## تمرين مرفق بالحل حول التحليل التمييزي

التمرين:

:سحبت عينة عشوائية من 12 رياضي من مجتمع طبيعي وأجري عليهم اختباري اللياقة والكفاءة وذلك لغرض تصنيفهم إلى مهرة أو غير مهرة، و كانت النتائج كما في الجدول أدناه حيث أن:

X1: تمثل إختبار اللياقة x2: تمثل إختبار الكفاءة:

| المهرة | | غير المهرة | |
|---|---|---|---|
| X2 | X2 | X1 | X2 |
| 33 | 60 | 35 | 57 |
| 36 | 61 | 36 | 59 |
| 35 | 64 | 38 | 59 |
| 38 | 63 | 39 | 61 |
| 40 | 65 | 41 | 63 |
| | | 43 | 65 |
| | | 41 | 59 |

**المطلوب:**

1- اختبر شرط تساوي متوسطي المتغيرين في المجموعتين.
2- أختبر شرط تساوي مصفوفتي التباين والتباين المشترك للمجموعتين؟
3- أوجد الدالة التمييزية؟
4- أوجد أهمية كل متغير في الدالة وخطأ التصنيف؟
5- استنتج إلى أي من المجتمعين تنتمي المشاهدة $x = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

حل التمرين:

من أجل تطبيق الاختبار نقوم أولا بإيجاد الحسابات التالية:

| المجموعة الأولى (المهرة) | | | | | |
|---|---|---|---|---|---|
| D | $X_1$ | $X_2$ | $X_1^2$ | $X_2^2$ | $X_1X_2$ |
| 1 | 33 | 60 | 1089 | 3600 | 1980 |
| 1 | 36 | 61 | 1296 | 3721 | 2196 |
| 1 | 35 | 64 | 1225 | 4096 | 2240 |
| 1 | 38 | 63 | 1444 | 3969 | 2394 |
| 1 | 40 | 65 | 1600 | 4225 | 2600 |
| MOY | 36,4 | 62,6 | | | |
| SOM | 182 | 313 | 6654 | 19611 | 11410 |

| المجموعة الثانية (غير المهرة) | | | | | |
|---|---|---|---|---|---|
| D | $X_1$ | $X_2$ | $X_1^2$ | $X_2^2$ | $X_1X_2$ |
| 2 | 35 | 57 | 1225 | 3249 | 1995 |
| 2 | 36 | 59 | 1296 | 3481 | 2124 |
| 2 | 38 | 59 | 1444 | 3481 | 2242 |
| 2 | 39 | 61 | 1521 | 3721 | 2379 |
| 2 | 41 | 63 | 1681 | 3969 | 2583 |
| 2 | 43 | 65 | 1849 | 4225 | 2795 |
| 2 | 41 | 59 | 1681 | 3481 | 2419 |
| MOY | 39 | 60,42 | | | |
| SOM | 273 | 423 | 10697 | 25607 | 16537 |

لدينا: n1=5 n2=7

المجموعة الأولى:

$$S_{11} = \sum x_1^2 - \frac{(\sum x_1)^2}{n} = 6654 - \frac{(182)^2}{5} = 29,2$$

$$S_{22} = \sum x_2^2 - \frac{(\sum x_2)^2}{n} = 19611 - \frac{(313)^2}{5} = 17,2$$

$$S_{12} = \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n} = 11410 - \frac{(182)(313)}{5} = 16,8$$

المجموعة الثانية:

$$S_{11} = \sum x_1^2 - \frac{(\sum x_1)^2}{n} = 10697 - \frac{(273)^2}{7} = 50$$

$$S_{22} = \sum x_2^2 - \frac{(\sum x_2)^2}{n} = 25607 - \frac{(423)^2}{7} = 45,71$$

$$S_{12} = \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n} = 16537 - \frac{(273)(423)}{7} = 40$$

حساب التباينات المدمجة:

$$V_{11} = \frac{S_{11}(1) + S_{11}(2)}{n_1 + n_2 - 2} = \frac{29.2 + 50}{5 + 7 - 2} = 7.92$$

$$V_{22} = \frac{S_{22}(1) + S_{22}(2)}{n_1 + n_2 - 2} = \frac{17.2 + 45.71}{5 + 7 - 2} = 6.291$$

$$V_{12} = \frac{S_{12}(1) + S_{12}(2)}{n_1 + n_2 - 2} = \frac{16.8 + 40}{5 + 7 - 2} = 5.68$$

$$S = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \begin{pmatrix} 7,92 & 5,68 \\ 5,68 & 6,29 \end{pmatrix}$$

$$|S| = 17.5623 \quad ,$$

$$S^{-1} = \frac{adj(S)}{|S|} = \frac{\begin{pmatrix} 6.291 & -5.68 \\ -5.68 & 7.92 \end{pmatrix}}{17.5623} = \begin{pmatrix} 0.358 & -0.323 \\ -0.323 & 0.451 \end{pmatrix}$$

1- اختبار شرط تساوي متوسطي المتغيرين في المجموعتين:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

يتم هذا الاختبار من خلال إختبار فيشر بالاعتماد على إحصاءة هوتلنك و قبلها نحسب القيمة $D^2$

$$D^2 = (\bar{x}_1 - \bar{x}_2)S^{-1}(\bar{x}_1 - \bar{x}_2)$$

$$= \begin{pmatrix} -2.6 & 2.18 \end{pmatrix} \begin{pmatrix} 0.358 & -0.323 \\ -0.323 & 0.451 \end{pmatrix} \begin{pmatrix} -2.6 \\ 2.18 \end{pmatrix} = 8.22$$

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 = \frac{(5)(7)}{5+7}(8.22) = 23.975$$

$$F = \frac{n_1 + n_2 - p - 1}{n_1 + n_2 - 2} T^2$$
$$= \frac{5+7-2-1}{5+7-2}(23.975) = 21.577$$

وبمقارنة القيمة المحسوبة مع القيمة المجدولة عند درجة حرية (2,9) ومستوى معنوية 5% والتي تساوي 4,26 نجد أن القيمة المحسوبة أكبر ومنه نقبل الفرضية البديلة التي تدل على وجود فرق معنوي في متوسطي المجموعتين.

2- إختبار شرط تساوي مصفوفتي التباين والتباين المشترك للمجموعتين:

$$\begin{cases} H_0: \Sigma_1 = \Sigma_2 \\ H_1: \Sigma_1 \neq \Sigma_2 \end{cases}$$

يتم اختبار هذه الفرضية من خلال الصيغة التالية:

$$Q = MC^* \sim \chi^2_{(k-1)(p-1)}$$

ومن المعطيات السابقة لدينا:

$$S = \begin{pmatrix} 7.92 & 5.68 \\ 5.68 & 6.29 \end{pmatrix}$$

$$S_1 = \frac{\begin{pmatrix} S_{11} & S_{12} \\ & S_{22} \end{pmatrix}}{n_1 - 1} = \begin{pmatrix} 7.3 & 4.2 \\ 4.2 & 4.3 \end{pmatrix}$$

$$S_2 = \frac{\begin{pmatrix} S_{11} & S_{12} \\ & S_{22} \end{pmatrix}}{n_2 - 1} = \begin{pmatrix} 8.33 & 6.66 \\ 6.66 & 7.61 \end{pmatrix}$$

$|S| = 17.56$ , $\ln|S| = 2.865$

$|S_1| = 13.7$ , $\ln|S_1| = 2.62$

$|S_2| = 19.035$ , $\ln|S_2| = 2.946$

$$M = 12(2.865) - [5(2.62) + 7(2.946)]$$
$$= 0.685$$

$$C^* = 1 - \frac{2(2)^2 + 3(2) - 1}{6(2+1)(2-1)}\left[\left(\frac{1}{3} + \frac{1}{7}\right) - \frac{1}{12}\right]$$
$$= 0.8131$$

$$MC^* = (0.685)(0.8131) = 0.534$$

$$\chi^2_{1, 0.05} = 3.841$$

بمقارنة القيمة المحسوبة مع المجدولة نجد أن القيمة المجدولة اكبر من المحسوبة، ومنه نقبل الفرضية الصفرية التي تقول أن هناك تساوي في مصفوفتي التباين و التباين المشترك للمجموعتين، ومنه هذه الفرضية محققة وتعد اساسية في دوال التحليل التمييزي الخطية.

**3- إيجاد الدالة التمييزية:**

يتم احتساب الدالة التمييزية من خلال الصيغة التالية :

$$y = x'S^{-1}(\bar{x}_1 - \bar{x}_2) = x'C^*$$

حيث:

$$C^* = S^{-1}(\bar{x}_1 - \bar{x}_2) = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.358 & -0.323 \\ -0.323 & 0.451 \end{pmatrix} \begin{pmatrix} -2.6 \\ 2.18 \end{pmatrix} = \begin{pmatrix} -1.634 \\ 1.822 \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

$$y = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} -1.634 \\ 1.822 \end{pmatrix}$$

$$y = -1.634x_1 + 1.822x_2$$

4- إيجاد أهمية كل متغيرة وخطأ التصنيف

يتم احتساب أهمية كل متغيرة من خلال العلاقة التالية:

$$C_i^* = C_i\sqrt{V_{ii}}$$

$$C_1^* = C_1\sqrt{V_{11}} = -1.634\sqrt{7.92} = -4.598$$

$$C_2^* = C_2\sqrt{V_{22}} = 1.822\sqrt{6.291} = 4.569$$

ومنه نستنتج أن المتغيرين لهما نفس الأهمية في التصنيف و ذلك ما ذل عليه الفرق الصغير جدا بين المتغيرين.

أما خطأ التصنيف فيمكن استنتاجه من خلال القيمة D حيث نجد أن:

**$D^2$=8,22 D=2,86**

بما أن القيمة D صغيرة نستنتج أن احتمال خطأ التصنيف صغير وهو يعبر عن الكفاءة العالية لدالة التصنيف.

**5- استنتاج إنتماء المشاهدة x**

نحسب أولا نقاط الفصل و القيمة z فنجد مايلي:

$$\bar{y}_1 = 54{,}5796$$
$$\bar{y}_2 = 46{,}359$$

$$Z = \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{54{,}5796 + 46{,}359}{2} = 50{,}469$$

بتعويض قيم المشاهدة x في دالة التمييز نجد:

**Y=-1,634(2)+1,822(2)=0,376**

بما أن قيمة Z أكبر من القيمة Y فنستنتج أن المشاهدة X تنتمي إلى المجتمع الثاني.