

## 統計学 I 2022 年度春学期 問題と解答

【注意】以下の設問の解答は解答用紙(別紙)に記入しなさい。**\*印の問題**では答えに至る途中過程も記しなさい(答えだけは 0 点)。**検定**では、仮説、帰無分布、検定統計量、臨界値、結論を明記しなさい。注記がない限り、数字の回答は割り切れるものは小数点、割り切れないものは分数で記しなさい。

### 【数表】

標準正規分布  $Z$  について  $P(Z < 1) = 0.84$ 、 $P(Z < 1.645) = 0.950$ 、 $P(Z < 1.96) = 0.975$ 、 $P(Z < 2) = 0.977$

1 第 1 回レポートでは、総務省「家計調査」の市町村別の家計支出額データを利用した統計分析をおこなった。(25)

- (1)「家計調査」の標本世帯抽出方法を簡潔に説明しなさい
- (2)(1)の調査方法は、単純な無作為抽出に比べて、どのような利点があるか
- (3)標本平均の欠点を補うオリンピックの採点競技でも利用されている平均を何というか
- (4)箱ひげ図の箱の下限と上限は、それぞれ何を表わしているのか
- (5)平均値の値が大きく異なる分布の散らばりを比較する指標を何というか

解答 各 5 点 正答率(1)37%(2)28%(3)35%(4)64%(5)55%

(1)第 1 段階で市町村を、第 2 段階で単位区を、第 3 段階で世帯を選ぶ層化 3 段階抽出法(説明がないと 3 点)(2)偏りなく、標本分散が小さくなる(偏りがないだけは 2 点、不偏性や効率性は各 1 点)(3)刈り込み平均(刈り上げ平均は 0 点)(4)第 1 四分位、第 3 四分位(片方のみ正解は 2 点)(5)変動係数(誤字は 3 点、 $cv$  や  $s/\bar{x}$  は 2 点)

2 ある学生の学力を筆記試験(100 点満点)の点数  $x$  で評価するとき、

①1 回だけの試験で評価する  $m_1 = x_1$  ②2 回試験の平均値で評価する  $m_2 = (x_1 + x_2)/2$  ③1 回の試験と 10 点の出席点をもとに評価する  $m_3 = 0.9 \cdot x_1 + 10$  の 3 とおりの評価方法を考えた。(25)

(1)筆記試験の点数  $x$  が母平均  $\mu$ 、母分散  $\sigma^2$  の正規分布に従うと考えられる理由を CLT(中心極限定理)にもとづいて説明しなさい

(2\*)不偏性の観点から 3 つの指標の順位(優れた順、例えば 1,2=3)とその根拠となる式をかきなさい

(3\*)効率性の観点から 3 つの指標の順位(優れた順、例えば 1,2=3)とその根拠となる式をかきなさい

解答 各 5 点 正答率(1)21%(2)42%(3)41%

(1)各設問の得点  $d$  が確率変数のとき、得点合計  $\sum d_i$  は問題数が大きいと CLT より正規分布に従う

(2)1=2,3  $\therefore E(m_1) = E(m_2) = \mu$ ,  $\mu = 100$  を除いて  $E(m_3) = 10 + 0.9\mu \neq \mu$

(3)2,3,1  $\therefore \text{Var}(m_1) = \sigma^2$ ,  $\text{Var}(m_2) = 0.5\sigma^2$ ,  $\text{Var}(m_3) = 0.81\sigma^2$

3 河口湖のマス資源量  $N$ (匹)を知るために、釣り人が釣った 100 匹のマスに目印をつけてから放流し、翌週に再度、釣り人が釣ったマス 100 匹を調べたら、10 匹に目印がついていた。(20)

(1)放流後に釣ったマス 100 匹中の目印がついたマスの割合  $p = 10/100 = 0.1$  は確率変数であると考えられるが、どのような確率分布に従うかをかきなさい(ただし  $\theta = 100/N$  で定義される  $\theta$  を使うこと)

(2)なぜ  $p$  が(1)の確率分布に従うのかを CLT にもとづいて説明しなさい

(3\*) $\theta$  の 95%信頼区間の下限と上限を求めなさい

(4\*)資源量  $N$  の 95%信頼区間の下限と上限を求める**計算式**を示しなさい

解答 各 5 点 正答率(1)29%(2)39%(3)57%(4)20%

(1)平均  $\theta$ 、分散  $\frac{\theta(1-\theta)}{100}$  の正規分布  $[N(\theta, \frac{\theta(1-\theta)}{100})]$  も可

(2)100 匹それぞれに目印が付いているか否かの  $d_i$  はベルヌイ分布に従うが、その 100 個の和  $x = \sum d_i$  は CLT より正規分布に従い、比率  $p = x/100$  も正規分布に従う

(3)下限:  $0.1 - 1.96 \frac{\sqrt{0.1(1-0.1)}}{\sqrt{100}} = 0.1 - 1.96 \cdot 0.03 = 0.0412$  上限:  $0.1 + 1.96 \cdot 0.03 = 0.1588$  (計算ミスは減点)

(4)下限:  $100/0.1588$  上限:  $100/0.0412$

4 梅雨時の降雨確率が例年 0.5 であったとき、以下の問いに答えなさい。(20)

(1\*)梅雨時の 7 日中、少なくとも 2 日は雨が降る確率  $P(x \geq 2)$  を求めよ(解答は分数で)

(2\*)梅雨時の 36 日中、雨が降る日が 15 日以下となる確率を正規近似より求めよ

(3\*)今年の梅雨時 36 日間の降雨確率  $p$  は 0.4 であったが、これは例年より低いと言えるか。有意水準 5%

で検定をおこないなさい

解答 (1)(2)5 点 (3)10 点 正答率(1)83%(2)49%(3)38%

(1) $P(0)+P(1)=(1+7)/2^7=8/128=1/16$   $P(x \geq 2)=1-P(0)-P(1)=15/16$

(2) $x \sim N(18,9)$ より  $P(x \leq 15)=P(z \leq (15-18)/3=-1)=1-0.84=0.16$

(3) $H_0: \theta = 0.5, H_1: \theta < 0.5$ とする。帰無仮説  $H_0: \theta = 0.5$  が真であるとする、 $p \sim N(0.5, \frac{0.5(1-0.5)}{36})$ となるが、 $p=0.4$ を標準化すると  $z = \frac{0.4-0.5}{\sqrt{0.025}/\sqrt{36}} = \frac{-0.1}{0.5/6} = -6/5 = -1.2 > -1.645$ より  $H_0$ は棄却できない。例年より低いとは言えない(仮説設定の誤り、帰無分布の有無、検定統計量の計算ミス、臨界値の誤り、結論の誤りの point で 2 点 づつ減点)

5 感染率 1%の Covid19 の PCR 検査の感度(感染者を陽性と判定する確率)が 70%、特異度(非感染者を陰性と判定する確率)が 99%であるとき、以下の問いに答えなさい (ともに解答は分数で) (10)

(1\*)PCR 検査で陽性だった人が、本当に Covid19 に罹患している確率を求めなさい

(2\*)逆に PCR 検査で陰性だった人が、本当は Covid19 に罹患している確率を求めなさい

解答 各 5 点 正答率(1)70%(2)61%

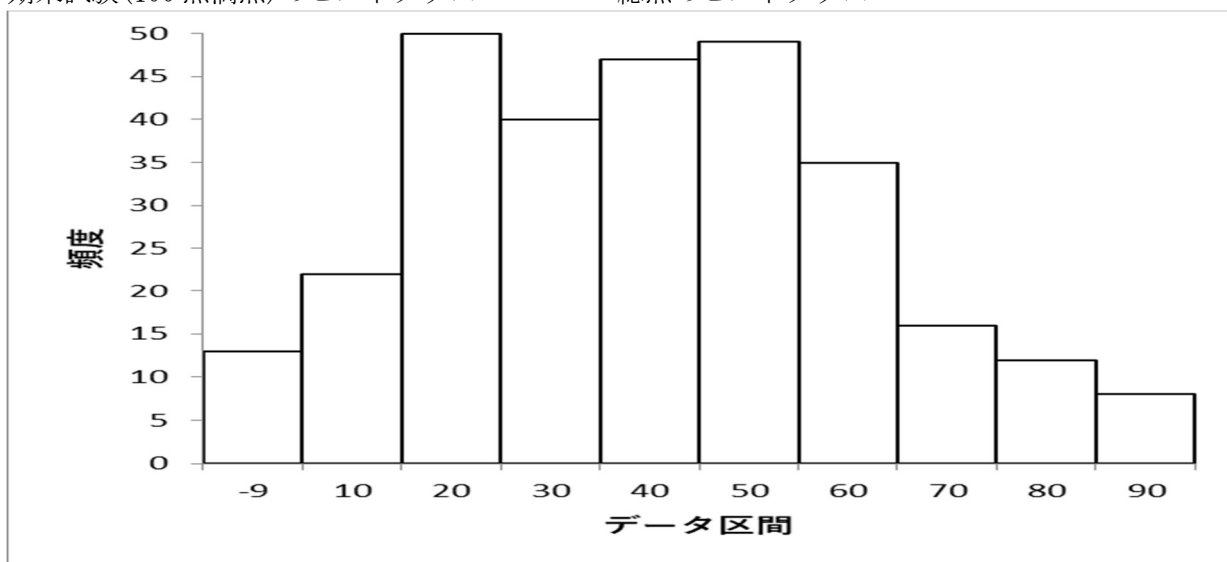
(1)  $70/(99+70)=70/169 (=0.414)$  (2)  $30/(9801+30)=30/9831=10/3277$

### 採点結果

受験者数 292 人 平均点 44.5 標準偏差 21.5 最高点 97 最低点 2

期末試験(100 点満点)のヒストグラム

総点のヒストグラム



成績は「総点＝レポート乗数の平均値×期末試験得点+小テスト得点」に基づき、S(総点 $\geq 90$ )、A( $90 >$ 総点 $\geq 70$ ) B( $70 >$ 総点 $\geq 45$ ) C( $45 >$ 総点 $\geq 25$ ) D(総点 $< 25$ )でつけます(経済学部ガイドラインに合わせるため、Sの割合を 10%、Dの割合を 10%となるように基準を変更しました)。