

t 検定と統計ソフトウェア

樋口さぶろお

龍谷大学工学部数理情報学科

確率統計☆演習 I L13(2015-01-16 Fri)

今日の目標

- p -値, 棄却域の意味が説明できる.
- Excel を使って t 検定ができる.



<http://hig3.net>

Quiz(母平均値の検定 (母分散未知)=t 検定)

あるドーナツ製造マシンが次々に製造するクロワッサンドーナツの重さ X_i g は, 母平均値が 57g であるはずだが, きょう 5 個製造したところ, 下のようだった.

52g, 52g, 53g, 48g, 50g.

本当にドーナツ製造マシンが次々に製造するクロワッサンドーナツの重さ X_i g の母平均値は 57g なのだろうか. 有意水準 5% で統計的検定を行って判定しよう.

L13-S1

Quiz 解答:母平均値の検定 (母分散未知)=t 検定

- ① 有意水準 5%で
- ② 正規分布の平均値に対する t 検定を行う.
- ③ 帰無仮説を「ドーナツの重さの母平均値 μ_1 が $\mu_0 = 57\text{g}$ に等しい」とする.
- ④ サイズ $n = 5$ の標本の標本平均値を \bar{X} , 不偏標本分散を S^2 とするとき,

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$$

は, 自由度 $5 - 1$ の t 分布に従う. これを検定統計量としても用いる.

- ⑤ この標本に対して, $t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}} = \frac{51 - 57}{\sqrt{\frac{1}{5} \frac{16}{5-1}}} = -3\sqrt{5} = -6.708$.
- ⑥ $t_{0.05/2}(4) = 2.776 < |t|$ だから, 帰無仮説は棄却される. 母平均値が異なると結論する.

L13-S2

Quiz 解答:正規分布の母平均値に関する t 検定

- ① 有意水準を 5%とする.
- ② 正規分布の母平均値の t 検定を行う.
- ③ 帰無仮説を「ドーナツ販売開始後の, 来店客数の母平均値 μ_1 は $\mu_0 = 196$ に等しい」とする.
- ④ サイズ $n = 4$ の標本の標本平均値を \bar{X} , 不偏標本分散を S^2 とすると,

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$$

は, 自由度 $4 - 1$ の t 分布に従う. これを検定統計量として用いる.

- ⑤ この標本に対して, $\bar{X} = 200, S^2 = \frac{224}{4-1} = 74.7$. よって,

$$t = \frac{200-196}{\sqrt{\frac{1}{4} \frac{224}{3}}} = 0.92582.$$
- ⑥ t 分布表より, $|t| > t_{0.05/2}(3) = 3.182 > |t|$ なので, 帰無仮説は棄却できない. 来店客数が変化したとは結論できない.

ここまで来たよ

- 1 略解:統計的仮説検定
- 2 t 検定と統計ソフトウェア
 - 棄却域と p 値
 - Excel で検定

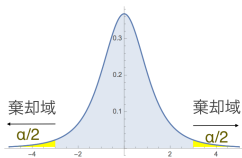
統計的仮説検定 (復習)

- ① 有意水準を書く
- ② (検定の名前があれば) 「…検定」を行う, と書く
- ③ 帰無仮説を書く
- ④ 選択した検定統計量 T と, それが (帰無仮説のもとで) 従う分布を書く
- ⑤ 標本に対する検定統計量の値 y_1 を書く.
- ⑥ T が t_1 より極端な値となる確率を求める ($=p$). それが α より大きい/未満なら, 帰無仮説を採択する/棄却する ($=$ 有意でなかった/有意だった) と書く.

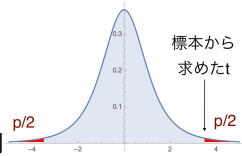
母平均値に関する t 検定するとき, 検定統計量は

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$$

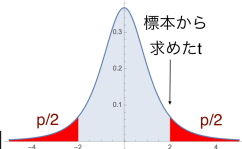
棄却域と p 値



帰無仮説棄却



帰無仮説採用



帰無仮説を棄却
 $\Leftrightarrow p$ 値 < 有意水準 α

過誤, 有意水準, 検出力

		真実	
		H_0 は真	H_0 は偽
判断	H_0 を棄却しない	正しい判断	第 2 種の過誤 (確率 β で起きる)
	H_0 を棄却	第 1 種の過誤 (確率 α で起きる)	正しい判断

$1 - \alpha$: 区間推定でいう に対応

p 値 (*p*-value): T がこの標本よりも
棄却域帰無仮説が棄却される, t の値の範囲.

採択域棄却域の補集合.

ここまで来たよ

- 1 略解:統計的仮説検定
- 2 t 検定と統計ソフトウェア
 - 棄却域と p 値
 - Excel で検定

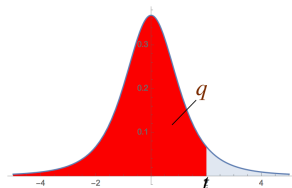
Excel 2010 で標本ナントカ

標本にまつわる Excel の関数

- 標本平均値 average
- 不偏標本分散 var
- 不偏標本標準偏差 stdev

注: 有限母集団の量は母平均値 average, 母分散 varp, 母標準偏差 stdevp. 要区別

Excel 2010 t 分布



n : 自由度

t 分布にまつわる Excel の関数

- `t.dist` $q = \text{t.dist}(t, n, \text{true})$
- `t.inv` $t = \text{t.inv}(q, n)$

ご注意

- Excel のバージョンで異なる
- Excel はバグがあるから信じない, という人も. → R 確率統計☆演習 II, 計算科学 II

2 標本 t 検定

確率統計☆演習 II

石山店のドーナツの重さは母平均値 μ_x , 瀬田店のドーナツの重さは母平均値 μ_y の確率分布にしたがうとする。

$\mu_x \neq \mu_y$ か? 2 個の標本があるとき, 有意水準 α % で検定しよう。

帰無仮説 2 個の標本は, 同じ母分布からとられたもの

対立仮説 2 個の標本は, 異なる母分布からとられたもの

石山店のサイズ m の標本 $\{X\}$, 瀬田店のサイズ n の標本 $\{Y\}$ があるとき,

検定統計量

$$T = \frac{\bar{X} - \bar{Y}}{\text{ある定数}}$$

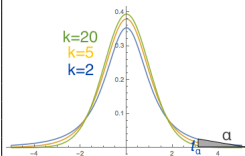
が自由度 $m + n - 2$ の t 分布にしたがうことを利用して検定する。

Excel: データ > 統計ツール > 2 標本 t 検定

t 分布表

$\alpha = P(T > t_{\alpha}(k))$ となる, $t_{\alpha}(k)$ の値の表. k は自由度.

$k \backslash \alpha$	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.00025
1	1.000	1.376	1.963	3.078	6.314	12.710	31.820	63.660	127.300	318.300	636.600
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.090	22.330	31.600
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.210	12.920
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
+∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



Maple T.A.

テスト使用にご協力ください.

Equation Editor で入力

Σ のアイコンからグラフィカルに入力できます.

Equation Editor のショートカット

^ キーを押すと、表示はされませんが上付き入力モードに入ります. 上付き入力を終了するには、Control と / の同時押し.

$x^n y$ は (コンマは不要です), x, ^, (, n,), Control+/, *, y とタイプ.

空欄に直接入力

C 言語みたいな感じ. 半角英数字だけで空欄に入力してください.

$\sin x \rightarrow \sin(x)$

$x^2 y \rightarrow x \ x \ y$ (スペースに注意) または $x*x*y$

ただし, $x^n \rightarrow \text{pow}(x,n)$ $x^{(n)}$

$\sin^2 x \rightarrow (\sin(x))^2$

連絡

- 今後、予習問題や非参照 Quiz はありません.
- 2014-11-17 から チューターは月火水木昼 (1-614).
- **2015-01-17 土 4 補講** 第 1/2 セルフラーニング室で e ラーニングから各自受講. <http://hig3.net> → RaMMoodle → 確率統計. 動画視聴のため各自でイヤフォンを用意してください. ただし, この日時場所でもなくとも, ファイナルトライアルまでに, 例えば自宅などからでも受講すればよいです.

ファイナルトライアル出題計画

外部記憶ペーパーが試用可能であることに注意. 電卓使用なし. 必要な表は印刷します.
Excel の問題はありません.

- 離散型確率変数の母平均値, 母分散, 母期待値, 確率 (L06, プチテスト問 6,7 と重なる)
- 連続型確率変数の母平均値, 母分散, 母期待値, 確率 (L08, プチテスト問 8 と重なる)
- 正規分布 $N(\mu, \sigma^2)$ にしたがうの確率変数が, ある条件を満たす確率を求める (L08)
- 標本から, 母平均値, 母分散を点推定する (L09)
- 標本から, 母平均値を区間推定する (母分散未知の小標本=t 分布を使う) (L10)
- 標本から, 母比率を区間推定する (L11)
- 標本から, 母分散を区間推定する (L14)
- 標本から, 母平均値に関する t 検定を行う (L12,L13)
- 標本から, 母分散に関するカイ 2 乗検定を行う (L14)
- 母集団, 標本, 標本抽出, 推定, 検定についての選択肢問題 × 数個