

データサイエンス

もくもく会

by  GRI

統計検定[®]2級合格を目指そう

問題1 (ベイズの定理)

ある会社の社員がどの応募媒体を経由して入社してきたかに関する調査結果がまとめられている

1) この会社の社員を1人選んできたときに、その人がインターン経由で入社している確率を計算せよ

※ 複数の媒体を利用した社員はいないとする

カバー範囲：事象と確率、加法定理、条件付き確率、乗法定理、ベイズの定理

応募の媒体	利用割合	利用した人の入社率
就職サイト	0.3	0.5
リファラル	0.2	0.7
インターン	0.2	0.2
ホームページ	0.2	0.4
エージェント経由	0.1	0.6

問題2 (統計量)

チョコクッキーを作っている会社の調査によると、1つの商品のパッケージに含まれるクッキーの数は以下のような確率で分布することが分かっている

	12個	13個	14個	15個
確率	0.1	0.3	0.45	0.15

- 1) 1つのパッケージに含まれる**クッキーの個数**の期待値、分散、標準偏差を求めよ
- 2) 包装容器、クッキーの1つあたりの重さがそれぞれ10 g、15 gと分かっている時、1つの**パッケージ全体の重さ**の期待値、分散、標準偏差を求めよ
- 3) パッケージを2つ開封したときの**クッキーの合計個数**の期待値、分散、標準偏差を求めよ
- 4) はじめにパッケージを1つ開封してその個数を記録し、その個数と同じかそれより多いクッキーが出るまで開封を続ける。このときの**1回目と最後に開封したパッケージのクッキーの個数**の共分散、相関係数を求めよ

カバー範囲：離散型確率変数、連続型確率変数、確率変数の期待値・分散・標準偏差、確率変数の和と差（同時分布、和の期待値・分散）、2変数の共分散・相関

問題3 (統計的推定)

次ページに示す小学校Aの小学6年生の身長と血液型のサンプルデータから学年全体の身長の情報を推定したい

ここで身長データは正規分布に従うと仮定し、以下信頼係数を $\alpha = 95\%$ とする

- 1) 小学校Aの小学6年生の身長の母分散が $\sigma^2 = 10 \text{ cm}^2$ と既知の場合、身長之母平均の信頼区間を推定せよ
- 2) 小学校Aの小学6年生の身長の母分散が未知の時、身長之母平均の信頼区間を推定せよ
- 3) 小学校Aの小学6年生の身長の母分散の信頼区間を推定せよ
- 4) 小学校Aの小学6年生の血液型がA型の母比率の信頼区間を推定せよ

ここからは2ページ後に示す小学5年生のサンプルデータも用いて推定を行う

- 5) 小学校Aの小学6年生と小学5年生の身長の母分散がそれぞれ $\sigma_1^2 = 6.2 \text{ cm}^2$ 、 $\sigma_2^2 = 5.6 \text{ cm}^2$ と既知であるとき母平均の差の信頼区間を推定せよ
- 6) 小学校Aの小学6年生と小学5年生の身長の母分散が未知だが等しいことがわかっているとき母平均の差の信頼区間を推定せよ
- 7) 小学校Aの小学6年生と小学5年生の身長の母分散が未知で等しいとは限らないとき母平均の差の信頼区間を推定せよ
- 8) 小学校Aの小学6年生と小学5年生の身長の母分散の比の信頼区間を推定せよ
- 9) 小学校Aの小学6年生と小学5年生の血液型がA型の母比率の差の信頼区間を推定せよ

カバー範囲：点推定、推定量と推定値、一性質、不偏性、信頼区間、信頼係数、大数の法則、中心極限定理、母集団、母数（母平均、母分散）、母平均・母分散の区間推定、母比率の区間推定、標準正規分布、t分布、カイ二乗分布、F分布、上側確率点（パーセント点）、分散（n-1で割る）、標準化（z得点）、標本平均の期待値・分散

小学校Aの小学6年生のサンプル8人の身長データ

ID	身長 [cm]	血液型
1	153	A
2	149	B
3	154	O
4	151	A
5	156	AB
6	149	A
7	151	O
8	153	O

小学校Aの小学5年生のサンプル10人の身長データ

ID	身長 [cm]	血液型
1	143	A
2	139	O
3	152	A
4	145	O
5	142	AB
6	138	B
7	151	B
8	150	A
9	143	O
10	136	A