統計学Ⅱ　2022 年度秋学期　問題と解答

【注意】解答は解答用紙に記しなさい。必ず、学年、組、学籍番号、氏名を記入しなさい。＊印の問題は答えに至る途中過程も記しなさい(**答えだけでは 0 点**)。<u>検定では</u>仮説、検定統計量とその分布、判定、結論を明記しなさい。数字の解答は割り切れるものは<u>小数点表示</u>で、割り切れないものは<u>分数</u>で記しなさい。**注意に従わない答案は減点します**

【数表】

標準正規分布の右側 a%点 Z_a $(\Pr(Z>Z_a)=a)$: $Z_5=1.645$ $Z_{2.5}=1.96$

自由度 m の χ^2 分布の右側 a%点 $W_a(m)$ $(\Pr(W>W_a)=a)$: $W_{97.5}(9)=3$、$W_{5.0}(9)=17$、$W_{2.5}(9)=20$

自由度 m,n の F 分布の右側 a%点 $F_a(m,n)$ $(\Pr(F>F_a)=a)$: $F_5(9,9)=3.18$、$F_5(10,10)=2.97$

自由度 m の t 分布の右側 a%点 $t_a(m)$ $(\Pr(t>t_a)=a)$: $t_5(9)=1.8$, $t_5(18)=1.7$, $t_{2.5}(9)=2.3$, $t_{2.5}(18)=2.1$

$\sqrt{2}\fallingdotseq 1.4$　$\sqrt{6}\fallingdotseq 2.5$

1 日吉駅の 20 件の家賃データを集めて家賃 Y(万円)を専有面積 X(m²)で説明するモデル $Y_i=\alpha+\beta X_i+\varepsilon_i$ を excel を用いて最小 2 乗法で推定し、推定式 $Y_i=a+bX_i+e_i$ について以下の結果を得た。(25)

$Y_i = 3.691+0.1205X_i$　　決定係数 0.1443　標準誤差 1.300

(2.992)　(1.742)　　(　　)内は t 値

(1)最小 2 乗法とは何を最小化する方法か。解答は 5 文字の漢字で書きなさい。

(2)最小 2 乗法の 1 階の条件式を書きなさい。

(3)最小 2 乗推定量は BLUE と呼ばれる望ましい性質をもつが、この性質がなぜ望ましいかを説明せよ。

(4)レポートではデータを等間隔で幅広い範囲で集めることで係数の標準誤差を小さくできた。その理由を説明しなさい。

(5)「専有面積が広いと家賃が高くなる」という仮説を有意水準 5%で検定しなさい。

解答　各 5 点　正答率(1:100% 2:37% 3:50% 4:22% 5:86%)

(1)残差二乗和

(2) $\Sigma e_i=0$, $\Sigma e_iX_i=0$

(3)推定量に偏りがなく、分散が最小の精度の高い推定量だから

(4)係数の標準誤差 $s_b=\sigma/\sqrt{\sum(x_i-\bar{x})^2}$ より X の散らばりが大きいほど、標準誤差が小さくなるから

(5)H_0: $\beta=0$, H_1: $\beta>0$ とする。帰無仮説が真だとすると $t=b/s_b$ は自由度 18 の t 分布に従うが、観測値においては t=1.742>1.7 なので帰無仮説は有意水準 5%で棄却され、専有面積は有意に正の効果をもつといえる　※仮説の誤り(H0:b=0 とか H1: β>1 など)は 3 点減点。検定統計量が示されていない、結論が示されていないものは 2 点減点。 t 値が明示されているのに不思議な t 値を計算しているものは 3 点減点

2　平均 10、分散 5 の正規母集団からの大きさ n の無作為標本の標本分散 s^2 について答えよ(25)

(1)この標本分散 s^2 から作られる χ^2 分布に従う確率変数 W の算式を書きなさい。

(2)この W の自由度が n−1 となる理由を説明しなさい。

(3*)n=10 のとき標本分散 s^2 が 10 より大きくなる確率は 5%より大きいか。

(4)n が十分大きいとき、W が正規分布で近似できる理由をカイ 2 乗分布の定義をもとに説明しなさい。

(5*)(4)の性質を利用して n=51 の時の s^2 が 6.5 を超える確率は 5%より大きいか。

解答　各 5 点　正答率(1:72% 2:34% 3:36% 4:20% 5:7%)

(1)$W=(n-1)s^2/5$

(2)$W=(n-1)s^2/5=\Sigma(x_i-\bar{x})^2/5=\Sigma z_i^2$ だが、$\Sigma(x_i-\bar{x})=0$ だから自由度が 1 減少する

(3)$P(s^2>10)=P(W=\frac{9}{5}s^2>\frac{9}{5}10=18)<P(W>17)=5\%$ <u>5%より小さい</u>　途中計算が正しいのに「大きい」は-2 点

(4)$W=\Sigma z^2$ で n が十分大きいとき、中心極限定理より正規分布で近似できる

(5) $P(s^2>6.5)=P(W=\frac{50}{5}s^2>\frac{50}{5}6.5=65)$。W～N(50,100)より $P(Z>\frac{65-50}{10}=1.5)>P(Z>1.645)=0.05$ <u>5%より大きい</u> W に変換してから、標準化します　途中計算が正しいのに「小さい」は-2 点

3　以下の問に答えなさい。(25)

(1*)ポアソン分布の積率母関数(MGF)を $M_x(t)=\exp\{\lambda(e^t-1)\}$ とするとき、ポアソン分布の平均を導出せよ。

(2*)ポアソン分布の分散を MGF から導出せよ。

(3*)確率変数 X、Y がともに平均 λ のポアソン分布に従うとき、X+Y もポアソン分布に従うことを示せ。

(4)2022FIFA ワールドカップの 1 試合の平均得点は 1 点だった。サッカーの得点がポアソン分布に従う

理由をベルヌイ分布、2 項分布との関係を書きながら説明しなさい。

(5*)このワールドカップで 3 点以上の得点が入る確率を求めよ。ただし、ポアソン分布の確率関数 $f(x)=e^{-\lambda}\frac{\lambda^x}{x!}$ を利用し、e=3 として近似しなさい。

解答　各 5 点　正答率(1:47% 2:32% 3:38% 4:38% 5:67%)

(1)$M_x(t)$の t に関する 1 階微分は $e^{\lambda(e^t-1)}\lambda e^t$ なので平均は $M'(0)=e^{\lambda(e^0-1)}\lambda e^0=\lambda$

(2) 2 階微分は $e^{\lambda(e^t-1)}\lambda e^t+e^{\lambda(e^t-1)}(\lambda e^t)^2$ なので分散は $M''(0)-\mu^2=\lambda+\lambda^2-\lambda^2=\lambda$

(3)X、Y の積率母関数をそれぞれ $M_x(t)=e^{\lambda(e^t-1)}$、$M_Y(t)=e^{\lambda(e^t-1)}$ とすると X+Y の積率母関数は $M_{X+Y}(t)=M_x(t)M_Y(t)=e^{\lambda(e^t-1)}e^{\lambda(e^t-1)}=e^{2\lambda(e^t-1)}$ となり、やはりポアソン分布となる。

(4) 1play で得点が入るか否かはベルヌイ分布だが、1 試合はそのプレーの積み重ねなので、2 項分布に従うが、得点確率が低く、play の数も数多いので、ポアソン分布に従う。1 試合あたりの得点はベルヌイ分布ではない。試合数が多いからではない。各分布の言及していないと 0 点

(5) $\lambda=1$ のとき $f(0)=1/3$, $f(1)=1/3$, $f(2)=1/6$ より $P(x\geqq 2)=1-f(0)-f(1)-f(2)=1/6$

4　近所のスーパー2 店(A 店,B 店)の LL サイズの卵をそれぞれ 1 パック(10 個)買って、卵 1 個あたり重量を測定したところ A 店、B 店それぞれの標本平均$\bar{x}$は 72.5g と 70g、標本標準偏差 s は$\sqrt{10}$g と$\sqrt{5}$g であった。(25)

(1*)この結果から A 店の卵 1 個あたり重量の母平均μ_Aの 95%信頼区間を求めよ。

(2*)この結果から A 店の卵 1 個あたり重量の母分散σ_A^2の 95%信頼区間を求めよ。

(3*)卵 1 個あたり重量の母分散は A 店＞B 店といえるか。有意水準 5%で検定を行いなさい。

(4*)卵 1 個あたり重量の母平均は A 店＞B 店といえるか。有意水準 5%で検定を行いなさい。

解答　(1)-(3)5 点(4)10 点　正答率(1:54% 2:74% 3:85% 4:72%)

ここで点数を稼いでいる答案が多かったが、仮説の誤り、検定統計量とその分布が示されない、判定がない場合は減点した

(1)$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}\sim t(n-1)$より $P[-2.3<t<2.3]=P[\bar{x}-2.3\frac{\sqrt{10}}{\sqrt{10}}<\mu<\bar{x}+2.3\frac{\sqrt{10}}{\sqrt{10}}]=P[\underline{70.2}<\mu<\underline{74.8}]=0.05$

(2)$W=(n-1)s^2/\sigma^2\sim\chi^2(9)$より $Pr[3<(n-1)s^2/\sigma^2<20]=Pr[(n-1)s^2/20<\sigma^2<(n-1)s^2/3]=Pr[9\cdot10/20<\sigma^2<9\cdot10/3]=Pr[\underline{4.5}<\sigma^2<\underline{30}]=0.05$　計算ミスは 5 点

(3)H_0: $\sigma_a^2=\sigma_b^2$　H_1: $\sigma_a^2>\sigma_b^2$　H_0 が真だとすると $F=s_a^2/s_b^2\sim F(9,9)$に従うが、調査結果より $F=10/5=2<F_5(9,9)=3.18$ なので有意水準 5%で H_0 は採択され、散らばりは等しいといえる

(4)H_0: $\mu_a=\mu_b$ H_1: $\mu_a>\mu_b$　H_0が真だとすると $t=(\bar{x}_a-\bar{x}_b)/(s\sqrt{\frac{1}{n}+\frac{1}{n}})$は自由度 2(n−1)の t 分布に従うが、調査結果より $s^2=\frac{1}{2}10+\frac{1}{2}5=15/2$, $s\sqrt{\frac{1}{n}+\frac{1}{n}}=\sqrt{\frac{15}{2}\frac{1}{5}}=\frac{\sqrt{6}}{2}\fallingdotseq1.25=\frac{2.5}{2}$。$t=(72.5-70)/(2.5/2)=2>t_5(18)=1.7$ なので H_0 は棄却される。よって有意水準 5%で A 店の方が B 店より重いといえる。

※片側検定なので臨界値は 1.7 です。計算ミスは適宜減点

総評

受験者数 305　平均 51.0　中央値 52　標準偏差 22　最低 0　最高 93

成績は、総点＝レポート乗数×期末得点+小テスト得点で、以下の基準で成績をつけます。

S：総点≧95　A：80≦総点＜95　B：50≦総点＜80　C：20≦総点＜50　D：総点＜20

（S と D の割合は学部指針に合わせるためにボーダーを決定しました。）

成績の人数分布は S：11%　A：17%　B：36%　C：27%　D：9%　となりました。