

مذكرة إضافية

456 صكل

الدكتور/ هشام أبو عودة

تقييم التحليل الإحصائي

للمنشورات الدوائية

Evaluation of
Statistical Analysis of
Drug Literature

تقييم التحليل الإحصائي للمنشورات الدوائية

Evaluation of Statistical Analysis of Drug Literature

قد يتساءل المرء، كيف يمكن لشخص غير متخصص في الإحصاء الحيوي Biostatistics أن يقوم بتقييم الاختبارات الإحصائية التي يتم إجراؤها في الدراسات الإكلينيكية؟ والجواب هو، أنك لا تحتاج إلى القيام بصناعة وتركيب سيارة لكي تتمكن من قيادة إحدى السيارات. ما تحتاج إلى معرفته عن الاختبارات الإكلينيكية هو كيفية اختيار الاختبار الإحصائي المناسب للدراسة إكلينيكية تم إجراؤها بطريقة معينة، ووصف ما يستطيع الاختبار فعله بالكلمات لا بالمعادلات، وما هي الأحوال التي لا يمكن تطبيق هذا الاختبار فيها.

من النادر في الأدبيات والمنشورات الطبية والصيدلانية أن تقرأ بحثاً دون أن تصادف استخدام إحدى الطرق الإحصائية فيه، إذا استخدم الإحصاء الحيوي لوصف عينة الدراسة أو مجتمع المرضى population باستخدام الإحصاء الوصفي Descriptive statistics، أو للخروج باستنتاجات عن مجتمع المرضى بالاعتماد على نتائج مجموعة صغيرة (عينة sample) من المرضى باستخدام الإحصاء الاستدلالي Inferential statistics. من المهم أن يكون لمن يقوم بتقييم المقالات البحثية الإكلينيكية بعض الدراية بكلا النوعين السابقين من الإحصاء لكن يتمكن من أداء مهمته بنجاح. وقد يقول قائل أن المجالات العلمية لا تقوم بنشر البحث العلمي إلا بعد التأكد من خلوه من الأخطاء في التحليل الإحصائي، فلماذا نتجرأ على التشكيك في صحة إجراء الاختبارات الإحصائية المنشورة في المجالات العلمية المرموقة؟ والجواب على هذا السؤال أتى من المجالات العلمية نفسها، ففي مراجعة للمقالات البحثية الإكلينيكية المنشورة بها، اعترفت بعض المجالات الطبية العالمية المتميزة بأن نصف المقالات البحثية المنشورة بها على الأقل تحتوي على أخطاء علمية سواءً في تصميم الدراسة نفسها أو في التحليل الإحصائي لبياناتها. ومن هذه المجالات، ثلاث مجالات طبية عالمية هي مجلة نيوانجلندا الطبية New England Journal of Medicine، والمجلة البريطانية British Medical Journal، ومجلة لانسييت Lancet. إن التطبيق الخاطئ للطرق الإحصائية هو أحد الأخطاء الشائعة في المنشورات الطبية والصيدلانية، فإذا تم إجراء تحليل إحصائي بطريقة خاطئة، فإن ذلك يبطل شرعية استنتاجات الدراسة.

يقوم المؤلف عادة بوصف الطرائق الإحصائية التي استخدمها في دراسته وذلك في طرائق الدراسة (منهجية الدراسة) مما يتيح للقارئ معرفة الخطوات الإحصائية المستخدمة. ونهدف في هذا الفصل إلى تزويد القارئ بمعلومات كافية عن الطرائق الإحصائية المستخدمة في الدراسات الإكلينيكية ليتمكن من تقييم هذه الدراسات، ولن نقوم هنا بمناقشة الحسابات الرياضية المستخدمة لإجراء هذه الاختبارات، فهذا ليس هدف هذا الفصل، وباستطاعة القارئ الرجوع إلى كتب الإحصاء المتخصصة للاستزادة في هذا المجال.

مستويات القياس Levels of measurement

يعتمد إجراء الاختبار الإحصائي المناسب على نوعية البيانات التي يتم جمعها وعلى نوعية الدراسة وخطتها العامة. وهناك أربعة أنواع من البيانات يمثل كل منها قياساً معيناً وهي:

- **البيانات الإسمية Nominal data**: وهي أضعف أنواع البيانات حيث يوضع الأشخاص في مجموعات فرعية subgroups أو فئات categories حسب الصفات والمميزات الخاصة بهم. وفي أبسط أشكالها تكون البيانات الإسمية ذات تفرع ثنائي dichotomous، فهي إما أن يكون المريض ذكراً أو أنثى، ولهذا فإن المريض لا يمكن أن يكون في المجموعتين الفرعيتين في نفس الوقت بل يقتصر على مجموعة واحدة فقط (اقتصار متبادل mutual exclusive).

- **بيانات الرتبة ordinal data**: وهي ذات مستوى أقوى للقياس من البيانات الإسمية، وفيها توضع البنود في مجموعات فرعية أيضاً ولكن البيانات تكون مرتبة ordered or ranked. ويستخدم هذا النوع في ترتيب استجابة الأشخاص مثل قياس مدى الألم، كأن نقوم بتصنيف الألم إلى عدة فئات مرتبة كالتالي: لا يوجد ألم No pain، وألم خفيف Mild، وألم متوسط Moderate، وألم شديد Severe، وألم لا يطاق Excruciating pain. وليس بالضرورة أن تكون المسافة بين الفئات متساوية.

- **بيانات الفترة Interval data**: وهي أعلى مستويات القياس السابقة إذ أنها من البيانات الكمية Quantitative data على عكس النوعين السابقين إذ أنهما من البيانات الكيفية (النوعية) Qualitative data. وفي هذا النوع من القياس يكون لكل شخص قراءة واحدة تمثل الصفة التي تتم دراستها مثل معدل النبض 70، 73، 75، وغيرها، نبضة في الدقيقة، والطول 175 سم، 185 سم، وغيرها. وهناك أمثلة كثيرة مثل ضغط الدم، ودرجة الحرارة، ومستوى الدواء في البلازما، وغيرها. وفي مثل هذه البيانات يوجد دائماً متغيران variables أحدهما للصفة مثل الطول أو الوزن وغيرها، والآخر للتكرار Frequency، ومثال ذلك أنه ضمن مدى الطول 162 إلى 164 يوجد عشرة أشخاص بين هذين الطولين.

- **بيانات النسبة Ratio data**: وهي أعلى مستويات القياس على الإطلاق، وتتميز بجميع الخواص السابقة بالإضافة إلى وجود الصفر المطلق الذي يعني عدم وجود الخاصية المطلقة مثل الوزن أو الطول، ولكن يمكن النظر إلى هذه الخواص من زاوية أخرى مثل أن نقول أن وزن الشخص يبلغ ضعف وزن شخص آخر.

كما ذكرنا سابقاً، يطلق على النوعين الأولين اسم البيانات الكيفية أو النوعية Qualitative، ويطلق عليهما أيضاً إسم البيانات المنفصلة discrete (أو غير المستمرة أو غير المترابطة). أما النوعان الآخران فيطلق عليهما اسم البيانات الكمية Quantitative، ويطلق عليهما اسم البيانات المستمرة continuous. ويمكننا تحويل البيانات الكمية إلى بيانات كيفية، ولكن لا يمكننا تحويل البيانات الكيفية إلى بيانات كمية.

أنواع العينات

هناك نوعان من العينات هما: العينات المستقلة Independent samples والعينات المرتبطة أو المزدوجة : Dependent or paired samples

- **العينات المستقلة** Independent samples : ومثالها تقسيم عينة الدراسة إلى مجموعتين وإعطاء الدواء A للمجموعة الأولى وإعطاء الدواء B للمجموعة الثانية، ففي هذه الحالة لا علاقة للمجموعتين ببعضهما البعض، وكل منهما مستقلة عن الأخرى، أي أنه لا توجد إلا درجة واحدة من القياس لكل فرد في المجموعة.
- **العينات المرتبطة أو المزدوجة** paired (dependent) sample : ومثالها تقسيم العينة إلى مجموعتين وإعطاء المجموعة الأولى الدواء A ثم إعطاء الدواء B للمجموعة الثانية. وبعد فترة غسيل (انقطاع) مناسبة washout تعطي المجموعة الأولى الدواء B ثم تعطي المجموعة الثانية الدواء A. وهكذا فإن كل فرد في المجموعتين يكون قد تناول الدواءين، وبالتالي يكون لكل فرد درجتين من القياس.

إن استخدام العينات المرتبطة أفضل من استخدام العينات المستقلة في كثير من الأحوال وكلما سمحت الدراسة بذلك، ففيها يتم دراسة الشخص الواحد تحت ظرفين مختلفين أو أكثر، وبالتالي يتم فيها التخلص من التباين الذاتي Intraindividual variability ولكن لا يتم التخلص من التباين بين الأفراد Interindividual variability أما في العينات المستقلة فلا يتم التخلص من هذين النوعين من التباين.

فيما يلي من سطور سوف نقوم بشرح أهم النقاط الأساسية في الإحصاء الحيوي المستخدمة في معظم الدراسات الإكلينيكية. وشرحنا هذا لن يكون بديلاً عن الرجوع إلى أحد الكتب المتخصصة في الإحصاء الحيوي، ونفترض أيضاً أن لدى القارئ إلمام مسبق بهذا الموضوع لأن هذا الفصل سوف يتجاهل الكثير من الأمور البسيطة والتي نفترض معرفة القارئ بها. لذلك فإن هذا الفصل لا يحتوي إلا على معلومات مختصرة من شأنها تذكير القارئ بالإحصاء الحيوي على شكل نقاط محددة.

ينقسم الإحصاء الحيوي إلى نوعين رئيسيين هما:

- الإحصاء الوصفي Descriptive statistics
- الإحصاء الاستدلالي Inferential statistics

أولاً: الإحصاء الوصفي Descriptive statistics

يستخدم الإحصاء الوصفي للحصول على ملخصات للبيانات من شأنها وصف بيانات العينة. وينقسم الإحصاء الوصفي إلى نوعين هما:

- مقاييس النزعة المركزية Measures of Central Tendency.

• مقاييس التشتت (التباين) (Measures of Dispersion (Variability))

أ - مقاييس النزعة المركزية (Measures of Central Tendency)

هي طريقة بسيطة ومناسبة لتلخيص البيانات للحصول على قيمة واحدة تمثل مجموعة من القياسات. وحيث أن هذه القيمة تنزع لأن تكون في منتصف مجموعة البيانات فقد أطلق عليها اسم مقاييس النزعة المركزية. وهذه المقاييس هي المتوسط والوسيط والمنوال.

1. المتوسط (Mean (Average) :

- يدعى المتوسط الحسابي في كثير من الأحيان، لتمييزه عن أنواع المتوسطات الأخرى.
- يتم حسابه بجمع جميع البيانات وقسمة المجموع على عدد البيانات كالتالي :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

ويستخدم هذا المقياس مع البيانات الكمية فقط، ولكن من أهم عيوبه أنه يتأثر بالقيم الناشئة outliers وهي القيم المتطرفة values extreme.

2. الوسيط Median : هو القيمة الوسطى في مجموعة من البيانات مرتبة ترتيباً تصاعدياً أو تنازلياً، أي أنها القيمة التي يسبقها في الترتيب نصف البيانات ويعقبها النصف الآخر. ويستخدم هذا المقياس مع البيانات الكمية في الغالب، ولكن يمكن استخدامه مع البيانات الترتيبية إذا كان عدد الفئات فيها كبير نسبياً، ومن أهم مميزاته أنه لا يتأثر بالقيم الناشئة.

3. المنوال Mode : هو القيمة الأكثر تكراراً في مجموعة من البيانات. وقد لا يوجد المنوال في إحدى المجموعات، ولكن قد يتوفر أكثر من منوال في مجموعات أخرى وتسمى المجموعة في هذه الحالة ثنائية المنوال Bimodal. ونظراً لعدم استقرار المنوال لا يتم استخدامه كمقياس للنزعة المركزية بشكل واسع.

ب - مقاييس التشتت (التباين) (Measures of Dispersion (variability))

بالرغم من أن مقاييس النزعة المركزية تمدنا بمعلومات مفيدة عن توزيع البيانات، إلا أنها لا تعطينا معلومات عن كيفية تباين البيانات، ومثال ذلك، إذا كان لدينا مجموعتين من البيانات كالتالي :

المجموعة الأولى: 70 80 100 120 130 المتوسط = 100

المجموعة الثانية: 80 90 100 110 120 المتوسط = 100

بالنظر إلى المتوسط في المجموعتين، قد يستنتج المرء أن المجموعتين غير مختلفتين. صحيح أن للمجموعتين نفس المتوسط، ولكن الفحص الدقيق للبيانات في المجموعتين يكشف أن تشتت أو تباين البيانات في المجموعتين مختلف. ولهذا من الضروري إيجاد مقياس لوصف التشتت أو التباين. والمقاييس المستخدمة لهذا الغرض هي التباين variance، والانحراف المعياري standard deviation، والمدى range، ومعامل التباين coefficient of variation.

1. المدى Range:

هو الفرق بين القيمة القصوى maximum والقيمة الصغرى minimum في مجموعة من البيانات.

2. الانحراف المعياري Standard Deviation:

هو أهم مقاييس التشتت على الإطلاق، وسوف تتضح أهميته في الإحصاء الاستدلالي. وهو مقياس لدرجة التشتت لبيانات العينة حول متوسط هذه العينة. ونستطيع حسابه باتباع الخطوات الستة التالية:

- إحسب المتوسط.
- إحسب قيمة الفرق بين البيانات والمتوسط، أي $(X - \bar{X})$.
- قم بتربيع هذا الفرق، أي $(X - \bar{X})^2$.
- إجمع مربعات الفروق لجميع البيانات، أي $\sum (X - \bar{X})^2$.
- إقسم المجموع من الخطوة السابقة على عدد البيانات مطروحاً منه واحد لتحصل على متوسط مجموع المربعات (وسوف نطلق على هذه القيمة فيما بعد اسم التباين Variance).
- خذ الجذر التربيعي لهذه القيمة فتحصل على الانحراف المعياري.

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

وهذه المعادلة هي معادلة اشتقاق الانحراف المعياري، ويمكننا تحويل هذه المعادلة إلى

معادلة أنسب وهي التي نستعملها مباشرة لحساب الانحراف المعياري، وهي كالتالي:

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}}$$

والجزء العلوي من المعادلة (البسط) يسمى مجموع المربعات sum of squares، والجزء السفلي (المقام) يسمى درجات الحرية degrees of freedom. ويعتبر الانحراف المعياري مفيداً في حالة البيانات ذات التوزيع الطبيعي normal distribution، حيث أنه في أي مجموعة بيانات موزعة توزيعاً طبيعياً تقع 68.27% من البيانات في مجال (\pm) انحراف معياري واحد، و 95.45% من البيانات في مجال (\pm) انحرافين معياريين، و 99.73% من البيانات في مجال (\pm) ثلاثة انحرافات معيارية).

3. المتباين Variance :

وهو مربع الانحراف المعياري كما ذكرنا سابقاً، ويستخدم بكثرة في الإحصاء الاستدلالي.

4. الخطأ المعياري للمتوسط Standard error of the mean :

ويرمز له بالرمز SEM وأحياناً $SE_{\bar{X}}$ ، ويمكن حسابه من الانحراف المعياري مباشرة وذلك بتقسيم هذا الأخير على الجذر التربيعي لحجم العينة، أي:

$$SEM = \frac{s}{\sqrt{n}}$$

وهو دائماً أصغر من الانحراف المعياري، ويمثل دقة أو مصداقية العينة حيث أنه يرتبط ارتباطاً وثيقاً بالمجتمع population الذي تم سحب العينة منه. وهو يقيس مقدار الخطأ في تقدير متوسط المجتمع بواسطة متوسط العينة. ويستخدم الخطأ المعياري للمتوسط في حساب فترة الثقة confidence interval، وفي الإحصاء الاستدلالي.

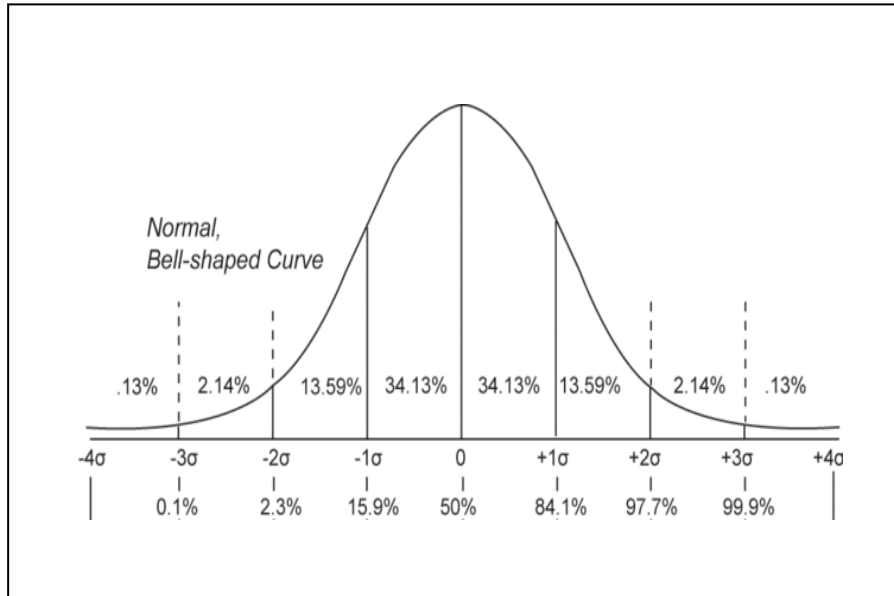
وقبل أن نتقل إلى الشرح المختصر للإحصاء الاستدلالي، يجب علينا التعرف على بعض المقاييس الأخرى وبعض النقاط الإحصائية الهامة مثل فترة الثقة، والمعنوية الإحصائية، والتوزيع الطبيعي، وقيمة مستوى الإحتمالية ألفا (α)، وقيمة p الإحتمالية p-value، وأنواع الخطأ الإحصائي.

التوزيع الطبيعي Normal Distribution

عندما يكون لدينا عدد كبير من البيانات في مجموعة واحدة مثل بيانات الطول أو الوزن أو قياسات ضغط الدم وغيرها من البيانات الكمية التي تم أخذها من مجتمع معين بطريقة عشوائية، ثم نقوم برسمها على شكل منحنى تكراري Frequency Curve فإننا نلاحظ أن بعض البيانات تكون أكبر من المتوسط وأخرى أقل من المتوسط، وسنجد أن معظم البيانات تكون بالقرب من المتوسط، ويكون شكل المنحنى الناتج مثل المنحنى الموضح في الشكل (1.8). ويسمى هذا التوزيع للبيانات بالتوزيع الطبيعي، ويطلق عليه أحياناً اسم التوزيع الجاوسي Gaussian Distribution، وهو على شكل الجرس Bell-shaped، ويقع فيه المتوسط والوسيط والمنوال في نقطة واحدة وجميعها تقسم المنحنى إلى نصفين متطابقين تماماً. ومن خواص هذا المنحنى أن 68.27% من المساحة تحت هذا المنحنى (أي حوالي الثلثين تقريباً) تقع في مدى (± 1) انحراف معياري من المتوسط، و 95.45% من المساحة تحت المنحنى تقع في مدى (± 2) انحراف معياري من المتوسط، و 99.73% من المساحة تحت المنحنى تقع في مدى (± 3) انحراف معياري من المتوسط. وللحصول على مساحة تحت المنحنى تعادل 95% بالضبط فإنها تكون بين مدى (المتوسط ± 1.96 انحراف معياري)، ومساحة تحت المنحنى تعادل 99% بالضبط تكون بين مدى (المتوسط ± 2.58 انحراف معياري).

وللتوزيع الطبيعي أهميته البالغة في الإحصاء الاستدلالي، حيث أن اختيار الاختبارات الإحصائية للبيانات الكمية يعتمد على توزيع هذه البيانات، فإذا كان توزيعها طبيعياً فإننا نختار اختبارات إحصائية معينة، أما إذا لم تكن موزعة توزيعاً طبيعياً فإننا نختار اختبارات إحصائية أخرى مختلفة، وهكذا فإن تحليل البيانات وبالتالي تقييم الدراسات يعتمد اعتماداً كبيراً على هذا الأمر.

الشكل (1.8): منحنى التوزيع الطبيعي



فترة الثقة Confidence Interval

هي وسيلة أخرى لوصف البيانات، وتوضح فترة الثقة كيف يمكن للمقاييس الوصفية مثل المتوسط والانحراف المعياري والخطأ المعياري للمتوسط أن تقوم بتمثيل المجتمع الحقيقي، تقوم الدراسات الإكلينيكية بعمل افتراضات assumptions حول المجتمع باستخدام عينة من هذا المجتمع، إن قيمة p الاحتمالية p-value تحدد ما إذا كانت إحدى المجموعات تختلف عن أخرى وذلك بالاعتماد على متوسط العينة، ولكنها تفشل في إظهار مدى هذا الاختلاف، وإذا قمنا بسحب عينة أخرى من نفس المجتمع فإننا سوف نحصل على متوسط مختلف، لذلك فإن فترة الثقة تعطينا تقديراً لمدى أكبر من القيم التي تمثل المتوسط الحقيقي للمجتمع، إن أكثر فترات الثقة شيوعاً هي فترة الثقة 95٪ وفترة الثقة 99٪ وفترة الثقة 90٪، ولكن الأولى هي الأكثر استخداماً في الدراسات الإكلينيكية، وتعني أنه لو تم دراسة المجتمع فإن متوسط المجتمع سوف يقع ضمن نطاق فترة الثقة هذه في 95٪ من الحالات وذلك باستخدام بيانات العينة. وتستخدم فترة الثقة وقيمة p الاحتمالية (p-value) في الدراسات الوبائية للمساعدة في حساب درجة الخطورة النسبية Relative Risk (RR) ونسبة الأرجحية Odds Ratio (OR)، حيث أن درجة الخطورة النسبية تحدد الفرق في خطورة تأثير غير مقصود للدواء على شخص لم يتعرض للدواء بالمقارنة مع شخص تعرض له، أما نسبة الأرجحية فهي أرجحية تأثير الدواء في حالة التعرض له أو عدم التعرض له عند نقطة زمنية محددة. وإذا تم حساب فترة الثقة لدرجة الخطورة النسبية RR أو لنسبة الأرجحية OR وكانت فترة الثقة تشتمل على الرقم 1 فإننا نستنتج في هذه الحالة أنه لا فرق بين من تعرض للدواء ومن لم يتعرض له (أي أن النسبية ليست أكبر من 1:1). وفي حالة كون فترة الثقة ضيقة فإن ذلك يعني أن البيانات دقيقة وموثوق بها، أما قيمة p الاحتمالية فإنها لا تقوم مقام فترة الثقة إذ أنها تدل على أن درجة الخطورة النسبية/ أو نسبة الأرجحية تختلف معنوياً Significantly different عن الرقم 1.

المعنوية الإكلينيكية والمعنوية الإحصائية

Clinical Significance and Statistical Significance

عند تطبيق البيانات على المرضى، فإن المعنوية الإحصائية لا تعني بالضرورة وجود معنوية إكلينيكية، فالمعنوية الإحصائية تستغل المؤشرات الإحصائية لإثبات وجود فرق معنوي بين المجموعات، أما المعنوية الإكلينيكية فإنها تحاول إثبات وجود تحسن لدى المرضى أم لا، ومثال ذلك إذا جربنا دوائين لخفض ضغط الدم على مجموعتين من المرضى ووجدنا أن هناك فرقاً معنوياً إحصائياً بين الدوائين في خفضهما لضغط الدم، أي أن درجة المعنوية أقل من 0.05 ($p < 0.05$)، وبمعنى آخر أن أحد الدوائين قام بخفض الضغط أكثر من الدواء الآخر، ولكننا في المقابل وجدنا أن المرضى في كلا المجموعتين لم يصل ضغط الدم لديهم إلى المعدل الطبيعي وما زال ضغط دمهم مرتفعاً فوق هذا المعدل، فإن المعنوية الإحصائية لا قيمة لها لأنه لا وجود للمعنوية الإكلينيكية.

ثانياً: الإحصاء الاستدلالي Inferential Statistics

ناقشنا فيما سبق أهم مؤشرات الإحصاء الوصفي ونوع البيانات والتوزيع الطبيعي وفترات الثقة، والإحصاء الوصفي يقوم كما يوحي الاسم بوصف مجموعة البيانات فقط ولكنه لا يعطينا أي استدلال (استنتاج) عن الفرضية الأساسية Null Hypothesis، أو بمعنى آخر لا يرينا الفرق بين المجموعات في حالة وجود فرق، ونادراً ما يتاح لنا عند إجراء البحوث الطبية والصيدلانية القيام بتقييم كل مجتمع population المرضى، ونستعيض عن ذلك بدراسة عينة sample من هؤلاء المرضى، وبعد ذلك تتم محاولة التقدير الاستقرائي extrapolation لنتائج العينة باستخدام الإحصاء الاستدلالي بواسطة القيام بإجراء الاختبارات الإحصائية المعنية Tests of Significance لمعرفة ما إذا كانت نتائج العينة تماثل النتائج التي يحتمل أن تصادفها في مجتمع المرضى بأكمله.

الفرضية الأساسية Null Hypothesis (H₀)

يمكننا القيام باستدلال إحصائي عن مجتمع population من مجتمعات المرضى باستخدام معلومات العينة sample، والقيم التي نحصل عليها من العينة نطلق عليها اسم "قيمة إحصائية" statistic ونرمز لها برموز مكونة من أحرف إنجليزية مثل المتوسط (\bar{X}) والانحراف المعياري "s" أما القيم التقديرية للمجتمع فنطلق عليها اسم "مؤشرات" أو "بارامترات" parameters ونرمز لها برموز إغريقية مثل متوسط المجتمع "μ" والانحراف المعياري "σ"، وتختلف القيمة الإحصائية من عينة لأخرى، حتى في العينات التي تم سحبها من نفس المجتمع بطريقة عشوائية، فعلى سبيل المثال، لو سحبنا عينة من 100 شخص من مجتمع كبير وحسبنا إحدى القيم الإحصائية كالتوسط مثلاً، ثم أعدنا سحب عينة أخرى من نفس المجتمع عشوائياً وقمنا بحساب المتوسط، وأعدنا هذه العملية عشرات المرات، فإننا لن نحصل على نفس قيمة المتوسط في كل مرة، ويمكننا الحصول على توزيع تكراري Frequency distribution من هذه المتوسطات.

عادةً ما نرغب في مقارنة مجموعتين أو أكثر من المجموعات العلاجية باستخدام عينات تمثل هذه المجموعات، فمثلاً قد نقوم بمقارنة معدل الشفاء من أحد الالتهابات البكتيرية باستخدام نوعين مختلفين من المضادات الحيوية، أو مقارنة درجة تخفيف الألم بعد تناول أحد ثلاثة أدوية مسكنة للألم. لهذا الغرض نستخدم أحد الاختبارات الإحصائية statistical tests المناسبة لكي نعرف ما إذا كان الفرق بين هذه المجموعات هو فرق ناجم عن الصدفة chance أو أنه فرق حقيقي real difference. والخطوة الأولى لإجراء أي اختبار إحصائي هو صياغة "الفرضية الأساسية" Null Hypothesis (H₀) والتي تنص دوماً على "أنه لا يوجد فرق حقيقي بين المجموعات، وأي فرق هو نتيجة للخطأ في أخذ العينة". أما الفرضية البديلة Alternative Hypothesis ونرمز لها بالرمز H_A أو H₁ فتتنص على "أن هناك فرقاً حقيقياً بين المجموعات"، وقد تم تصميم الاختبارات الإحصائية لرفض Reject الفرضية الأساسية مما يسمح بقبول Accept الفرضية

البديلة ، فعلى سبيل المثال ، لو أن باحثاً رغب في مقارنة دواء مسكن analgesic بدواء وهمي "دواء شبيه" placebo فإن فرضيته الأساسية ستكون أن الدواء المسكن والدواء الوهمي متكافئان في تأثيرهما المسكن (أي لا فرق بينهما). وبعد إجراء الدراسة وأخذ البيانات تبين أن تأثير الدواء المسكن كان أكبر من تأثير الدواء الوهمي في خفض الألم ، وكانت نتيجة إجراء الاختبار الإحصائي المناسب هي رفض الفرضية الأساسية ، لذلك نستنتج أن الفرق بين الدوائين هو فرق حقيقي وليس نتيجة للصدفة. أما إذا فشل الاختبار الإحصائي في رفض الفرضية الأساسية ، فإن ذلك يعني أن الفرق هو نتيجة للصدفة وليس فرقاً حقيقياً. ولا يوجد اختبار إحصائي بمقدوره قبول الفرضية الأساسية بتأكيد مطلق إذ أن أي اختبار إحصائي لا يزودنا بتقييم مطلق لوجود أو عدم وجود العلاقة بين السبب cause والتأثير effect ، ولكنه يتيح لنا تقدير احتمالية probability اتخاذ القرار الصحيح من ناحية رفض أو عدم رفض الفرضية الأساسية. وهناك نوعان من أنواع الخطأ error يمكن ارتكابهما في محاولة الوصول إلى القرار الصحيح بشأن الفرضية الأساسية (الجدول 1.8) ، وهما كالتالي :

- **خطأ النوع الأول Type I error** : ويتمثل هذا الخطأ في أننا قد نرفض الفرضية الأساسية في الوقت الذي تكون فيه صحيحة وكان يجب علينا عدم رفضها. ويدعى هذا الخطأ أيضاً "خطأ ألفا" α -error.
- **خطأ النوع الثاني Type II error** : ويتمثل هذا الخطأ في أننا قد لا نرفض الفرضية الأساسية في الوقت الذي كان يجب علينا رفضها لأنها غير صحيحة. ويدعى هذا الخطأ أيضاً "خطأ بيتا" β -error. وينتج الخطأ الأول لأسباب عديدة وأهمها سبب رئيسي واحد وهو استخدام الاختبار الإحصائي الخاطئ لتحليل البيانات ومعرفة الفروقات بين المجموعات. أما الخطأ الثاني فإنه ينتج عن استخدام حجم عينة sample size صغير. كما أن بعض الاختبارات الإحصائية أقوى من غيرها في التعرف على الفروقات بين المجموعات. ويمكننا تعريف "قوة الاختبار الإحصائي Statistical power" على أنها "قدرة الاختبار الإحصائي على رفض الفرضية الأساسية عندما تكون غير صحيحة ويجب رفضها". ويمكن تعريف "القوة" أيضاً بالمعادلة التالية :

$$\text{Power} = 1 - \beta$$

حيث أن β كما أسلفنا هي مقدار خطأ النوع الثاني وتبلغ في العادة (0.8). وكقاعدة عامة ، فإن قوة الاختبار الإحصائي تزداد بزيادة حجم العينة sample size.

الجدول 1.8: تفسير النتائج المحتملة من التحليل الإحصائي :

النتيجة الحقيقية		النتيجة التي أظهرها الاختبار الإحصائي
يوجد فرق معنوي بين المجموعات	لا يوجد فرق معنوي بين المجموعات	
خطأ النوع الأول Type I error (نتيجة موجبة زائفة)	نتيجة صحيحة (قرار سليم)	لا يوجد فرق معنوي بين المجموعات لذلك لا نرفض الفرضية الأساسية
نتيجة صحيحة (قرار سليم)	خطأ النوع الثاني Type II error (نتيجة سلبية زائفة)	يوجد فرق معنوي بين المجموعات لذلك نرفض الفرضية الأساسية

أنواع الإحصاء الاستدلالي

تتطلب جميع اختبارات الإحصاء الاستدلالي معرفة الأمور التالية قبل الشروع في اختيار الاختبار الإحصائي المناسب :

- نوع البيانات data type (أي ، بيانات اسمية Nominal ، أو ترتيبية ordinal ، أو بيانات الفترة Interval ، أو بيانات النسبة Ratio).
 - التوزيع distribution الذي تنتمي إليه البيانات (هل هي موزعة توزيعاً طبيعياً Normally distributed أو غير ذلك؟).
 - عدد مجموعات أو عينات الدراسة Number of groups or samples.
 - نوع العينات (المجموعات) ، (هل هي عينات مستقلة independent أو عينات مرتبطة dependent or paired).
- ويمكن التعرف على جميع النقاط السابقة بسهولة تامة ما عدا معرفة نوع التوزيع الذي تنتمي إليه البيانات ، إذا نحتاج في بعض الأحيان إلى إجراء بعض الاختبارات الإحصائية الخاصة بذلك ، ولن نخوض في تفاصيل كثيرة هنا حول هذا الموضوع إذ أن بإمكان القارئ الرجوع إلى الكتب المتخصصة في هذا المجال ، ولكن أهم الاختبارات الإحصائية التي يمكننا إجراؤها لمعرفة ما إذا كان توزيع البيانات توزيعاً طبيعياً هي ما يلي :

- اختبار "شاييرو- ويلك" Shapiro-Wilk test.
- اختبار "كولموغوروف- سميرنوف" Kolmogorov-smirnov test.
- اختبار ليليفورز Lilliefors test.

- إختبار أندرسون- دارلنج Anderson-Darling test.
- إختبار راين- جوينر Ryan-Joiner test.
- إختبار جاك- بيرا Jacque-Bera test.
- تحليل منحني الإحتمالية الطبيعية Probit analysis.

هناك نوعان من أنواع الإحصاء الاستدلالي هما :

1. الاختبارات الإحصائية المقياسية Parametric tests :

تستخدم هذه الاختبارات عندما تكون البيانات من النوع الكمي (أي بيانات الفترة أو النسبة) وتكون هذه البيانات موزعة توزيعاً طبيعياً.

2. الاختبارات الإحصائية اللامقياسية Nonparametric tests :

تستخدم هذه الاختبارات عندما تكون البيانات الكمية غير موزعة توزيعاً طبيعياً، كما يمكن استخدام بعضها في حالة البيانات الإسمية والبيانات الترتيبية والتي لا تنطبق عليها معايير الإحصاء المقياسي.

ماذا تعني قيمة p:

دائماً ما نطالع في المنشورات الصيدلانية والطبية ذكراً لقيمة p (p-value) مع اهتمام كبير بهذا الحرف الصغير وهذه القيمة دون فهم حقيقي لمعناها الصحيح. لقد جرى العرف على أخذ قيمة (0.05) كأكبر قيمة لها. وتمثل هذه القيمة احتمالية probability ارتكاب خطأ النوع الأول Type I error وذلك عند رفض الفرضية الأساسية في الوقت الذي كان يجب فيه عدم رفضها، وعند الحصول على قيمة p أصغر من 0.05 (أي $p < 0.05$) فإننا نستنتج في هذه الحالة أن الفرق بين مجموعات الدراسة له دلالة إحصائية هام (أي فرق معنوي) Statistically Significant Difference وعندما تكون قيمة p أكبر من 0.05 (أي $p > 0.05$) فإننا نستنتج أن الفرق بين مجموعات الدراسة ليست له دلالة إحصائية وأنه غير هام (أي أن الفرق غير معنوي إحصائياً) Statistically Insignificant، وأن هذا الفرق الملحوظ بين المجموعات ليس فرقاً حقيقياً وإنما حدث بالصدفة.

إن الحصول على قيمة p أقل من 0.05 (أي أن الفرق بين المجموعات هو فرق حقيقي) لا يثبت بالضرورة أن دواء معيناً أفضل من غيره إذا كنا نعقد هذه المقارنة بين دوائين، فقيمة p الصغيرة هذه تعني وبمنتهى البساطة أنه من غير المحتمل أن الفرق الملحوظ بين الدوائين كان نتيجة للصدفة chance، ولذلك فإنه من المحتمل أن يكون الفرق نتيجة لعوامل أخرى. لذلك يصبح من الضروري القيام بتقييم دقيق لأهداف الدراسة والطرائق التي اتبعتها الباحثة، فإذا تأكدنا أن الفرق الملحوظ لم يكن نتيجة للصدفة ($p < 0.05$) وأن

هذا الفرق لم يكن نتيجة لبعض الهفوات في طريقة الدراسة، فإننا في هذه الحالة نستطيع افتراض وجود فرق بين الدوائين.

خطوات إجراء الاختبارات الإحصائية:

قبل أن نقوم بمناقشة اختبارات الإحصاء الاستدلالي المختلفة يجب علينا الإطلاع على الخطوات التي يجب أتباعها لإجراء أي اختبار إحصائي من هذه الاختبارات، وهذه الخطوات كالتالي:

1. أذكر الفرضية الأساسية التي تحدد اتجاه اختيار العينة وطريقة التحليل الإحصائي.
2. حدد مجتمع الدراسة population الذي سيتم سحب العينة sample منه، ويجب تحديد حجم العينة sample size، وقد سبق لنا مناقشة أهمية أن يكون حجم العينة كافياً للحصول على نتائج موثوقة وذات مصداقية.
3. حدد مستوى المعنوية Level of Significance (أي 0.05، أو 0.01، أو غيرها)، وحدد الاختبار الإحصائي Statistical test المناسب الذي يجب إجراؤه في هذه الحالة.
4. قارن القيمة الإحصائية التي حصلت عليها نتيجة إجراء الاختبار الإحصائي بالقيمة الحرجة critical value المذكورة في الجداول الإحصائية Statistical Tables الخاصة بالاختبار الإحصائي الذي تم إجراؤه، وعليك أن تقرر بعد ذلك رفض أو قبول الفرضية الأساسية. ومن الجدير بالذكر فإن معظم برامج الإحصاء بالكمبيوتر تعطيك قيمة p مباشرة دون الحاجة إلى الاستعانة بالجدول الإحصائية.

اختيار الاختبار الإحصائي المناسب

يعتمد اختيار الاختبار الإحصائي المناسب على عوامل عدة سبق لنا مناقشتها، مثل نوع البيانات، وعدد المجموعات، والتوزيع الذي تنتمي إليه البيانات وحجم العينة، وغيرها من العوامل الأخرى، ويوضح الشكلان 1.8 و 2.8 أهم اختبارات الإحصاء الاستدلالي الشائعة الاستعمال في مجال الصيدلة والطب. وسوف نتطرق إلى إعطاء نبذة مختصرة عن كل اختبار من هذه الاختبارات دون الخوض في تفاصيل كثيرة، فالهدف هنا هو التعريف بكل اختبار منها، ولمعرفة المزيد فإننا نحيل القارئ إلى الكتب المتخصصة في هذا المجال. وفيما يلي عرض سريع لهذه الاختبارات الإحصائية.

الإختبارات الإحصائية المقياسية Parametric tests

1. إختبار t (t-test):

يستخدم هذا الاختبار في حالة البيانات الكمية (بيانات الفترة أو النسبة)، وهناك من ينادي باستخدامه، وهم قلة، في حالة البيانات الترتيبية ordinal (مثل البيانات الناتجة عن مقياس لايكرت Likert scale) رغم

أنها لا تتوفر فيها شروط استخدام هذا الاختبار. وفي العادة يستخدم هذا الاختبار لمقارنة عينتين (أو مجموعتين) تكون البيانات فيهما موزعة توزيعاً طبيعياً. وهناك عدة اختبارات تحت هذا الإسم ولها معادلات مختلفة ، ولكن أشهرها اختباران هما :

• اختبار t للعينات المستقلة Independent t-test

• اختبار t للعينات المرتبطة Paired t-test

ويقوم كلا من الاختبارين بمقارنة متوسطي المجموعتين لمعرفة ما إذا كان الفرق بينهما فرقاً حقيقياً وليس نتيجة للصدفة. ولا يجوز استخدام اختبار t في المقارنات المتعددة multiple comparisons أي المقارنة بين ثلاث مجموعات مثلاً بحيث تتم مقارنة كل مجموعتين على حدة، إذ أن ذلك قد يؤدي إلى ارتكاب خطأ النوع الأول. والاختبار المناسب في هذه الحالة هو "تحليل المتباين ذو الاتجاه الواحد (ANOVA).

2. تحليل المتباين ذو الاتجاه الواحد : ANOVA

يستخدم هذا الاختبار لمقارنة ثلاث مجموعات أو أكثر تكون البيانات فيها من النوع الكمي وموزعة توزيعاً طبيعياً. ونظرياً يمكننا مقارنة عدد لا نهائي من المجموعات ، وهناك أنواع مختلفة من اختبار تحليل المتباين ، ويعتمد استخدام كل نوع منها على نوع البيانات المكونة للمجموعات.

يقوم اختبار تحليل المتباين بمقارنة متوسطات جميع المجموعات لمعرفة ما إذا كان الاختلاف بينها هو اختلاف حقيقي أم هو نتيجة للصدفة المحضة ، وفي حالة ما إذا توصل الباحث إلى أن هناك فرقاً حقيقياً بين المتوسطات (فرق معنوي إحصائياً Statistically Significant) فإن عملية القيام بإجراء اختبارات إضافية للتعرف على المجموعة أو المجموعات التي تسببت في هذا الفرق ، وتسمى هذه الاختبارات الإضافية "الاختبارات اللاحقة Post-Hoc test أو *aposteriori tests*" ، ولا يتم إجراؤها إلا بعد إجراء اختبار تحليل المتباين والاستنتاج بأن هناك فرق معنوي إحصائياً بين المتوسطات ، ومن أهم الاختبارات اللاحقة ما يلي :

• اختبار فيشر لأقل فرق معنوي (LSD) Fisher's Least Significant Difference

• اختبار بونفيروني Bonferroni

• اختبار ستودنت - نيومان - كولز Student-Neuman-Keuls (SNK)

• اختبار دنكان المتعدد المدى Duncan Multiple Range

• اختبار توكي Tukey test

• اختبار شوفيه Scheffé test

• اختبار دانيت Dunnett test (ويستخدم هذا الأخير في حالة وجود مجموعة ضابطة control بين

المجموعات)

ومن الجدير بالذكر أننا لا نحتاج إلا لإجراء اختبار واحد من هذه الاختبارات فقط. هناك أنواع عديدة أخرى من تحليل المتباين نلخصها فيما يلي :

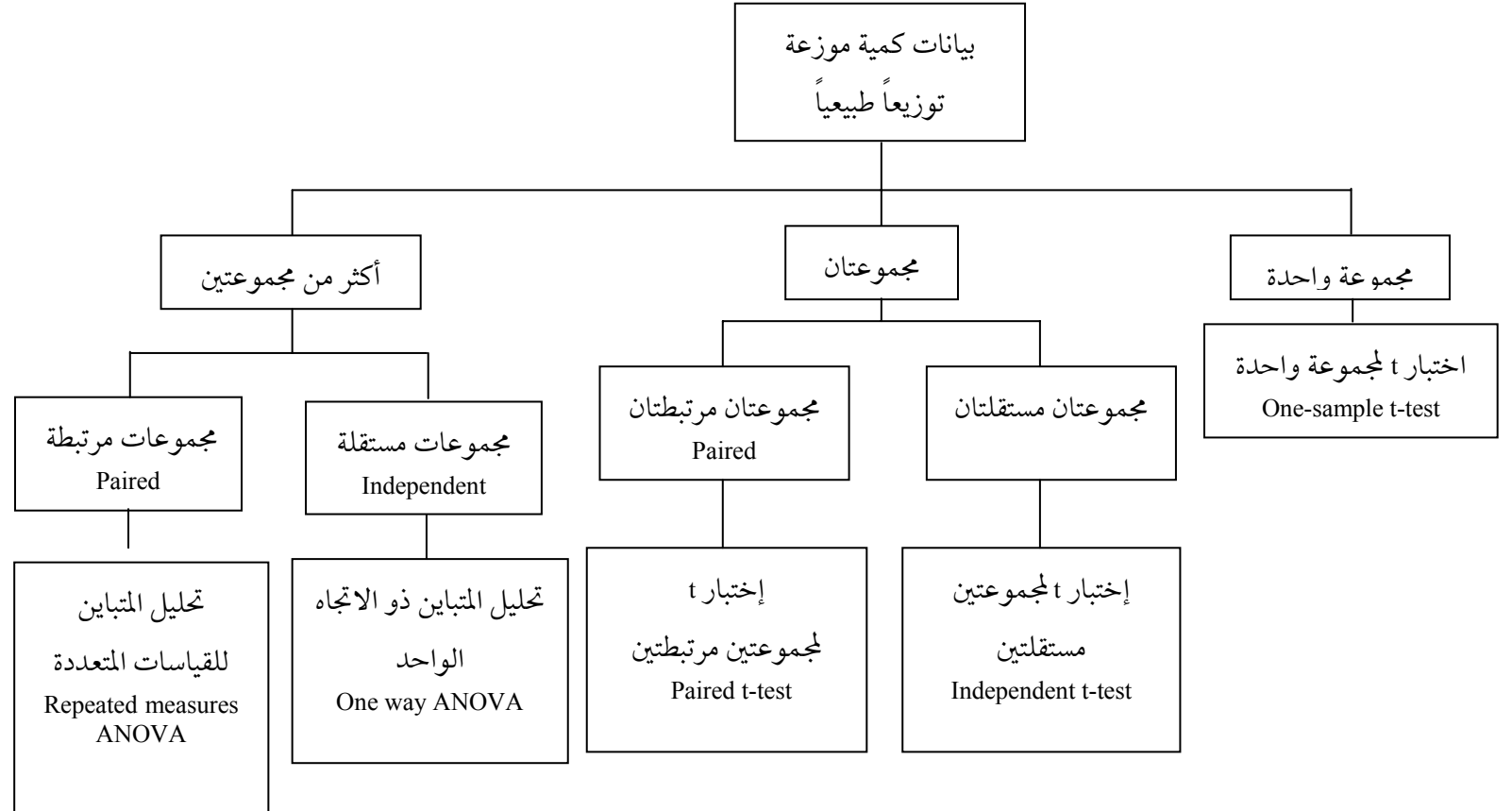
- **تحليل المتباين ذو القياسات المتكررة Repeated Measures ANOVA :**

يستخدم في حالة وجود قيمتين على الأقل لكل فرد في المجموعات التي تتم مقارنتها، وهذا يعني أن تكون العينات مرتبطة dependent أو مزدوجة paired. وهو بذلك يشبه اختبار t للعينات المرتبطة paired t-test ما عدا أن هذا الأخير يستخدم لمقارنة عينتين مرتبطتين فقط. ومثال تحليل المتباين ذو القياسات المتكررة هو مقارنة ثلاثة أدوية على مجموعة من المرضى بحيث يتناول كل مريض الثلاثة أدوية عشوائياً، كل دواء على حدة، ويتم تناول الأدوية بطريقة تبادلية crossover، ويفصل بين فترات تناول الأدوية فترة انقطاع زمنية مناسبة (فترة غسيل Washout period).

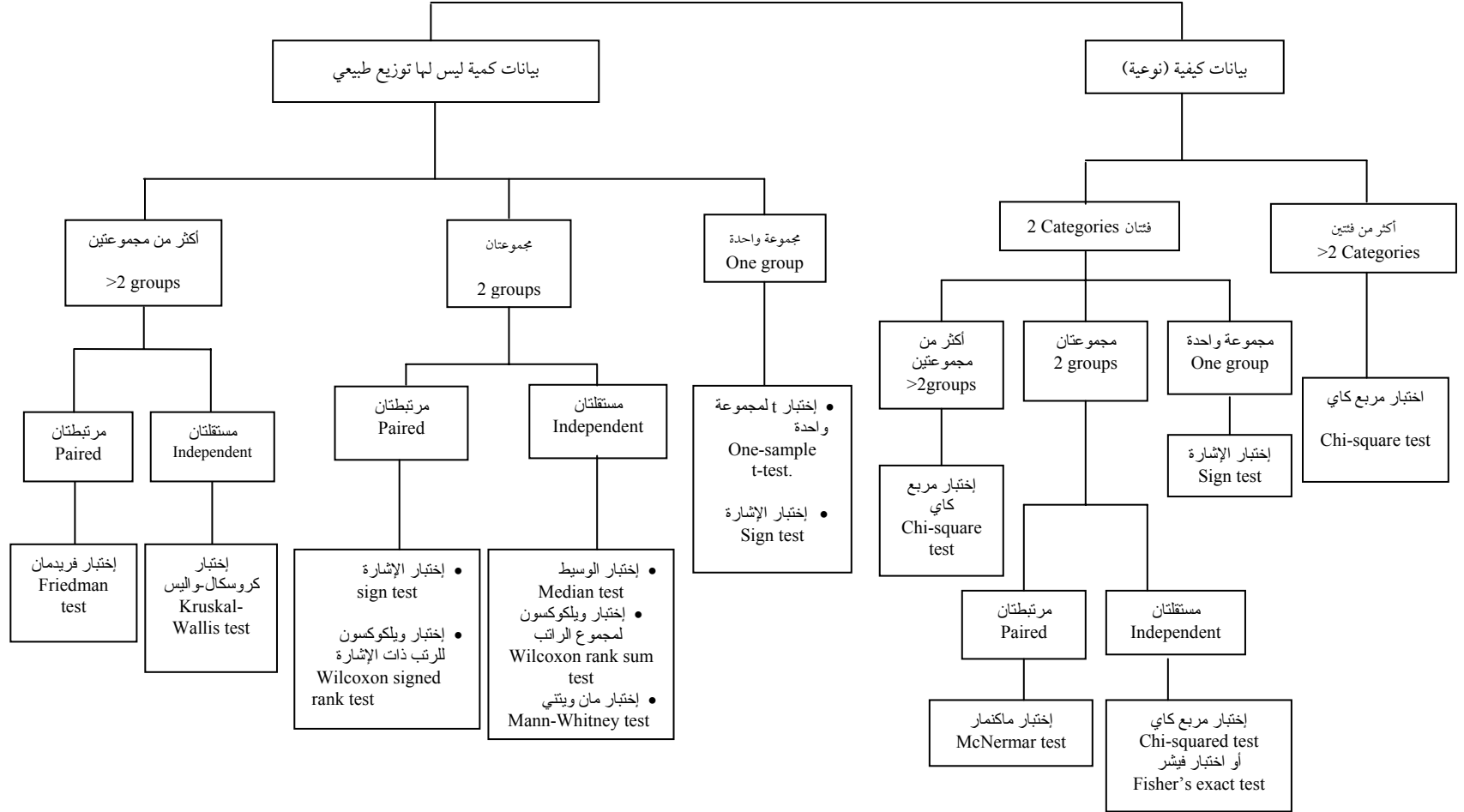
- **تحليل المتباين ذو العوامل المتعددة Factorial ANOVA :**

يستخدم لمعرفة تأثير متغيرات مستقلة متعددة على متغير غير مستقل واحد. ومثال ذلك إذا تم عمل أكثر من قياس لنفس المتغير لكل فرد في المجموعة. وكما رأينا في تحليل المتباين ذي الاتجاه الواحد، نحتاج إلى إجراء اختبارات لاحقة Post-Hoc tests في حالة ثبوت وجود فروقات معنوية بين المجموعات لمعرفة المجموعة أو المجموعات التي تسببت في حدوث هذه الفروقات.

الشكل 2.8 اختيار الاختبار الإحصائي المناسب للبيانات الكمية ذات التوزيع الطبيعي :



الشكل 2.8 اختيار الاختبار الإحصائي المناسب للبيانات الكيفية (والكمية بدون توزيع طبيعي):



الإختبارات الإحصائية الالامقياسية Nonparametric tests

يوضح الشكل 2.8 أهم الاختبارات الالامقياسية التي غالباً ما تستخدم في العلوم الصيدلانية والطبية. هناك اختبارات لا مقياسية خاصة بالبيانات الكيفية (البيانات الاسمية nominal)، وأخرى خاصة بالبيانات الكمية التي لا تكون موزعة توزيعاً طبيعياً.

أ. الإختبارات الالامقياسية للبيانات الإسمية:

Nonparametric tests for nominal data

1. إختبار مربع كاي (χ^2) Chi-square:

يتم في هذا الاختبار تحليل بيانات التكرار frequency data فقط أو البيانات الكمية التي يتم تحويلها إلى بيانات تكرارية أو إلى فئات categories، ويجب أن تكون كل فئة مستقلة عن الفئة الأخرى. ولا يمكن استخدام هذا الاختبار إذا كان حجم العينة sample size أقل من 20 في حالة جدول كاي 2×2 ، كما يجب ألا يقل التكرار المتوقع expected frequency عن 5 في كل خلية من خلاياه الأربعة، وإذا حدث ذلك يبطل استخدام اختبار مربع كاي، ويجب في هذه الحالة استخدام اختبار فيشر المضبوط Fisher's Exact Test. أما إذا كان جدول كاي أكبر من 2×2 كأن يكون مثلاً 2×3 أو 3×3 أو أكبر من ذلك، فيجب ألا يقلل التكرار المتوقع عن 5 في أقل من 20% من الخلايا حتى يمكن استخدام اختبار مربع كاي، وفي حالة حدوث ذلك، لا يصلح اختبار فيشر المضبوط لأن يكون بديلاً له، والحل هو زيادة حجم العينة لتلافي هذا النقص. ويوضح الجدول (2.8) هذه الشروط.

الجدول 2.8: معايير استخدام اختبار مربع كاي واختبار فيشر المضبوط.

إختبار فيشر المضبوط Fisher's Exact test	إختبار مربع كاي Chi-square test	نوع العينة
❖ جدول 2x2 (أي مجموعتين)	❖ حجم العينة أكبر من 20 ❖ التكرار المتوقع expected frequency في كل خلية < 5	مقارنة واحدة مستقلة
لا يمكن استخدام اختبار فيشر المضبوط، والحل هو دمج الخلايا لزيادة التكرار بطريقة منطقية، أو زيادة حجم العينة.	❖ حجم العينة > 20 ❖ لا يوجد أي خلية لها تكرار ملحوظ observed frequency أقل من 1. ❖ لا تزيد نسبة الخلايا التي يكون فيها التكرار المتوقع ألق من 5 عن 20٪ من الخلايا.	ثلاث عينات مستقلة أو أكثر

لو أننا قمنا بإجراء دراسة لمقارنة تأثير الكالسيوم في التقليل من الإصابة بالكسور فأعطينا الكالسيوم لمجموعة من الأشخاص وأعطينا الدواء الوهمي (اللدواء) placebo لمجموعة أخرى من الأشخاص وأعطينا ثم تتبعنا حدوث الكسور لديهم خلال فترة زمنية محددة، وحصلنا على البيانات الموضحة في الجدول 3.8، فإن بإمكاننا معرفة الاختبار الإحصائي المناسب كالتالي:

الجدول 3.8: مثال

المجموعات العلاجية	كسور	بدون كسور	مجموع الصفوف
كالمسيوم Calcium	4	20	24
دواء وهمي placebo	0	20	20
مجموع الأعمدة	4	40	N=44

لو افترضنا أننا استخدمنا كشف مربع كاي لتحليل هذه البيانات، فهل كان اختيارنا لهذا الاختبار سليماً؟ دعونا نرى ذلك:

القاعدة الأولى: N=44 وهي أكثر من 20 وبذلك لا نخالف قانون استعمال اختبار مربع كاي.

القاعدة الثانية: لنقم بحساب التكرار المتوقع في الخلايا الأربعة لمعرفة ما إذا كان يقل عن 5 في أي خلية من الخلايا.

- خلية الكالمسيوم/الكسور = $4 \times 24 \div 44 = 2.18$ *
- خلية الكالمسيوم/بدون كسور = $20 \times 24 \div 44 = 21.81$
- خلية الدواء الوهمي/الكسور = $4 \times 20 \div 44 = 1.81$ *
- خلية الدواء الوهمي/بدون كسور = $20 \times 20 \div 44 = 18.18$

إن خليتين من الخلايا الأربعة (أشرنا إليهما بالرمز *) قيمة التكرار المتوقع فيها أقل من 5، وهكذا فإن القاعدة الثانية قد تم مخالفتها وفي هذه الحالة لا يجوز استخدام اختبار مربع كاي، ويجب استخدام اختبار فيشر المضبوط بديلاً عنه.

2. اختبار فيشر المضبوط: Fisher's Exact test

يستخدم هذا الاختبار كاختبار بديل لاختبار مربع كاي وذلك لمعرفة "التوافق" association في جدول 2×2 عندما يكون التكرار المتوقع في الخلايا أقل من خمسة، ولا يمكن استخدامه في أي جداول أكبر من 2×2 .

3. اختبار ماكنمار: McNemar test

يستخدم لتحليل البيانات الإسمية المرتبطة paired أو dependent أي عندما تجري التجربة على نفس الشخص مرتين. فمثلاً، قد يرغب الباحث في دراسة تأثير دواء على مرض معين، ففي هذه الحالة يقوم بتسجيل شدة المرض قبل تناول الدواء وشدته بعد إعطاء الدواء، ويتطلب الاختبار أن تشمل الدراسة نفس المرضى في الحالتين.

4. اختبار كوكران: Cochran test

هو مثل اختبار ماكنمار ما عدا أنه يستخدم لعقد ثلاث مقارنات أو أكثر.

ب. الإختبارات الالامقياسية للبيانات الكمية

Nonparametric tests for quantitative data

تستخدم هذه الاختبارات لمقارنة البيانات الكمية (بيانات الفترة أو النسبة) والتي تكون غير موزعة توزيعاً طبيعياً وتتلخص في معظم هذه الاختبارات في ترتيب البيانات من الأصغر إلى الأكبر في كل مجموعة وإعطاء كل منها رتبة Rank رقمية، ثم بعد ذلك معالجة الرتب حسب نوع المجموعات وعددها، وأهم هذه الاختبارات مايلي:

1. كشف الإشارة: Sigh test

يستخدم هذا الكشف في حالة المجموعات المرتبطة paired. يفترض أن أحد الباحثين يرغب في معرفة ما إذا كان أحد الأدوية المنومة يساعد المرضى على النوم بالفعل، فقام

بإعطاء المرضى الدواء في إحدى الليالي ، وأعطى نفس المرضى دواءً وهمياً placebo في ليلة أخرى ، ثم سجل في كل حالة الزمن الذي استغرقه كل مريض ليغط في النوم. إذا لم يكن للدواء أي تأثير فسوف تكون نسبة المرضى الذين ينامون أسرع عند تناول الدواء هي 0.5 أي (50%) ، وبمعرفة العدد الحقيقي للمرضى الذين ناموا في زمن أسرع من نومهم تحت تأثير الدواء الوهمي نستطيع معرفة ما إذا كان الفرق حقيقياً بين الدواء المنوم والدواء الوهمي أو أن هذا الفرق نتيجة للصدفة.

2. اختبار الوسيط: Median test

يستخدم في حالة العينات المستقلة لمقارنة عينتين أو أكثر ، ويتم بتسجيل عدد القيم التي تقل عن الوسيط median وعدد القيم التي تزيد عن الوسيط ، وبعد ذلك نطبق اختبار مربع كاي لمعرفة الفرق.

3. اختبار ويلكوكسون لمجموع الراتب: Wilcoxon Rank-Sum test

يستخدم لمقارنة مجموعتين مستقلتين تتكونان من بيانات كمية غير موزعة توزيعاً طبيعياً.

4. اختبار مان- ويتني: Mann-Whitney test

هو نفس اختبار ويلكوكسون لمجموع الرتب ويعطي نفس النتائج.

5. اختبار ويلكوكسون للرتب ذات الإشارة: Wilcoxon Signed-Rank test

يستخدم لمقارنة مجموعتين مرتبطتين أو مزدوجتين Independent أو paired بياناتهما غير موزعة توزيعاً طبيعياً.

مقاييس الارتباط والتراجع الخطي

Measures of Correlation and Linear regression

الارتباط Correlation هو خطوات إحصائية تتيح لنا قياس درجة ارتباط متغيرين ببعضهما البعض بحيث إذا تغير أحدهما تغير الآخر بالزيادة أو النقصان ، فإذا ازداد أحدهما وازداد الآخر نتيجة لذلك يقال أن بينهما علاقة طردية وأن ارتباطهما موجب ، أما إذا ازداد أحدهما فنقص الآخر فالعلاقة عند ذلك تكون عكسية والارتباط بينهما سالب. وعادة ما نقوم بحساب معامل الارتباط Correlation Coefficient بين متغيرين حسب نوع البيانات والتوزيع الذي تنتمي إليه تلك البيانات. وهناك أنواع متعددة من مقاييس الارتباط أهمها "معامل بيرسون للارتباط" Pearson Correlation Coefficient ويرمز له بالرمز (r) ويتم حسابه لمتغيرين موزعة بياناتهما توزيعاً طبيعياً للحصول على قيمة رقمية عن قوة العلاقة بين هذين المتغيرين ، وهذه القيمة تتراوح بين $(+1)$ و (-1) ، بينما تمثل القيمة صفر عدم وجود أي ارتباط بينهما. أما الإشارة الموجبة فتعني أن الارتباط موجب والعلاقة طردية ، والإشارة السالبة تعني أن الارتباط سالب والعلاقة عكسية ، وكلما اقتربت قيمة (r) من الرقم (1) كلما كانت العلاقة قوية بغض النظر عن الإشارة.

من ناحية أخرى ، عندما تكون البيانات غير موزعة توزيعاً طبيعياً فإننا نقوم بحساب معامل ارتباط آخر هو "معامل سبيرمان للارتباط" Spearman Correlation Coefficient ويرمز له بالحرف الإغريقي (ρ) "ويلفظ رو" ويتم حسابه عن طريق الرتب ranks وينطبق على القيمة الناتجة منه ما ينطبق على (r) .

أما التراجع الخطي Linear regression (ويسمى أحياناً الانحدار الخطي) فإنه يختلف قليلاً عن الارتباط ، ففي الارتباط يحاول الباحث التعرف على الدرجة التي يتغير بها متغيران بالنسبة لبعضهما البعض ، أما في حالة التراجع الخطي ، فإن الباحث يدرس مقدار

التغير في أحد المتغيرين عندما يتغير المتغير الآخر بالزيادة أو النقصان، ونستطيع معرفة العلاقة relationship بين المتغيرين حسب معادلة الخط المستقيم التالية:

$$Y = a + b X$$

حيث Y هو المتغير المرتبط (التابع) dependent variable، وX هو المتغير المستقل independent variable، و a هو القاطع intercept و b هو الميل (أو الانحدار) slope. وهناك أنواع أخرى كثيرة من التراجع regression مثل التراجع الخطي المتعدد multiple linear regression، وهو بين متغير مرتبط واحد وعدة متغيرات مستقلة، والتراجع اللاخطي Nonlinear regression وغيرها من الأنواع الأخرى والتي لا يتسع المجال لمناقشتها هنا.

تقييم الجزء الإحصائي للدراسات الإكلينيكية

هناك الكثير من الأسئلة التي يجب سؤالها عند تقييم الجزء الإحصائي من أي دراسة إكلينيكية، وسوف نلخص بعض هذه الأسئلة في النقاط التالية:

- هل قام المؤلف بتحديد ما إذا كانت مجموعات الدراسة متجانسة homogeneous، دون تحيز bias لمجموعة معينة، وتم اختيار المجموعات بطريقة عشوائية random؟ وهل قام المؤلف بتعديل الخط القاعدي baseline للمجموعات؟
- ما هي نوعية البيانات data type التي حصل عليها المؤلف؟ وهل قام باستخدام الاختبار الإحصائي المناسب لنوعية هذه البيانات؟
- إذا كانت الاختبارات الإحصائية غامضة أو غير معروفة، فما الذي دفع الباحث إلى استخدامها؟ وهل ذكر مرجعاً بهذا الخصوص؟

- هل تم تحديد البيانات حسب بروتوكول الدراسة Study protocol الأصلي المذكور في جزء الطرائق Methods؟
- هل تم التعرف على القيم الناشئة Outliers أو المتطرفة extreme values بطرق إحصائية متعارف عليها أو بأي طريقة أخرى؟
- هل تم التفريق بين الترابط Correlation وبين التراجع الخطي regression linear؟ وهل تم حساب معامل الترابط Correlation Coefficient وتم تفسيره بشكل صحيح؟
- هل تم تعيين "الترافق" association بين المتغيرات في حالة وجوده؟
- هل تم حساب قيمة p الإحصائية (p-value) وتم تفسيرها بشكل صحيح؟
- هل تم حساب فترة الثقة Confidence Interval؟ وهل تعكس استنتاجات المؤلف هذه الفترة؟
- هل توجد معلومات عن طريقة تعيين حجم العينة Sample size وهل حجم عينة الدراسة مناسب؟ وإذا لم يكن مناسباً، هل تم تحديد "القوة" power لهذه العينة؟
- هل تم حساب المقاييس الإحصائية الوصفية descriptive statistics مثل المتوسط والانحراف المعياري وغيرها؟
- بوجه عام، هل هناك أي ملاحظات أخرى على الطرائق الإحصائية التي تم تحليل البيانات بواسطتها.

الدكتور/ هشام سليمان أبوعودة

Copyright © Hisham S. Abou-Auda 2007