

### 第3章 t検定 (pp. 33-42)

#### §3-1 統計的検定

● 統計的検定とは、  
設定した仮説を検証する場合に、仮説に基づいて集めた標本を、確率論の観点から分析・検証すること。  
使用する標本は、母集団から無作為抽出されたものでなければならない。

● パラメトリック検定とノンパラメトリック検定  
パラメトリック検定は母集団が正規分布に従う間隔尺度あるいは比率尺度の連続データを対象とする。  
ノンパラメトリック検定は母集団に特定の分布を仮定しない。名義尺度、順序尺度に使用可能。

表 1. 2群、あるいは3群以上の標本の差を検定する方法

分布	群の数	用いる代表値	データの対応	検定方法
正規分布	2群	平均値	あり・なし	t検定 (第4章)
	3群以上			分散分析 (第5章)
正規分布ではない	2群	中央値・順位など	なし	マン・ホイットニーのU検定 (本章第2節)
			あり	ウィルコクソンの符号付順位和検定 (本章第5節)
	3群以上		なし	クラスカル・ウォリスの順位和検定 (本章第3節)
			あり	フリードマン検定 (本章第5節)

三浦他, (2004) 『英語教師のための教育データ分析入門』 p.61, 大修館書店より引用

#### 3-1-1 統計的検定の手順

##### (1) 仮説の設定

例. **対立仮説 (alternative hypothesis,  $H_1$ )**

「異なった教授法で指導を受けたグループ A とグループ B の英語力に差がある」

**帰無仮説 (null hypothesis,  $H_0$ )**

「グループ A とグループ B の英語力に差がない」

→平均の差がない ( $A-B=0$ ) であることを仮定している。

##### (2) 有意水準の決定

帰無仮説を棄却して対立仮説を採択するかどうかを判断する基準の設定を行う。

この基準が**有意水準(significance level:  $\alpha$ )**=危険率(critical value)であり、通常 5% ( $\alpha=.05$ ) に設定される。

→100回のサンプリング中、5回おこることの確率のこと。

(場合によっては 1%有意水準を設定する場合もある。)

(3) 検定統計量の有意確率にもとづく仮説の採否

データから**有意確率(significant probability, p値)**を求め、有意水準と照合する。

→有意確率とは、データの分析によって得られた統計値が**偶然おこる確率**のこと。

あらかじめ設定した有意確率より低い場合は、帰無仮説を棄却して対立仮説を採択する。

有意確率が5%より高くなった場合には、2群の平均差は偶然に起こる確率の範囲内の差であるとし、帰無仮説を採択し「有意差がなかった」と報告する。

→偶然起きたものではない=比較する過程で与えられた刺激が何らかの影響を与えたと考える。

3-1-2 統計的検定における過誤と問題点

(1) 統計的検定における過誤

表 3.1 統計的検定における過誤

真実	判定	帰無仮説を採択	帰無仮説を棄却し 対立仮説を採択
	帰無仮説が真	正しい判断 $1-\alpha$	第1種の過誤 $\alpha$ (有意水準)
対立仮説が真	第2種の過誤 $\beta$		正しい判断 $1-\beta$ (検定力)

有意水準を5%に設定するという事は、本当は差がないのに誤って差があると判断してしまう (=第一種の過誤: Type I error) 可能性を5%含んでいることを意味する。

→Type I error は有意水準を低く設定 (1%有意水準など) で回避できるが、その場合第二種の過誤 (= Type II error:  $\beta$ )が起きることがある。

→Type II error=「本当は有意差があるのに、有意さがないと判断してしまうこと」

有意確率が  $p=.06$  である場合、帰無仮説を採択するが、サンプルサイズが原因となった可能性がある。(こういった  $.05 < p < .10$  の範囲の場合、有意傾向とみなす場合がある。)

「本当に有意差があり、有意だと判断する」のであれば問題ないが、その判断を行う場合には検定力 (power:  $1-\beta$ )を検出する。

サンプルサイズが小さい場合は、検定力が低く、有意になりにくいいため、検定力分析(power analysis)によって十分な検定力を確保するのに必要なサンプルサイズを求めておくことお必要である。

→検定力分析は R など可能。

```
> power.t.test(n=xx, delta=xx', sd=2.5)
```

で計算できるが、対応するサンプルの数等によって、検定力があるかどうか、基準値が変わってくる。検定力分析のツールとして **iphone/ ipad アプリとして Power Analysis** がある。

(2) 有意性検定の問題点

統計的検定は確率的に有意かどうかを判断するため、**有意性検定(significance testing)**を行う。

→必ずしも有意確率だけに結果の解釈を頼るのはよくない。

①標本が母集団を代表するように無作為抽出ができていないかという、ほとんどの場合、純粋な意味の

無作為抽出は行われていない。

→サンプリングによって結果が変わってくるという誤差を含む。

②有意性検定の結果はサンプルサイズに大きく左右され、サンプルサイズが小さいと検定力右側で第二種の過誤を犯す可能性が高くなる。逆に、サンプルサイズがかなり大きいと、少しの差でも有意になってしまう。

サンプルサイズに左右されにくい、効果の大きさを表す統計量である効果量(effect size: § 3-5 参照)も併せて求めることが推奨される。

### 3-1-3 標本分布

母集団の分布は正規分布を仮定している。

正規分布する性質から導きだされた**標本分布(sampling distribution)**と呼ばれる確率分布がある。

母集団から理論的に無限回ランダム・サンプリングをした場合に求めた統計量が、どのような確率でどのような値をとるのか、1回にサンプリングするサンプルサイズ（正確には自由度, df）別に分布したものである。

$\chi^2$ 値 :  $\chi^2$ 分布(chi-square distribution)、F値 : F分布(F distribution)、t値 : t分布 (t distribution) などがある。

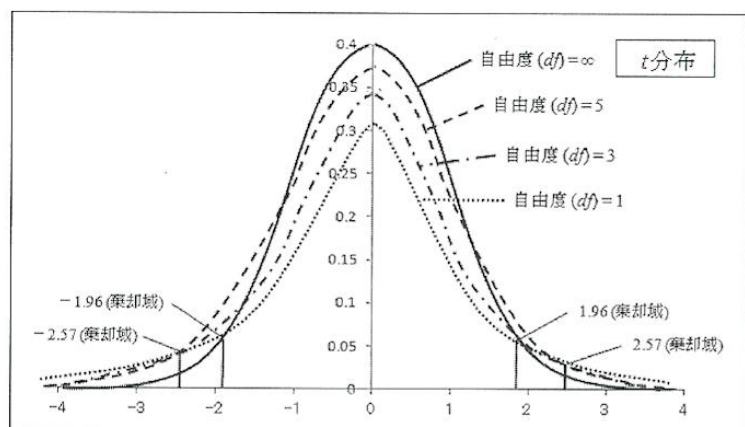
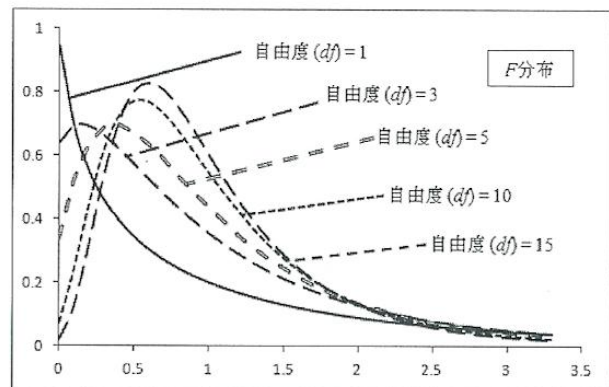
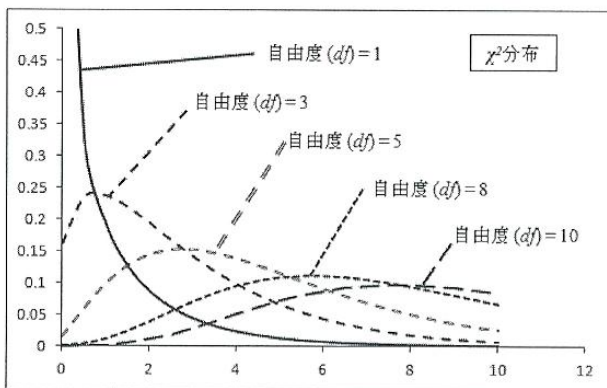


図 3.1  $\chi^2$  分布, F 分布, t 分布  
(横軸はそれぞれ $\chi^2$  値, F 値, t 値,  
縦軸は確率密度を示す)

サンプルサイズが大きくなるほど標準誤差が小さくなり、母集団の真の値（母数）に集中した分布になる。

- ◆  $\chi^2$  分布は標準正規分布から抽出した標本の 2 乗値の分布で、ノンパラ、因子分析、構造方程式モデリングにおけるモデルの適合度の検定などで利用される。自由度 1 のとき、 $\chi^2$  分布の形状は F 分布に類似し、ほかの標本分布同様、自由度が高くなるにつれて、徐々に正規分布の形状に近づく。
- ◆ F 分布は、分散分析などの分散比の検定に用いられる分布で、2 つの異なる正規分布からの標本の 2 乗値の分布を  $\chi^2_\alpha$  と  $\chi^2_\beta$  分布とすると、その比  $\chi^2_\alpha / \chi^2_\beta$  に従う。そのため、この分布は自由度を 2 つ持つことになる。

### 3-1-4 両側検定と片側検定

有意であるかどうかを決定する棄却域（有意水準）は、分布の上側と下側の両方に設定して行う両側検定（two-tailed test）が一般的である。

5% 有意水準で棄却域を設定すると、両側 2.5% ずつ設定されることになる。

有意差が片側一方にしか起こらないと予測がつく場合に、分布の片側だけに基準を設定する片側検定（one-tailed test）がある。これは 5% 有意水準をそのまま片側に設定する。方向性が明らかな場合以外は、通常両側検定を行う。

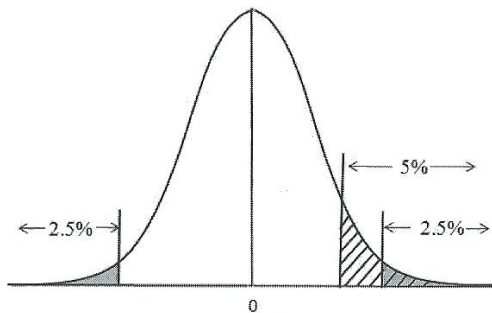


図 3.2 両側検定と片側検定

### § 3-2 t 検定とは

t 分布に照らし合わせて、2 群の平均の差を検証する場合に用いるパラメトリック検定。

例. 中学生と高校生のテレビを見る時間の長さに違いがあるのか

男子生徒のほうが女子生徒より理系科目が強いのか

など、2 群間を比較する際に使用。

平均値の大小だけを比較するのではなく、それぞれの群の得点分散を考慮する必要がある。

#### 3-2-1 t 検定の実験計画と前提

- (1) t 検定では、対応あり (repeated-measures) と 対応なし (independent-measures) の 2 種類の実験計画を立てることができる。

対応ありの検定では、同じ被験者に異なる 2 つの条件を与え、その条件間の差を検討する。

対応なしの検定では、異なる性質をもった被験者に同じ条件を与えて、グループ間を比較する場合と、同じ性質をもった2群に異なる条件を振り分ける方法の2種類がある。

後者のデザインでは、それぞれの群を統制群(control group)と実験群(experimental group)と呼ぶ。

実験群と統制群や男女など、被験者を分ける条件や変数を独立変数 (independent variable) と呼ぶ。また、得点や時間など、独立変数の条件をもとに集めたデータを扱った変数を従属変数(dependent variable)と呼ぶ。

(2) t 検定を使用する際には、以下の前提のもとで分析が行われる。

①データの種類：連続性のある間隔尺度、または比率尺度の量的データであること。

②ランダム・サンプリング：サンプル母集団からランダム・サンプリングされ、母集団を十分代表していること

③正規性：標本平均の分布が正規分布に従うこと。

\*正規分布から少々外れている場合でも、t 検定は正規性に対して頑健(robust)で結果が影響を受けにくいいため、そのまま t 検定を利用する。

対応なし t 検定で加わる前提

④等分散性(homogeneity of variance)

比較する2群のデータ分散が等しいこと。

SPSS で t 検定を行う場合、2つの母集団が等しいことを帰無仮説としたルビーン(Levene)の検定が行われ、この検定で有意でなければ前提を満たしているといえる。

有意だった場合、「等分散性を仮定しない」と表示されるウェルチの方法による結果を参照する。そのほかにコ克蘭・コックスの方法もある。

⑤観測値の独立性：異なった被験者からのデータが独立していること。データがお互いに影響しあい相関が高い場合、第1種の過誤が起きやすくなる。

### 3-2-2 t 検定の設定と t 値の算出

t 検定では、 $t = \frac{\text{観測された標本分散の差 } (\bar{x}_1 - \bar{x}_2)}{\text{標本平均の差の標準誤差}}$  という式が使用され、これは何らかのある効果あるいは原因に

よる標本平均の差がその標準誤差のいくつ分ゼロから離れているかを計算することにより、偶然おこる誤差よりどの程度大きいか調べる。

t 値は「A の標本数 + B の標本数 - 2」で求めた自由度の t 分布に従うので、そこから t 値の偶然に起こる確率を求める。

(1) 対応なし t 検定 (2 群のサンプルサイズが同じ場合)

条件によって被験者が異なる対応なし t 検定の場合、2群は、集団として受けた条件による違いに加えて、個人の性質の違いも誤差として含める。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

(2) 対応なし t 検定 (2 群のサンプルサイズが異なる場合)

異なる人数の 2 群間を比較する場合には、各群のサンプルサイズの違い=を考慮するために、サンプルサイズから 1 を引いた自由度をそれぞれの分散にかけることで、サンプルサイズの大きいほうの値がより大きくなるように重みづけした式を求め、別式に代入し、t 値を求める。

$$\textcircled{1} \quad S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad \textcircled{2} \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

(3) 対応あり t 検定の場合

同じ被験者に 2 条件が割り当てられるので、分子には 1 つの集団の 2 条件の差を置き、分母の差の標準誤差は 1 集団内の 2 条件の分散から求める。

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\frac{S_D}{\sqrt{n}}} \quad (df = n - 1)$$

補足

検定力分析をフリーで行うことのできる HP

- Russ Lenth's and sample-size page (<http://www.math.uiowa.edu/~rlenth/Power/>)  
(t 検定だけでなく、ANOVA などの検定力分析も可能)
- Power Analysis for ANOVA Designs (<http://www.math.yorku.ca/SCS/Online/power/>)  
(ただし多元配置分散分析用に設計されているため、処理には注意が必要)