比率の区間推定, 平均値の差の区間推定

樋口さぶろお http://hig3.net

龍谷大学理工学部数理情報学科

生活の中の統計技術 L07(2018-11-05 Mon)

最終更新: Time-stamp: "2018-11-05 Mon 14:58 JST hig"

今日の目標

- 比率の区間推定ができる
- 平均値の差の区間推定ができる

L06-Q1

Quiz 解答:平均値の区間推定

平均値は $m = 80$.
分散は $S^2 = \frac{1}{5}[0^2 + 1^2 + 1^2 + 3^2 + 3^2] = 4$.
標準偏差は $S = 2$.
よって, 信頼係数 0.95 で, 本来の平均値 μ は,

$$80 - 1.96 \times \frac{2}{\sqrt{5}} < \mu < 80 + 1.96 \times \frac{2}{\sqrt{5}}$$

.

L06-Q2

Quiz 解答:区間推定

標本の平均値で与えられ, $\bar{x} = \frac{1}{5}[10 + 20 + 30 + 30 + 110] = 40$(分)
標本の分散は
$S^2 = \frac{1}{5}[(10-40)^2 + (20-40)^2 + (30-40)^2 + (30-40)^2 + (110-40)^2] = 1280$(分^2), 標本の標準偏差は $\sqrt{1280} = 37.78$(分).

よって, 母集団の平均値 μ の信頼係数 0.99 の信頼区間は,

$$40 - 2.58 \times \sqrt{\frac{1280}{5}} < \mu < 40 + 2.58 \times \sqrt{\frac{1280}{5}}$$

ここまで来たよ

5 略解:母集団と標本抽出, 平均値の推測

6 比率の区間推定, 平均値の差の区間推定
- 母比率
- 平均値の差の区間推定

母比率の信頼区間高校 数学 B

- 高校 1 年生に, この内容を 1 時間でこうやって教えたとき, 何%がこの問題をできるようになる? 10 人にやってみて推定したい
$\rightsquigarrow$ 比率 0 以上 1 以下. 正解 1 点, 不正解 0 点と思った平均点
- 龍大生のうち何%が喫煙者? 10 人に質問してみて推定したい
$\rightsquigarrow$ 比率 0 以上 1 以下. 正解 1 点, 不正解 0 点と思った平均点

- 候補者 A の得票率は何% ? N 人に質問しただけで推定したい.
- 出荷する製品の何% が不良品? N 個だけ抜き出して調査したい.
- このコインの表が出る確率は? N 回投げるだけで推定したい.

母比率の信頼区間 (母分散未知)

サイズ N の標本で, 標本の比率が $\hat{p} = y/N$ のとき, 母集団の比率 p の信頼係数 $1-\alpha = 0.95$, 信頼係数 $1-\alpha = 0.99$ の信頼区間は

$$\frac{y}{N} - 1.96 \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}} < p < \frac{y}{N} + 1.96 \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}},$$

$$\frac{y}{N} - 2.58 \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}} < p < \frac{y}{N} + 2.58 \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}}.$$

覚え方: 標本の平均値 $\hat{p} = \frac{y}{N}$, 分散 $\hat{p}(1-\hat{p})$.

L07-Q1

Quiz(母比率の区間推定)

選挙で出口調査をしたところ, 50 人中 35 人が A 候補に投票したと答えた. 母集団を投票した人全体とする. そのうち A 候補に投票した人の母比率 (得票率) を考える.

1. A 候補の得票率を, (点) 推定しよう
2. A 候補の得票率を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう.
3. A 候補の得票率を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう.

L07-Q2

Quiz(母比率の区間推定)

受講生 500 人の科目で期末試験で科目の合否を決めるが, 合格率があまり低いのもよくないと思っている．そこで, 期末試験に公欠届を出した 10 人に内証で事前に受験してもらうことにした．10 人中合格者は 6 名だった．

1. 500 人のうちの不合格者の人数を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう．
2. 500 人のうちの不合格者の人数を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう．

ここまで来たよ

5 略解:母集団と標本抽出, 平均値の推測

6 比率の区間推定, 平均値の差の区間推定
- 母比率
- 平均値の差の区間推定

実験群と統制群の平均値の差

ある高校で, 英語の従来の自習用問題集を新しい自習用問題集に変更することが提案された. 新しい教科書を使うと, 点数は上がるだろうか?

- 新しい問題集を使う生徒=母集団 1,
- 従来の問題集を使う生徒=母集団 2

の平均値の差はどのくらいか (正か負か?)
それぞれから標本 1, 標本 2 をランダムに抽出する. この高校のクラス編成がもともとランダムなら, 新しい問題集を使うクラスをランダムに 1 つ決めればいい. 問題集以外は, 同じことをおこなう.

- 標本 1=実験群 (experimental group)
- 標本 2=統制群, 対照群 (control group)

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

$$\overline{X_1} - \overline{X_2} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{N}} < \mu_1 - \mu_2 < \overline{X_1} - \overline{X_2} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{N}}$$

$\overline{X_1}, \overline{X_2}$ 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

N は N_1, N_2 の調和平均: $\frac{1}{N} = \frac{1}{N}_1 + \frac{1}{N}_2$
S^2: 標本 1,2 の合併した (プールした) 標本分散

2 標本の合併した (プールした) 標本分散

標本 $X_{1\ i}$ $(i = 1, 2, \ldots, N_1)$, 標本 $X_{2\ j}$ $(j = 1, 2, \ldots, N_2)$.

$$\begin{aligned} S^2 =& \frac{1}{N_1 + N_2}[(X_{11} - \overline{X_1})^2 + \cdots + (X_{1N_1} - \overline{X_1})^2 \\ &+ (X_{21} - \overline{X_2})^2 + \cdots + (X_{2N_2} - \overline{X_2})^2] \\ =& \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2} \end{aligned}$$

第 1, 第 2 の標本の分散 S_1^2, S_2^2 を, 標本サイズで重み付けして平均したもの.
標本 1 がある値に集中し, 標本 2 が別の値に集中したとき, 標本 1,2 共通の平均値 $\overline{X}$ から分散を作ると大きくなる. そういう値が差の精度に関係するのはへんでしょ. 標本 1 と標本 1 の平均値, 標本 2 と標本 2 の平均値, の差がでてくるのがもっとも.

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_1 は本当は自由度 $N_1 - 1$ 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.

L07-Q3

Quiz(母平均値の差の区間推定 (母分散未知))

ドーナツ製造マシン 1 号,2 号が製造するドーナツの重さ X, Y g は, 独立で, それぞれ正規分布 $\mathrm{N}(\mu_1, \sigma^2), \mathrm{N}(\mu_2, \sigma^2)$ にしたがう.

- 1 号で製造したドーナツの標本は, サイズ 5, 標本平均値 90g, 不偏標本分散 10g^2,
- 2 号で製造したドーナツの標本は, サイズ 7, 標本平均値 80g, 不偏標本分散 20g^2

だった.

1. 合併した不偏標本分散 S^2 を求めよう.
2. 母平均値の差 $\mu_1 - \mu_2$ を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう. 整理したり計算したりしなくてよい. ただし, 表の $t_\bullet(k)$ の値は小数に直して書くこと.

L07-Q4

Quiz(母平均値の差の区間推定 (母分散未知))

ドーナツ製造マシン 1 号,2 号が製造するドーナツの重さ X, Yg は, 未知の母平均値 μ_1, μ_2 の独立同分布にしたがう確率変数である. 母分散も未知だが, 1 号と 2 号で等しいことがわかっている.
1 号,2 号で製造したドーナツの重さは次のようだった.
1 号: $51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}.$
2 号: $55\mathrm{g}, 56\mathrm{g}, 51\mathrm{g}, 52\mathrm{g}, 56\mathrm{g}, 54\mathrm{g}.$

1. $\overline{X}, \overline{Y}$, 合併した不偏標本分散 S^2 を求めよう.
2. 母平均値の差 $\mu_1 - \mu_2$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

お知らせ

- 中間試験計画
  - 30 ピーナッツ/科目 100 ピーナッツ
  - 60 分
  - 2018-11-12 月
  - 出題計画

    60% 計算問題. データが与えられたときに, 平均値, q-分位数, 中間値, 四分位数, 分散, 標準偏差, 共分散, 相関係数, 単回帰の回帰直線, データ中の 1 個の数値の偏差値が求められる.

    30% これらの量の性質や意味についての正誤判定問題

    10% 上記にあてはまらないかもしれない問題 (ワイルドカード)

    ★ Excel の操作方法については出題しない
  - 持込 紙はコピーを含め何でも. 電子機器は単機能電卓 (平方根まで) のみ.