2003 年度　統計工学期末試験問題

テキスト、ノート、プリント持ち込み可。電卓使用可。解答は結果だけでなく、導出過程を要領よく記述すること。

１．　一般健康診断では、たとえば肝機能の特性値である GPT については 22 以下というように、各計測項目について正常範囲が規定され、これに基づき「問題なし」、「要再検査」の判定が下される。この論理を統計的仮説検定の立場から記述し、その場合の帰無仮説、対立仮説、および第１種の誤りと第２種の誤りを示せ。また、現行法の合理性・非合理性についても考察せよ。

２．　正規分布 $N(\mu,\sigma^2)$ を仮定し、分散についての帰無仮説

H_0：$\sigma^2=1$

と片側対立仮説

H_1：$\sigma^2>1$

を設定し、有意水準 $\alpha=0.05$ の仮説検定を行う。標本数は $n=21$ とする。
真の値が $\sigma^2=2$ のときの検出力を求めよ。数表から適当に補間すればよい。

３．　説明変数が２つの重回帰モデル

$$y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\varepsilon_i \quad (i=1,2,\cdots,n)$$

を想定し、n 組のデータを採取した。基本統計量の値は

$$\sum y_i=485.3 \qquad \sum x_{1i}=112 \qquad \sum x_{2i}=180$$

$$\sum y_i^2=24168.87 \qquad \sum x_{1i}^2=1368 \qquad \sum x_{2i}^2=3378$$

$$\sum x_{1i}y_i=5551.2 \qquad \sum x_{2i}y_i=9002.4 \qquad \sum x_{1i}x_{2i}=2028$$

であった。このとき、パラメータ $\alpha,\ \beta_1,\ \beta_2$ の最小２乗推定値を求めよ。

４．　寿命を表す確率変数 T が形状母数 $m=2$ のワイブル分布にしたがうとする。このとき、２つの正の実数 $t,\ s$ において

$$P\{T>t+s \mid T>t\} \quad と \quad P\{T>s\}$$

の大小関係を示せ。ヒント：$m=1$（指数分布）のとき両者は一致する。

基準解答

１．　帰無仮説：正常である。

対立仮説：正常でない（何らかの疾患の疑いがある)。

第１種の誤り：正常であるにもかかわらず「要再検査」と判定され、結果として無駄な時間と費用を要すること。

第２種の誤り：何らかの疾患であるにもかかわらず「問題なし」と判定され、最悪の場合には手遅れになり死に至ること。

解説:「正常でない」ことを帰無仮説にとる考え方もあろう。しかし、これは正しくない。それは、判定の根拠が正常人における検査項目の分布をもとにしているからである。正常でない人の分布を設定することは難しい。それは第一に、正常でないパターンが無数にあること、第二に各パターンの患者数が少ないことである。

このように帰無仮説と対立仮説は対称でなく、判定のもとになっている分布の背後にある状態が帰無仮説になる。この点では一般健康診断のロジックは合理的といえる。ただし問題もある。それは第一に検査項目が数多くあり、健康診断全体での有意水準が不明な点である。第二に検査項目間には互いに相関があるのだが、それが全く考慮されていない点である。そのため、正常人でも、いくつかの項目で正常範囲から外れ、再検査の結果「問題なし」と判定されることが多い。しかし、第１種の誤りと第２種の誤りによる損失を秤にかければ、有意水準をかなり大きめに設定する現行法にはそれなりの妥当性があるといえる。

２．　検定統計量はテキスト p.99、問題 4.2 にあるように

$$\chi^2 = \frac{\sum (X_i - \overline{X})^2}{1}$$

で、片側（上側）検定なので棄却判定値は数表より $\chi^2(20,0.05)$ ＝31.41 である。

母分散の真値が σ^2 ＝2 のとき、上記の $\sum (X_i - \overline{X})^2$ が 31.41 以上になる確率を求めればよい。このとき

$$\frac{\sum (X_i - \overline{X})^2}{2}$$

が自由度 20 のカイ２乗分布にしたがう。求める確率は

$$\mathrm{Pr}\{\sum (X_i - \overline{X})^2 > 31.41\} = \mathrm{Pr}\{\frac{\sum (X_i - \overline{X})^2}{2} > 15.705\}$$

である。数表では上側 75%点が 15.45 だから、補間すると約 0.733 となる。

３．　各変数の偏差平方和と変数間の偏差積和を次式で求める。

$$S_{11} = 1368 - (112)^2 / n$$
$$S_{22} = 3378 - (180)^2 / n$$
$$S_{12} = 2028 - 112 \times 180 / n$$
$$S_{1y} = 5551.2 - 112 \times 485.3 / n$$
$$S_{2y} = 9002.4 - 180 \times 485.3 / n$$

これより β_1 と β_2 の最小２乗推定値は

$$\hat{\beta}_1 = \frac{S_{1y}S_{22} - S_{12}S_{2y}}{S_{11}S_{22} - S_{12}^2}$$
$$\hat{\beta}_2 = \frac{S_{2y}S_{11} - S_{12}S_{1y}}{S_{11}S_{22} - S_{12}^2}$$

で与えられる。問題文では標本数 n が記載されていないので、数値としては定まらない。さらに、各変数の標本平均を

$$\bar{y} = 485.3 / n, \quad \bar{x}_1 = 112 / n, \quad \bar{x}_2 = 180 / n$$

と求めておいて、α の最小２乗推定値は

$$\hat{\alpha} = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

で与えられる。これも数値は定まらない。

４．　形状母数 m ＝2 のワイブル分布の分布関数は

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\eta} \right)^2 \right]$$

であるから、まず

$$P\{T > s\} = \exp\left[-\left(\frac{s}{\eta} \right)^2 \right] = \exp\left[-\frac{s^2}{\eta^2} \right] \qquad (1)$$

である。

一方、条件付き確率の定義より

$$P\{T > t + s | T > t\} = \frac{P\{T > t + s\}}{P\{T > t\}}$$

であるから、これにワイブル分布をあてはめれば

$$P\{T > t + s | T > t\} = \frac{\exp\left[-\left(\frac{t+s}{\eta}\right)^2\right]}{\exp\left[-\left(\frac{t}{\eta}\right)^2\right]}$$

である。これを変形して整理すると

$$P\{T > t + s | T > t\} = \exp\left[-\frac{2ts + s^2}{\eta^2}\right] \qquad (2)$$

となる。ここに、(1)式と(2)式において

$$2ts + s^2 > s^2$$

の関係があるので

$$P\{T > t + s | T > t\} < P\{T > s\}$$

という結論が得られる。これはすなわち、形状母数 $m = 2$ のワイブル分布では、時間とともに個体が劣化していくことを意味する。