母分散の片側カイ二乗検定・p 値・統計ソフトウェア

樋口さぶろお　http://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習 I　L14(2019-01-16 Wed)

最終更新: Time-stamp: "2019-01-16 Wed 13:24 JST hig"

今日の目標

- 母分散の片側カイ二乗検定ができる 前園確率統計 §6.2
- Excel で p 値が計算できる

L13-Q1
Quiz 解答:母比率の両側二項検定の正規近似

1. 有意水準 $\alpha = 0.05$ で, 母比率の両側検定を行う.
2. 帰無仮説を「瀬田学舎生のうち, 滋賀県の高校を卒業した人の母比率は $p = 0.4$」,「対立仮説を $p \neq 0.4$」とする.
3. サイズ $n = 68$ の標本の標本比率を $\hat{p}$ とすると, 検定統計量
$$Z = \frac{\hat{p} - 0.4}{\sqrt{0.4(1-0.4)/68}}$$
は, 標準正規分布に近似的にしたがう.
4. この標本に対して, $\hat{p} = 20/68 = 0.2941$ より, $z = -1.782$.
5. 標準正規分布表より境目の値は $z_{0.05/2} = 1.960$. (または $t(\infty; 0.05/2)$) 棄却域は $|z| > 1.960$.

6. $1.960 > |-1.782|$ なので, z は棄却域に含まれない. 帰無仮説は棄却できない. 瀬田学舎生のうち, 滋賀県の高校を卒業した人の母比率は $p \neq 0.4$ である, とは結論できない.

L13-Q2

Quiz 解答:母比率の片側二項検定の正規近似

1. 有意水準 0.05 で, 母比率の片側検定を行う.
2. 帰無仮説を「瀬田学舎生のうち, 滋賀県の高校を卒業した人の母比率は $p = 0.4$」,「対立仮説を $p > 0.4$」とする.
3. サイズ $n = 68$ の標本の標本比率を $\hat{p}$ とすると, 検定統計量

$$Z = \frac{\hat{p} - 0.4}{\sqrt{0.4(1-0.4)/68}}$$

は, 標準正規分布に近似的にしたがう.

4. この標本に対して, $\hat{p} = 20/68 = 0.2941$ より, $z = -1.782$.

5. 標準正規分布表より境目の値は $-z_{0.05/2} = -1.645$. (または $-t(\infty; 0.05/2)$) 棄却域は $z < -1.645$.
6. $z < -1.645$. なので, z は棄却域に含まれる. 帰無仮説を棄却する. 瀬田学舎生のうち, 滋賀県の高校を卒業した人の母比率は $p < 0.4$ である, と結論する.

ここまで来たよ

13 略解:母比率・母分散の両側/片側検定

14 母分散の片側カイ二乗検定・p 値・統計ソフトウェア
- 母分散のカイ二乗検定
- p 値=有意確率, Excel を利用した検定

カイ二乗分布 前園確率統計 p.36

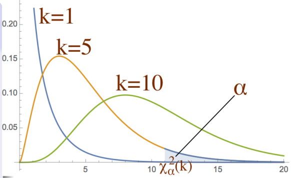
カイ二乗分布

$Z_1, \ldots, Z_k$, を標準正規分布 $\mathrm{N}(0, 1^2)$ に従う独立な確率変数とするとき,
確率変数 $Y = Z_1^2 + \cdots + Z_k^2$ とおく.
Y は, 自由度 k のカイ二乗分布 $\chi^2(k)$ に従う.

言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

$\chi^2(k)$ の確率密度関数 前園確率統計 p.36

$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (\text{他}) \end{cases}$$



$\mathrm{E}[Y] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{V}[Y] = \text{積分} = 2k.$

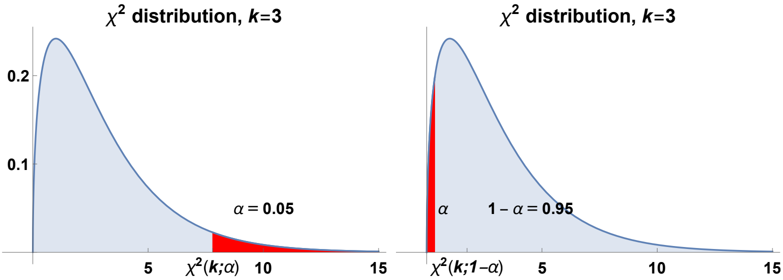
母分散の片側カイ二乗検定 前園確率統計 §6.2 I

未知の正規分布からの標本に基づき, 母分散 σ^2 が $\sigma^2 > \sigma_0^2$ と言いたい, (または $\sigma^2 < \sigma_0^2$ と言いたい)

母分散の片側カイ二乗検定

- 前提 母集団が正規分布にしたがう
- 帰無仮説 母分散 $\sigma^2 = \sigma_0^2$, 対立仮説 $\sigma^2 > \sigma_0^2$ (または $\sigma^2 < \sigma_0^2$)
- 検定統計量 $Y = (n-1) \times \frac{s^2}{\sigma_0^2}$ は自由度 $n-1$ のカイ二乗分布にしたがう. 前園確率統計定理 2.3
- 棄却域 $Y > \chi^2(n-1;\alpha)$ (または $Y < \chi^2(n-1;\alpha)$)

母平均値, 母比率の検定統計量は差, 正常値は 0.
母分散の検定統計量 Y は比, 正常値は $1 \times (n-1)$.


χ² distribution, k=3
0.2
0.1
α = 0.05
5
χ²(k;α)
10
15
χ² distribution, k=3
α
1 − α = 0.95
χ²(k;1−α)
5
10
15

L14-Q1

TA Prob and Sol:母分散の片側カイ二乗検定

あるファーストフードチェーンのポテトフライ S の重さは, 母分散が $4\mathrm{g}^2$ であることが定められているという.
トレーニング中のアルバイトの人に, ポテトフライ S サイズを 9 個作ってもらったところ, 重さは下のようだった (単位は g).

$$76, 76, 76, 76, 80, 84, 84, 84, 84.$$

このアルバイトの作るポテトフライ S の重さの母分散 σ^2 は, $2^2\mathrm{g}^2$ より大きいか? 重さが正規分布にしたがうと仮定し, 有意水準 $\alpha = 0.05$ で, 母分散のカイ二乗検定で判定しよう.

略解

1. 有意水準 $\alpha = 0.05$ で, 母分散の片側カイ二乗検定を行う.

2. 帰無仮説を，「アルバイトの…重さの正規分布の母分散 σ^2 は，$2^2\mathrm{g}^2$ に等しい」対立仮説を $2^2\mathrm{g}$ より大きい」とする．
3. サイズ n の標本の不偏標本分散を s^2 とすると，量 $Y = (n-1) \times \frac{s^2}{2^2}$ は，自由度 $n-1$ のカイ二乗分布に従う．この量を検定統計量として用いる．
4. この標本に対して $Y = (n-1) \times \frac{s^2}{2^2} = (9-1) \cdot \frac{16}{2^2} = 32$.
5. カイ二乗分布表より，棄却域の境い目は，$\chi^2(n-1;\alpha) = 15.5073$，棄却域は $Y > 15.5073$.
6. 不等式 $32 > 15.5073$ が成立するので，帰無仮説を棄却する．母分散は $2^2\mathrm{g}^2$ より大きいと結論する．

不等式が逆のとき「帰無仮説は棄却できない．母分散は $2^2\mathrm{g}^2$ より大きいとは結論できない.」

前園確率統計例題 6.2

L14-Q2

Quiz(母分散の片側カイ二乗検定)

あるファーストフードチェーンのポテトフライ S の重さは, 母分散が $4\mathrm{g}^2$ であることが定められているという.
アルバイトのリーダーの人に, ポテトフライ S サイズを 11 個作ってもらったところ, 重さは下のようだった (単位は g).

$$73, 74, 74, 75, 75, 75, 75, 76, 76, 76, 76$$

このリーダーの作るポテトフライ S の重さの母分散 σ^2 は, $2^2\mathrm{g}^2$ より小さいか? 重さが正規分布にしたがうと仮定し, 有意水準 $\alpha = 0.01$ で, 母分散のカイ二乗検定で判定しよう.

ここまで来たよ

p 値=有意確率による棄却する/しない判定 前園確率統計 p.90

検定の Step6 では, s^2 を α と比べてる. 2 つの比べ方.

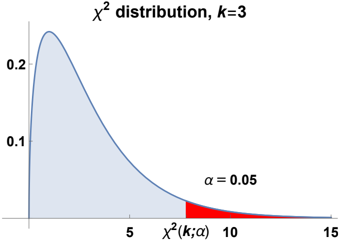
		棄却する	しない	
統計量の値勝負 (Step6)	$\chi^2(n-1; 1-\alpha)$	$>$	$<$	$y = (n-1)\frac{s^2}{\sigma^2}$
	$\chi^2(n-1; \alpha)$	$<$	$>$	
	↑ 表 or chisq.inv.rt			↓ chisq.dist.rt
確率=面積勝負	有意水準 α	$>$	$<$	p 値

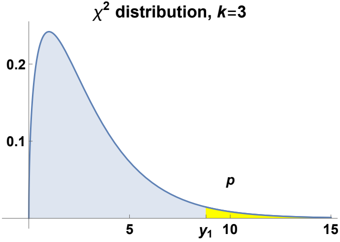
標本の p 値 (p-value)=有意確率 前園確率統計 p.90

帰無仮説のもとで, 検定統計量がこの標本よりも極端な値をとる確率=端側の面積

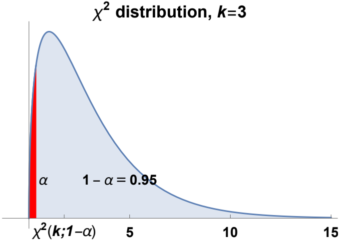
$\alpha > p$ のとき帰無仮説を棄却

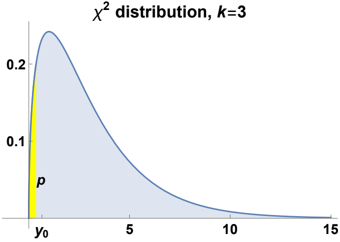
片側検定の対立仮説「母分散は…より大きい」





片側検定の対立仮説「母分散は…より小さい」





Excel 2016 で標本ナントカ

標本にまつわる Excel の関数

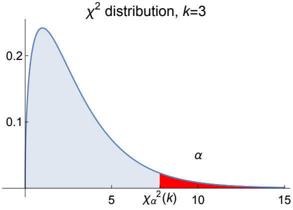
- 標本平均値 `average`
- 不偏標本分散 `var.s`
  - s for sample (標本), p for population (母集団)
- 不偏標本標準偏差 `stdev.s`

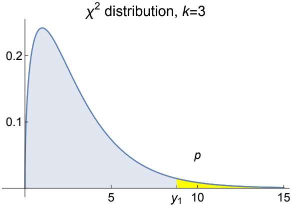
要区別: 有限母集団の量は母平均値 `average`, 母分散 `var.p`, 母標準偏差 `stdev.p`.

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

Excel 2016 でのカイ二乗分布





k: 自由度

カイ二乗分布にまつわる Excel の関数

- Y_1 より上の面積 $= \int_{Y_1}^{\infty} f(y)\,\mathrm{d}y =$`chisq.dist.rt`$(Y_1, k)$
  - $p =$ `chisq.dist.rt`(Y_1, k) 片側検定の対立仮説が「母分散が大きい」
  - $1 - p =$`chisq.dist.rt`(Y_0, k) 片側検定の対立仮説が「母分散が小さい」
- 数表の代わり `chisq.inv.rt`$(\alpha, k)$$= \chi^2(k;\alpha)$

`chisq.test` は別用途. 適合度, 独立性のカイ二乗検定のための関数.

Excel を用いた母分散のカイ二乗検定の手順

1. 不偏標本分散を計算 s^2 `var.s`
2. s^2 からカイ二乗検定統計量 Y を計算
3. 確率変数の値勝負
    1. α から境い目の値 $\chi^2(k;\alpha)$ を計算 `chisq.inv.rt`
    2. $\chi^2(k;\alpha)$ より Y が極端なら帰無仮説棄却
4. 面積勝負
    1. Y に対する p-値を計算 `chisq.dist.rt`
    2. $p < \alpha$ なら帰無仮説棄却

連絡

Moodle
`https://learn.math.ryukoku.ac.jp/`

Moodle モバイルアプリ
`https://download.moodle.org/mobile`
起動後, URL `https://learn.math.ryukoku.ac.jp/moodle` を登録

GeoGebra 確率電卓
`https://www.geogebra.org/classic#probability`

- 今週も予習復習問題はあります.
- 予習復習問題を, 期限後も (再/初) 受験できます. 点数にはカウントしないけど, プチテスト準備に活用してね.
- 配布資料は 1-503 向かいの引出, `http://hig3.net` で再配布.
- 樋口オフィスアワー火昼 (1-539) 金 14:40-15:40(1-502), Math ラウンジ月-木昼 (1-614)
- インターンシップ (学外実習) おすすめ中. 担当教員の面接にどうぞ.
- 学力認定試験参加おすすめ中 (2018-02-16 土). 1 月中に樋口または他の教員まで.

ファイナルトライアル出題計画 別紙参照.
外部記憶ペーパー使えます.
過去問題を公開していますが, 出題傾向は毎年変わります. 去年のものに対応するより, 下の出題計画と Trial を参照することをお奨めします.