الإحصاء الحيوي

وصف وعرض البيانات

Describing and Displaying Data

المحاضرة السابعة : تلخيص البيانات الكمية

الدكتور

محمد يونس حَجير

أستاذ تقويم الأسنان والفكين بكلية طب الأسنان بجامعة دمشق

دكتوراه دولة (PhD) في تقويم الأسنان والفكين من جامعة غلاسكو Glasgow – بريطانيا

ماجستير في أصول البحث العلمي في بحوث العناية الصحة والاجتماعية من جامعة شيفيلد Sheffield - بريطانيا

عضو مجموعة الدراسة البحثية الوجهية القحفية في جامعة غلاسكو Glasgow - بريطانيا

محكم دولي في المجلة البريطانية لتقويم الأسنان و مجلة تقويم الأسنان والبحث الوجهي القحفي

ومجلة «آنجل التقويمية السنية» ومجلة «التطور في تقويم الأسنان»

وصف وعرض البيانات الكمية

تلخيص البيانات الكمية

عرض البيانات الكمية

التباين ضمن الأفراد

نقاط يجب أن تؤخذ بالحسبان عند قراءة الأدبيات الطبية

تلخيص البيانات الكمية

- تحتوي القياسات الكمية على معلومات اكثر من البيانات الفئوية ، وبالتالي تلخيص هذه البيانات هو أكثر تعقيدا ..
- يختار الباحث مقياسا تلخيصيا من أجل أن يضغط كما هائلا من المعلومات في أرقام صغيرة يمكن التواصل فيها شفويا
- أكثر ما يهمنا هو « أي تقع البيانات؟» و «ما هو مقدار التبيانية والاختلافية فيها ؟» ..
- وهذا ما نسميه : «مقاييس التوضع أو مقاييس النزعة المركزية» ، «مقاييس التشتت أو التباين» ...

تلخيص البيانات الكمية

مقاييس التوضع

- الوسط أو المعدل Mean : وهو الوسط الحسابي لمجموعة من القيم ويسمى إكس بار ، مجموعة قيم المشاهدات مقسومة على عددها.

$$\bar{x} = \frac{\text{Sum of all sample values}}{\text{Size of sample}} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

In the above equation, x_i represents the individual sample values and $\Sigma_{i=1}^{n} x_i$ their sum. The Greek letter 'Σ' (sigma) is the Greek capital 'S' and stands for 'sum' and simply means 'add up the n observations x_i from the 1st to the last (nth)'.

تلخيص البيانات الكمية

Example: Calculation of the mean – birthweights

Consider the following five birthweights in kilograms recorded to 1 decimal place selected randomly from the Simpson (2004) study of low birthweight babies which is described in more detail in Chapter 2.

$$1.2, 1.3, 1.4, 1.5, 2.1$$

The sum of these observations is $(1.2 + 1.3 + 1.4 + 1.5 + 2.1) = 7.5$. Thus the mean $\bar{x} = 7.5/5 = 1.50$ kg. It is usual to quote 1 more decimal place for the mean than the data recorded.

تلخيص البيانات الكمية

- ميزة الوسط الحسابي أنه يأخذ كل القيم بالحسبان وبالتالي فهو «فعال» بالمفهوم الإحصائي .....
- السيئة هو أن عرضة للتأثر بما يسمى «المتطرفات» القيم المتطرفة ، النواشز ، خارجيات التوضع Outliers ..
- القيم المتطرفة هي عبارة عن مشاهدات لو أبعدت من الدراسة لكان لها تأثير ملاحظ على النتائج ...

- لكن يجب أن لا يفهم أنه ينبغي على الباحث استبعاد القيم المتطرفة (الناشزة) أو أنه يجب النظر إليها أنها تسببت من قراءات خاطئة أو قياسات خاطئة ...

تلخيص البيانات الكمية

- الوسيط Median : يتم أولا ترتيب البيانات ترتيبا تصاعديا من الأقل إلى الأعلى قيمة ، ونعد حتى نصل إلى وسط هذه القيم فيسمى الوسيط .
- لو كان عدد المشاهدات فرديا ، فهو يقع في المنتصف تماما بعد ترتيبها ..
- لو كان عدد المشاهدات زوجيا ، فهو الوسط الحسابي للمشاهدتين اللتين تقعان في المنتصف بعد ترتيب البيانات .

تلخيص البيانات الكمية

Example: Calculation of the median – birthweights

Consider the following five birthweights in kilograms selected randomly from the Simpson (2004) study.

Rank order	Weight (kg)	
1	1.2	
2	1.3	
3	1.4	← median
4	1.5	
5	2.1	

If we had observed an additional value of 3.5 kg in the birthweight the median would be the average of the 3rd and the 4th observation in the ranking, namely the average of 1.4 and 1.5, which is 1.45 kg.

تلخيص البيانات الكمية

- ميزة الوسيط أنه لا يتأثر بالقيم المتطرفة أو الناشزة مثلما هو الحال مع الوسط الحسابي ، فلو حدث أي خلل طارئ باستبدال الرقم ٢.١ كغ بالرقم ٢١ كغ .. لن يتأثر .. ولكنه ليس «فعالا» بالمفهوم الإحصائي لأنه لا يأخذ بالحسبان جميع القيم الموجودة بالدراسة ...

تلخيص البيانات الكمية

- المنوال Mode : هو ثالث نوع من مقاييس النزعة المركزية . إنها القيمة الأكثر تكرارا أو الأكثر مشاهدة في الدراسة . وإذا ما جمعت البيانات وقسمت إلى مجموعات ، فإنها المجموعة ذات التكرارية الأعلى .

- عندما نقول «توزع ثنائئ المنوال» فإن هذا يشير إلى التوزع الذي يحوي على ذروتين بداخله ، مثلما يحدث عندما يتم الدمج بين مجتمعين معا ، مثلما هو الحال عندما ندرس الأطوال لمجموعة تحتوي على ذكور وإناث ..

تلخيص البيانات الكمية

مقاييس التشتت أو التباين

- المدى أو المدى ما بين الربيعي . المدى هو الفرق بين قيمة المشاهدة الصغرى والكبرى . هذا أبسط مقياس تشتت.

- ولكن المشكلة في حال وجود قيم ناشزة ، فإن الصورة قد تصبح مشوهة .. لأننا نشمل فقط نقطتين من نقاط البيانات عند حساب المدى .

تلخيص البيانات الكمية

- الربيعات : الربيع السفلي والوسيط والربيع العلوي تقسم البيانات إلى أربعة أقسام متساوية وبالتالي سوف تكون أعداد متساوية تقريبا من المشاهدات في كل ربع .

- تحسب الربيعات بطريقة مشابهة لطريقة حساب الوسيط ..
- يعتبر المدى ما بين الربيعي من المقاييس الجيدة لدراسة التباين (التشتت)... وهو الفرق بين الربيع السفلي والربيع العلوي .. وهو لا يتأثر بالقيم الناشزة .. ونوع التوزيع الذي بين أيدينا حيث نعلم أن ٥٠% من المشاهدات تقع حتما ضمن هذا المدى ...

Illustrative example: Calculation of the range, quartiles and interquartile range

Suppose we had 10 birthweights arranged in increasing order from the Simpson (2004) study.

Order	Birthweight (kg)	
1	1.51	
2	1.55	
3	1.79	← Lower quartile (25th percentile)
4	2.10	
5	2.18	
		← Median (50th percentile)
6	2.22	
7	2.37	
8	2.40	← Upper quartile (75th percentile)
9	2.81	
10	2.85	

Inter quartile range

The range of birthweights in these data is from 1.51 kg to 2.85 kg (simply the smallest and largest birthweights).

The median is the average of the 5th and 6th observations (2.18 + 2.22)/2 = 2.20 kg. The first half of the data has five observations so the first quartile is the 3rd ranked observation, namely 1.79 kg, and similarly the third quartile would be the 8th ranked observation, namely 2.40 kg. So the interquartile range is from 1.79 to 2.40 kg.

تلخيص البيانات الكمية

- الانحراف المعياري Standard Deviation :
- وهو يعطى بالمعادلة التالية :

$$SD = s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}.$$

- ويتأثر هذا المقياس بشكل بالغ بوجود القيم المتطرفة .

Illustrative example: Calculation of the standard deviation

Consider the five birthweights (in kg): 1.2, 1.3, 1.4, 1.5, 2.1. The calculations to work out the standard deviation are given in the following table.

Subject	Weight (kg) x_i	Mean Weight (kg) $\bar{x}$	Differences from mean $x_i - \bar{x}$	Square of differences from mean $(x_i - \bar{x})^2$
1	1.2	1.5	–0.30	0.09
2	1.3	1.5	–0.20	0.04
3	1.4	1.5	–0.10	0.01
4	1.5	1.5	0.00	0.00
5	2.1	1.5	0.60	0.36
Totals (Sum)	7.5		0	0.50 kg^2
Mean	1.50		**Variance**	0.13 kg^2
n	5		**Standard**	
$n-1$	4		**deviation**	0.35 kg

Variance = 0.50/4

SD = square root of the Variance

تلخيص البيانات الكمية

لماذا يعتبر الانحراف المعياري مفيدا ؟

- من البيانات المتعلقة بدراسة سيمبوس ، كان الوسط الحسابي للوزن الولادي ١.٣١ كغ ، والانحراف المعياري هو ٠.٤٢٤ كغ ..
- من المعلوم أنه في ٩٥% من المشاهدات تقع في مدى محصور بالوسط الحسابي زائد ناقص انحرافين معياريين .
- هذا ما نسميه الفاصل المرجعي Reference Interval وهو من مزايا العمل مع الانحرافات المعيارية ...

تلخيص البيانات الكمية

هل نستخدم الأوساط الحسابية أم الوسطاء ؟

- إن كل من هذين المقياسين يعطيان انطباعات مختلفة حول توضع البيانات ، وكلاهما مفيد .
- إذا كان توزع البيانات متناظرا فإن الوسط الحسابي هو أفضل إحصائي تلخيصي Summary Statistic .
- إذا كان توزع البيانات ملتويا ، فإن الوسيط أفضل .
- إذا كان متوسط الدخل السنوي للمواطن السوري ٦٥٠ ألف ل س ، وكان الوسيط ٥٥٠ ألف ل س ، فإن دخل معظم الأفراد سوف يكون أقرب للمقياس الإحصائي التلخيصي الثاني .

تلخيص البيانات الكمية

هل يمكن استخدام الوسط الحسابي مع البيانات غير الكمية؟

- مع البيانات الاسمية الإثنانية Dichotomous ، والتي تكون فيها العلامة إما ٠ أو ١ ، فإن الوسط الحسابي هو ببساطة هو نسبة هؤلاء الأشخاص الذي حصلوا على العلامة ١ .

- أما إذا كانت البيانات فئوية مرتبة (رتبية Ordinal) حيث تقسم استجابات المرضى إلى العلامات التالية : ١ ، ٢ ، ٣ إلخ .. فإن الوسط الحسابي يمكن أن يحسب لهذه البيانات .

تلخيص البيانات الكمية

- هل يمكن استخدام الوسط الحسابي مع البيانات غير الكمية؟
- إن استخدام **الوسط الحسابي** مع البيانات الفئوية المرتبة (الرتبية) يمكن أن يعطي معلومات ربما أكثر فائدة مما لو استخدمنا الوسيط في مثل هذه البيانات ..
- ولكن يجب استخدام الوسط الحسابي هنا بحذر ، بسبب **الافتراض الضمني** أن التغير من العلامة ١ إلى العلامة ٢ ، مثلا ، يعادل في قيمته (أو مقداره أو معناه) التغير من العلامة ٢ إلى العلامة ٣ وهَلُمّ جرا ...

شكرا جزيلا لحسن استماعكم

myhajeer@gmail.com