متوفرة على الموقع:http://www.basra-sciencejournal.org

ISSN –1817 –2695

أسلوب تحليل البيانات المصنفة المترابطة (غير المستقلة)

فاضل عبدالعباس العابدي

المعهد التقني_ كوفة

Fadhil_ab2000@yahoo.com

الاستلام 26-3-2012، القبول 13-11-2012

مـلخص

أهتم هذا البحث بدراسة الأسلوب الأحصائي لتحليل البيانات المصنفة التي تكون فيها الأستجابات عبارة عن أزواج مرتبطة فيما بينها حيث تم التطرق الى الأسس الرياضية المتبعة في تحليل البيانات من هذا النوع. وقد اعتمد البحث على جانب تطبيقي من خلال دراسة معدلات الطلبة المتخرجين وعلاقتها بمعدلات قبولهم في المعهد بأعتماد تصنيف هذه المشاهدات تحت صفتي (Positive,Negative) بالأعتماد على معدل الطالب مقارنة مع المعدل العام للظاهرة، ومن التطبيق أتضح للباحث زيادة نسبة الترجيح (Odds ratio) تكون أعلى لصالح معدلات التخرج وهناك ترابط بين الظاهرتين(Odd=0.60).

الكلمات المفتاحية: البيانات المعدودة، البيانات المصنفة ، نسبة الترجيح ، المخاطرة النسبية

1. مقدمــة *Introduction*

تهتم البيانات المصنفة بشكل عام بدراسة المتغيرات التي يمكن أن تصنف فيها الأستجابات بشكل صفات مثل دراسات أستطلاع الرأي (موافق/غير موافق) الدراسات البايولوجية بعد أعطاء معالجة (Pos./Neg.)كذلك في التخصصات الطبية حول درجة الأصابة (خفيفة /متوسطة / شديدة)وغيرها من الدراسات، حيث أن مشاهدات الأستجابة في الوحدات التجريبية تسجل بشكل عدد (عدد المشاهدات التي تحقق كل الصفة) ولذلك تسمى في بعض المراجع العلمية بالبيانات المعدودة (Count Data)، ولهذه البيانات توزيعات أحصائية مختلفة (*Binomial dist., Poisson dist., Multinomialdist.*).

أهتم هذا البحث بالبيانات المصنفة المعتمدة (غير المستقلة) التي تكون فيها أزواج الصفات معتمدة أي أن الصفة(i) ضمن الأستجابة (B) تتأثر الصفة (j) ضمن الأستجابة(A).

ولتسهيل المهمة على القارئ الكريم فقد تضمنت منهجية البحث ملخص وكذلك قُسم البحث الى مقدمة عن موضوع البيانات المصنفة مع توضيح هدف البحث، كذلك تضمن الجانب النظري فكرة عن طبيعة البيانات المعتمدة وأسلوب تنظيمها داخل جداول توافقية والأسلوب الرياضي المستخدم في التحليل الأحصائي .

ولغرض تحقيق هدف البحث طُبق الأسلوب الرياضي على بيانات حقيقية تمثل معدلات كافة الطلبة المتخرجين من الأقسام الطبية في المعهد التقني / كوفة (عينة البحث) مقارنة مع معدلات قبولهم في المعهد (معدلات الأعدادية) حيث تصنف هذه البيانات على أنها بيانات

معتمدة(غيرمستقلة)، وأخيرا" أهم الأستنتاجات التي توصل اليها الباحث والأفاق المستقبلية لتطوير البحث.

2. الهـدف*Objective*

يهدف الباحث من بحثه الى المساهمة الجادة والصادقة في تسليط الضوء على أسلوب تحليل البيانات المصنفة غير المستقلة وذلك من خلال التطرق الى الأسلوب الرياضي المتبع في تحليل هذه البيانات مع أجراء دراسة تطبيقية عن بيانات مترابطة تمثل معدلات الطلبة المتخرجين وعلاقتها مع معدلات قبولهم في هيئة التعليم التقني.

3. الجانب النظــري

1.3 مفهوم البيانات المصنفةCategorical Data[1,3]

تظهر البيانات المصنفة في المشاهدات الأحصائية عندما تكون متغيرات الدراسة (الأستجابة) بشكل صفات (جيد/غير جيد)، (ناجح/فاشل)، مما يؤدي الى تسجيل المشاهدات في الوحدة التجريبية(تقاطع صفتين) بشكل عدد ولتبسيط الفكرة يمكن تلخيص ذلك من خلال الجدول الأتي:-

جدول(1):مخطط لتجربة ذات بيانات مصنفة

		Response (B)						Total
		1	2		j		c	
Response (A)	1	n_{11}	n_{12}		n_{1j}		n_{1c}	n_{1+}
	2	n_{21}	n_{22}		n_{2j}		n_{2c}	n_{2+}
		.	.					
	i	n_{i1}	n_{i2}	.	n_{ij}	.	n_{ic}	n_{i+}
		.	.					
	r	n_{r1}	n_{r2}	.	n_{rj}	.	n_{rc}	n_{r+}
Total		n_{+1}	n_{+2}	.	n_{+j}	.	n_{+c}	n

حيث أن n_{ij} تمثل عدد المشاهدات الناتجة من تأثير الصفة (العامل i) ومن ثم تأثير الصفة (العاملj)

n_{i+} : تمثل مجموع المشاهدات للأستجابة الصفة (i)

n_{+j} : تمثل مجموع المشاهدات للأستجابة للصفة (j)

أماالبيانات المصنفة المعتمدة (*Categorical dependent Data*) تظهر عندما تكون الأستجابات للمتغيرات غير مستقلة أي أن الأستجابة الثانية تعتمد على الأستجابة الأولى أو بمعنى آخر العينتين لهما الأستجاباتنفسها مما يؤدي أن تكون المشاهدات في الوحدة التجريبية ناتجة من التفاعل بين الأستجابيتين، ويظهر هذا النوع من البيانات في الكثير من التطبيقات الطبية والبايولوجية ودراسات أستطلاعات الرأي.

فعلى سبيل المثال في دراسة الأستطلاع لأخذ رأي عينة من الأشخاص في الفترة الزمنية (t) ومن ثم أعادة الأستطلاع على نفس العينة في الفترة الزمنية (t+1) مما يؤدي أن أجابة الشخص في الفترة اللاحقة تتأثر بدرجة كبيرة بأجابته في الفترة السابقة، وكذلك أعطاء سؤالين مترابطين لشخص واحد مما تكون الأستجابة (أجابة الشخص) متأثرة بترابط السؤالين مما يولد ترابط بين أستجابة السؤال الأول والأستجابة على السؤال الثاني، في الجانب الطبي لها تطبيقات في تجارب القياسات المكررة (*Repeated Measurements*) حيث الأستجابة في الفترة اللاحقة (t+1) تتأثر بظهور الأستجابة في الفترة السابقة(t).

من الأمثلة أعلاه فأن كل مشاهدة من الأستجابة الأولى ترتبط مع مشاهدة من الأستجابة الثانية وبالتالي فأن عدد

المشاهدات (المتغير المعتمد) يتأثر بالعلاقة بين الصفات للتجربة.

ونظرا" لوجود ترابط بين أزواج الأستجابات أي مايعرف (*Matched Pairs*) مما يؤدي أن تكون الجداول مربعة (2x2) وسيتم في الجانب النظري والتطبيقي للجدول ونترك الجداول المتعددة للأبحاث المستقبلية، حيث يمكن تمثيل مثل هذه الجداول في الشكل الأتي:-

جدول(2):مخطط لتجربة ذات بيانات مصنفة مترابطة

		Response (B)		Total
Respons e (A)		B_1	B_2	
	A_1	n_{11}	n_{12}	n_{1+}
	A_2	n_{21}	n_{22}	n_{2+}
Total		n_{+1}	n_{+2}	n

وبدلالة الأحتمالات المشتركة (π_{ij})(*Joint Probabilities*) يمكن توضيحها حسب الجدول الأتي:

جدول(3): الأحتمالات المشتركة للبيانات المصنفة

		Response (B)		Total
Respons e (A)		B_1	B_2	
	A_1	π_{11}	π_{12}	π_{1+}
	A_2	π_{21}	π_{22}	π_{2+}
Total		π_{+1}	π_{+2}	

حيث أن π_{ij} تمثل الأحتمال المشترك لظهور الأستجابة (i,j) وتحسب الأحتمالات الفرعية والكلية من خلال الصيغة التالية :-

$$\pi_{ij} = \frac{n_{ij}}{n} \quad , \quad \pi_{i+} = \frac{n_{i+}}{n} \quad , \quad \pi_{+j} = \frac{n_{+j}}{n} \qquad \dots\dots(1)$$

2.3 الأساليب الرياضية

يمكن الأعتماد على عدة أساليب لتحليل البيانات المصنفة المعتمدة منها معلمية (*Parametric*) مثل النماذج الخطية ، نماذج أنحدار لوجستك وهناك أساليب لامعلمية والتي تعتمد على أختبارات النسب (*Proportions* Tests) وسنعتمد في هذا البحث على الأساليب اللامعلمية لأختبار الفرضية التي تنص على "**عدم وجود فروق أحصائية بين نسب الأستجابات**"، ومن هذه الأساليب هي:

1. نسبة الترجيح Odds Ratio

تعرف نسبة الترجيح (نسبة الفرصة) بأنها ترجيح أحتمالية حدوث (ظهور) تصنيف في متغير (أستجابة) مقارنة مع تصنيف آخر في نفس المتغير(أستجابة)، بأفتراض أن π هي أحتمالية النجاح هي $\pi = \frac{n_{ij}}{n}$ فأن أحتمالية الفشل ($1-\pi$) وبالتالي فأن نسبة الترجيح تكتب من خلال الصيغة التالية:[3]

$$\theta = \frac{\pi}{1-\pi} \qquad \dots\dots\dots\dots(2)$$

ومن الجدول (3) فأن النسبة للصفوف هي ($\theta_i = \frac{\pi_i}{1-\pi_i}$; $i = 1,2$) وبالتالي فان نسبة الترجيح بين الصفين $\theta = \frac{\theta_1}{\theta_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$ ، وكذلك ممكن أن تكون نسبة الترجيح للمقارنة بين الأعمدة أو المقارنة بين نسب صف ونسب عمود، ومن خلال مراجعة مثل هذه البيانات فأن عدم الأستقلالية يكوّن لدينا توزيعات مشتركة (*Joint distributions*)فيها (π_{ij}) أحتماليةالخلية (i,j)وبالتالـي يمكن أن تكتب نسبة الترجيح بين الصفين($\theta_i = \frac{\pi_{i1}}{\pi_{i2}}$; $i = 1,2$)ومنها نستنتج أن نسبة الترجيح بين الأستجابتين(المتغيرين)يمكن أن تكتب بدلالة

عناصر القطر للأحتمالات (Cross−product ratio) من خلال الصيغة التالية:[3]

$$\hat{\theta} = \frac{\theta_1}{\theta_2} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \quad \dots\dots\dots(3)$$

أو بدلالة التكرارات المشاهدة (*Observed Frequencies*, F's,$F_{ij}=n_{ij}$)بالصيغة التالية :

$$\hat{\theta} = \frac{F_{11}F_{22}}{F_{12}F_{21}} \quad \dots\dots\dots(4)$$

وفي حالة الجداول المتعددة (r*c) فتظهر لدينا نسب ترجيح فرعية ممكنة بعدد $\left[\binom{r}{2}, \binom{c}{2}\right]$حيث يمكن أن تكتب بالصيغة العامة التالية:

$$\hat{\theta}_{ij} = \frac{F_{ij}\,F_{(i+1)(j+1)}}{F_{i(j+1)}F_{(i+1)j}} \quad ;\mathrm{i} = 1,2,\dots,\mathrm{r}-1 \;;\mathrm{j} = 1,2,\dots,\mathrm{c}-1 \quad \dots.(5)$$

ويمكن تلخيص خصائص هذا المقدر بالشكل الأتي :[1]

$$\begin{cases} independent & if\ \hat{\theta} = 1 \\ first\ categorices\ is\ higher\ (postive) & if\ \hat{\theta} > 1 \\ second\ categorices\ is\ higher\ (postive) & if\ \hat{\theta} < 1 \end{cases}$$

وبالتالي يمكننا الأستدلال من خلال هذه الخاصية اذا كانت $\hat{\theta} = 1$ فأن الأستجابات مستقلة ويؤدي أن $\log(\hat{\theta}) = 0$، وعندما يكون حجم العينة كبير فأن توزيع $\log(\hat{\theta})$ يقترب من التوزيع الطبيعي لذلك فأن يمكن الأستدلال على حدود الثقة للمعلمة $\hat{\theta}$من خلال الصيغة التالية:

$$\log\hat{\theta} \pm Z_{\alpha/2} * S_{\hat{\theta}} \quad \dots\dots..(6)$$

حيث أن $S_{\hat{\theta}}$ تمثل تباين المعلمة $\hat{\theta}$ وتحسب من خلال الصيغة $S_{\hat{\theta}} = \sqrt{\sum_{i,j=1}^{2}\left(\frac{1}{n_{ij}}\right)}$ و$Z_{\alpha/2}$ قيمة جدولية من توزيع Z−Normal.

وبعد الحصول على حدود المعلمة يمكن أخذ معكوسها من $\log(\hat{\theta})$أي $exp^{log\hat{\theta}}$ لأيجاد حدود $\hat{\theta}$.

2. المخاطرة النسبية (Relative Risk) [1,3]

المخاطرة النسبية تبين درجة المخاطرة في ظهور صفة معينة من صفات الأستجابة أي يمكن معرفة عدد مرات ظهور الصفة مقارنة مع صفة أخرى ويستخدم هذا المقياس خصوصا" في تجارب الحياة (الحياة/الموت) (النجاح/الفشل) وليس في دراسات الأستطلاع أو جداول المعالجة والسيطرة [5].

بأفتراض أن $i,j = 1,2 \;;\; \hat{\pi}_i = \frac{n_{ij}}{n}$وبالتالي فأن نسبة المخاطرة النسبية $r = \frac{\hat{\pi}_1}{\hat{\pi}_2}$والعلاقة بين odds و Relativeيمكن توضيحها من الصيغة التالية:−

$$r = \theta * \left(\frac{1-\hat{\pi}_1}{1-\hat{\pi}_2}\right) \dots\dots\dots(7)$$

حيث يلاحظ أنه عندما $\hat{\pi}_i \to 0$ فأن نسبة المخاطرة تكون مساوية الى نسبة الترجيح، ويمكن بناء حدود الثقة لنسبة المخاطرة النسبية وبأحتمال ثقة $(1-\alpha)$ كالأتي :−

$$r \pm Z_{\alpha/2}\hat{\sigma}_{\log(r)} \quad \dots\dots\dots(8)$$

حيث يمكن أحتساب الصيغ السابقة اعتمادا" على التطبيقات الجاهزة ومنها التطبيق المشار اليه في [7].

3. أختبار McNamara's Test [2,3,6]

لمعرفة تساوي أحتمالية الصف والعمود يمكن الأعتماد على أختبار (McNamara's) حيث يعتمد هذا الأختبار على التعامل مع النسب (الأحتمالات المشتركة) المذكورة في الجدول (3)، ويمكن صياغة فرضية أختبار تساوي أحتمال ظهور متغير الأستجابة الأولى مع أحتمال ظهور متغير الأستجابة الثانية للصفة الأولى من خلال الصيغةالتالية:-

$$H_o: \pi_{1+} = \pi_{+1} \quad or \quad d = \pi_{1+} - \pi_{+1} = 0 \qquad \ldots\ldots\ldots(9)$$

وبنفس الأسلوب للحالة الثانية

$$H_o: \pi_{2+} = \pi_{+2}$$

من خلال الفرضية الأولى أو الثانية يمكن أعادة صياغتهما بالشكل التالي:-

$$\boldsymbol{H_o: \pi_{1+} = \pi_{+1};\ \pi_{1+} - \pi_{+1} = 0;\ \pi_{11} + \pi_{12} - (\pi_{11} + \pi_{21})}$$
$$\boldsymbol{d = \pi_{12} - \pi_{21}} \qquad \ldots\ldots\ldots\ldots(10)$$

ونستنتج من الفرضية أعلاه أعتمادها على أحتمالات القطر الثانوية وهي المؤثرة في تجانس جدول الأستجابة .

ويمكن الأستدلال على هذه النسبة من خلال التوزيع الرياضي لها، حيث كما هو معلوم أن النسبة (π_{ij})تتبع التوزيع الثنائي [(*binomial distribution*($n\pi, n\pi(1-\pi)$] وبملاحظة الفرضية (1) فأن تباين الفرق (d) يمكن التوصل اليه من خلال الخطوات التالية:[3]

$$var(d) = var(\pi_{1+} - \pi_{+1}) = var(\pi_{1+}) + var(\pi_{+1}) - 2cov(\pi_{1+}, \pi_{1+})$$
$$var(\sqrt{n}\, d) = \pi_{1+}(1 - \pi_{1+}) + \pi_{+1}(1 - \pi_{+1}) - 2\pi_{1+}, \pi_{1+} \quad \ldots\ldots(11)$$

بعد التعويض عن الحدود الأحتمالية للمعادلة (11) يمكن أعادة كتابة مقدر التباين التقديري حسب الصيغة التالية:-

$$\hat{\sigma}^2(d) = \frac{[(\pi_{12} + \pi_{21}) - (\pi_{12} + \pi_{21})^2]}{n} \qquad \ldots\ldots(12)$$

عندما حجم العينة كبير فان توزيع *binomial* يقترب من التوزيع الطبيعي *Normal* وتحت فرضية العدم Ho فان توزيع d يكون $d \sim N(o, \sigma^2)$لذا فأن حدود الثقة (C.I.) باحتمال ثقة $(1-\alpha)$تكون وفق الصيغة الأتيه:-

$$d \pm z_{\alpha/2}\hat{\sigma}(d) \qquad \ldots\ldots\ldots(13)$$

وأن أحصاءة الأختبار للفرضية $H_o: \pi_{1+} = \pi_{+1}\ or\ d = 0$ تأخذ الصيغة التالية:-

$$Z_c = \frac{d}{\hat{\sigma}(d)} \qquad or \qquad Z_c = \frac{n_{12} - n_{21}}{S(d)} \qquad \ldots\ldots\ldots(14)$$

وتحت تحقق فرضية العدم فأن$Z_c^2 \sim \chi_{(\alpha,1)}$حيث ترفض الفرضية Hoاذا كانت $Z_c^2 > \chi_{(\alpha,1)}$أوالأحتمال المشاهد $P_value \to 0$ أو على الأقل($p_value \leq 0.05$).

4. الجانب التطبيقي *Application Part*

أجريت دراسة تطبيقية بالأعتماد على بيانات الطلبة المتخرجين من الأقسام الطبية في المعهد التقني/كوفة للعام الدراسي 2011/2010حيث تتميز هذه الأقسام بتجانس مدخلاتها وذلك من خلال المقارنة بين معدل التخرج بأعتبارة أستجابة لاحقة ومعدل القبول في المعهد بأعتباره عامل تفسير (أستجابة سابقة)، حيث تم الأعتماد على تصنيف الطلبة حسب عاملي الدراسة على أساس معدل القبول حيث أعتبر متوسط المعدلات المقبولة في بداية التسجيل في المعهد (70) حيث أعطيت صفتي (*Positive*)و *(Negative)*ويمكن توضيح ذلك من خلال الدالةالتالية:-

$$f(mean) = \begin{cases} positive(pos.) & \bar{x} \geq 70 \\ Negative(Neg.) & \bar{x} < 70 \end{cases}$$

وفرزت البيانات حسب تطبيق *Excel* بالأعتماد على الصيغة الرياضية رقم(1) وكذلك حساب عدد المشاهدات الدالة أعلاه حيث وضحت أستجابة العاملين ضمن حالتي التي تحقق الشروط المركبة للدالة أعلاه حيث أعطيت الأستجابة (*Pos., Neg.*) ولحساب النسب طبقت النتائج في الجدول الأتي:-

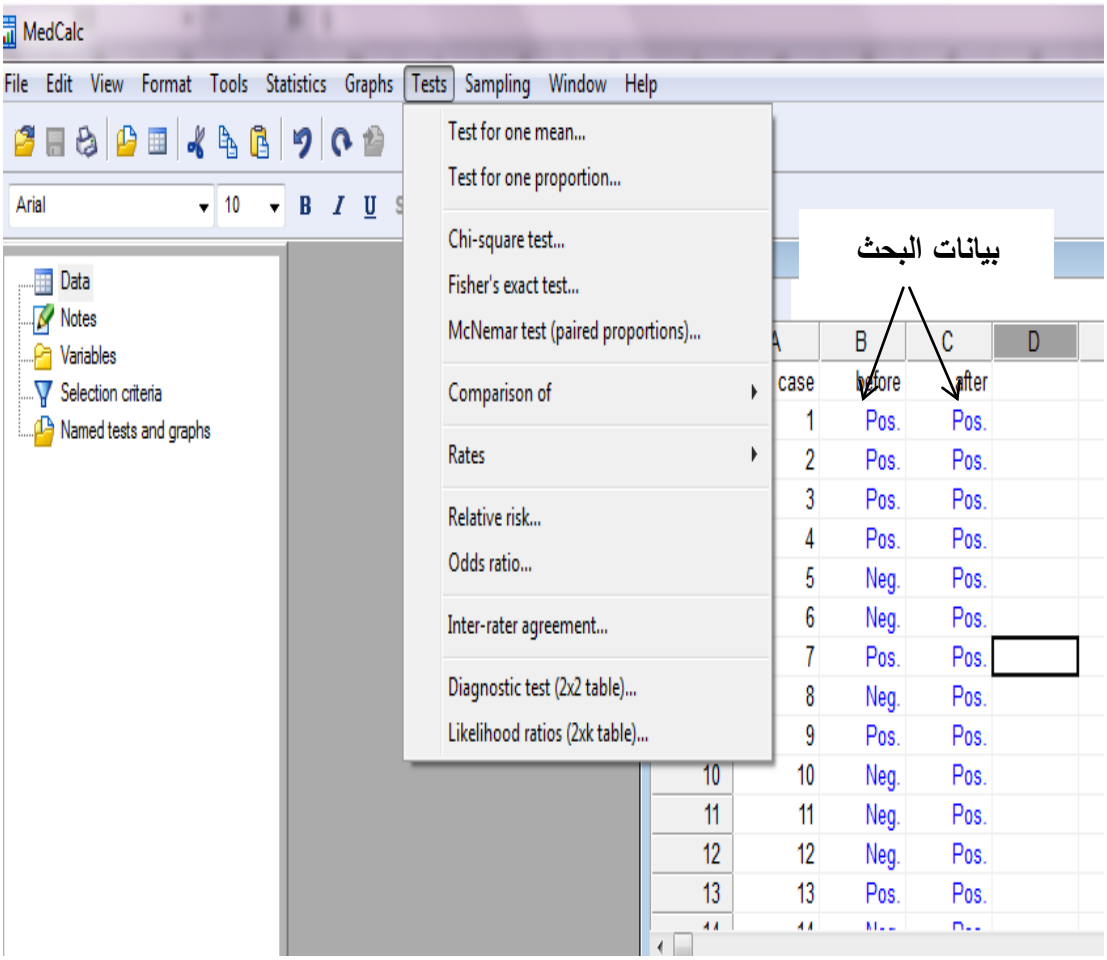
جدول (4):- تجميع بيانات البحث حسب الصفات

		Response_ after		Total	Ratio
Response_ before		Pos.	Neg.		
	Pos.	13	34	47	0.14
	Neg.	65	215	280	0.86
Total		78	249	327	
Ratio		0.24	0.76		

ومن خلال طبيعة البيانات في الجدول أعلاه يتضح لنا أنها بيانات مصنفة (معدودة) غير مستقلة أي أن الأستجابة اللاحقة (Response after)(معدل التخرج) تعتمد بدرجة معينة على (Response before)(معدل القبول)

1.4التحليل الأحصائي (Statistical Analysis)

بعد تجميع البيانات الخاصة بالبحث أعتمد على التطبيق الأحصائي *MedCalc Statistical Software ver 2011* وهومن التطبيقات الأحصائية الحديثة والمعرف على الموقع WWW.medcalc.org ولغرض فائدة زملائي الباحثين الكرام في تطبيق هذا البرنامج يمكن توضيح شاشة الأوامر الرئيسة له وأسلوب أدخال البيانات من خلال الشاشة التالية :[7]



شكل رقم (1): نافذة تنفيذ الأوامر الرئيسة والفرعية مع أسلوب تسجيل البيانات

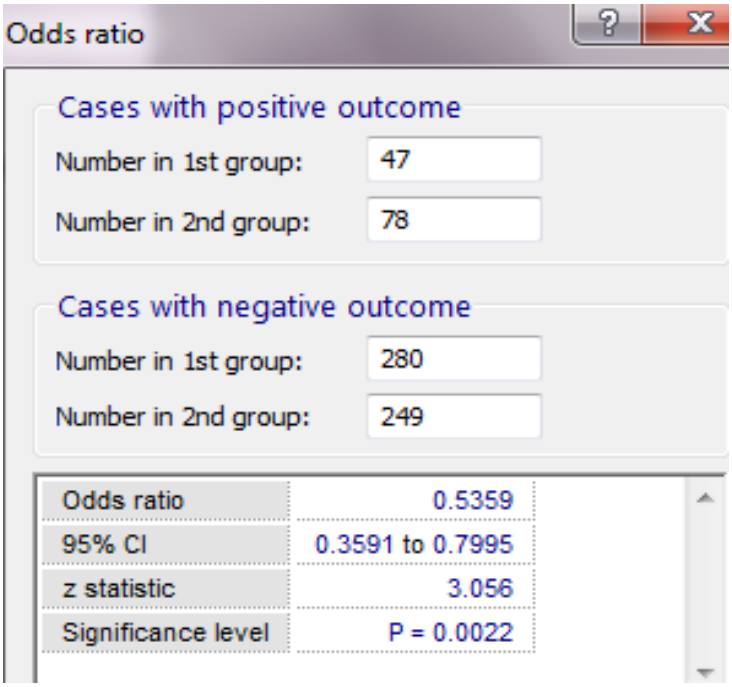
ومن متابعة واجهة العمل والأوامر الموضحة في الشاشة أعلاه يتضح لنا:

1 –أن جداول البيانات متوافقة مع جداول بيانات *EXCEL* مما يسهل العمل بين التطبيقين .

2 –يمكن تنفيذ أيعازات الأختبار (Test)من خلال أختيار الأمر الرئيسي (Test) ومن ثم أختيار الأمر الفرعي المطلوب.

3 –يمكن تنفيذ تحليلات أحصائية أخرى من خلال أختيار الأمر الرئيسي (*Statistics*) ومن ثم الأمر الفرعي المطلوب على سبيل المثال لبحثنا الأمر (*Categorical*)ولكون الأختيار الثاني (Test) يضم كافة أختبارات البيانات المصنفة فقد تم الأعتماد عليه.

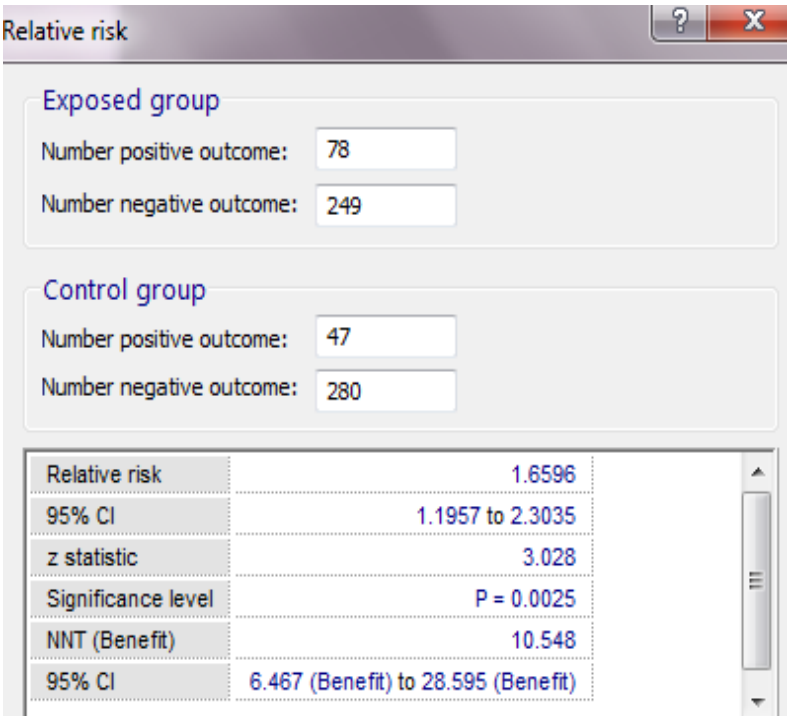
1. نسبة الترجيح والمخاطرة النسبية



شكل رقم (2): نافذة نتائج تنفيذ برنامج *MedCal* لمقياس *Odds ratio*

لغرض المقارنة بين نسب الطلبة من صفة (pos.) لكلا الأستجابتين(π_{1+} ,π_{+1}) أي معرفة نسبة الترجيح للطلبة المقبولين في المعهد من ذوي المعدلات الجيدة مقارنة مع معدلاتهم عند التخرج لنفس الصفة من خلال $postive\ Ratio = \frac{47}{78} = 0.602$وهي تعني أن نسبة ترجيح أن يكون الطلبة المتخرجين من ذوي المعدلات الجيدة مقارنة مع معدلات الطلبة المقبولين في المعهد متوسطة.

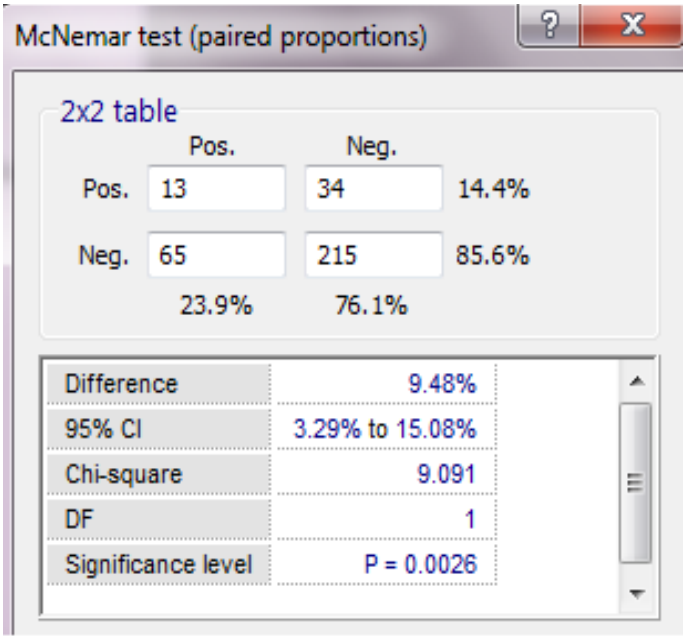
ولكن كانت العكس لصفة (Neg.) لكلا الأستجابتين (π_{2+} ,π_{+2})حيث بلغت نسبة الترجيح لنفس الصفة من خلال $Negative\ Ratio = \frac{280}{249} = 1.120$ وهي تميل لأستجابة الطلبة المتخرجين وذلك بأنخفاض نسبة الطلبة المتخرجين من صفة (Neg.)، وكانت هذه النتائج الفرعية متطابقة مع حساب نسبة الترجيح الكلية حيث بلغت حسب نتائج البرنامج الموضحة في النافذه أعلاه بلغت(0.5359) وهناك فرق معنوي بين نسب الترجيح للأستجابين واضح من خلال أختبار (Z-test)وبأحتمال خطأ صغير جدا" (P_value=0.0022) كذلك معرفة الحدود المسموح بها لتغير النسبة بثقة (95% C.I.).



شكل رقم (3): نافذة نتائج تنفيذ برنامج *MedCal* لمقياس *Relative Risk (RR)*

يمكن حساب المخاطرة النسبية حول انخفاض معدلات الطلبة المتخرجين مقارنة مع معدلات قبولهم من خلال الصيغة رقم (3) قد بلغت (0.603)أي ان بعض المعدلات ستبقى منخفضة رغم أنخفـــاض نسبة ترجـيح المعدلات من صفة (Neg.) عند التخـــرج في أختبـــــار Odds ratioوهذا يدل على أن الطلبة المنخفضة معدلاتهم لايوجد لديهم أندفاع في تحسين معدلاتهم ويمكن متابعة أختبارات هذه النسبة من خلال أختبـــار (Z-test) وأحتماليـــة الخطأ (P_value=0.003)وحسب النتائج الموضحة في النافذة أعلاه .

1. **أختبار** *McNamara's Test*



شكل رقم (4): نافذة نتائج تنفيذ برنامج *MedCal* لمقياس *McNemar test*

حيث يمكن من خلال النافذة أعلاه أختبار الفرضيات الموضحة في الصيغة رقم (9) ويتضح لنا أن أختبـــار (*Chi-square*) أثبت وجـــــود فروق معنوية بين الأستجابة اللاحقـــــــــة(نسب معدلات التخرج) والأستجابة السابقة(نسب معدلات القبول) مما يدل على رفض فرضيةالعدم التي تنص أن الفرق بين نسب الأستجابتين غير مؤثر (غير معنوي) حيث الفرق 9.5وأن الأستجابتين غير مستقلتين حيث بلغ أحتمال الخطأ (p_value=0.0026).

5. الأستنتاجات والتوصيات

1.5 الأستنتاجات

من خلال الجانب التطبيقي للبحث والتحليل الأحصائي للبيانات تم التوصل الى الأستنتاجات التالية:-

1 -من أختبار نسبة الترجيح (Odds ratio) نستنتج أن حدوث (ظهور) صفة المعدلات الجيدة لدى المتخرجين هي أكبر مقارنة مع نفس الصفة للطلبة المقبولين وذات فروق معنوية وبأحتمال خطأ قليل جدا" (p_value=**0.0022**) .

2 -أرتفاع نسبة المتخرجين من صفة (pos.) 24% مقارنة مع نفس الصفة عن قبولهم في المعهد (14%)، أما بخصوص صفة (Neg.) فأنها انخفضت من 85.6% عند قبولهم في المعهد الى 76% مما يدل على الرغبة الحقيقية لدى طلبة الأقسام الطبية للمنافسة من أجل تحسين مستواهم العلمي لغرض قبولهم في الكليات .

3 -عدم أستقلالية مستويات الطلبة في التخرج مقارنة مع مستوياتهم عند القبول في المعهد مما يدل على حرص الطلبة ذات معدلات القبول العالية في المحافظة عليها خصوصا" أن أغلب هؤلاء الطلبة يقبلوا برغبتهم لغرض أجتهادهم من أجل الحصول على كليات مناظرة لأختصاهم (الصيدلة ، الكلية التقنية الصحية ، ...) ضمن قناة الأوائل.

2.5التوصيـــات

من خلال الأستنتاجات أعلاه يوصي الباحث بما يأتي:-

1 -الأهتمام بدراسة البيانات المصنفة (وخصوصا" المعتمدة)لأنها ذات تطبيقات مهمة في العلوم الطبية وعلوم الحياة ودراسة أستطلاعات الرأي.

2 - أستخدام نماذج اللوجستك والمقدرات الحصينة في دراسة البيانات المصنفة.

REFRENCES

1. Alan Agresti , 2007, *An Introduction to Categorical Data Analysis*, 2nd Edition. John Wiley & Sons, Inc.
2. Bland J.M. & Altman D.G., 2000,*The odds ratio. British MedicalJournal* 320, 1468, New York: John Wiley & Sons.
3. Daniel, A. Yu Xie, 2008,*Statistical Methods for Categorical Data*, 2nd Edition, Emerald Group, UK.
4. Deeks J. , 1998, *When can odds ratios mislead? Odds ratios should be used only in case-control studies and logistic regression analyses* [letter]. British Medical Journal:317(7166);1155-6; discussion 1156-7.
5. Elizabeth S., 2002, *Point Estimation , Odds Ration, Risk differences*, Clinical Trials in 20 Hours, Oncology Biostatistics

6. Fleiss, J. L., Levin, B, 2003, *Statistical Methods for Rates and Proportions*, 3th Edition, John Wiley & Sons, Inc.
7. MedCalc®,2011,*StatisticalSoftwarever sion12.1.4.0*, WWW.medcalc.org
8. Pearce N , 1993, *What Does the Odds Ratio Estimate in a Case-Control Study*?. Int J Epidermal:22(6);1189-92.

Analysis Procedure of correlated Categorical Data (Dependent Data)

Fadhil Abdul Abbas AlAbidi
Kufa Technical institute
Fadhil_ab2000@yahoo.com

Summary

This research interested in study the method of statistical analysis of categorical Data(Count Data) where the responses is a pairs associated with each other, where they were addressed to the mathematical foundations used in the analysis of the data of this type. The research has adopted on the application by studying the rates of graduating students and their relationship acceptance rates at the Institute of their adoption of the classification of these observations under ***(Positive , Negative***) by depending on the student average (rate) compared with the overall of rate of the phenomenon ,it turns out the application for the researcher to increase the proportion of Odds Ratio for the benefit of higher graduation rates and there is correlationbetweenthe two phenomena(Odd = 0.60).

Keywords: *Categorical Data, Classification Data, Odds ratio, Relative Risk*