経済統計概論試験問題（第１部用）　　　　　＜９０分＞

持ち込み：電卓のみ（携帯電話，PHS，パソコン，電子辞書，PDA など，通信機能を持つもの，あるいは，それと一見によっては区別がつけがたいものを電卓の代わりに持ち込み，使用することは認めない）

注意：解答用紙には答えのみを記入すること．

１．以下の問に答えよ．

（1）６００世帯の無作為抽出調査によって，２００４年のある番組の視聴率は３９．３％であった．母集団における実際の視聴率の９９％信頼区間を求めよ（%表示で計算せよ．この表示方法で０．１％まで計算せよ）．（１５点）

＜解説＞

ｎを標本数，ｍを調査時に支持すると回答した数とすると，９９％信頼区間は $\frac{m}{n}-2.58\times\frac{\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$ から $\frac{m}{n}+2.58\times\frac{\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}}{\sqrt{n}}$ であるから，$n=600, m/n=0.393$ を代入して，０．３４２から０．４４４を得る．

（答え）

３４．２％から４４．４％（±0.1%ポイントを正答と見なす）

＜答え＞

（2）９９％信頼区間の幅が０．５％ポイントになるようにするには，何世帯を対象に調査を行えばよいか．（５点）

＜解説＞

９９％信頼区間の幅の最大値は，$2\times2.58\times\frac{\sqrt{0.5\times0.5}}{\sqrt{n}}=\frac{2.58}{\sqrt{n}}$．これが，0.005 より小さければよいので，$2\times2.58\times\frac{\sqrt{0.5\times0.5}}{\sqrt{n}}=\frac{2.58}{\sqrt{n}}\le0.005$ より，$n\ge266257$ 以上．（万以上の桁の数字が一致すればよい．）

ただし，（1）の場合に合わせて，$m/n=0.393$ として計算した場合も正答とする．この場合は，$2\times2.58\times\frac{\sqrt{0.393\times(1-0.393)}}{\sqrt{n}}\le0.005$ より，254064 万以上の桁の数字が一致すればよい．）

＜答え＞２６６２５７（万以上の桁の数字が一致すればよい．）

２．母集団が正規分布に従っているとする．その中から５つの標本を取り出した場合に，その標本平均は７，データの値の合計は１０，データの値の２乗の合計は，１６４であっ

た．このとき，母平均の９５％信頼区間を求めよ．（小数点以下第１位まで求めよ）　（２０点）

<お詫び>

標本平均とデータの値の合計が矛盾していました．標本平均７をそのままとると，$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}\overline{X}^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2=164-5\times7^2=-81$となって，データが実数値ではあり得なくなります．そのことを指摘した場合は，それだけで，２０点満点とします．データの値の合計１０から計算した２が標本平均の正しい値です．

<解説>

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}\overline{X}^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2=164-5\times2^2=144$$

$S^2=\dfrac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}=\dfrac{144}{4}=36$．$\dfrac{\sqrt{n}\left(\overline{X}-\mu\right)}{S}$ が自由度４のｔ分布に従うので，$\overline{X}-t\dfrac{S}{\sqrt{n}}\le\mu\le\overline{X}+t\dfrac{S}{\sqrt{n}}$．ただし，ｔは自由度４のｔ分布の９７．５％点である．教科書の数表によると，２．７７７である．したがって，９５％信頼区間は，

$2-2.777\dfrac{6}{\sqrt{5}}\le\mu\le2+2.777\dfrac{6}{\sqrt{5}}$．－5.45 から 9.45.

また，標本平均７を信じて$7-2.777\dfrac{6}{\sqrt{5}}\le\mu\le7+2.777\dfrac{6}{\sqrt{5}}$で，―０．４５から１４．４５とした場合も正解として２５点とする．

<答え>－5.45 から 9.45<ここまでできたら２５点とする>

３．標本数が４００，標本平均が４０，分散の推定値は６４であった．なお，分散の推定値は，$\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ で求めた．母集団平均が４５であるかどうかについて，仮説検定を行う．有意水準は１％とする．以下の問いに答えよ．（各４点）

ア．母集団平均をμとしたとき，帰無仮説を述べよ．

<答え>$\mu=45$

イ．母集団平均をμとしたとき，対立仮説を述べよ．

<答え>$\mu\neq45$

ウ．検定統計量の分子が（標本平均－....）となるように，検定統計量Ｔを定めたとき，棄却域と境界値をＴと実際の数字の不等号，等号で示せ．

<解説>有意水準１％で両側検定なので，標準正規分布の９９．５％点をもとめ，それを元に棄却域を構成する．標準正規分布の９９．５％点は，２．５８である．

＜答え＞$T < -2.58$ または $T > 2.58$

エ．検定統計量の分子が（標本平均－．．．．）となるように，検定統計量 T を定めたときの検定統計量値を計算せよ．

＜答え＞$\dfrac{\sqrt{n}\left(\overline{X}-45\right)}{\sigma} = \dfrac{\sqrt{400}\left(40-45\right)}{\sqrt{64}} = -12.5$

オ．帰無仮説は棄却されるか答えよ．

＜答え＞棄却

４．ある集団は，A群とB群に分けることができる．A群について，標本数を１００について調査したところ，標本平均が５０，分散の推定値は１００であった．また，B群について，標本数が４００について調査したところ，標本平均は４５，分散の推定量は８１であった．なお，分散の推定値は，$\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ で求めた．A群の母平均がB群の母平均より大きいかどうかについて，仮説検定を行う．有意水準は５％とする．母分散はA群とB群で等しいものと考えること．以下の問いに答えよ．（各４点）

ア．A群の母平均を μ_A，B群の母平均を μ_B としたとき，帰無仮説を述べよ．

＜答え＞$\mu_A = \mu_B$

イ．A群の母平均を μ_A，B群の母平均を μ_B としたとき，対立仮説を述べよ．

＜答え＞$\mu_A > \mu_B$

ウ．検定統計量の分子が（μ_A －．．．．）となるように，検定統計量 T を定めたとき，棄却域と境界値を T と実際の数字の不等号，等号で示せ．

＜解説＞「より大きい」という検定なので，片側で上方に棄却域がとられる．有意水準が５％なので，標準正規分布の９５％点，すなわち１．６４が境界値．

＜答え＞$T > 1.64$

エ．検定統計量の分子が（μ_A －．．．．）となるように，検定統計量 T を定めたときの検定統計量値を計算せよ．

＜解説＞

$\hat{\sigma}^2_{\text{母集団}} = \dfrac{n_A\hat{\sigma}^2_A + n_B\hat{\sigma}^2_B}{n_A + n_B}$．ただし，$n_A, n_B$ は A 群，B 群の標本数，$\hat{\sigma}^2_A, \hat{\sigma}^2_B$ は A 群，B 群の分散推定値とする．$\hat{\sigma}^2_{\text{母集団}} = \dfrac{n_A\hat{\sigma}^2_A + n_B\hat{\sigma}^2_B}{n_A + n_B} = \dfrac{100\times100 + 400\times81}{100+400} = 84.8$．

$$\frac{\overline{X}_A - \overline{X}_B}{\hat{\sigma}_{\text{母集団}}\sqrt{\dfrac{1}{n_A}+\dfrac{1}{n_A}}} = \frac{50-45}{\sqrt{84.8}\sqrt{\dfrac{1}{100}+\dfrac{1}{400}}} = 4.86$$

（ちなみに A 群と B 群の分散が異なるとした場合の検定統計量値は，

$$\frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_A}}} = \frac{50-45}{\sqrt{\frac{100}{100} + \frac{81}{400}}} = 4.56）$$

＜答え＞４．８６

オ．帰無仮説は棄却されるか答えよ．

＜棄却される＞

５．以下の問いに答えよ．

（１）データ数が２０で標本平均が４，$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ で推定した分散は５であった．チェビシェフの不等式によって，データが平均より $k\hat{\sigma}$ 以上離れたデータの数の全データ数に対する比率の上限を計算し，これが１／２０を下回るｋの下限を求めよ．この関係からわかる限りでのこのデータの最大値の上限を求めよ．（１０点）

＜解説＞

標本平均より $k\hat{\sigma}$ 以上離れたデータ数の比率は $1/k^2$ 以下．これが，１／２０を下回っていると言うことは，この範囲にはデータがないということ．$1/k^2 < 1/20$ をといて，$k > \sqrt{20} = 4.47$．従って，$標本平均 + \sqrt{20}\hat{\sigma} = 4 + \sqrt{20} \times \sqrt{5} = 14$ より大きいところにはデータは絶対存在しない．従って，チェビシェフの不等式からわかる最大値の上限は，１４．

＜答え＞１４

（２）５．全国の男子大学生 10000 人の体重測定を実施した．平均５５kg，標準偏差１０kg の正規分布をしていた．以下の問に答えよ．

ア．体重 58kg 以上の人は何人いるか（３点）

＜解説＞

X を体重，Z を標準席分布に従う確率変数として

$$P(X \geq 58) = P\left(\frac{X-55}{10} \geq \frac{58-55}{10}\right) = P(Z \geq 0.3) = 1 - P(Z \leq 0.3) = 1 - 0.6179 = 0.3821$$

＜答え＞３８２１人

イ．体重 40kg 以上の人は何人いるか（３点）

＜解説＞

$$P(X \geq 40) = P\left(\frac{X-55}{10} \geq \frac{40-55}{10}\right) = P(Z \geq -1.5) = P(-Z \geq -1.5)$$
$$= P(Z \leq 1.5) = 0.9332$$

＜答え＞９３３２人

ウ．体重 42kg 以上 52kg の人は何人いるか（４点）

＜解説＞

$$P(42 \le X \le 52) = P\left(\frac{42-55}{10} \le \frac{X-55}{10} \le \frac{52-55}{10}\right) = P(-0.7 \le Z \le -0.3)$$

$$= P(-1.3 \le -Z \le -0.3) = P(0.3 \le Z \le 1.3)$$

$$= P(Z \le 1.3) - P(Z \le 0.3) = 0.9032 - 0.6179 = 0.2853$$

＜答え＞２８５３人