

科目名	統計の手法	担当者名	淵野 昌	所要時間	80分	2007年7月18日(水) 施行
持込	すべて可					
添付する 解答用紙	1 枚配付 (問題用紙の回収要・ <input type="checkbox"/>)		計算用紙 0 枚配付			

このテストの回答例と解説は、試験後に

<http://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-ss07-kimatsu.pdf>

として掲示する予定です。

1) サイズ N のデータ x_1, x_2, \dots, x_N の平均 \bar{x} , 分散 s^2 , 不偏分散 u^2 を N, x_1, x_2, \dots, x_N を用いた式であらわしてください。

2) 確率変数 X と Y がそれぞれ正規分布 $N(20, 30)$ と $N(10, 19)$ に従うとき, (a) $X + Y$ の従う確率分布が何になるか教えてください。 (b) $P(25 \leq X + Y \leq 40)$ を計算してください。

3) 次の から にあてはまる語句を教えてください:

“推測統計”は多数の個体からなる全体の統計量を知るために、全体のごく一部についてのデータをとって、この測定値から全体での値を結論する方法である。このときの、全体(の特性値の全体)のことを母集団とよび、この全体での、そこで着目している特性値の平均と分散を、それぞれ , とよぶ。これに対し、この全体の一部としてとられたデータを とよび、その平均と分散を、, とよぶ。また、このような全体の一部をできるだけ偏りのないようにとることを 抽出するという。

4) ある母集団からサイズが N のデータをとったときのデータの k 番目の数値を返す確率変数を X_k とする。3) での意味で、データのとりかたが偏りのないものになっていることに対応する、 X_1, \dots, X_N の満たすべき性質は何かを教えてください。

5) 分散が 49 であることがわかっている母集団から、サイズが N の標本をとったとき、標本平均が a なら、母平均 μ の信頼度 95% の信頼区間は何になるか教えてください。また、母平均を信頼度 99% での信頼区間の幅が 10 以下になるように推測するために必要な標本のサイズは何かを教えてください。

6) ある地方のカラスの成鳥 25 羽の体重を測定したところ、平均が 554.2g で分散は 36.0g² だった。

(a) このデータの不偏分散を求めてください。

(b) t -分布を用いて、この地方のカラスの成鳥の平均体重 μ の信頼度 95% の信頼区間を求めてください。

統計の手法 解答例と解説

1): 平均 \bar{x} , 分散 s^2 , 不偏分散 u^2 はそれぞれ,

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{k=1}^N x_k, \\ s^2 &= \frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2 = \frac{1}{n} \sum_{k=1}^N (x_k)^2 - \bar{x}^2 \\ u^2 &= \frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2\end{aligned}$$

である.

2): 正規分布の再生性 (教科書 p.96) から, $X + Y$ は $N(20 + 10, 30 + 19) = N(30, 49)$ に従う. したがって $Z = (X + Y - 30)/\sqrt{49} = (X + Y - 30)/7$ は標準正規分布 $N(0, 1)$ に従う. ここで

$$P(25 \leq X + Y \leq 40) = P\left(\frac{25 - 30}{7} \leq Z \leq \frac{40 - 30}{7}\right) = P\left(0 \leq Z \leq \frac{5}{7}\right) + P\left(0 \leq Z \leq \frac{10}{7}\right)$$

で, $\frac{5}{7} \approx 0.71$, $\frac{10}{7} = 1.43$ だから, 教科書の p.201 の数表により $I(0.71) = 0.2612$, $I(1.43) = 0.4236$ となっていることを使って, $P(25 \leq X + Y \leq 40) \approx 0.2612 + 0.4236 = 0.6848$ となる.

3): ア: 母平均 イ: 母分散 ウ: 標本 (サンプル) エ: 標本平均 オ: 標本分散 カ: 無作為 (抽出)
(カについては教科書には標本抽出という用語も見られるが, 講義では, 偏りのない標本抽出をすることを無作為抽出と言う, という説明をしていた)

4) X_1, \dots, X_N の試行の結果のデータが偏りのないものになっている (つまり理想的な無作為抽出がなされている) ということは, X_1, \dots, X_N がすべて母集団の分布に従い, X_1, \dots, X_N が互いに独立であることに対応する.

5) 母標準偏差 σ の判っている母集団について, サイズ N の標本の標本平均を \bar{x} とすると, 母平均 μ の信頼度 99% の信頼区間は,

$$\bar{x} - 2.58 \times \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + 2.58 \times \frac{\sigma}{\sqrt{N}}$$

として計算できる (教科書 p.114, ポイント). したがって, 信頼区間のこのきの幅は

$$2 \times 2.58 \times \frac{\sigma}{\sqrt{N}}$$

となる. ここでは $\sigma = \sqrt{49} = 7$ だから,

$$2 \times 2.58 \times \frac{7}{\sqrt{N}} \leq 10$$

この幅が ≤ 10 になることを主張する式となる. この不等式を解くと

$$2 \times 2.58 \times \frac{7}{10} \leq \sqrt{N} \Leftrightarrow 13 \approx \left(2 \times 2.58 \times \frac{7}{10}\right)^2 \leq N$$

したがって丸め誤差を考慮して標本のサイズが 14 より大きければ, この場合の信頼区間は 10 以内になる, と結論できる.

6) (a): サイズ N のデータの不偏分散 U^2 は S^2 をデータの分散とすると $\frac{n}{n-1} S^2$ で与えられるのだったので (教科書 p.117), ここでの不偏分散の値は

$$\frac{25}{24} \times 36 = 37.5$$

である.

(b): 母平均 μ の信頼度 95% の信頼区間は, 標本平均 $\bar{x} = 554.2$ と不偏分散 U^2 を用いて

$$\bar{x} - t_{N-1}(0.025) \frac{U}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{N-1}(0.025) \frac{U}{\sqrt{N}}$$

とあらわされる. ここで $U = \sqrt{U^2} = \sqrt{37.5} \approx 6.12$, また, 教科書 p.202 の数表から $t_{24}(0.025) = 2.064$ なので, これらの値を用いると,

$$554.2 - 2.064 \times \frac{\sqrt{37.5}}{\sqrt{25}} \leq \mu \leq 554.2 + 2.064 \times \frac{\sqrt{37.5}}{\sqrt{25}}$$

となり, 数値計算をすると, $551.6 \leq \mu \leq 556.8$ (g) という結果になる.