2005年度　統計工学　期末試験問題

[1]　$X_1, X_2, \cdots, X_n$ が標準偏差<u>既知</u>の正規分布 $N(\mu, \sigma^2)$ に独立にしたがうとする．

帰無仮説と対立仮説がそれぞれ

H$_0$：$\mu = \mu_0$

H$_1$：$\mu \neq \mu_0$

のときの尤度比検定を導け（テキスト p.80，例 4.2 は n＝1，μ_0＝0，σ＝1）．

[2]　ある化学会社では，最近需要が増加してきている化学製品を A と B の２つの工場で製造している．２つの工場の生産性を比較するため，単位時間当たりの生産量のデータを，A 工場より n_1＝10，B 工場より n_2＝9 それぞれ採取した．データに正規性と等分散性を仮定し，２つの工場での平均値の差の検定を行え．

工場 A：78, 80, 79, 83, 82, 85, 78, 74, 76, 84

工場 B：81, 84, 82, 88, 86, 83, 78, 84, 89

なお，基本統計量として工場 A では，標本平均 $\bar{x}_1$＝79.9，平方和 S_1＝114.9

工場 B では，標本平均 $\bar{x}_2$＝83.9，平方和 S_2＝94.9　である．

[3]　マンション建築における鉄筋工事に投入する鉄筋工の工数を予測する回帰式を作成した．目的変数 y は工数（単位は人・日），説明変数は消費鉄筋量（単位はトン）を x_1，延床面積（単位はm^2）を x_2 とした．最近の n＝40 実績データより推定された回帰式は次のようになった．

$$\hat{y} = 138.91 + 4.134x_1 - 0.098x_2$$

この回帰式の解釈として，次の２つの文はいずれも誤りである．その理由を述べよ．

(1)　x_1 の係数は 4.134 で，x_2 の係数－0.098 よりも絶対値が大きい．よって，消費鉄筋量の工数への影響は，延床面積の工数への影響よりも大きい．

(2)　x_2 の係数は－0.098 と負値である．よって，目的変数 y と x_2 の単相関係数も必ず負値になっている．

[4]　ある週にある地域で発生した食中毒に対して，食品 A を摂取したかどうかを，患者 50 名と健常者 100 名について尋ねた結果を左下の分割表にまとめた。

	患者	健常者	計
摂取	42	13	55
非摂取	8	87	95
計	50	100	150

(1)　食品 A の摂取と食中毒発生に関連性があるかどうかを χ^2 検定せよ．

(2)　食品 A を摂取した人で食中毒になった人の割合を 42／55 で推定するのは不適切であることの理由を述べよ．

基準解答

[1]　尤度比検定の棄却域は

$$\frac{\max_{\mu} f(x_1, x_2, \cdots, x_n; \mu)}{f(x_1, x_2, \cdots, x_n)} \geq c$$

の形で与えられる．左辺の分子を最大にする μ は最尤推定値 $\bar{x}$ であるから，上式左辺は

$$\frac{\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left[-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}\right]}{\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left[-\frac{\sum (x_i - \mu_0)^2}{2\sigma^2}\right]} = \exp\left[-\frac{\sum (x_i - \bar{x})^2 - \sum (x_i - \mu_0)^2}{2\sigma^2}\right]$$

$$= \exp\left[\frac{n(\bar{x} - \mu_0)^2}{2\sigma^2}\right]$$

となる．これは $U = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ の実現値 u の 2 乗の単調増加関数であるから，結局

$$|u| \geq c'$$

という形の棄却域を与えることになる．

[2]　共通の分散 σ^2 に対する（十分統計量に基づく）不偏推定値は

$$V = \frac{S_1 + S_2}{n_1 + n_2 - 2} = \frac{114.9 + 94.9}{10 + 9 - 2} = 12.34$$

である．よって t 検定統計量の値は

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{V(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{79.9 - 83.9}{\sqrt{12.34 \times (\frac{1}{10} + \frac{1}{9})}} = -2.48$$

$t(17; 0.05) = 2.11$ なので，有意差がある．

[3]　(1)　偏回帰係数の値は説明変数の単位に依存するので、絶対値の大きさで影響の大きさを比べることは適切でない．標準偏回帰係数や t 検定統計量のような単位に依存しない統計量で評価する必要がある．

(2)　目的変数 y に対して，x_2 のみを説明変数としたときの単回帰係数 $\hat{\beta}_2$ の符号と両

者の単相関係数 r_{2y} の符号は必ず一致する．一方，目的変数 y に対して，x_1 と x_2 を説明変数としたときの偏回帰係数をそれぞれ $\hat{\beta}_{1\bullet 2}$，$\hat{\beta}_{2\bullet 1}$ としたとき

$$\hat{\beta}_2 = \hat{\beta}_{2\bullet 1} + \frac{S_{12}}{S_{22}}\hat{\beta}_{1\bullet 2}$$

の関係があるから，$\hat{\beta}_{2\bullet 1}$ が負値であっても，x_1 と x_2 の間に強い正の相関があり，$\hat{\beta}_{1\bullet 2}$ も正値であれば，$\hat{\beta}_2$ は正値になりえる．

[4]　(1)　2×2 分割表の χ^2 適合度検定統計量は

$$\chi^2 = n\frac{(n_{11}-n_{22})^2}{n_{1+}n_{2+}n_{+1}n_{+2}} = 150\frac{(42\times 87-13\times 8)^2}{55\times 95\times 50\times 100} = 72.36$$

となり，$\chi^2(1;0.05)$＝3.84 なので 5%有意である．

(2) セル度数／行和は，患者と健常者の数に依存するため，このような列和を所与としたサンプリングでは「食品 A を摂取した人で食中毒になった人の割合」を推定することはできない．