統計検定「データサイエンスエキスパート」サンプル問題の解答及び解説

2022 年 3 月

問題 1

正解

[1] ③

[2] ③

[3] ⑤

[4-1] ④，[4-2] ①

解説

[1] ③ SELECT 性別, AVG(国語) AS 国語平均点, AVG(英語) AS 英語平均点 FROM 成績 GROUP BY 性別;

は国語と英語に対して GROUP BY を使うことで,性別ごとに平均値を求めていることから，妥当である。それ以外の選択肢は男女合わせた平均値を求めており，妥当ではない。

[2] 2 群の平均値の差の検定を行う場合，2 群で分散が異なるかをＦ検定により確認する必要がある。①～③はともに分散が等しいことを仮定しているが，①と②は良く調べずにとりあえず等分散性がなりたつとしてスチューデントの t 検定を使用しているのに対し，③は性別ごとの英語得点の分散に差がないことをＦ検定で検定し，p 値が 0.06793 であることから 2 群の分散に差があるとは言えないことを確認している。そのため，③のほうが妥当な操作を行っている。また，①も③も等分散性が仮定できる 2 標本スチューデントの t 検定を用いて p 値を求め，0.03506 を得たことから，5%有意水準で「2 群での平均値は等しい」という帰無仮説を棄却し，英語平均点は性別で異なるという結論を得ていることは妥当である。②は等分散性を仮定して片側スチューデント t 検定を使っているが，平均が明らかに女性のほうが男性よりも大きいという仮定のもと，片側検定の p 値を両側検定に対して用いており妥当ではない。④は英語試験に男女で対応は出来ないはずなのに，対応付けをしてスチューデント t 検定を行っており，適切な検定の操作とは言えない。⑤は片側 p 値を両側検定で用いており，妥当ではない。

[3] 辞書型を使って 20 ごとに階級値を求める問題である。関数 `funcx()`の最初の `for` 文では，0 から 24 までの整数を 20 倍することで辞書型の連想値を 0 から 480 まで 20 ごとに設定しているので，【ア】は 20 が適切である。また，これは 2 番目の `for` 文の中で `int(v/20)*20` とすることで与えられたデータ `v` を 0 から 480 までの階級のどこで加算するべきかを決定していることから,【ア】は 20 であることが確認できる。一つのデータの度数は 1 を加算するべきであるので,【イ】は 1 が正しい。

［4］
［4-1］ 上位 20%は 10 名中 2 名のことである。これに該当する合計得点は，度数分布から階級値 300～320 に 1 名と 260～280 に 1 名存在していることがわかる。よって足切り点として適切な点数は選択肢のうち 260 点である。
［4-2］ SQL 文としては合計得点に対して上位からソートして 2 名に限定して出力すればよいので，
① SELECT 氏名, 合計 FROM 成績 ORDER BY 合計 DESC LIMIT 2;
が正解となる。DESC がない②では，下位からソートされて 2 名に制限されるため最下位の 2 名が取り出されるため，題意に一致しない。③は一見正しいように思えるが，合計を順番に大きい順に並べて一致するものを取り出す事は，全てのレコードを並び替えずに取り出すことに対応しているので不適切である。また④の選択肢では最大得点の 80%得点より大きいものが出力されるが，307 点の 0.8 倍は 245.6 点であり，足切り点よりも小さな値でかつ，そのような得点に合致するのは 4 名存在するため不適切である。⑤は合計点ごとにレコードを合計の大きい順に全て表示するだけなので題意に適合しない。

問題 2

正解

［1］①

［2］⑤

［3］②

［4］②

解説

［1］ 題意よりデータ(x_i, y_i) (i=1,…,n)が与えられている場合，確率$F(\alpha+\beta x)$で 1， 確率$1-F(\alpha+\beta x)$で 0 を取る場合ベルヌーイ型確率変数yに対する対数尤度関数$l(\alpha,\beta)$はその定義より

$$l(\alpha,\beta)=\sum_{i=1}^{n}\left\{y_i \log F(\alpha+\beta x_i)+(1-y_i)\log\bigl(1-F(\alpha+\beta x_i)\bigr)\right\}$$

となる。

［2］

【ア】: 対数尤度関数の 1 階偏微分を 0 と置いた方程式は尤度方程式と呼ばれる。

【イ】: 一般に多変数関数に対する 2 階偏微分の組合せにより構成される行列はヘッセ行列と呼ばれる。

【ウ】: 対数尤度関数の微分はスコア関数と呼ばれる。

［3］ 対数尤度関数を偏微分することにより，以下を得る。

$$\frac{\partial l}{\partial \alpha}=\sum_{i=1}^{n}\left\{y_i\frac{1}{1+\exp(\alpha+\beta x_i)}-(1-y_i)\frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)}\right\}$$

$$\frac{\partial l}{\partial \beta}=\sum_{i=1}^{n}\left\{y_i\frac{x_i}{1+\exp(\alpha+\beta x_i)}-(1-y_i)\frac{x_i\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)}\right\}$$

$$\frac{\partial^2 l}{\partial \alpha^2}=-\sum_{i=1}^{n}\frac{\exp(\alpha+\beta x_i)}{\bigl(1+\exp(\alpha+\beta x_i)\bigr)^2}$$

$$\frac{\partial^2 l}{\partial \alpha\partial \beta}=-\sum_{i=1}^{n}\frac{x_i\exp(\alpha+\beta x_i)}{\bigl(1+\exp(\alpha+\beta x_i)\bigr)^2}$$

$$\frac{\partial^2 l}{\partial \beta^2}=-\sum_{i=1}^{n}\frac{x_i^2\exp(\alpha+\beta x_i)}{\bigl(1+\exp(\alpha+\beta x_i)\bigr)^2}$$

［4］ここで問われているのは，2 変量のニュートン・ラフソン法

$$\begin{bmatrix}\alpha_{k+1}\\ \beta_{k+1}\end{bmatrix}=\begin{bmatrix}\alpha_k\\ \beta_k\end{bmatrix}-\begin{bmatrix}\dfrac{\partial^2 l}{\partial\alpha^2} & \dfrac{\partial^2 l}{\partial\alpha\partial\beta}\\ \dfrac{\partial^2 l}{\partial\alpha\partial\beta} & \dfrac{\partial^2 l}{\partial\beta^2}\end{bmatrix}_{\alpha=\alpha_k,\beta=\beta_k}^{-1}\begin{bmatrix}\dfrac{\partial l}{\partial\alpha}\\ \dfrac{\partial l}{\partial\beta}\end{bmatrix}_{\alpha=\alpha_k,\beta=\beta_k}$$

の反復計算である。一般に 2×2 の行列

$$\boldsymbol{D}=\begin{bmatrix}a & b\\ b & c\end{bmatrix}$$

の逆行列は

$$\boldsymbol{D}^{-1}=\frac{1}{\Delta}\begin{bmatrix}c & -b\\ -b & a\end{bmatrix}$$

$$\Delta=ac-b^2$$

により与えられる。ここで，$a=\frac{\partial^2 l}{\partial\alpha^2}$， $b=\frac{\partial^2 l}{\partial\alpha\partial\beta}$， $c=\frac{\partial^2 l}{\partial\beta^2}$ と置くと，

$$\Delta=\frac{\partial^2 l}{\partial\alpha^2}\frac{\partial^2 l}{\partial\beta^2}-\left(\frac{\partial^2 l}{\partial\alpha\partial\beta}\right)^2 \qquad \text{(A)}$$

であり，

$$\begin{bmatrix}\frac{\partial^2 l}{\partial\alpha^2} & \frac{\partial^2 l}{\partial\alpha\partial\beta}\\ \frac{\partial^2 l}{\partial\alpha\partial\beta} & \frac{\partial^2 l}{\partial\beta^2}\end{bmatrix}^{-1}=\frac{1}{\Delta}\begin{bmatrix}\frac{\partial^2 l}{\partial\beta^2} & -\frac{\partial^2 l}{\partial\alpha\partial\beta}\\ -\frac{\partial^2 l}{\partial\alpha\partial\beta} & \frac{\partial^2 l}{\partial\alpha^2}\end{bmatrix} \qquad \text{(B)}$$

となる。【ア】は(A)式を問うており，右辺第 2 項が対応するので `lab` が適切である。【イ】と【ウ】は(B)式を問うており，これらも `lab` が適切である。