

仮説検定

2つの差の検定の分類

	標本標準偏差	サンプル数	標本分散	適用分布	注記
対応の有る データ	—	$n \geq 100$	—	z分布	
		$n < 100$	—	t分布	
対応の無い データ	既知	$n_1 + n_2 \geq 100$	—	z分布	
	不明	$n_1 + n_2 < 100$	等分散	t分布	
			非等分散	ウェルチ t 分布	
		$n < 100$	—	t分布	

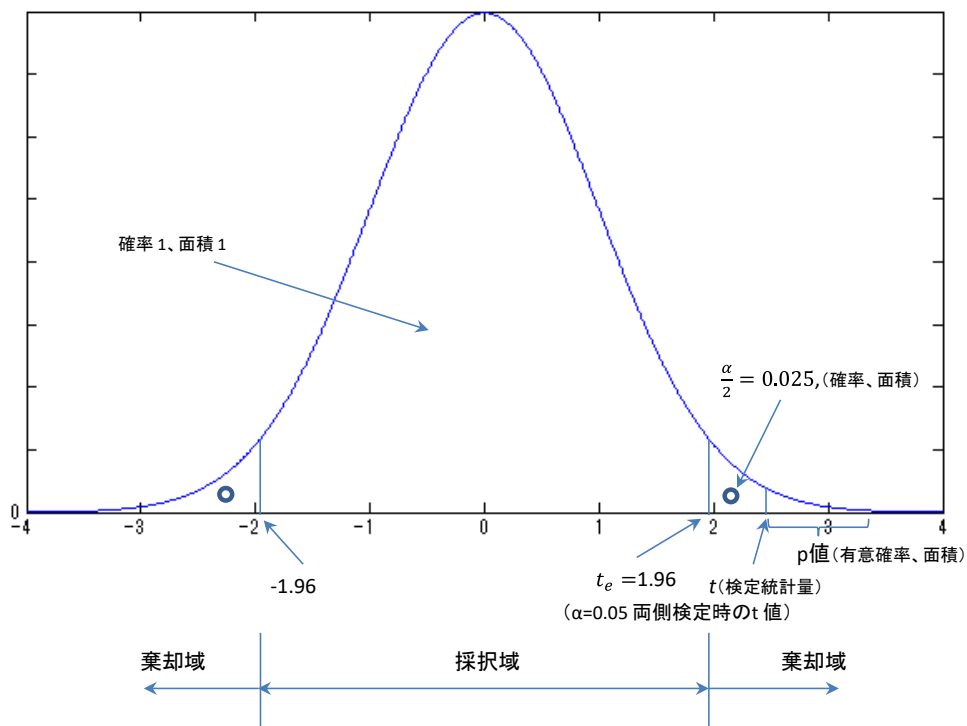
(t 検定は、外れ値が有る場合には適さないと言われています。)

標本：母集団から部分的に抽出されたデータの集まりです。

α ： 有意水準 p ： 有意確率

t 分布 (自由度が100以上になるとz分布に近づく。)

$\alpha=0.05$ 両側検定



グループ統計量

- ① 標本数： データ数 n
- ② 標本の平均値：
データの合計をデータ数 n で割った値
- ③ 標本の標準偏差 s
分散 s^2 の平方根が s

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

自由に選べるデータの個数、自由度は $(n-1)$

(We don't actually have n independent pieces of information about the variance.)

- ④ 標本平均の標準誤差

$$\frac{s}{\sqrt{n}}$$

等分散性の検定

- ① F 値 (検定統計量)

2つの標本分散 s_A^2, s_B^2 を基に、母分散 σ_A^2 と σ_B^2 の等分散性を判定します。

$$F = \frac{s_A^2}{s_B^2}$$

F 分布表から α 、自由度に対応する F_e を算出し、 $F_e < F$ の場合は不等分散と判定します。

- ② 有意確率、p 値 (後述の「p 値について」を参照願う。)

検定統計量の外側の確率を意味します。

2つの母平均の差の検定

- ① H

帰無仮説(2つの平均に差は無いという仮説)ベースで、 $H=1$ では棄却される (有意差が有る)。 $H=0$ では棄却されない (有意差が有るとは言えない)。

- ② 平均値の差

グループ1の標本数 n_1 とグループ2の標本数 n_2 とすると、

グループ1の平均値 \bar{x}_1 とグループ2の平均値 \bar{x}_2 の差

- ③ 標準偏差

$$\text{偏差平方和} : S = \sum(x_i - \bar{x})^2$$

$$\text{標本分散} : s^2 = \frac{\text{偏差平方和}}{(n-1)}$$

グループ1の標本標準偏差 s_1

グループ2の標本標準偏差 s_2

④ 標本平均の標準誤差

グループ 1 の標本平均の標準誤差 $\frac{s_1}{\sqrt{n_1}}$

グループ 2 の標本平均の標準誤差 $\frac{s_2}{\sqrt{n_2}}$

⑤ 信頼区間 (区間推定、two-sample)

例えば、母集団の真の母数が含まれる確率が、一定の確率以上と保証される区間を信頼区間といいます。

対応有り	対応無し	
	等分散 ($\sigma_1^2 = \sigma_2^2$)	非等分散 (Welch の検定)
$\mu_d \sim \bar{d} \pm t\left(f, \frac{\alpha}{2}\right) \frac{s_d}{\sqrt{n}}$ $\bar{d} = \frac{\sum d}{n}$ $\sum d = \sum (x_i - y_i)$ <p> x_i : データ1 y_i : データ2 s_d : 対になったデータの差の標準偏差 f : 自由度 t : t 分布表 </p>	$(\mu_1 - \mu_2) \sim (\bar{x}_1 - \bar{x}_2) \pm t\left(n_1 + n_2 - 2, \frac{\alpha}{2}\right)$ $* \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left\{ \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right\}}$ <p> n_1, n_2 : データ数 μ_1, μ_2 : 母平均 \bar{x}_1, \bar{x}_2 : 標本平均 </p>	$(\mu_1 - \mu_2) \sim (\bar{x}_1 - \bar{x}_2) \pm$ $t\left(f, \frac{\alpha}{2}\right) * \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$ <p>ここに、自由度 f は、</p> $f = \frac{1}{\frac{C^2}{n_1 - 1} + \frac{(1 - C)^2}{n_2 - 1}}$ $C = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

⑥ t 値 (= (標本平均の差) / (標本平均の差の標準誤差))

対応有り	対応無し	
	等分散	非等分散 (Welch の検定)
$t = \frac{\sum (x_i - y_i)}{\frac{n s_d}{\sqrt{n}}}$ <p> ・2つの対応有るデータ x_i と y_i ・s_d : 対応の有るデータの差の標準偏差 </p>	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} - \frac{1}{n_2}} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

対応の有る t 検定では差の標準誤差は小さくなり、対応の無い t 検定では差が大きくなるため標準誤差も大きくなります。

標準誤差が小さくなるという事は、信頼区間の幅が狭くなり、真の値を推測できます。

⑦ 自由度 m

対応有り	対応無し	
	等分散	非等分散 (Welch の検定)
$m = n_1 - 1 = n_2 - 1$	$m = n_1 + n_2 - 2$	$m = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$

多重比較検定の補正

多重比較の問題点は、検定数が増えると type I エラーの確率が増えることです。「表. 2 種類の過誤」を参照願います。

このエラー回避のため、多重比較検定の補正を行います。

fMRI での補正方法 (357 ページ、文献 4) :

①有意水準 $\alpha=0.05$ で t 検定しますと、スクリーンには脳以外の場所でも有意差のポイントが検出されます。これはおかしい検定です (独立した検定数は、voxel 数と対応します)。

②一つの断面が $64 \times 64 \times 34$ (slices) で $140,000$ voxels が有ったとします。刺激が無い状態でも、 $\alpha=0.05$ の場合、 $140,000 \times 0.05$ で 6800 個の仮説検定における第 1 種の誤りが発生する確率があります。

第 1 種の誤りが発生しない確率 p (脳全体での検定で、1 voxel たりとも間違いでない確率) は、

$$p(\text{no Type I error}) = (1 - \alpha)^N \quad \text{ここに、} N \text{ は voxel 数.}$$

例えば、single slice 断面で、 $\alpha=0.01$ で $N=4096$ の場合、確率 p は 1.3×10^{-18} となり、全脳の偽陽性出現率は、ほぼ 100% となります。

③第 1 種の誤りを補正するために、単純な Bonferroni correction を適用しますと、 $\alpha=\alpha/V$ で、例えば、 $0.01/4096=0.000002$ となり、第 1 種の誤りは補正されます。が、逆に第 2 種の誤りの確率が増大します。第 2 種の誤りは、被験者の症状を見落とす深刻な問題となります。

④Bonferroni では、すべての検定が独立していると仮定しています。しかし実際は異なります。実際の image は空間的に相関しています。つまり、しばしば voxels 毎に独立しておらず同じ voxel 族です。そこで SPM では random field theory に基づいた補正 FWE (family wise error) が選択されます。

⑤補正手法として他に、FDR, small-volume correction, permutation, cluster-size thresholding, GRF (Gaussian random field) が有ります。

OEG-16/SpO2 での補正方法 (案)

OEG-16/SpO2 での t 検定評価についても、上記 fMRI の多重検定と同じ問題が存在します。

そこで、その対処として、次の案が有ります。

①Bonferroni をそのまま適用しますと、 $\alpha=\alpha/16$ で、低い厳しい閾値となり、実際に有意差が有る賦活を捉えない場合が有ります。

②Bonferroni の代替案は数多く提案されています。e. g. Dubey/Armitage-Parmar algorithm

③ここでは、1 例として下記手法を提案します。

i) 賦活の無いチャンネルは無視するというスクリーニングをまず行います。

ii) 続いて、有意な賦活が認められたチャンネルを ROI とします。例えば、CH2, 3, 13, 14 の 4 チャンネルに賦活が見られたとした場合、有意水準 $\alpha/4$ という値で検定します。

p 値について

- p 値は inference であって、estimation では無い。
- p 値は、測定データから得られる確率（有意確率）
- α (False positive)値は、経験則から得られる値（有意水準）。（例えば、 $\alpha=0.05$ or 0.01 , - - ）
- β (False negative)値は、経験則から得られる値。
- $(1-\beta)$ の値は、power と言われ、80%か 90%にセットされるところ（文献5 pp76）。
- 信頼区間は信頼度に関係しています。信頼区間が狭いほど信頼度は増します。

例えば、標準正規分布の信頼度 95%の区間は（one sample、両側検定）、

母平均 μ , 母標準偏差 σ , 標本平均 \bar{x} , サンプル数 n 標準誤差 SE

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \rightarrow \bar{x} - 2 * SE < \mu < \bar{x} + 2 * SE$$

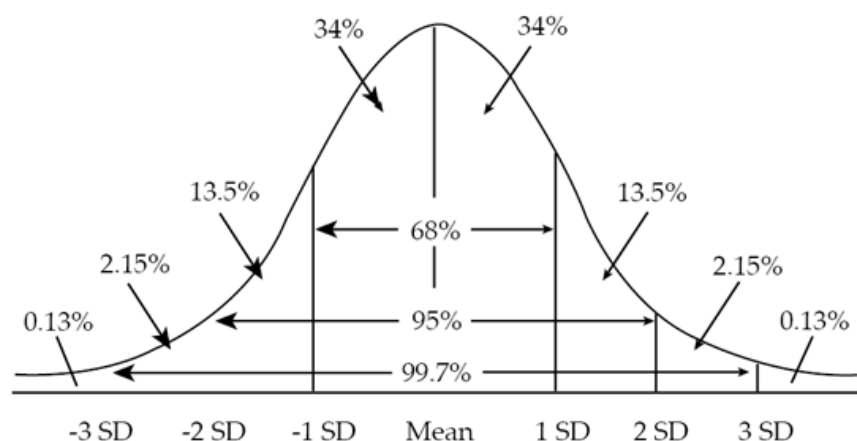


Fig. 1: Normal distribution

(Normal Distribution, "p" Value and Confidence Intervals NJ Gogtay, SP Deshpande, UM Thatte)

この図の SD (Standard Deviation) は SE (Standard Error) の誤り。

- 2つの標本（two sample）の母平均の差の信頼区間は、前の⑤表に記載されています。
- 標準誤差は信頼区間と関係しています。

$$SE = \frac{Variation}{\sqrt{Sample\ size}}$$

精度を上げる一つのやり方は、sample 数を増やすことが、この式より判ります。

仮説検定における 2 種類の過誤

仮説検定は、確率的な事象に基づいているため、誤った判断も有り得ます。2 種類の過誤について下図に説明します。(文献 5)

表. 2 種類の過誤

帰無仮説 H_0 「○○は□□である」

		Truth (仮説の真偽)	
		Null hypothesis false H_1 (active) (偽の仮説)	Null hypothesis true H_0 (inactive) (正しい仮説)
Result of statistical test (検定に拠る判定)	Statistically significant: reject null hypothesis Reject H_0 (active) (仮説を棄却)	True positive $(1 - \beta)$:power	False positive Type I error α (有意水準) 偽陽性
	Not statistically significant: fail to reject null hypothesis Accept H_0 (inactive) (仮説を棄却しない)	False negative Type II error β 偽陰性	True negative $(1 - \alpha)$

(注) この $(1 - \alpha)$ が、多重比較検定の補正の説明と対応します

(注) 陽性とは、例えばインフルエンザの検査を受けて、インフルエンザに感染しているケース

[参考文献]

1. 「入門はじめての統計解析」石村貞夫、東京図書（2013）
2. 「44 の例題で学ぶ統計的検定と推定の解き方」上田拓治、オーム社（2009）
3. 「Modern Statistics for the Life Sciences」,Alan Grafen & osieHails,Oxford ,2002
4. 「Functional Magnetic Resonance Imaging」、Huettel SA,Song AW,McCarthy G,
Sinauer Associates Inc,2nd edition,2009
5. 「What is a p-value」, Andrew Vickers,Addison-Wesley,2010
6. 「SPM12 STARTER' GUIDE」, Erno Hermans