

平均値, 平均値の差の検定

樋口さぶろお <http://hig3.net>

龍谷大学工学部数理情報学科

生活の中の統計技術 L08(2018-11-19 Mon)

最終更新: Time-stamp: "2018-11-19 Mon 10:04 JST hig"

今日の目標

- Excel で平均値の区間推定ができる
- 統計的仮説検定の考え方を説明できる
- Excel で平均値の統計的仮説検定ができる



L07-Q1

Quiz 解答:母比率の区間推定

A 候補に投票したを $X = 1$, しなかったを $X = 0$ とする.

- ① 標本比率は $\hat{p} = \frac{35}{50} = 0.7$. 母比率 p を 0.7 と推定する.
- ② X の母分散は $0.7 \times (1 - 0.7) = 0.21$ と推定する.
母比率 p の信頼係数 $1 - \alpha = 0.95$ の信頼区間は,

$$0.7 - 1.96 \times \sqrt{\frac{1}{50} \cdot 0.21} < p < 0.7 + 1.96 \times \sqrt{\frac{1}{50} \cdot 0.21}$$

$$0.7 - 0.13 < p < 0.7 + 0.13$$

$$0.57 < p < 0.83$$

信頼係数 0.95 では当選ってことですね (放送用語「当選確実」で, 後であやまらなきゃいけない確率は 0.05 以下).

- ③ 母比率 p の信頼係数 0.99 の信頼区間は,

$$0.7 - 2.58 \times \sqrt{0.0042} < p < 0.7 + 2.58 \times \sqrt{0.0042}$$

$$0.7 - 0.17 < p < 0.7 + 0.17$$

$$0.53 < p < 0.87$$

信頼係数 0.99 のほうが慎重な判断基準ですが, それでも当選ってことですね.

L07-Q2

Quiz 解答:母比率の区間推定

$$0.6 - 1.96 \times \sqrt{\frac{0.6 \cdot 0.4}{10}} < p < +0.6 - 1.96 \times \sqrt{\frac{0.6 \cdot 0.4}{10}}.$$

$$0.3 < p < 0.9.$$

ここまで来たよ

6 略解:比率の区間推定, 平均値の差の区間推定

7 平均値, 平均値の差の検定

- Excel による平均値の区間推定
- (平均値の) 統計的仮説検定
- Excel による平均値の統計的仮説検定

平均値, 平均値の差の区間推定 I

平均値の区間推定

母集団の平均値 μ の, 信頼係数 $1 - \alpha = 0.95, 0.99$ の信頼区間は,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{N}} < \mu < \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{N}}.$$

\bar{x} : 標本の平均値

s : 標本の標準偏差

N : 標本のサイズ

平均値, 平均値の差の区間推定 II

母平均値の差の区間推定

母集団 1 と母集団 2 の平均値の差 $\mu_1 - \mu_2$ の信頼係数 $1 - \alpha$ の信頼区間は

$$\bar{X}_1 - \bar{X}_2 - z_{\alpha/2} \frac{S}{\sqrt{N}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + z_{\alpha/2} \frac{S}{\sqrt{N}}$$

\bar{X}_1, \bar{X}_2 標本 1, 2 の平均値

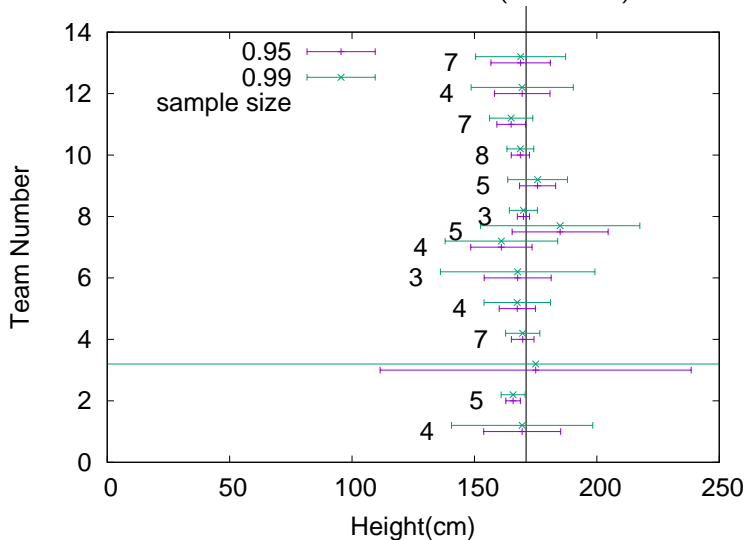
N_1, N_2 標本 1, 2 のサイズ

S^2 'プールした' 分散

$$\text{係数 } z_{\alpha/2} = \begin{cases} 1.96 & (1 - \alpha = 0.95) \\ 2.58 & (1 - \alpha = 0.99) \end{cases}$$

区間推定のイメージ

クラス全体の平均身長を, 少人数の標本 (サンプル) を抽出して推定した.



チーム	標本サイズ	滋賀県 $\sum_i Y_i$	標本平均値 \bar{X} (cm)	不偏標本分散 s^2 (cm ²)
1	4	2	169.5	97.7
2	5	1	165.8	5.7
3	2	2	175.0	50
4	7	4	169.7	24.9
5	4	1	167.5	21.7
6	3	2	167.7	30.3
7	4	1	161.0	62
7.5	5	2	185.0	250
8	3	1	170.0	1.0
9	5	1	175.8	35.2
10	8	3	168.8	19.6
11	7	3	165.0	39.7
12	4	1	169.5	51.0
13	7	1	168.9	171.8

ここまで黙っていたこと

ここでの説明は正確さより単純さ重視で書いてます.

- 母集団は, 共通の分散, 異なる平均値で正規分布していることを仮定しています.
- ○ 本来は, 標本サイズ N_i は本当は自由度 $N_i - 1$ で考えるべきです.
- ○ 本来は, 標本の分散のかわりに不偏標本分散を使うべきです.
- ○ 本来は, $z_{\alpha/2}$ でなく, t -分布表を見て $t_{\alpha/2}(N_i - 1)$ を使うべきです.

○ = Excel は正しくやってくれる点.

Excel による区間推定

準備:統計ツールを有効化 [動画](#)

ファイル > オプション > アドイン > Excel のアドイン > 設定 > 分析ツール に
チェックを入れて OK する.

平均値の信頼区間

データ > データ分析 > 基本統計量 > 統計情報, 平均の信頼区間の出力
先頭行をラベルとして使用: 指定する範囲の先頭が, 量の名前ならチェックする, 使う数値ならチェックしない.

Excel の出力と読み方

1	列1	
2		
3	平均	41
4	標準誤差	1.914854216
5	中央値 (メジアン)	41
6	最頻値 (モード)	#N/A
7	標準偏差	6.055300708
8	分散	36.66666667
9	尖度	-1.2
10	歪度	0
11	範囲	18
12	最小	32
13	最大	50
14	合計	410
15	標本数	10
16	信頼区間(95.0%)	4.331701179

95%信頼区間 ($1 - \alpha = 0.95$) は,

$$41 - 4.331701179 < \mu < 41 + 4.331701179.$$

ここまで来たよ

6 略解:比率の区間推定, 平均値の差の区間推定

7 平均値, 平均値の差の検定

- Excel による平均値の区間推定
- (平均値の) 統計的仮説検定
- Excel による平均値の統計的仮説検定

統計的仮説検定とは

観測のある科学 (心理学, 社会学, 生物学…) でいちおう合意されている, データから「異常である」ことを主張する手法.

実行する人の定義する '正常な状態' (=帰無仮説) からの異常を, ときどきは検出できる試験紙 (変色したら異常っぽい)

- 変色した = 帰無仮説の棄却 = 「差があったと結論する」
- 変色しなかった = 何もいえない = 「差があったと結論する」 \neq 「差がなかったと結論する」
- 異常なのに変色しないこと = 偽陰性 = 第2種の過誤
- 異常なのに変色しない確率 = β . 検出力 $1 - \beta$
- 正常なのに変色してしまうこと = 偽陽性 = 第1種の過誤
- 正常なのに変色してしまう確率 = 有意水準 α 信頼水準 $1 - \alpha$

α と β を両方小さくするのは難しい.

方針: α の値が 0.01 や 0.05 になるように調整する. β はできる範囲で小さくする (けど難しい)

		真実	
		帰無仮説 は真	帰無仮説 は偽
判断	帰無仮説 を棄却しない	正しい判断	第 2 種の過誤 (確率 β で起きる)
	帰無仮説 を棄却	第 1 種の過誤 (確率 α で起きる)	正しい判断

統計的仮説検定の例

- あるテストは, 授業を受ける前 (**事前 pre-**) は平均点が 100 点満点で 50 点であることがわかっているとしよう (まったく未知の知識で, ○×問題だから).
- ある方式の授業を受けた後 (**事後 post-**), 成績があがることは確か, と主張したい.

正常状態=事後の平均点が 50 点 (正常状態としては, 数値がはっきりわかっているものを選ぶ)

異常状態=事後の平均点が 50 点より大きい (これを言いたい)

平均値の場合の, 試験紙の仕組み

平均値の統計的仮説検定

信頼係数 $1 - \alpha$ の信頼区間が, 正常値 (実行する人が設定する) にかかっていなければ, 変色.

平均値の差の統計的仮説検定

差の信頼係数 $1 - \alpha$ の信頼区間が, 正常値 0 にかかっていなければ, 変色.

ここまで来たよ

6 略解:比率の区間推定, 平均値の差の区間推定

7 平均値, 平均値の差の検定

- Excel による平均値の区間推定
- (平均値の) 統計的仮説検定
- Excel による平均値の統計的仮説検定

Excel による平均値の統計的仮説検定

準備:統計ツールを有効化 動画

ファイル > オプション > アドイン > Excel のアドイン > 設定 > 分析ツール に
チェックを入れて OK する.

平均値の差の検定

データ > データ分析 > t 検定: 等分散を仮定した 2 標本による検定 >

- 仮説平均との差異: 0 (この説明の範囲では).
- α : 有意水準 $\alpha = 0.05, 0.01$ など.

平均値の検定

データ > データ分析 > t 検定: 等分散を仮定した 2 標本による検定 >

- 2 個目の標本には, すべて, 帰無仮説の値を手で入れておく.
- 仮説平均との差異: 0 (このやり方では)
- α : 有意水準 $\alpha = 0.05, 0.01$ など.

Excel の出力の読み方

1	t-検定: 等分散を仮定した2標本による検定		
2			
3		変数1	変数2
4	平均	34.6	30
5	分散	35.8	0
6	観測数	5	5
7			
8	プールされた分散		17.9
9	仮説平均との差異		0
10	自由度		8
11	t		1.719100713
12	P(T<=t) 片側		0.061957246
13	t 境界値 片側		1.859548038
14	P(T<=t) 両側		0.123914492
15	t 境界値 両側		2.306004135

変色判定: $|t| > t$ 境界値片側.

お知らせ

- 来週 2018-11-26 月 2 休講させていただきます
 - ▶ 補講たぶん 2019-01-22 火 集中補講日
- 期末試験計画
 - ▶ 30 ピーナッツ/科目 100 ピーナッツ
 - ▶ 60 分
 - ▶ 2019-01-28 月
- レポート計画