　付表7を使用
  - $q(p,\phi_E;\ \alpha_4)$ と c_3 の大きい方がp=4の棄却限界値c_4
    - c_4 = (2.730)

テューキー・ウェルシュの方法 検定手順(4)

- **ステップ数p=4の統計量Q（5.654）と棄却限界値c_4 = 2.730 を比較**
- **Q≧c_4なので、帰無仮説を棄却**
  →ステップ数p=3の検定へ進む
  - もし、Q＜c_4ならば、帰無仮説を保留し、コヒーレンスからその他の帰無仮説も保留されるため、検定を終了

テューキー･ウェルシュの方法 演習③

- ステップ数p=3の検定
  - {A,B,C}　統計量Q = (　　　　)　(棄却 or 保留)
  - {A,B,D}　統計量Q = (　　　　)　(棄却 or 保留)
  - {A,C,D}　統計量Q = (　　　　)　(棄却 or 保留)
  - {B,C,D}　統計量Q = (　　　　)　(棄却 or 保留)
  - 棄却限界値c_3 = (　　　　)

テューキー･ウェルシュの方法 演習③解答

- ステップ数p=3の検定
  - {A,B,C}　統計量Q = 5.654　棄却
  - {A,B,D}　統計量Q = 5.219　棄却
  - {A,C,D}　統計量Q = 3.044　棄却
  - {B,C,D}　統計量Q = 5.654　棄却
  - 棄却限界値c_3 = 2.474

  ⇒棄却されたものが誘導するステップ数p=2の検定を行う
  - もし、棄却されず、保留された場合
    - 例えば、{A,C,D}が保留されたとしたら
    - {A,C,D}が誘導するA vs C, C vs D, D vs Aも保留され、この3つの対比較はこれ以上検定を行わない

テューキー･ウェルシュの方法 演習④

- {A,B,C} 棄却
  - {A,B}　統計量Q = (　　　　)　(棄却 or 保留)
  - {B,C}　統計量Q = (　　　　)　(棄却 or 保留)
  - {C,A}　統計量Q = (　　　　)　(棄却 or 保留)
- {A,B,D} 棄却
  - {A,B}
  - {B,D}　統計量Q = (　　　　)　(棄却 or 保留)
  - {D,A}　統計量Q = (　　　　)　(棄却 or 保留)
- {A,C,D} 棄却
  - {A,C}
  - {C,D}　統計量Q = (　　　　)　(棄却 or 保留)
  - {D,A}
- {B,C,D} 棄却
  - {B,C}
  - {C,D}
  - {D,B}
- 棄却限界値 c_2= (　　　　)

テューキー･ウェルシュの方法 演習④解答

		テューキー･ボンフェローニ
– {A,B,C} 棄却		
• {A,B}　統計量Q = (2.609　)	棄却	
• {B,C}　統計量Q = (5.654　)	棄却	棄却
• {C,A}　統計量Q = (3.044　)	棄却	棄却
– {A,B,D} 棄却		
• {A,B}		
• {B,D}　統計量Q = (5.219　)	棄却	棄却
• {D,A}　統計量Q = (2.609　)	棄却	
– {A,C,D} 棄却		
• {A,C}		
• {C,D}　統計量Q = (0.434　)	保留	
• {D,A}		
– {B,C,D} 棄却		
• {B,C}		
• {C,D}		
• {D,B}		

- 棄却限界値　c_2= 2.362

その他のステップワイズ法（1）

- Newman-Keulsの方法
  Duncanの方法
  - テューキー･ウェルシュの方法と有意水準の設定の仕方が違うだけで、後の手順は同じ
    - Newman-Keulsの方法の有意水準
      - すべてのステップで$\alpha_p=\alpha$を使用
    - Duncanの方法の有意水準
      - $\alpha_p=1-(1-\alpha)^{p-1}$と設定して使用

↓

- どちらの方法の有意水準も、少なくとも1つは間違う確率をコントロールできない
- 使用してはいけない

その他のステップワイズ法（2）

- テューキー(Tukey)のb法
  - テューキー･ウェルシュの方法とは違うもの
  - Newman-Keulsの方法とTukeyの方法を折衷したステップダウン法
  - 棄却限界値に、各ステップ数pの棄却限界値とテューキーの方法の棄却限界値の平均を用いる
  - 有意水準は、すべてのステップで$\alpha_p=\alpha$を使用
  - 検定力は、次のライアン法とほぼ同じ
    - 詳しくは森･吉田(1990)

その他のステップワイズ法 (3)

- オリジナルのライアン(Ryan)の方法 (S30より)
  - 特定の統計量に依存しない
    - 平均値差だけでなく、比率、中央値、分散、相関係数ノンパラメトリックなど様々な統計量の多重比較に適用可能
  - ステップ数の大きなものから検定していく手順はテューキー・ウェルシュと同じ
  - 各ステップにおける有意水準の設定の仕方が異なる
    - α：全体の有意水準 (自分で最初に設定したもの)
    - a：全体の処理水準数
    - p：ステップ数

$$\alpha_p = \frac{2\alpha}{a(p-1)} = \frac{\alpha}{\frac{a(p-1)}{2}}$$

  - ステップ数が大きい程、厳しい有意水準 (小さくなる程ゆるくなる)
    - 最大ステップ数のときにはボンフェローニと同じ有意水準

その他のステップダウン法 (4)

- ペリ (Peritz) の方法
  - Tukey-Welschの方法とNewman-Keulsの方法を組み合わせて検出
  - もっとも検出力が良いとされる
    - 詳しくは永田・吉田(1997)
  - しかしSPSSにはなし
    - SPSSで、全ての対比較を行う場合、最も検出力が高いのはR-E-G-WのFとなる
  - ステップダウン法の欠点
    - (要因効果の差の) 区間推定が実行しにくい
    - 方向性の決定についての過誤の大きさを特定しにくい

対比についての検定

- 対比(contrast)
  - 母平均に対し、線形の式を作り、その係数によって様々な帰無仮説を表現する
  - 例
  - 仮説1：A vs B
    - $\mu_A - \mu_B = (+1)\mu_A + (-1)\mu_B + (0)\mu_C + (0)\mu_D$
  - 仮説2：C vs D
    - $\mu_C - \mu_D = (0)\mu_A + (0)\mu_B + (+1)\mu_C + (-1)\mu_D$
  - 仮説3：(A and B vs C and D)
    - $(\mu_A + \mu_B)/2 - (\mu_C + \mu_D)/2$
      $= (1/2)\mu_A + (1/2)\mu_B + (-1/2)\mu_C + (-1/2)\mu_D$
  - 1つの仮説の係数の和が0となるもの⇒対比
  - 各々の仮説間の係数の積の和が0⇒直交対比 (S21)

対比についての検定方法

- シェフェ(Scheffe)の方法
  - 対比で表現できる帰無仮説を検定可能
  - 詳しくは永田・吉田 (1997)を参照
  - 注意点
    - 対比較も、対比で表現できるので、シェフェの方法で検定可能
    - しかし、検出力から対比較ではテューキーの方法の方が良い

対照群との対比較

- 全ての群と対比較する必要がなく、ある一つの対照群と残りの群との対比較を検定したいとき
- ダネット(Dunnett)の方法
  - 検定統計量はテューキーの方法と同じ
  - しかし、対比較の個数がテューキーよりも少なくなるため、検定で用いる表が異なる
  - 詳しくは永田・吉田 (1997)を参照

まとめ (多重比較のタイプ)

適用状況	手法の特徴		検定方法
全ての群間の対比較	パラメトリック	シングルステップ	Tukey, Bonferroni
		ステップワイズ	Tukey-Welsch(R-E-G-WのQ, F), オリジナルRyan, Tukey-b
	ノンパラメトリック	シングルステップ	Bonferroni
		ステップワイズ	Tukey-Welsch, オリジナルRyan
対照群との対比較	パラメトリック	シングルステップ	Bonferroni, Dunnett
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Bonferroni
		ステップワイズ	
対比についての検定	パラメトリック	シングルステップ	Scheffe, Bonferroni
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Scheffe, Bonferroni
		ステップワイズ	

もう一つの適用状況

- 適用状況としては
  - 単調性を想定した対比較というものもあり
    - 対照群との対比較において、さらに群間に単調増加、または単調減少を想定できる場合に適用
    - 例　$\mu_A < \mu_B < \mu_C < \mu_D$ (Aを対照群として、単調増加)
    - 詳しくは永田・吉田(1997)を参照
    - また、今回、紹介していない検定法についても永田・吉田(1997)に多数掲載（ただし、SPSSにはない多重比較）

行ってはいけないといわれている検定方法（永田・吉田, 1997）

- 2標本t検定の繰り返し、無制約LSD法、Dunncan法
- 4群以上のときは
  - Newman-Keuls法（Student-Newman-Keuls）
  - 制約付LSD法（ANOVAを行い、それからt検定）
    - 3群では行っても良い(詳細は永田・吉田, 1997, p31)

4群で制約付LSD法をやってはいけない理由(1)

- 制約付LSD法では、ANOVAを先に行う
- 母平均の配置が $\mu_A = \mu_B = \mu_C = \mu_D$ のとき
  - ANOVAで誤って棄却される確率5%
  - この場合、その後にどんな検定を行っても5%を超えず、問題なし

少なくとも1つ間違う（第Ⅰ種の過誤を犯す）確率

= P(E(A,B,C,D) and [G(A,B) or G(B,C) or ･･or G(C,D)])

=P(E(A,B,C,D))×P [G(A,B) or G(B,C) or ･･or G(C,D)])

=　0.05　×　0.185

= 0.00925 (0.925%)

4群で制約付LSD法をやってはいけない理由(2)

- 母平均の配置が $\mu_A = \mu_B = \mu_C \neq \mu_D$ のとき
  - 1つの条件だけが他の条件より大きく離れているとき
  - この場合、ANOVAではほぼ100%で棄却される

少なくとも1つ間違う（第Ⅰ種の過誤を犯す）確率

= P(E(A,B,C,D) and [G(A,B) or G(B,C) or ･･or G(C,D)])

=P(E(A,B,C,D))×P [G(A,B) or G(B,C) or ･･or G(C,D)])

1　×P [G(A,B) or G(B,C) or G(C,A)])

=　1　×　0.143

= 0.143 (14.3%)

4群で制約付LSD法をやってはいけない理由(3)

- (1)(2)では、母平均の配置を想定して計算したが、実際には、どの母平均の配置になるのかは不明
- もしかしたら、母平均の配置が $\mu_A = \mu_B = \mu_C \neq \mu_D$ となっている可能性もあるので、4群で制約付LSD法はやってはいけない

3群で制約付LSD法を行っても良い理由(1)

- 母平均の配置が $\mu_A = \mu_B = \mu_C$ のとき
  - 4群と同じで問題なし

少なくとも1つ間違う（第Ⅰ種の過誤を犯す）確率

= P(E(A,B,C) and [G(A,B) or G(B,C) or G(C,A)])

=P(E(A,B,C))×P [G(A,B) or G(B,C) or G(C,A)])

=　0.05　×　0.143

= 0.00715 (0.715%)

3群で制約付LSD法を行っても良い理由(2)

- **母平均の配置が**$\mu_A = \mu_B \neq \mu_c$**のとき**
  少なくとも1つ間違う（第Ⅰ種の過誤を犯す）確率
  = P(E(A,B,C) and [G(A,B) or G(B,C) or G(C,A)])
  =P(E(A,B,C))×P [G(A,B) or G(B,C)])
  1 × 0.05
  = 0.05 (5%)
  E(A,B,C,D):ANOVA**で棄却される事象**
  G(A,B)：A**と**B**の対比較で棄却される事象**

61

3群で制約付LSD法を行っても良い理由(3)

- **3群の場合には、どの母平均の配置になっても、少なくとも1つ間違える確率（第Ⅰ種の過誤を犯す確率）は、有意水準(0.05)以下に抑えられる**

62

どれを使えばいいか(1)

(Tabachnick & Fidell, 2001；**小野寺・山本,** 2004)

- **各群のケース数が等しい場合**
  - **分散が等質ならば** Tukey**の方法**
  - **分散が等質でなければ**
    - **誤差自由度が75以上のとき**
      - Dunnett**の**C
    - **誤差自由度が75未満のとき**
      - Dunnett**の**T3
- **各群のケースが等しくない場合**
  - **分散が等質ならば** Tukey-Kramer(**ケー数が異なる場合の**Tukey**の方法**)
  - **分散が等質でなければ**
    - **誤差自由度が75以上のとき**
      - Dunnett**の**C
    - **誤差自由度が75未満のとき**
      - Dunnett**の**T3

63

どれを使えばいいか(2)

(Fidell, 2000；**小野寺・山本,** 2004)

- **全ての対比較をする場合**
  - **各群のケース数が等しい**
    - R-E-G-W**の**Q（Ryan, Einot, Gabriel and Welsch**の範囲検定**）
  - **各群のケース数がわずかに違う**
    - Gabriel
  - **各群のケース数が大きく違う**
    - Hochberg**の**GT2
  - **分散の等質性に確信が持てない場合**
    - Games-Howell

64

ＳＰＳＳではどれを使えばいいか

個々の検定の具体的な内容は小野寺・山本(2004)**を参照**

ＳＰＳＳではどれを使えばいいか（1）

(**小野寺・山本,** 2004)

- **使用は勧められないもの**
  - Student-Newman-Keuls
  - Dunncan
  - Waller- Dunncan
  - LSD
- **厳しく第Ⅰ種の過誤を統制したいとき**
  - Bonferroni
  - Tamhane**の**T2
  - Scheffe
  - Dunnett

66

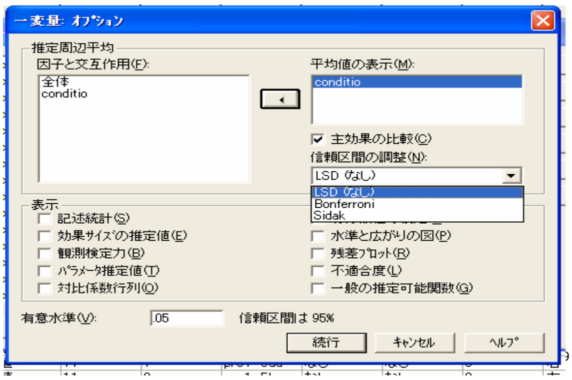
ＳＰＳＳではどれを使えばいいか（2）

- 統制群と比較したいとき
  - Dunnett
- 全ての対比較（群のケース数が等しいとき）
  - 各群の分散が等質のとき
    - Tukey
    - 実質科学的な観点から、さらに有意差が見出せるはずだというとき（ただし、このときは、区間推定は諦める）
      - R-E-G-WのQ（Ryan, Einot, Gabriel and Welschの範囲検定）
  - 各群の分散が等質でないとき
    - 誤差自由度が75以上のとき
      - DunnettのC
    - 誤差自由度が75未満のとき
      - DunnettのT3

ＳＰＳＳではどれを使えばいいか（3）

- 全ての対比較（群のケース数が等しくないとき）
  - 各群の分散が等質のとき
    - Tukey-Kramer (SPSSではTukeyと区別していない)
    - Gabriel（ケース数の違いがわずかな場合）
    - HochbergのGT2（ケース数の大きく違う場合）
  - 各群の分散が等質でないとき
    - 誤差自由度が75以上のとき
      - DunnettのC
      - Games-Howell
    - 誤差自由度が75未満のとき
      - DunnettのT3

「一般線形モデル」→「オプション」→「主効果の比較」とした場合



「一般線形モデル」→「オプション」→「主効果の比較」とした場合

- 主効果についての多重比較を行う
  - 「その後の検定」
    - 多重比較は観測平均（記述統計で出力される平均）を基に計算が行われる
  - 「主効果の比較」
    - 多重比較が、修正されたグループ平均値（推定平均）を基に計算が行われる
    - 共分散分析では、観測平均ではなく、推定平均が検定に使われるので、SPSSでは自動的に「その後の検定」が使えなくなり、「主効果の比較」でしか多重比較ができない

「主効果の比較」では何を使うか？

- 多重比較の調整には「信頼区間の調整」で「LSD」「Bonferroni」「Sidak」の3つが利用可能（おそらく特定の統計量を必要としないため、推定平均でも使用可能）
  - LSDは調整はないに等しく、LSDの選択は勧められない
  - Bonferroniは検出力が低い
  - 検出力の低さが心配なら、残るSidakが選択肢となるのかも
    - 詳しくは、小野寺・山本（２００４）p86

Sidak法とは

- Bonferroni法の検出力を高めるため改良したもの
  - 検定方法は、ほぼBonferroni法と同じ
  - 使用する有意水準が異なる
    - Bonferroni法は、設定する有意水準を比較のペア数で割って、各比較における有意水準を算出
    - Sidak法は、次の式を用いて、各比較における有意水準を算出
      - α：設定する有意水準
      - h：比較のペア数
      - α_i：各比較における有意水準

$$\alpha_i = 1-(1-\alpha)^{1/h}$$

SPSSにあるもので 紹介しなかった検定方法（1）

- HochbergのGT2
  - 保守的になり過ぎるというボンフェローニの方法を、検定力を高めるために改良したもの
  - Tukeyの方法と類似しているが、個々で用いるのはスチューデント化された最尤法
  - 一般的にはTukeyの方が強力
- Gabriel の方法
  - スチューデント化された最大法（最尤法？）を使用
  - 一般的に、セルの大きさが均等でない場合は、HochbergのGT2より強力。セルの大きさのばらつきが大きい場合には、 Gabriel の方法の方が公平になることがある。

SPSSにあるもので 紹介しなかった検定方法（2）

- Waller-Duncanの方法
  - T統計量に基づいた多重比較検定
  - ベイズ派アプローチを使用
- 等分散性を仮定しない多重比較は（小野寺・山本, 2004, pp52-53)を参照

まとめ（青：SPSSにあるもの、茶：使用の勧められないもの）

適用状況	手法の特徴		検定方法
全ての群間の対比較	パラメトリック	シングルステップ	Tukey, Bonferroni, HochbergのGT2, Gabriel, Sidak, LSD, Scheffe
		ステップワイズ	Tukey-Welsch (R-E-G-WのQ, F), オリジナルRyan, Duncan, Student-Newman-Keuls
	ノンパラメトリック	シングルステップ	Bonferroni , Sidak
		ステップワイズ	Tukey-Welsch, オリジナルRyan
対照群との対比較	パラメトリック	シングルステップ	Bonferroni, Dunnett , LSD
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Bonferroni, Sidak
		ステップワイズ	
対比についての検定	パラメトリック	シングルステップ	Scheffe, Bonferroni, Sidak, LSD
		ステップワイズ	
	ノンパラメトリック	シングルステップ	Scheffe, Bonferroni, Sidak
		ステップワイズ	

参考文献

- 永田靖・吉田道弘 (1997) 統計的多重比較法の基礎, サイエンティスト社
- 森敏昭・吉田寿夫 (1990) 心理学のためのデータ解析テクニカルブック, 北大路書房
- 石井秀光 (2005) 統計分析のここが知りたい, 文光堂
- 小野寺孝義・山本嘉一郎(2004) SPSS事典BASE編, ナカニシヤ出版
- Tabachnick, B. G., & Fidell, L. S. (2001) Computer-Assisted Research Design and Analysis. Allyn & Bacon.
- Fidell, L. S. (2000) Discovering Statistics: using SPSS for Windows. SAGE Publications.
- 狩野裕(2004) 集中講義「心理・教育測定特論」講義資料
- spssときどき記
  - http://www.ec.kagawa-u.ac.jp/~hori/spss/tokidoki3.html