標本と推定

樋口さぶろお

龍谷大学理工学部数理情報学科

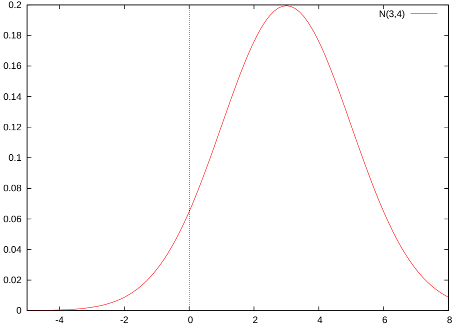
使える統計! L12(2014-01-08 Wed)

今日の目標

1. 母集団, 標本, 標本抽出, 推定を説明できる
2. 標本から, 母平均値, 母分散, 母比率を推定できる.

http://hig3.net

L11-S1
Quiz 解答:正規分布のグラフ



L11-S2

Quiz 解答:正規分布の確率

1. $Z \geq \frac{5-3}{2}$ となる確率だから, 標準正規分布の表をひいて, $Q(1.00) = 0.1587$. (これは図からも求められる $(1 - 0.6827)/2 = 0.1587$).
2. $-1 \leq Z \leq 2$ となる確率だから, 標準正規分布の表をひいて, $1 - Q(2.00) - Q(1.00) = 0.8186$. (これは図からも求められる. $0.6827/2 + 0.9545/2 = 0.8186$)

L11-S5

Quiz 解答:中心極限定理

各日の価格の分布は独立で, 30 は十分大きいので, 中心極限定理より, 母平均値 $2000 \times 30 = 60000$ 円, 母分散 $40000 \times 30 = 1200000$ 円 ^{2} の正規分布に近似的に従う.

L11-Q2

Quiz(正規分布の確率)

あるお店で, 琵琶湖特産瀬田シジミ 500g パックは店主の気まぐれの '時価' で販売される. 長年の調査から, その価格は, 母平均値 2000 円, 母分散 40000 円 ^{2} の正規分布に従うことがわかっている.
500g 買うためには, お金をいくら財布に入れておけば十分か答えよう. ただし, 絶対に買えるように, というと何億円あっても足りないので, 40 回に 1 回は足りなくて買えなくてもかまわないとする.

ここまで来たよ

1. 復習:中心極限定理
2. 標本と推定
   - 母集団と標本
   - 標本抽出
   - 母平均値・母分散・母比率の推定

母集団と標本

AKB48 の身長ふたたび

- AKB48 メンバー全員 (=$\boxed{\qquad\qquad}$) の身長の平均を求めたい!
- メンバー全員分のデータがあれば公式使うだけ
- しかし, データ非公開になった今, 握手会でメンバー 1 人ずつに質問しなければいけないとしたら?
- 握手会参加券 74 枚集めないで何とかすませたい.

$\rightsquigarrow$

握手会参加券がゲットできて質問できたメンバー 5 人 (=$\boxed{\qquad\qquad}$) の答えから$\boxed{\qquad\qquad}$したい.

5 人を '無作為に' 選ぶ (=$\boxed{\qquad\qquad}$)

$\boxed{\qquad\qquad\qquad}$=5.

- 母集団=考えたい集団. どんな分布, 平均値, 分散, などわかっていないことがあるが, 全体を調べるわけにはいかない集団.
- 標本=母集団から ‘無作為に’ とってきた一部分
- 推定=標本を調べて母集団について正しそうな事実を見つけること.

もちろん推定には誤差あるかも. もともと, 標本の選び方ごとに答えは違うし.

ここまで来たよ

1. 復習:中心極限定理
2. 標本と推定
   - 母集団と標本
   - 標本抽出
   - 母平均値・母分散・母比率の推定

標本抽出

決まったやり方で, 母集団の性質を反映した標本を作ること.
L12-Q1

Quiz(標本抽出)

悪い標本抽出の例はどれ? また母集団は何か考えよう.

1. このクラスの平均身長を知りたい $\rightsquigarrow$ 教室の前から 5 人に身長を教えてもらう
2. ゆがんだサイコロで, 1 の目がでる確率を知りたい $\rightsquigarrow$ 100 回ふってみる
3. 龍谷大学生の週平均学習時間を知りたい $\rightsquigarrow$ このクラス全員にきく
4. 民主党の得票率を開票前に知りたい $\rightarrow$ 投票所の出口で 100 人おきに質問する.
5. 日本で飼われている犬の平均年齢を知りたい $\rightarrow$ 保健所に頼んで登録リストをもらって, 1000 番おきに飼い主に電話してきく

大注意

- よい標本抽出には実はいろいろなハイテクがある $\rightsquigarrow$ 社会調査法, 社会統計学
- 本当は, 母集団が無限でないとき, 標本を作るのに重複選出を禁止するとき, 数学的にいろいろ複雑なことが起きるのだが, この授業では知らん顔する. $\rightsquigarrow$ 確率・統計
- 要所で '無作為' を使うことがある. サイコロ, 乱数表, $\cdots$

ここまで来たよ

平均値の推定

標本平均値

$$標本平均値\ \overline{X} = \frac{1}{n}[X_1 + \cdots + X_n]$$

が, 母平均値のよい推定値になっている.

母平均値はひとつに定まっているが, 標本平均値は, 標本を抽出するたびにかわる (それ自体が確率分布をもつ)

分散の推定

標本 (不偏) 分散

$$\text{標本 (不偏) 分散 } s^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

が, 母分散のよい推定値になっている.
ここで, $\overline{X}$ は母平均値でなく, 上のように求めた標本平均値.

なぜ $n-1$? だって…

$n = 1$ のときに, 母分散が $\boxed{}$ って
おかしくない?
$n = 2$ のときに, 母分散は

$\boxed{}$

標本サイズが小さいときには大事な差. 標本サイズ $n \to +\infty$ では差はなくなる

標本抽出と推定の実験 I

statcrs

or

http://hig3.net→ (最初 or 左上) 生活の中の統計技術データ収集

L12-Q2

Quiz(推定)

瀬田学舎の龍大生の通学時間の分布を知るために, 無作為に 5 人を選んで質問したところ次のようだった. 母平均値, 母分散, 母標準偏差を推定しよう.
10 分, 20 分, 30 分, 30 分, 110 分.

L12-Q3

Quiz(推定)

ある確率分布に従うスピードくじを 10 回ひいたところ, 賞金は,
0 円, 0 円, 0 円, 0 円, 0 円, 0 円, 10 円, 10 円, 30 円, 100 円
だった. 確率分布の母平均値と母分散と母標準偏差を推定しよう.

母比率の推定

- クラスの中で, 血液型 A 型の人の比率は? n 人に質問しただけで推定したい.
- 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.
- 工場から出荷する製品のうち, 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=1 人とりだしたとき A 型である確率$=\frac{\text{母集団のうち A 型の人数}}{\text{母集団の人数}}$

母比率=表の出る確率 (ベルヌーイ分布の確率 p)

比率の推定

n 人中 k 人が A 型だったとき

$$\frac{k}{n}$$

が母比率 p のよい推定値になっている.

理由

確率分布 (ベルヌーイ分布) に従う量 X を考える.

X	どんなとき	確率
1	A 型である	p
0	A 型でない	$1-p$

確率分布

X の母平均値 $\mathrm{E}(X) = 1 \times p + 0 \times (1-p) = p$.
X の母分散 $\mathrm{V}(X) = (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$.

標本

標本 (サイズ n) で k 人が A 型のとき,
母平均値の推定値=標本平均値 $\overline{X} = \frac{1}{n}[\underbrace{1+\cdots+1}_{k}+\underbrace{0+\cdots+0}_{n-k}] = \frac{k}{n}$.
母分散の推定値=($p(1-p)$ に p の推定値 $\frac{k}{n}$ を代入し)= $\frac{k}{n}(1-\frac{k}{n})$

L12-Q4

Quiz(比率の区間推定)

クラスからサイズ (　　　) の標本をとったところ, (　　　) 人中 (　　　) 人が血液型 A であると答え, 残りが A 型でないと答えた.

1. クラスの, 血液型 A である比率を推定しよう

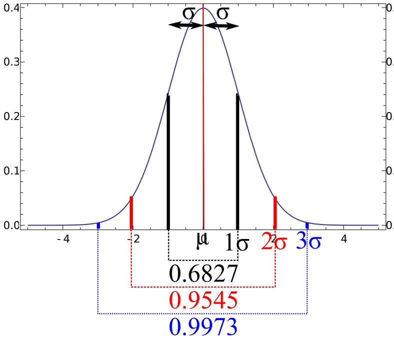
statcrs

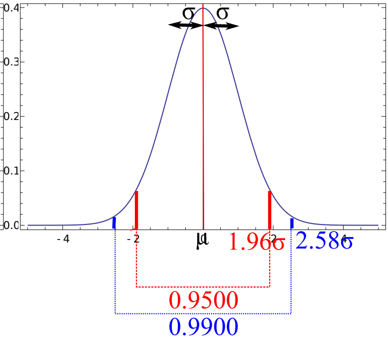
or

`http://hig3.net`→ (最初 or 左上) 生活の中の統計技術データ収集

標準正規分布 (ガウス分布) のグラフに関係した面積をおぼえよう!

いまは $\mu = 0, \sigma = 1$ と思ってね.



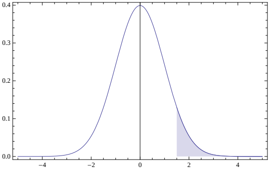


次のページの表さえあれば, 導ける.

標準正規確率表 (上側確率 $Q(z)$)

$z \geq z_0$ となる確率 $= Q(z_0) = \frac{1}{2}\mathrm{erfc}(z_0/\sqrt{2})$. よく統計の教科書の付録に表が載ってる.

z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



連絡

- きょうは紙 1 枚+アンケート 1 枚提出+携帯データ送信
- 近く台風の分の補講 (補講期間内ではない). 2014-01-15 水 以降 ファイナルトライアル以前に e ラーニングで実施します. 出題計画に含まれます.
- 加減乗除と平方根 (ルート) の使える電卓持ってきてね. 関数電卓でも関数電卓でなくてもかまいません. 携帯電話の機能・アプリでもかまいません.
- ファイナルトライアルでは携帯電話 (のアプリの電卓) は使えません.

ファイナルトライアル計画 I

日時 2014-01-29 Wed 水 3 (60 分)

場所 ??

形式 ペーパーテスト．計算問題が多いけど選択肢問題群もあるかも．関数電卓持込可．正規分布の確率の表は問題とともに印刷して配ります．

参照 A4 $\times$ 1 枚両面持込可．

- この後につけている専用用紙を使ってください．コピー可．
- **縮小**コピー, 貼り付けは不可．
- 回収します．

配点 100 点 50 ピーナッツ

公欠 通常の追試規定に従います．

出題計画 2014-01-15 水 の授業で修正・詳細化します．

ファイナルトライアル計画 II

- (離散的) 確率分布が与えられたとき, 母平均値, 母分散, 母標準偏差をもとめる (L08)
- (連続的) 確率分布 (のグラフ) が与えられたとき, 確率, 母平均値を求める (L09)
- 標準とは限らない正規分布について, どの範囲ならこんな確率, こんな確率ならこの範囲, を求める (L11)
- 標本が与えられたとき, 母集団の母平均値, 母分散, 母比率を推定する (L12)
- 区間推定 (L13)
- 未定 (補講=L14 の内容)