応用統計学 2017　第10回　正規線形モデル

2017 年 12 月 13 日 (水)
清 智也　sei@mist.i.u-tokyo.ac.jp
http://ur0.pw/yTzt

- 正規線形（回帰）モデル，F 検定，両側検定，片側検定，t 検定，分散分析[1]。
- 次回は一般化線形モデル，最終回は情報量規準を扱う予定。

演習問題

問題 10-1. データ $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ が正規線形モデルに従うと仮定する：

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n), \quad \boldsymbol{\mu} \in M, \quad \sigma^2 > 0, \tag{1}$$

ただし M は $\mathbb{R}^n$ の線形部分空間である。いま M_0 を M の線形部分空間とし，帰無仮説 $\mathrm{H}_0 : \boldsymbol{\mu} \in M_0$ を考える。この場合の対数尤度比検定統計量 $T(\boldsymbol{y})$ を求めよ。また，$T(\boldsymbol{y})$ は次の統計量と単調増加の関係にあることを示せ：

$$F(\boldsymbol{y}) = \frac{\|\boldsymbol{P}\boldsymbol{y} - \boldsymbol{P}_0\boldsymbol{y}\|^2/(p - p_0)}{\|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}\|^2/(n - p)}.$$

ただし $\|\cdot\|$ はユークリッドノルムを表す。また $\boldsymbol{P}$ と $\boldsymbol{P}_0$ はそれぞれ M と M_0 への直交射影行列を表し，p と p_0 はそれぞれ M と M_0 の次元である。

補足：先週出題したレポート課題 2 と同様の手順により，帰無仮説のもとで $F(\boldsymbol{Y})$ が F 分布に従うことが示される。この分布に基づく検定方式が F 検定である。

問題 10-2. 単回帰モデル

$$Y_i = a + bx_i + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2), \quad i = 1, \ldots, n,$$

を正規線形モデル (1) の形で表せ。また帰無仮説 $b = 0$ に対応する部分空間 M_0 も求めよ。

問題 10-3. 次の 2 つの検定の違いを説明せよ。統計モデルや帰無仮説は適宜設定せよ。

(i) 対応のある二群の平均値の差の検定。

(ii) 対応のない二群の平均値の差の検定。

[1] normal linear (regression) model, F-test, two-sided test, one-sided test, t-test, analysis of variance.

問題 10-4. 重回帰モデル $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, を考える。回帰係数 β_1 の検定について以下の問いに答えよ[2]。ただし，$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$，$\boldsymbol{X}_0 = (\boldsymbol{x}_2, \ldots, \boldsymbol{x}_p)$ とおき，$\boldsymbol{X}$ の列空間，$\boldsymbol{X}_0$ の列空間への直交射影行列をそれぞれ $\boldsymbol{P}$，$\boldsymbol{P}_0$ とおく。

(i) 最小二乗推定量 $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ の共分散行列は $\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$ で与えられることを示せ。このことから回帰係数 $\hat{\beta}_1$ の標準誤差は $\mathrm{se}(\hat{\beta}_1) = \hat{\sigma}\sqrt{((\boldsymbol{X}'\boldsymbol{X})^{-1})_{11}}$ と見積もられる。ここで $\hat{\sigma}^2 = \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}\|^2/(n-p)$ は σ^2 の不偏推定量である（問題 6-2 参照)。

(ii) $\boldsymbol{r} = \boldsymbol{x}_1 - \boldsymbol{P}_0 \boldsymbol{x}_1$，$\boldsymbol{e} = \boldsymbol{r}/\|\boldsymbol{r}\|$ とおくとき，次の等式を示せ：

$$t(\boldsymbol{y}) := \frac{\hat{\beta}_1 - \beta_1}{\mathrm{se}(\hat{\beta}_1)} = \frac{\boldsymbol{e}'(\boldsymbol{y} - \beta_1 \boldsymbol{x}_1)}{\hat{\sigma}}.$$

補足：レポート課題 2 と同様の手順により，$t(\boldsymbol{Y})$ は t 分布に従うことが示される。この結果に基づく検定方式が t 検定であり，普通は帰無仮説を $\beta_1 = 0$ として $t(\boldsymbol{y})$ の値を検討する。同時に，母回帰係数 β_1 の厳密な信頼区間を求めることもできる。

宿題 10

問題 10-5 (1 元配置分散分析). 下のデータは，ミニ四駆の走行タイムを測定したものである。モーターを A_i としたときの走行タイムが平均 a_i，分散 σ^2 の正規分布に従うと仮定して，帰無仮説を $a_1 = a_2 = a_3$ とする F 検定を行いたい。

モーター	走行タイム [秒]			
トルクチューン (A_1)	16.66	15.28	14.19	15.95
レブチューン (A_2)	15.60	15.99	15.73	15.55
プラズマダッシュ (A_3)	13.53	13.47	14.96	13.47

統計ソフトウェア R で実行したところ，以下の結果を得た。この結果を解説せよ。

```
> y1 = c(16.66, 15.28, 14.19, 15.95)
> y2 = c(15.60, 15.99, 15.73, 15.55)
> y3 = c(13.53, 13.47, 14.96, 13.47)
> y = c(y1, y2, y3)
> A = factor(c(1,1,1,1,2,2,2,2,3,3,3,3))
> lm.1 = lm(y ~ A)
> anova(lm.1)

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
A          2 8.3500  4.1750  7.4401 0.01239 *
Residuals  9 5.0504  0.5612
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[2] もちろん β_1 以外の回帰係数についても同様である。