統計手法一覧

質的指標の検定

	対応の有無	検定手法
想定比率との比較		適合度検定①
標本間の比較	対応なし	χ^2検定②
	対応あり	McNemar 検定③

①適合度検定

n回の観測の中で、事象１はn_1回、事象２はn_2回、・・・、事象kはn_k回起こるとする。確率$p_1, p_2, \cdots, p_k$に基づく予想値$m_1, m_2, \cdots, m_k$（$m_i = np_i$）と等しいといえるか、有意水準αで検定する。

$$\chi^2 = \frac{(n_1 - m_1)^2}{m_1} + \frac{(n_2 - m_2)^2}{m_2} + \cdots + \frac{(n_k - m_k)^2}{m_k} \sim \chi^2_{k-1} \text{分布}$$

$p = chidist(\chi^2, k-1)$として、$p < \alpha$ならば、理論値と差があると判定する。

②χ²検定

2×2 表の場合

ある事象の出現、非出現を要因の有無により分けると以下のようになった。
出現、非出現の間に要因の有無による差があるか、有意水準αで検定する。

	出現	非出現	計
要因有り	a	b	$a+b$
要因無し	c	d	$c+d$
計	$a+c$	$b+d$	$a+b+c+d=n$

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2_1 \text{分布}$$

$p = chidist(\chi^2, 1)$として、$p < \alpha$ならば、要因により差があると判定する。

r×s 表の場合

ある事象（r種）の出現状況を要因（s種）により分けると以下のようになる。
出現、非出現の間に要因の有無による差があるか、有意水準αで検定する。

	事象１	事象２	$\cdots$	事象s	計
要因１	x_{11}	x_{12}	$\cdots$	x_{1s}	$x_{1\cdot}$
要因２	x_{21}	x_{22}	$\cdots$	x_{2s}	$x_{2\cdot}$
：	：	：		：	：
要因r	x_{r1}	x_{r2}	$\cdots$	x_{rs}	$x_{r\cdot}$
計	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot s}$	n

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(x_{ij} - x_{i\cdot} x_{\cdot j} / n)^2}{x_{i\cdot} x_{\cdot j} / n} \sim \chi^2_{(r-1)(s-1)} \text{分布}$$

$p = chidist(\chi^2, (r-1)(s-1))$として、$p < \alpha$ならば、差があると判定する。

統計手法一覧

③ McNemar 検定

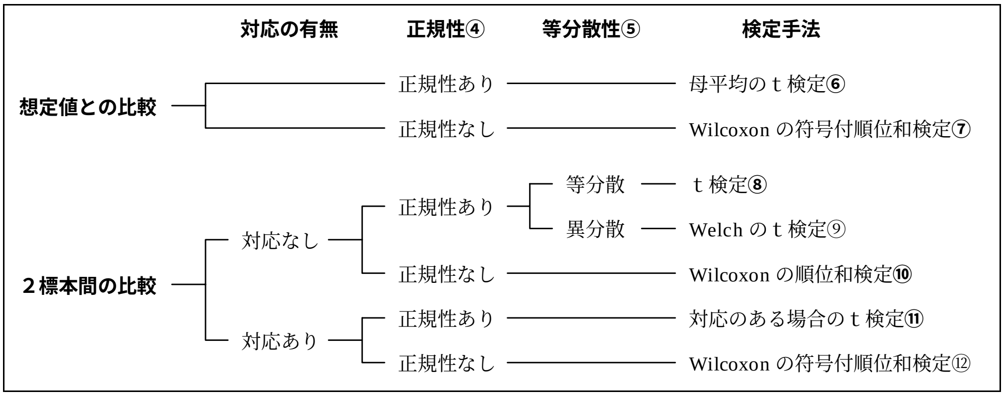
データと対照データとをある条件でマッチさせて、要因の有無で分類したところ以下の表を得た。要因はデータと対照データの違いに影響があると考えられるか、有意水準αで検定する。

	要因あり	要因なし
要因あり	a	b
要因なし	c	d

$$\chi^2 = \frac{(b-c)^2}{b+c} \sim \chi_1^2 \text{ 分布}$$

$p = chidist(\chi^2, 1)$ として、 $p < \alpha$ ならば、要因により差があると判定する。

量的指標の検定



④正規性の検定

ヒストグラム　　省略

正規確率紙的手法

1) データを入力する。（データ数 n）
2) データを小さい順に並べ替える。（[データ－並べ替え]）
3) データに 1 から番号を振る。
4) 累積比率を求める。　$p_i = i/(n+1)$　i は番号
5) $x = normsinv(p)$ 関数を用いて x 値を求める。
6) データと x 値を用いて散布図を描く。
7) グラフに近似曲線を加える。（[グラフ－近似曲線の追加]）
8) 直線に近く並んでいるようなら正規分布

⑤等分散性の検定

正規分布する２つの標本の母分散に差があるかどうか、有意水準αで検定する。

データ数 n_1, n_2, 不偏分散 u_1^2, u_2^2　（$u_1^2 \geq u_2^2$ とする）

$$F = \frac{u_1^2}{u_2^2} \sim F_{n_1-1,\, n_2-1} \text{ 分布}$$

$p = fdist(F, n_1 - 1, n_2 - 1)$ として、 $p < \alpha$ ならば、差があると判定する。

統計手法一覧

⑥母平均のｔ検定

正規分布する標本について、標本の母平均と母集団の平均μとを比較し、差があるかどうか有意水準αで判定する。データ数n，平均$\bar{x}$，不偏分散：u^2

$$t = \frac{\sqrt{n}(\bar{x}-\mu)}{u} \sim t_{n-1} \text{分布}$$

$p = tdist(|t|, n-1, 2)$として、$p < \alpha$のとき差があると判定する。

⑦ Wilcoxon の符号付順位和検定

標本と母集団の中間値μを比較し、差があるかどうか有意水準αで検定する。

$z_i = x_i - \mu$として、$|z_i|$の小さい順に 0 を除いて順位r_iを付け、z_iの正負で 2 群に分ける。（同数値の場合は、順位平均をとる。）それぞれの群の順位和をとり、このうち小さい方を選ぶ。$R = \min(R_r, R_s)$，　$n = r + s$　　データ数n

データ数が少ないとき

数表（$p = \alpha / 2$）を参照し、$R \le R_n(\alpha / 2)$のとき、差があると判定する。

データ数が多いとき

$$z = \frac{|R - n(n+1)/4| - 1/2}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1) \text{分布（正の部分）}$$

$p = 2 \times (1 - normsdist(z))$として、$p < \alpha$のとき、差があると判定する。

⑧ ｔ検定

正規分布する等分散の２つの標本について母平均を比較し、差があるかどうか有意水準αで検定する。データ数 n_1, n_2，平均 $\bar{x}_1, \bar{x}_2$，不偏分散 u_1^2, u_2^2

$$t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1-1)u_1^2 + (n_2-1)u_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1+n_2-2} \text{分布}$$

$p = tdist(|t|, n_1 + n_2 - 2, 2)$として、$p < \alpha$ならば差があると判定する。

⑨ Welch のｔ検定

正規分布する分散の異なる２つの標本について母平均を比較し、差があるかどうか有意水準αで検定する。データ数 n_1, n_2，平均 $\bar{x}_1, \bar{x}_2$，不偏分散 u_1^2, u_2^2

$c = \dfrac{u_1^2/n_1}{u_1^2/n_1 + u_2^2/n_2}$、自由度を　$d = \left(\dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}\right)^{-1}$　（切り捨て）とし、

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{u_1^2/n_1 + u_2^2/n_2}} \sim t_d \text{分布}$$

$p = tdist(|t|, d, 2)$として、$p < \alpha$ならば差があると判定する。

統計手法一覧

⑩ Wilcoxon の順位和検定

一般的分布の２つの標本について母集団の中間値を比較し、差があるかどうか有意水準αで検定する。

両群のデータの小さい順に順位を付ける（同じ値には順位の平均値）。

データ数$n_1 \le n_2$の小さい群の順位合計を$W = \sum_{i=1}^{n_1} r_i$として、

$n_2 \le 20$ の場合

数表を参照し、データ数n_1, n_2の組で$W \le U_{1-\alpha/2}$または$W \ge U_{\alpha/2}$であれば、中間値に差があると判定する。

$n_2 > 20$ の場合

$$z = \frac{|W - n_1(n_1+n_2+1)/2| - 1/2}{\sqrt{n_1 n_2 (n_1+n_2+1)/12}} \sim N(0,1) \text{分布}$$

$p = 2 \times (1 - normsdist(z))$として、$p < \alpha$のとき、中間値に差があると判定する。

⑪ 対応のある場合の t 検定

正規分布し、対応のある２標本の母平均を比較し、差があるかどうか有意水準αで検定する。各標本の差z_i，データ数n，差の平均$\bar{z}$，差の不偏分散u_z^2

$$t = \frac{\sqrt{n}\bar{z}}{u_z} \sim t_{n-1} \text{分布}$$

$p = tdist(|t|, n-1, 2)$として、$p < \alpha$ のとき、差があると判定する。

⑫ Wilcoxon の符号付き順位和検定

一般の分布に従う対応のある２つの標本の母平均を比較し、差があるかどうか有意水準αで検定する。データ数n

対応する各標本の差をz_iにとり、後は⑦の手順に従う。

統計的推定

⑬ 母比率の区間推定

データ数n、比率$\hat{p}$の標本から、母比率pを信頼区間$(1-\alpha)\times 100\%$で推定する。

$$z_0 = normsinv(1-\alpha/2) \text{として、} \quad \hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_0 \le p \le \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_0$$

⑭ 母平均の区間推定

データ数n、平均$\bar{x}$、不偏分散u^2の標本から、正規分布する母集団の母平均μを信頼区間$(1-\alpha)\times 100\%$で推定する。

$$t_0 = tinv(\alpha, n-1) \text{として、} \quad \bar{x} - \frac{u}{\sqrt{n}} t_0 \le \mu \le \bar{x} + \frac{u}{\sqrt{n}} t_0$$