2016.6.20

統計検定　準１級
2016年6月19日試験の正解一覧

問			解答番号	正解
問1	1		記述１	0.2 （20%）
	2		記述２	12(kg)
問2	1		記述３	11/9
	2		記述４	7/36
問3	1		記述５	5.67
	2		記述６	12.59，　帰無仮説は棄却できない
問4	1		1	2
	2		2	1
問5	1		3	3
	2		4	5
	3		5	5
問6	1		6	1
	2		7	5
問7			8	3
問8	1		9	4
	2		10	5
	3		11	2
問9	1		12	4
	2		13	4
問10	1		14	2
	2		15	4
問11	1		16	4
	2		17	3
	3		18	1
問12	1		19	1
	2		20	4
問13	〔1〕	(1)	21	2
		(2)	22	2
	〔2〕	(1)	23	3
		(2)	24	5
問14	1		記述７	部分記述用解答用紙参照
	2		記述８	
	3		記述９	

統計検定 準1級 部分記述用　解答用紙

問 1

記述1	記述2	得点１
0.２（20%）	12（ｋｇ）	

問 2

記述3	記述4
11/9	7/36

問 3

記述5
5.67

記述6	
棄却限界値　12.59	結論　帰無仮説は棄却できない

得点２

問14

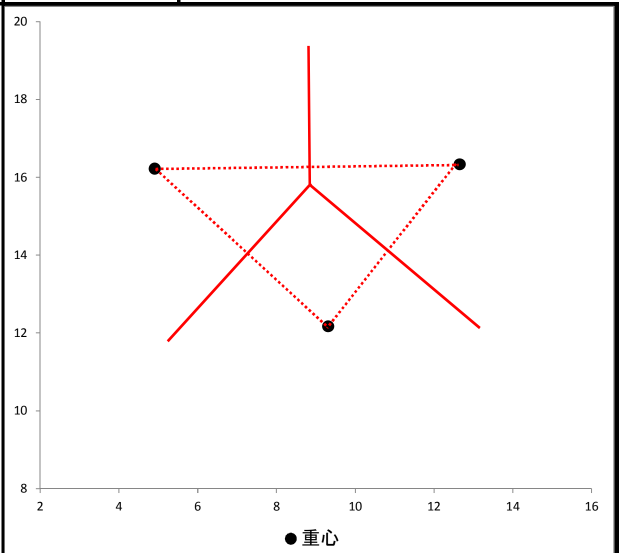
記述7

	手法の名前	その手法であると考えた理由	誤判別率
（ア）	ＬＤＡ 線形判別分析	解答のポイント ・2つの判別領域を分ける境界が明確な直線である。 ・群Aに属する個体が散らばりの小さい群Bに分類される個数が多くなる。	20/170
（イ）	ＳＶＭ	解答のポイント ・2つの判別領域に凹凸が見える。 ・2群の散らばりが大きく異なるため，LDAと比較して誤判別が少なくなる。	9/170

記述8

解答のポイント
・『交差検証法』とは，モデル選択の際に，そのモデルがどの程度適正であるかを検証する方法である。
・判別分析では，モデルのパラメータ数が多い場合は『過学習』が生じ，『誤判別率』を小さくすることが知られている。
・過学習を防ぐために判別分析に対応する交差検証法について記述する。
　(1)観測値を学習用と検証用にわけ，学習用で判別手法を構築，検証用で誤判別率を求め評価する。
　(2)データをK個のグループに分割し，(K-1)個のグループから判別手法を構築，残りの1個のグループを使って誤判別率を求める。これをK回繰り返し評価する。
（2つの交差検証法を示したが，解答には1つ示されていればよい。）

記述9



（上の図の書き方）互いの重心を結び，それぞれの垂直２等分線を引く。

準1級 論述問題 解答（略解）

問1

〔1〕

(1) 分散／平均 $= 1.67^2/2.84 \fallingdotseq 1$ からポアソン分布に従うと考えている。

(2)

表1. ポアソン分布を仮定した場合の表

上陸数（回）	観測度数（年）	確率	期待度数（年）
0	4	0.058	3.74
1	7	0.166	10.62
2	17	0.236	15.08
3	18	0.223	14.28
4	10	0.158	10.14
5	5	0.090	5.76
6	2	0.043	2.72
7	0	0.017	1.11
8	0	0.006	0.39
9	0	0.002	0.12
10以上	1	0.001	0.05
計	64	1.000	64.00

(3) 適合度の χ^2 統計量の値は 24.28。自由度9の上側5%の境界値は 16.92。帰無仮説は有意水準5%で棄却される。「10以上」に対する乖離度が他と比較して大きい。

(4) 「上陸数6回以上をまとめる場合」の適合度の χ^2 統計量の値は 3.01。自由度5の上側5%の境界値は 11.07。帰無仮説は有意水準5%で棄却されない。ポアソン分布の当てはまりは悪いとはいえない。

〔2〕上陸数が6回以上の実際の比率は $3/64 \fallingdotseq 0.047$ 。ポアソン分布近似より6回以上となる確率は 0.069。この程度の確率となることが予想される。2004年は外れ値と言えなくもないが，実際に起こった事象であるので除外はできない。

問2

〔1〕

(1) 単回帰式は $y = 7.3 + 0.12z = 6.82 + 0.024x$。

(2) 残差平方和は 1.10。

〔2〕

(1) $\tilde{Z} = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$,

重回帰式は $y = 6.79 + 0.12z + 0.26z^2 = 10.47 - 0.392x + 0.0104x^2$。

(2) 残差平方和は 0.17。

〔3〕$n \log S_e + 2(p+2)$ の大小のみを考えればよい。単回帰モデルの値は 6.46，重回帰モデルの値は -0.85 となり，重回帰モデルのほうが AIC を小さくする。

問 3

〔1〕

(1) モデル M_1 の式は $\log p_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + (\alpha\beta)_{ab} + (\alpha\gamma)_{ac} + (\alpha\delta)_{ad} + (\beta\gamma)_{bc} + (\beta\delta)_{bd} + (\alpha\beta\gamma)_{abc} + (\alpha\beta\delta)_{abd}$，逸脱度の自由度は 4。

(2) モデル M_1 の下での期待度数は $m_{abcd} = n_{abc+}n_{ab+d}/n_{ab++}$，(no, no, no, < 140) の期待度数は $(79+67)(79+197)/(79{+}67{+}197{+}179) = (146 \times 276)/522 = 77.195$。

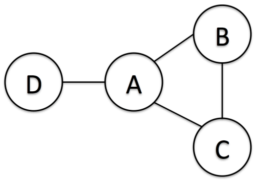
(3) 自由度 4 のカイ二乗分布より，$\chi^2_{0.90}(4) = 1.06$，$\chi^2_{0.10}(4) = 7.78$。P-値 $= (0.1 - 0.9) \times (2.062 - 1.06)/(7.78 - 1.06) + 0.9 = 0.78$ からモデル M_1 の当てはまりは悪くないと判断してよい。

〔2〕

(1) モデル M_2 の式は $\log p_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + (\alpha\beta)_{ab} + (\alpha\gamma)_{ac} + (\alpha\delta)_{ad} + (\beta\gamma)_{bc} + (\alpha\beta\gamma)_{abc}$，逸脱度の自由度は 6。

(2) モデル M_2 の下での期待度数は $m_{abcd} = n_{abc+}n_{a++d}/n_{a+++}$。逸脱度は 3.082 ($= 2\sum_{a,b,c,d} n_{abcd} \log(n_{abcd}/m_{abcd}) = 2\,(79\log(79/78.24) + \cdots + 31\log(31/28.68))$)。
自由度 6 のカイ二乗分布より，$\chi^2_{0.90}(6) = 2.20$，$\chi^2_{0.10}(6) = 10.64$。P-値 $= 0.75$ からモデル M_2 の当てはまりも悪くないと判断してよい。

〔3〕対応する無向独立グラフは



となる。喫煙者のみ，非喫煙者のみのそれぞれの集団の中では，職業と血圧の間には関連が見られない。