入門 統計学　第9章

多重比較法

『入門　統計学　第2版　ー検定から多変量解析・実験計画法・ベイズ統計学までー』（オーム社）

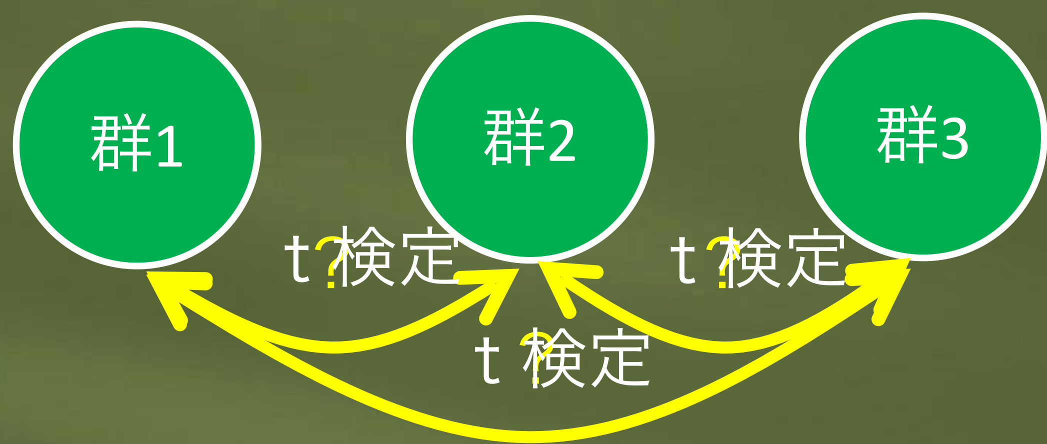
※注：本書を購入された方へのサービスですので，教科書指定（参考図書は不可）していない授業での使用はお控えください。

検定の多重性
（本章で学ぶこと）

- 内容：同じデータを使って検定を複数実施すると，どれかは有意になりやすくなること
- 場面：多群あるときに，どの2群間に差があるのかを見つけたいときなどに発生しやすい
- 対処：簡単には有意にならないように厳しく調整しておく必要がある（多重比較法）
- 注意：多重比較法を複数回実施するとやはり多重性が発生する

多重比較

分散分析で要因効果があるとわかっても…



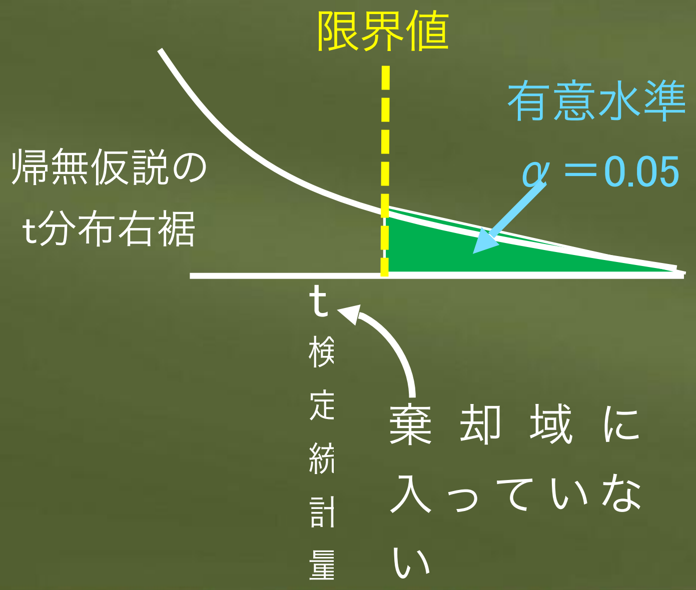
どの対（2群間）に差があるかはわからない…

一対ずつt検定を実施（多重比較）すれば良い?

多重比較は有意水準を増大させる

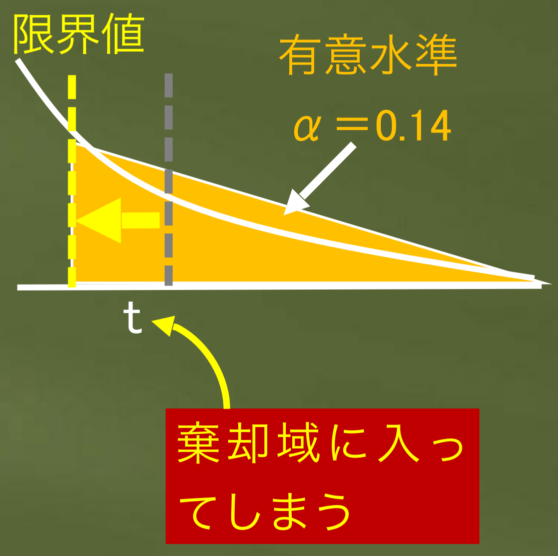
（検定の多重性）

個別の検定



3回繰り返すと…

全体としてみた一番甘い検定



多重性が発生する理由

- 検定では，第一種過誤を犯さない確率（1-α）を重視
- 検定をn回行うということは，全体でn個の帰無仮説が同時に棄却されることが求められる
- （同時は確率論では乗算であるため）検定をn回行うということは，個別には第一種過誤を犯さない確率を（1-α）としていても，全体ではそのn乗（1-α）^{n}となる
- 有意水準αはその補数であるため，$1-(1-\alpha)^{n}$となり，当初設定しておいた厳しさ（α）よりも大幅に甘くなってしまう

→α=0.05の検定3回で$1-(1-0.05)^{3}=0.14$と3倍近くになる

宝くじの例で解説

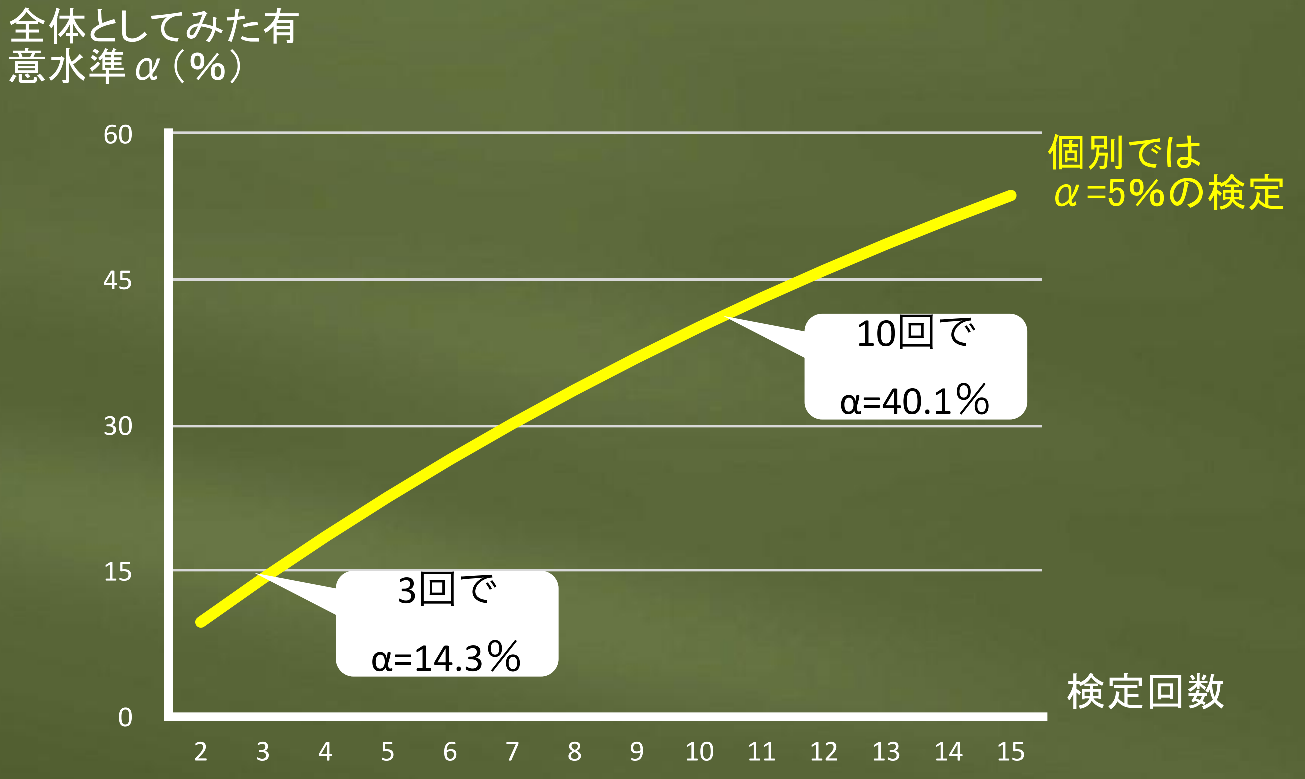
問：なぜ宝くじを複数枚買うの？

答：1枚1枚はなかなか当たらなく（厳しく）ても，複数買えばどれかは当たりやすくなる

検定をたくさん実施して，有意となったのがあれば（論文で）取り上げようとするならば（有意差探し），検定の数を増やすほど，どれかは有意となりやすくなる

→個別には厳しい検定でも全体では甘くなっている

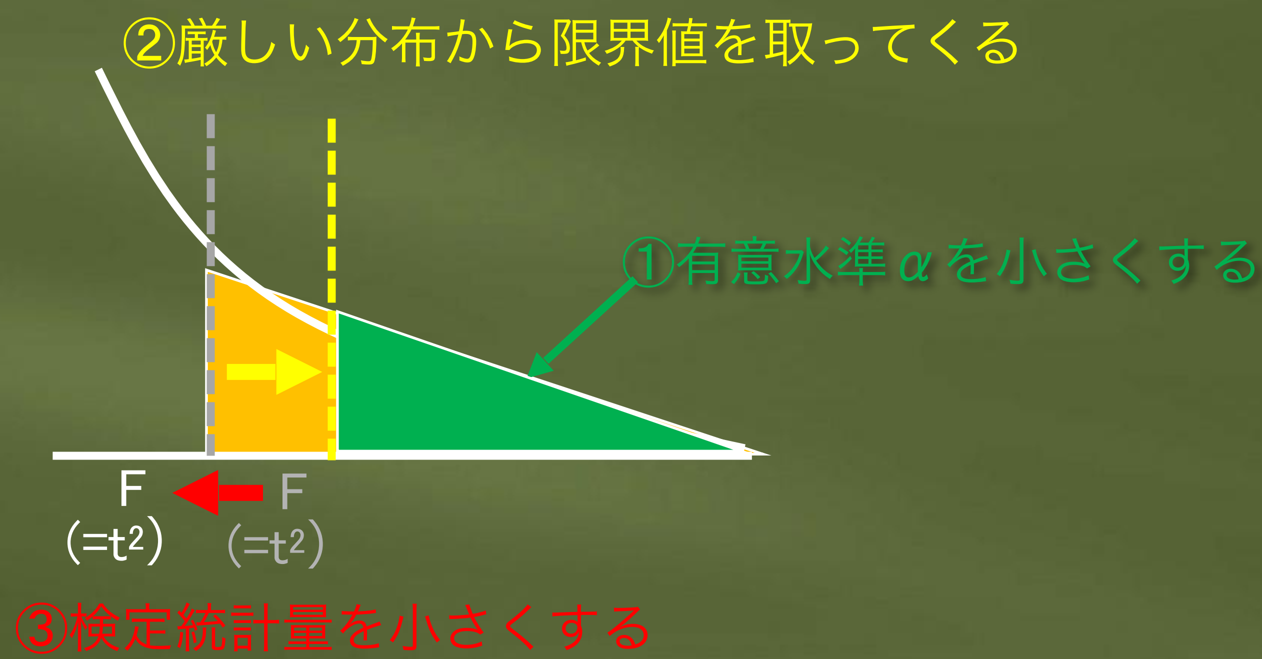
検定回数と有意水準の増え方



多重比較法

(3種類の多重性調整法)

解決策：帰無仮説が棄却され難くなるよう，検定を厳しく調整
（調整法は大きく分けて3種類）



調整型別多重比較法

多重性の調整法

- ①有意水準調整型：Bonferroni, Sidak, P-E-G-WのF(Q), TamhaneのT2など
- ②分布調整型：Tukey(-Kramer), Hochberg, Gabriel, Dunnett(T3, C)など
- ③検定統計量調整型：Scheffeなど

※注：各調整型ごとに色々な手法（計20種類以上）がありますが，以降では黄色の3手法を解説します。

ボンフェローニ

9.2　Bonferroni法（有意水準調整型）

α=5％の検定を3回繰り返す事例

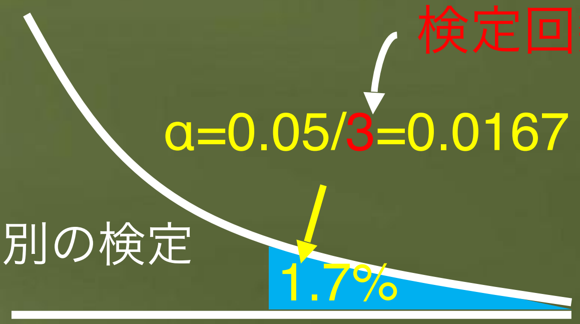
あらかじめ繰り返す数で割った有意水準で検定すれば…

検定を繰り返しても，有意水準が当初想定した基準内に収まる

検定回数

α=0.05/3=0.0167

個別の検定



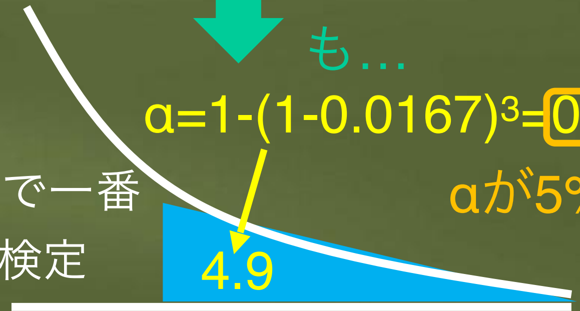
厳しいαで検定

検定を繰り返しても…

α=1-(1-0.0167)^3=$\boxed{0.049}$

αが5%に収まる

全体で一番甘い検定



Bonferroni法の長所と短所

- 長所①：有意水準αを小さくするだけなので，ノンパラなど，どのような検定にも適用可
- 長所②：有意水準αの値を操作できるソフトウェア（分析ツールなど）ならば利用可
- 欠点：厳しく調整し過ぎて検出力が落ちやすいため，5〜6回までの検定（t検定ならば4群程度）にしておく（ただしHolm法やSidak法など改良型も開発されている）

多重比較の事例①

飼料添加物と肉牛成長速度（g/日）

処理	添加物無N	添加物A	添加物B	添加物C
	470	520	510	530
	480	510	530	570
	490			550
平均	480	515	520	550

対応のない一元配置分散分析
（Excel分析ツール，α=0.05）

分散分析: 一元配置

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	7390	3	2463.3	11.82	0.006	4.76
グループ内	1250	6	208.3			
合計	8640	9				

要因効果あり＝飼料添加物を変えたことによって成長速度が変化したことが有意水準5%で統計的に確認できた。

どの群間（飼料間）で差があるのかを特定するため多重比較を実施してみよう

※注：多重比較は分散分析とセットの必要はありません。

多重比較の事例②
（分析ツールを使ったBonferroni法）

t検定を6回実施した場合（多重性調整前）

	添加物A	添加物B	添加物C
添加物無N	5%有意	5%有意	1%有意
添加物A		有意でない	有意でない
添加物B			有意でない

3対に有意差あり？

N−A群間のt検定（Bonferroni法）

t 検定: 等分散を仮定した 2 標本による検定

入力元

変数 1 の入力範囲(1): $B$2:$B$5

変数 2 の入力範囲(2): $C$2:$C$5

仮説平均との差異(Y):

☑ ラベル(L)

α(A): 0.00833

0.05を6で割る

Bonferroni法で多重性を調整後

	添加物A	添加物B	添加物C
添加物無N	有意でない	有意でない	5%有意
添加物A		有意でない	有意でない
添加物B			有意でない

有意差があるといっても良いのはN−C間だけだった

t−検定: 等分散を仮定した2標本による検定

	添加物なしN	添加物A
平均	480	515
分散	100	50
観測数	3	2
プールされた分散	83.33	
仮説平均との差異	0	
自由度	3	
t	−4.2	
P(T<=t) 片側	0.01	
t 境界値 片側	4.86	
P(T<=t) 両側	0.02	
t 境界値 両側	6.23	

差があるとはいえない

2群しかなかったことにすれば良い？

疑問：添加物Bと添加物Cの処理群はなかったことにすればN−A間も統計的に有意となるのでは？

添加物無N	添加物A	添加物B	添加物C
470	520	510	530
480	510	530	[illegible]
490			550

無かったことにすれば良い？

やっと有意になった！

答え：（行き当たりばったりではなく）事前に立てた計画に従って観測や分析を行うのが実験です。
結果を見てから都合の良いようにデータを削除することは“ごまかし行為”です。

9.3　Tukey法（分布調整型）

テューキー

- 全ての対を比較する多重比較専用の手法
- 検定回数が多くても検出力が低下しないため，もっともよく使われる（迷ったらこれ）
- ソフトウェアが必要（手計算は面倒）
- 2群の標本サイズがアンバランスでも使えるようにKramer氏が改良したため，Tukey-Kramer法とも呼ばれる

検定統計量

（注：Tukey専用ではありません）

普通のt検定の統計量
（対応のない群1と群2の平均の差）

$$t_{\bar{x}_1-\bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

対象となる2群の分散のみが影響

2群の分散（加重平均）

多重t検定の統計量
（Tukey法以外の多重比較でも利用）

$$t_{\bar{x}_1-\bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}_e^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

対象となる2群以外の処理群の影響も受ける

全群の分散（F値の分母である誤差分散）

群数が増えたり，他の群の変動が大きくなったり，他の群の反復数が少なくなった場合には検定統計量が小さくなり保守的な検定結果となる

事例の検定統計量を計算

Bonferroni法と同じ肉牛飼料の事例

添加物無N	添加物A	添加物B	添加物C
470	520	510	530
480	510	530	570
490			550
480	**515**	**520**	**550**

分散分析: 一元配置

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	7390	3	2463.3	11.82	0.006	4.76
グループ内	1250	6	208.3			
合計	8640	9				

誤差分散

N-A間の検定統計量 $t_{\bar{x}_N-\bar{x}_A} = \dfrac{\bar{x}_N - \bar{x}_A}{\sqrt{\hat{\sigma}_e^2\left(\dfrac{1}{n_N}+\dfrac{1}{n_A}\right)}} = \dfrac{480-515}{\sqrt{\boxed{208.33}\left(\dfrac{1}{3}+\dfrac{1}{2}\right)}} = \boxed{-2.66}$

t検定では，t分布自身から限界値を読み取って検定統計量と比較するが…

限界値を別の分布から取ってくる

（↑全く別というわけでもありません）

Tukey法では，スチューデント化された範囲の分布から限界値を取ってくる

全ての対の中で最大となる平均の差（範囲）を不偏標準誤差で割った統計量qが従う分布

（大きい値から小さな値を引くので正値になる）

最大の群平均　　最小の群平均

スチューデント化※された範囲

$$q = \frac{\bar{x}_{\max} - \bar{x}_{\min}}{\sqrt{\hat{\sigma}_e^2 \frac{1}{n}}}$$

不偏標準誤差

1群あたり（2群に同じ数を想定）

従う

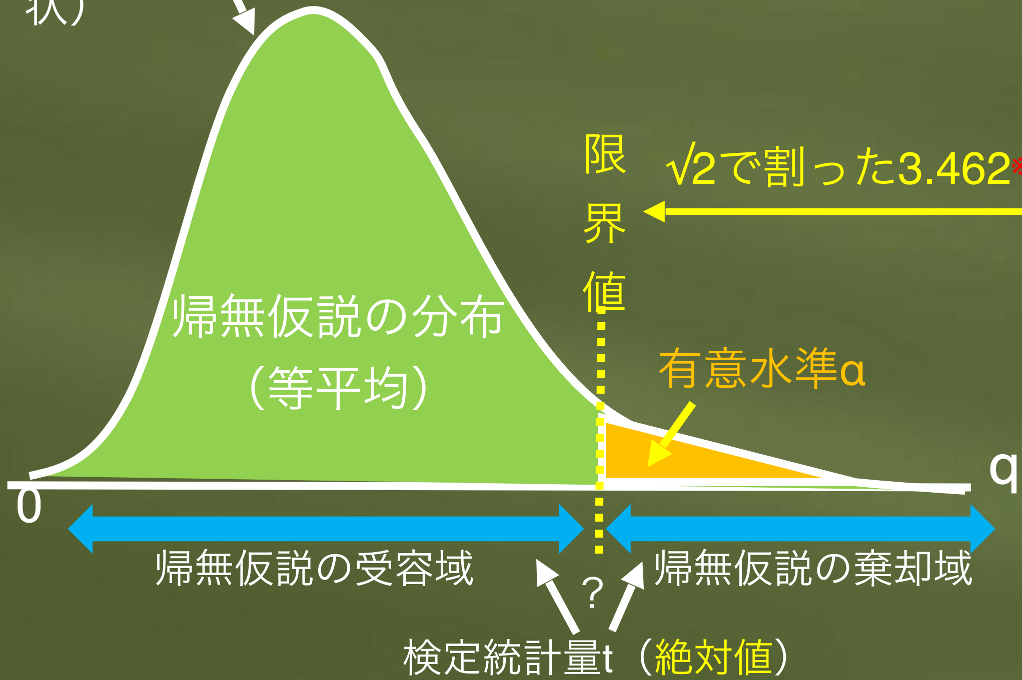
対の数が増えるほど限界値が厳しくなる（＝多重性の問題を逆手に取る）

検定統計量が最大になる対の統計量の分布を他の対の限界値にも用いれば，（他の対にとって）最も厳しい検定となる

※注：不偏標準誤差で割ることをスチューデント化（準標準化）という

スチューデント化された範囲（q）の分布

自由度vによって形状は変化（t分布の負の下側半分が上側に折り重なるのでF分布の形状）



q分布表（上側5%）（全体は付録VI）（v：データ総数-群数，j：群数）

v \ j	2	3	4	5
2	6.085	8.331	9.798	10.881
3	4.501	5.910	6.825	7.502
4	3.927	5.040	5.757	6.287
5	3.635	4.602	5.218	5.673
6	3.460	4.339	4.896	5.305

肉牛飼料の事例

※注：アンバランスでも使えるようにKramerが統計量を修正した（次掲）

qを√2で割った値を限界値とする理由

（同サイズの場合を考える）

	検定統計量	有意条件	限界値
Tukeyが最初に考えた2群が同標本サイズ(n)の場合の検定統計量	$\dfrac{\lvert\bar{x}_1-\bar{x}_2\rvert}{\sqrt{\hat{\sigma}_e^2\dfrac{1}{n}}}$	$>$	$q=\dfrac{\bar{x}_{\max}-\bar{x}_{\min}}{\sqrt{\hat{\sigma}_e^2\dfrac{1}{n}}}$

√2で割った値

比較する限界値も√2で割らなければならない

Kramerがアンバランスな場合でも使えるように変形した検定統計量で2群が同サイズ($n_1=n_2=n$)の場合

$$\frac{\lvert\bar{x}_1-\bar{x}_2\rvert}{\sqrt{\hat{\sigma}_e^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}=\frac{\lvert\bar{x}_1-\bar{x}_2\rvert}{\sqrt{\hat{\sigma}_e^2\dfrac{2}{n}}} > \frac{q}{\boxed{\sqrt{2}}}=\frac{\bar{x}_{\max}-\bar{x}_{\min}}{\sqrt{\hat{\sigma}_e^2\dfrac{2}{n}}}$$

仮説の判定

以下の場合に対象2群間で統計的に有意差があると判定

$$\left|\ \text{多重t検定の統計量}\ \right| > \frac{q}{\sqrt{2}}$$

qは必ず正となるため検定統計量も絶対値

q分布表から読み取ってきた限界値

肉牛飼料の事例をTukey法で多重比較した結果

	添加物A	添加物B	添加物C
添加物無N	2.66 (0.129)	3.04 (0.082)	5.94** (0.004)
添加物A		0.35 (0.984)	2.66 (0.129)
添加物B			2.28 (0.204)

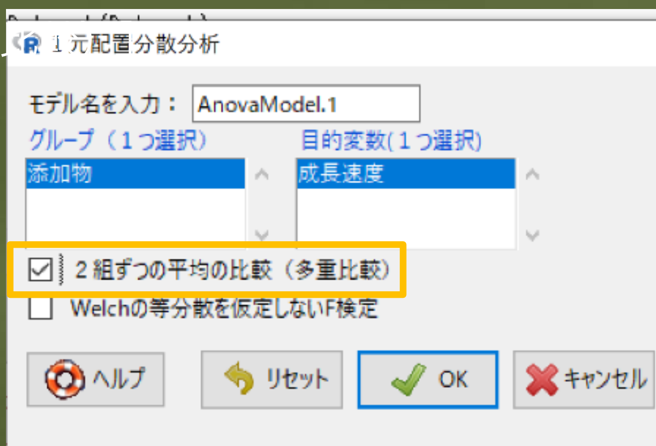
注：上の段が検定統計量で下段括弧内がp値，
**は1%水準で有意を示す。

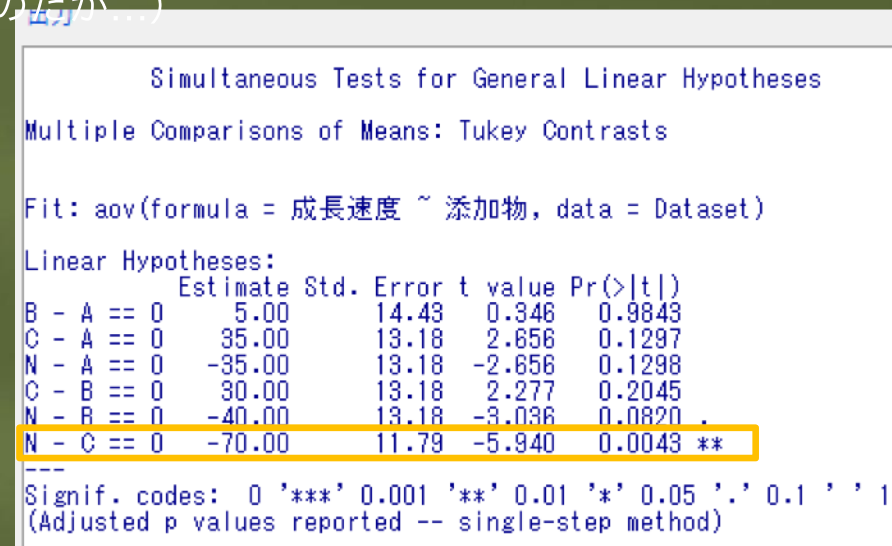
成長速度に統計的な有意差があると言えるのは添加物Nと添加物Cの対のみだった（Bonferroni法と同じ結果）

Rコマンダーによる Tukey法

［統計量］→［平均］→［1元配置分散分析］の［2組ずつの平均の比較（多重

このように，多重比較法を分散分析のメニュー内においてあるソフトが多い（セットの必要は無いのだが…）





```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = 成長速度 ~ 添加物, data = Dataset)

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
B - A == 0     5.00      14.43   0.346   0.9843
C - A == 0    35.00      13.18   2.656   0.1297
N - A == 0   -35.00      13.18  -2.656   0.1298
C - B == 0    30.00      13.18   2.277   0.2045
N - B == 0   -40.00      13.18  -3.036   0.0820 .
N - C == 0   -70.00      11.79  -5.940   0.0043 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

注：データはオーム社HPから入手可
（多重比較法.RData）

N-C群間のみ1%水準で有意であることがわかる
（なお，Rコマンダーの多重比較法はTukey法だけです）

9.4　Scheffe法（検定統計量調整型）

シェッフェ

- 2群の対を多重比較するのではなく，複数群を2グループに分けて比較する（対比）
  →探索的な検定が可能
- 検定統計量を“群数-1”で割って厳しく調整
  →ノンパラ検定など多重比較以外にも汎用可（本書では使わない）
- 検定統計量が分散分析と一部共通している
  →結果が整合するため，セットで使用可

たいひ

対比の設定

対比(Contrast)：群jの母平均μ_jと対比係数c_jの積を全群足し合わせたもの

$$C=\Sigma c_j\mu_j \quad ただし，\Sigma c_j=0$$

↓3群に設定する帰無仮説と対比の事例
（群1が偽薬で群2と3が真薬とか，群2と3の差が小さい場合など）

群1の母平均

μ_1
$(c_1=1)$

群2の母平均と群3の母平均の平均

μ_2
$(c_2=-\frac{1}{2})$

μ_3
$(c_3=-\frac{1}{2})$

$$H_0: \mu_1=\frac{\mu 2+\mu 3}{2}$$

$$対比C=c_1\mu_1+c_2\mu_2+c_3\mu_3=\mu_1-\frac{\mu 2}{2}-\frac{\mu 3}{2}$$

$=0$

群1の対比係数を1，群2・3を−1/2とすればC=0

どんな2グループ比較でも
帰無仮説が対比C=0となることが便利

↓（Tukey法のように）2群間の対を比較することも可能

群1の母平均と群2の母平均

比較しない群の対比係数を0にする

μ_1 (c_1=1) ⟷ μ_2 (c_2=-1)　μ_3 (c_3=0)

H_0: $\mu_1=\mu_2$

対比$C=c_1\mu_1+c_2\mu_2+c_3\mu_3=\mu_1-\mu_2=0$

群1の対比係数を1，群2を−1，群3を0とすれば，やはりC=0

対比係数次第で興味のある対比をいくらでも設定でき，どの帰無仮説もC=0となるため，どの対比においてもC=0さえ棄却できれば良い
→対比較のみの多重比較法よりも検定数が多くなるため検出力は弱い
→ソフトウェアのScheffe法は対比較しか対応していない

検定統計量

帰無仮説が誤っている(Cが0でない)ときに大きくなる検定統計量を考え，
限界値と比較する（限界値か検定統計量のどちらかで多重性を調整）

検定統計量　限界値

対比の不偏推定量

たくさん設定した対比の最大値
（シュワルツの不等式より）

H_0が棄却される条件:
（限界値調整パターン）

$$\frac{\left(\sum c_j \bar{x}_j\right)^2}{\hat{\sigma}_e^2 \sum (c_j^2 / n_j)} > \boxed{(j\text{-}1) \times F}$$

不偏誤差分散

限界値をいちいち修正しなければならない

検定統計量も限界値も“j−1”で割る

Scheffe法の検定統計量

H_0が棄却される条件:
（検定統計量調整型）

$$\boxed{F = \frac{\left(\sum c_j \bar{x}_j\right)^2 / (j-1)}{\hat{\sigma}_e^2 \sum (c_j^2 / n_j)}} > F$$

分子の自由度:j−1
分母の自由度:データ総数−j

F分布表から限界値を直接読める

検定統計量で多重性を調整

2群の場合を考えると，検定統計量Fがt^2と等しくなる（2群の平均の差の検定統計量となる）ことが確認できる

群数が増えるほどF値が小さくなる

$$F = \frac{\left(\sum c_j \bar{x}_j\right)^2 / \boxed{(j-1)}}{\hat{\sigma}_e^2 \sum (c_j^2 / n_j)} \xrightarrow[c_1=1, c_2=-1]{2群の場} \frac{(\bar{x}_1 - \bar{x}_2)^2}{\hat{\sigma}_e^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t_{\bar{x}_1 - \bar{x}_2}^2$$

検定統計量Fを“群数−1”で割ることで多重性を調整している

まとめ：限界値を毎回修正するのは面倒なので，検定統計量を“群数−1”で小さくすることで，検定作業を簡易化している

仮説判定の事例

（2群対の多重比較の場合）

肉牛飼料の例題でN−A群間を比較

添加物無N	添加物A	添加物B	添加物C
470	520	510	530
480	510	530	570
490			550
480	**515**	**520**	**550**

Scheffe法の結果（上段がF値，下段がp値）

Scheffe法	添加物A	添加物B	添加物C
添加物無N	2.35 (0.171)	3.04 (0.112)	11.76** (0.006)
添加物A		0.04 (0.988)	2.35 (0.171)
添加物B			1.73 (0.264)

N−C群間のみ1%有意

$$F=\frac{\left(\sum c_j\bar{x}_j\right)^2/(j-1)}{\hat{\sigma}_e^2\sum\left(c_j^2/n_j\right)}=\frac{(1\times\bar{x}_N-1\times\bar{x}_A)^2/(j-1)}{\hat{\sigma}_e^2\left(\frac{1^2}{n_N}+\frac{(-1)^2}{n_A}\right)}=\frac{(480-515)^2/(4-1)}{208.33\left(\frac{1}{3}+\frac{1}{2}\right)}=\boxed{2.35}$$

棄却できない

分子の自由度（群数j−1），分母の自由度（データ総数−j）のF値（5%のとき4.76）と比較

仮説判定の事例

（2群の対でない対比を設定した場合）

Tukey法では不可能な，添加物無と添加物有
の2グループの対比（N−A＆B&C）を比較

添加物無N	添加物A	添加物B	添加物C
470	520	510	530
480	510	530	570
490			550
480	**515**	**520**	**550**
$c_1=1$	$c_2=-\frac{1}{3}$	$c_3=-\frac{1}{3}$	$c_3=-\frac{1}{3}$

$$F=\frac{\left(\sum c_j\bar{x}_j\right)^2/(j-1)}{\hat{\sigma}_e^2\sum(c_j^2/n_j)}$$

$$=\frac{\left(1\times480-\frac{1}{3}\times515-\frac{1}{3}\times520-\frac{1}{3}\times550\right)^2/(4-1)}{280.33\left(\frac{1^2}{3}+\frac{\left(-\frac{1}{3}\right)^2}{2}+\frac{\left(-\frac{1}{3}\right)^2}{2}+\frac{\left(-\frac{1}{3}\right)^2}{3}\right)}$$

$$=\boxed{5.77}$$

帰無仮説は棄却できる
（母平均に差はある＝添加
物の成長促進効果あり）

自由度は先ほど同じため，5%有意水準の限界値は4.76

注意点

- 多様な対比を設定することで，探索的な検定が可能だ（実験結果を見てから設定してもよい）が，一般のソフトでは対比較しか対応していない
- 単なる対比較ならばTukey法の方が検出力が高い
- 分散分析とは結果が整合しているため，Scheffe法だけは分散分析とセットで使用可（他の手法は結果が異なることがあり，多重性の問題が生じるため，どちらかだけを用いる）

9.5　いろいろな多重性
有意差探しが多重性を生む

- 多重比較以外でも検定の多重性は発生！
- 代表的な事例：
  - ・多重比較法（Tukey法など）を繰り返す←多いので次掲
  - ・分散分析を繰り返す
  - ・相関行列表でたくさんの無相関の検定結果を掲載する
  - ・時系列データにおいて多時点で検定を繰り返す
  - ・異なる検定をいろいろ試す
  - ・水準を分割・統合する
  - ・データを新たに追加して再度検定する

多重比較法を繰り返すことで
多重性が発生している事例

属性（項目）別旅行回数の仮想データ

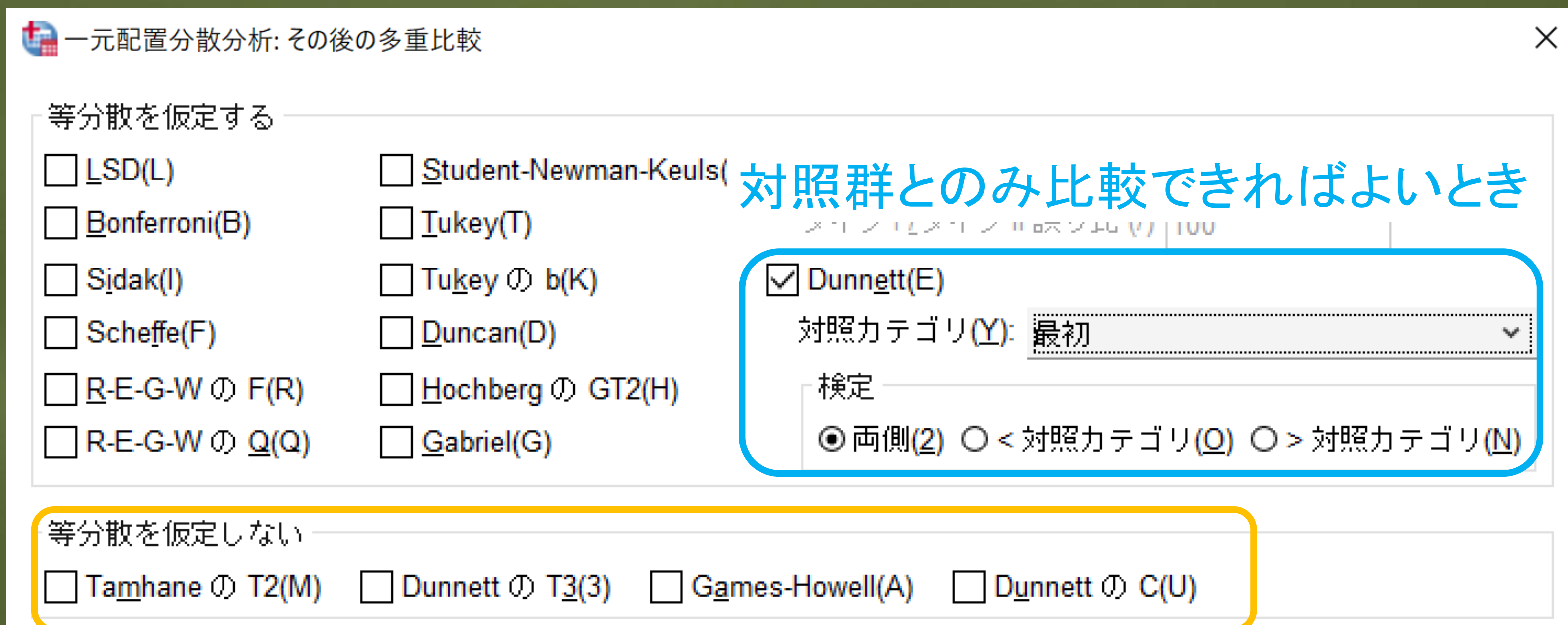
項目	水準	平均旅行回数/月	Tukey法
年齢	30歳未満 a	0.5	ac*, ad**
	30歳代 b	1.0	
	40歳代 c	1.5	
	50歳以上 a	2.0	
所得	300万未満 a	0.3	ad**, bd*, cd*
	～600万未満 b	1.0	
	～900万未満 c	1.0	
	900万以上 d	4.0	
職業	会社員 a	2.0	bc**
	主婦 b	3.0	
	学生 c	0.2	

属性内では多重性が発生しないようにTukey法を用いているが，Tukey法自体を繰り返すことで多重性が発生している。

対処法：Tukey法を行う際に有意水準αを検定回数で割る（Bonferroni法）

目的に沿った使い分け方
（SPSSのメニューを使って）

SPSS27の多重比較法のメニュー画面



群間に等分散を前提できないとき

多重比較法の一覧

	高検出力 ⟷ 弱検出力				
対象群とのみの比較	Williams	Dunnett	Holm	Scheffe	Bonferroni
異分散	TamhaneのT2	Games-Howell	DunnettのT3(C)		
対応あり	Sidak	Holm	Bonferroni		
順位データ(ノンパラ)	Shirley-Williams	Steel(-Dwass)	Holm	Scheffe	Bonferroni

※SPSSを使う場合の注意：Dunncan, Waller-Dunncan, Student-Newman-Keulsの3つは，多重性が調整されていないため使用しないこと。

多重比較法が適さない場面

- 第二種の過誤が致命的なとき

→事例：投与量を数段階設定して，どこまでが無毒かを見つけようとする場合

→理由：多重比較法を用いると，毒性が発揮される最低投与量を見逃してしまう

→対応：普通のt検定や多重t検定を繰り返す

以上で第９章は終了です。