الفصل الخامس

الاستنتاج الإحصائي
Statistical Inference

الآن وبعد أن جمعنا البيانات وقدّرنا بعض الإحصائيات الوصفية الأساسية وافترضنا نموذج احتمال للمجتمع الإحصائي أو العملية فإننا جاهزون لإجراء بعض التحليلات الإحصائية التي من شأنها أن تسمح لنا بمقارنة المجتمعات الإحصائية واختبار الفروض. سنناقش في هذا الكتاب فقط التحليل الإحصائي الذي يكون فيه التوزيع الطبيعي نموذج احتمال جيد للمجتمع (المجتمعات الإحصائية). إذا كانت البيانات أو العينات تشير إلى أن المجتمع الإحصائي أو العملية ليست مُنمذجة بشكل جيد بواسطة التوزيع الطبيعي فإننا بحاجة عندئذ إلى اللجوء إلى أنواع أخرى من التحليل الإحصائي مثل التقنيات غير البارامترية (nonparametric) [6] التي لا تفترض نموذج احتمال أساسي. وتشمل بعض من هذه الاختبارات اختبار مجموع مرتبة ويلكوكسون (Wilcoxon rank sum test) واختبار *U* لمان ويتني (Mann-Whitney *U* test) واختبار كروسكال واليس (Kruskal–Wallis test) واختبار التشغيل (the runs test) [6, 12]. توفّر الكتب الأكثر تقدماً تفاصيل عن إدارة هذه الاختبارات غير البارامترية. يجب أن لا ننسى أنه إذا تم استخدام التحليل الإحصائي الوارد في هذا الكتاب للمجتمعات الإحصائية أو العمليات التي لم يتم توزيعها بشكل طبيعي فإن النتائج قد تكون ذات قيمة قليلة وقد يفقد الباحث نتائج هامة من البيانات.

بافتراض أن بياناتنا تمثل مجتمعًا إحصائيًّا أو عملية مُوزِّعة بشكل طبيعي فإننا جاهزون عندئذ لإجراء مجتمع إحصائي متنوع من الاختبارات الإحصائية التي تسمح لنا باختبار الفروض فيما يتعلق بمساواة المتوسطات والتباينات عبر مجتمعين إحصائيين أو أكثر. لنتذكر أنه من أجل التوزيع الطبيعي هناك حاجة فقط إلى المتوسط والانحراف المعياري للتوصيف التام للطبيعة الاحتمالية للمجتمع الإحصائي أو العملية. وهكذا إذا كان علينا مقارنة مجتمعين مُوزَّعين بشكل طبيعي فإننا بحاجة فقط إلى مقارنة المتوسطات والتغيرات لهذين المجتمعين الإحصائيين. وإذا لم تكن المجتمعات الإحصائية أو العمليات مُوزَّعة بشكل طبيعي فقد تكون هناك بارامترات أخرى مثل عدم التماثل والتفرطح (kurtosis) تميز بين اثنتين أو أكثر من المجتمعات الإحصائية أو العمليات.

(٥,١) مقارنة متوسطات المجتمعات الإحصائية
COMPARISON OF POPULATION MEANS

إن أحد الأسئلة الأساسية التي طرحها العلماء والمهندسون الذين يقومون بعمل التجارب هو ما إذا كان مجتمعان إحصائيان أو طريقتان أو علاجان مختلفيْن حقًّا في النزعة المركزية. وبشكل أكثر تحديداً هل متوسطًا المجتمعين المنعكسين في متوسطي العينتين اللتين تم جمعهما تحت شرطين تجريبيين مختلفان اختلافاً كبيراً؟ أو هل الفرق الذي تم ملاحظته بين متوسطين هو ببساطة بسبب الصدفة وحدها؟

بعض الأمثلة أو الأسئلة التي طرحها المهندسون الطبيون الحيويون التي تتطلب المقارنة بين متوسطّي مجتمعين إحصائيين تشمل ما يلي :

١- هل أحد أدوية العلاج الكيميائي أكثر فعالية من دواء علاج كيميائي آخر في تقليص حجم الورم السرطاني؟

٢- هل هناك فرق في وضعية الراحة بين مجتمع الشباب ومجتمع كبار السن؟

٣- هل هناك فرق في كثافة العظام للنساء اللواتي قبل سن اليأس مقابل أولئك اللواتي بعد سن اليأس؟

٤- هل التيتانيوم مادة أقوى من الفولاذ لعملية زرع العظام؟

٥- بالنسبة للتصوير بالرنين المغناطيسي هل أداء أحد أنواع متتاليات النبضة أفضل من الآخر في الكشف عن مساحات (سُبُل) المادة البيضاء في الدماغ؟

٦- هل تمنع الدعامات داخل الأوعية الشاطفة للأدوية (drug-eluting intravascular stents) عودة التضيق على نحو أكثر فعالية من الدعامات غير المُغلَّفة (noncoated stents)؟

للإجابة على هذا النوع من الأسئلة يجمع مهندسو الطبية الحيوية في كثير من الأحيان عينات من مجتمعين مختلفتين للأشخاص أو من مجتمع واحد من الأشخاص ولكن تحت ظرفين مختلفين. وبالنسبة لقياس محدد يمثل العينات يتم عادة تقدير متوسط العينة لكل مجتمع أو تحت كل ظرف من الظرفين. ربما تكون مقارنة متوسطي مجتمعين إحصائيين منعكسين في مجموعتين من البيانات التحليل الإحصائي الأكثر ذكراً في المراجع العلمية والهندسية ويمكن إنجازه باستخدام ما يعرف بالاختبار t.

(٥,١,١) الاختبار t The t Test

عند إجراء الاختبار t نطرح السؤال التالي "هل المتوسطان لمجتمعين إحصائيين مختلفان حقًا؟" أو "هل سنرى الفروق الملحوظة فقط بسبب الصدفة العشوائية؟" إن المجتمعين منعكسان في بيانات تم جمعها تحت ظرفين مختلفين. وقد تشمل هذه الظروف علاجين أو عمليتين مختلفتين.

لمعالجة هذه المسألة نستخدم أحد الاختبارين التاليين اعتماداً على ما إذا كانت مجموعتا البيانات مستقلتين أو تعتمدان على بعضهما:

١- اختبار t غير مزدوج لمجتمعين إحصائيين من البيانات المستقلة.

٢- اختبار t مزدوج لمجتمعين إحصائيين من البيانات التابعة.

قبل وصف كل نوع من أنواع الاختبار t نحن بحاجة إلى مناقشة فكرة اختبار الفروض.

(٥,١,١,١) اختبار الفروض Hypothesis Testing

كلما أجرينا التحليل الإحصائي نختبر بعض الفروض. في الواقع حتى قبل جمع البيانات نقوم بصياغة فرضية ومن ثم تصميم تجربتنا بعناية وجمع وتحليل البيانات لاختبار الفروض. إن نتيجة الاختبار الإحصائي إذا كان نموذج الاحتمال المُفترض صحيحًا تسمح لنا بقبول أو رفض الفروض وفعل ذلك مع مستوى معين من الثقة.

هناك أساساً نوعان من الفروض نختبرهما عند إجراء التحليل الإحصائي. هما فرضية العدم (الصفرية) (null hypothesis) والفروض البديلة (alternative hypothesis).

يتم التعبير عن فرضية العدم التي يُرمز لها بـ H_0 على النحو التالي بالنسبة للاختبار t مقارناً متوسطيّ مجتمعين μ_1 و μ_2 :

$$H_0 : \mu_1 = \mu_2$$

يتم التعبير عن الفرضية البديلة التي يُرمز لها بـ H_1 على أنها أحد الاختبارات التالية بالنسبة للاختبار t مقارناً متوسطيّ مجتمعين μ_1 و μ_2 :

$H_0 : \mu_1 \neq \mu_2$ (الاختبار t ثنائي الذيل)

$H_0 : \mu_1 < \mu_2$ (الاختبار t أحادي الذيل)

أو

$H_0 : \mu_1 > \mu_2$ (الاختبار t أحادي الذيل)

إذا كانت H_1 من الشكل الأول حيث لا نعرف مسبقاً من مجتمع البيانات أي متوسط سيكون أكبر من الآخر فإننا سوف نجري الاختبار ثنائي الذيل وهو ما يعني

ببساطة أن مستوى الأهمية الذي من أجله نقبل أو نرفض فرضية العدم سيكون ضعف مستوى الحالة التي نتنبأ فيها مقدماً من التجربة أن أحد المتوسطين سيكون أقل من الثاني استناداً إلى الفسيولوجيا أو العملية الهندسية أو الظروف الأخرى التي تؤثر على النتائج التجريبية. إن الطريقة الأخرى للتعبير عن ذلك هي أنه بالاختبار أحادي الذيل سيكون لدينا ثقة في رفض أو قبول فرضية العدم أكبر منها في حالة الاختبار ثنائي الذيل.

(٥,١,١,٢) تطبيق الاختبار t Applying the t Test

الآن بعد أن حددنا فرضياتنا نحن جاهزون لإجراء الاختبار t. وبالأخذ في الاعتبار مجتمعين لهما n_1 و n_2 عينة نستطيع مقارنة متوسطيّ المجتمعين باستخدام الاختبار t. ومن المهم أن نتذكر الافتراضات الأساسية الواردة في استخدام الاختبار t لمقارنة متوسطيّ المجتمعين:

١- التوزيعات الأساسية لكلا المجتمعين طبيعية.

٢- تغيرات المجتمعين متساوية تقريباً: $s_1^2 = s_2^2$.

هناك افتراضات كبيرة نقوم بها عندما تكون n_1 و n_2 صغيرة. وإذا كانت هذه الافتراضات ضعيفة بالنسبة للبيانات التي يجري تحليلها فإننا بحاجة إلى إيجاد إحصائيات مختلفة لمقارنة المجتمعين الإحصائيين.

وبالأخذ في الاعتبار مجتمعين من البيانات التي تم أخذ عيناتها x_i و y_i فإن المتوسطين للمجتمعين أو العمليتين المنعكستين في البيانات التي تم أخذ عيناتها يمكن مقارنتهما باستخدام الإحصاء t التالي:

(٥,١,١,٣) الاختبار t غير المزدوج Unpaired t Test

يمكن تقدير الإحصاء t غير المزدوج باستخدام:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

حيث n_1 هو عدد الملاحظات x_i و n_2 هو عدد الملاحظات y_i S_x^2 هو تباين العينة لـ x_i S_y^2 هو تباين العينة لـ y_i $\overline{x}$ هو متوسط العينة لـ x_i و $\overline{y}$ هو متوسط العينة لـ y_i.

حالما يتم حساب الإحصاء T يمكننا مقارنة القيمة T المُقدَّرة مع القيم t الواردة في جدول للتوزيع t. وبشكل أكثر تحديداً إذا أردنا رفض فرضية العدم بمستوى مقداره $1-\alpha$ من الثقة فإننا بحاجة عندئذ إلى تحديد ما إذا كانت القيمة T المُقدَّرة أكبر من مدخل القيمة t في الجدول t المرتبطة بمستوى كبير لـ α (اختبار t أحادي الجانب) أو $\alpha/2$ (اختبار t ثنائي الجانب). وبعبارة أخرى نحن بحاجة إلى معرفة القيمة t من التوزيع t الذي تكون من أجله المساحة تحت المنحنى t إلى اليمين من t مساوية إلى α. يجب أن تكون T المُقدَّرة أكبر من هذه الـ t لرفض فرضية العدم بمستوى $1-\alpha$ من الثقة.

وهكذا نحن نقارن القيمة T مع مدخل جدول التوزيع t من أجل:

$$t(\alpha, n_1+n_2-2) \qquad \text{(أحادي الجانب)}$$

أو

$$t(\alpha/2, n_1+n_2-2) \qquad \text{(ثنائي الجانب)}$$

حيث α هو مستوى الأهمية (يساوي 1− مستوى الثقة) و n_1 و n_2 هما عدد العينتين من المجتمعين اللذين تُجري مقارنتهما.

لاحظ أن α هو مستوى الأهمية الذي نريد عنده قبول أو رفض فرضيتنا. نرفض فرضية العدم بالنسبة لمعظم البحوث عندما $\alpha \le 0.05$. وهذا يتوافق مع مستوى ثقة مقداره 95%.

على سبيل المثال إذا أردنا رفض فرضية العدم بمستوى ثقة مقدارها 95% فإنه عندئذ بالنسبة لاختبار t أحادي الجانب يجب أن تكون T المُقدَّرة أكبر من $t(0.05, n_1+n_2-2)$ لرفض H_0 وقبول H_1 بثقة مقدارها 95% أو مستوى أهمية تساوي 0.05 أو أقل. علينا أن نتذكر أن ثقتنا = $(1-0.05)\times 100\,\%$. وكان هذا الاختبار من أجل $H_1: \mu_1 < \mu_2$ أو $H_1: \mu_1 > \mu_2$

(أحادي الجانب). إذا كانت H_1 هي $H_1: \mu_1 \neq \mu_2$ فإن T المُقاسة يجب أن تكون عندئذ أكبر من $t(0.05/2, n_1+n_2-2)$ لرفض فرضية العدم بنفس الثقة 95%.

وهكذا لاختبار الفرضية البديلة التالية :

١- بالنسبة لـ $H_1: \mu_1 \neq \mu_2$: استخدم اختبار t ثنائي الذيل. بالنسبة لـ $\alpha < 0.05$ فإن $T > t(0.025, n_1+n_2-2)$ المُقدرة لرفض H_0 بثقة مقدارها 95%.

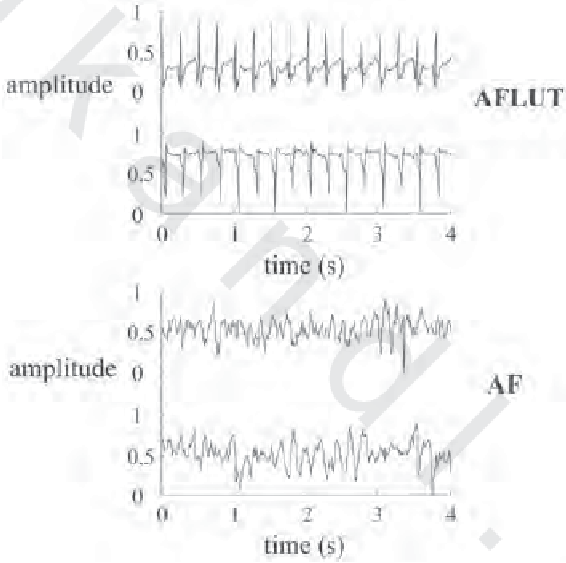
٢- بالنسبة لـ $H_1: \mu_1 < \mu_2$ أو $H_1: \mu_1 > \mu_2$: استخدم اختبار t أحادي الذيل. بالنسبة لـ $\alpha < 0.05$ فإن $T > t(0.05, n_1+n_2-2)$ المُقدرة لرفض H_0 بثقة مقدارها 95%.

يُشار أيضاً إلى α بأنها نوع الخطأ الأول (I). بالنسبة للاختبار t فإن α هي احتمال ملاحظة قيمة t مُقاسة أكبر من إدخال الجدول $t(\alpha; df)$ إذا كان المتوسطان الحقيقيان لمجتمعين x و y متساويين فعلاً. وبعبارة أخرى لا يوجد فرق كبير $(\alpha < 0.05)$ في متوسطيّ المجتمعين ولكن تحليل البيانات التي تم أخذ عيناتها أدت بنا إلى الاستنتاج بأن هناك فرقاً وبالتالي رفض فرضية العدم. يُشار إلى هذا بأنه نوع الخطأ الأول.

مثال (٥,١)

إن أحد الأمثلة على تحدٍّ في الهندسة الطبية الحيوية حيث يمكننا استخدام الاختبار t لتحسين عمل الأجهزة الطبية هو تطوير أجهزة إزالة الرجفان القابلة للزرع لاكتشاف وعلاج نُظم القلب غير الطبيعي [13]. تعد أجهزة إزالة الرجفان القابلة للزرع أجهزة إلكترونية صغيرة يتم وضعها في الصدر ولها أسلاك رفيعة يتم وضعها في حجرات القلب. تحتوي هذه الأسلاك على إلكترودات تكشف التيار الكهربائي الصغير العابر لعضلة القلب. تتبع هذه التيارات في ظل الظروف العادية نمطاً منظماً جداً للتوصيل ضمن عضلة القلب. يبين الشكل (٥,١) مثالاً على مخطط كهربائي أو تسجيل لنشاط كهربائي يتم تحسسه بواسطة أحد الإلكترودات في ظل ظروف طبيعية. عندما لا يعمل القلب بشكل طبيعي ويدخل في حالة رجفان (الشكل ٥,١) حيث لا يتقلص

القلب بشكل طبيعي أو يضخ الدم إلى بقية أجزاء الجسم ينبغي للجهاز صدم القلب بتيار كهربائي كبير من الجهاز في محاولة لإعادة نَظم القلب إلى طبيعته. ومن المهم لعدد من الأسباب أن يكشف الجهاز بدقة بداية اضطرابات النّظم المهددة للحياة مثل الرجفان وإعطاء الصدمة المناسبة. لإعطاء الصدمات يجب على الجهاز استخدام بعض أنواع خوارزميات معالجة الإشارات لكي يحدد بشكل آلي أن المخطط الكهربائي غير طبيعي وأنه صفة للرجفان.

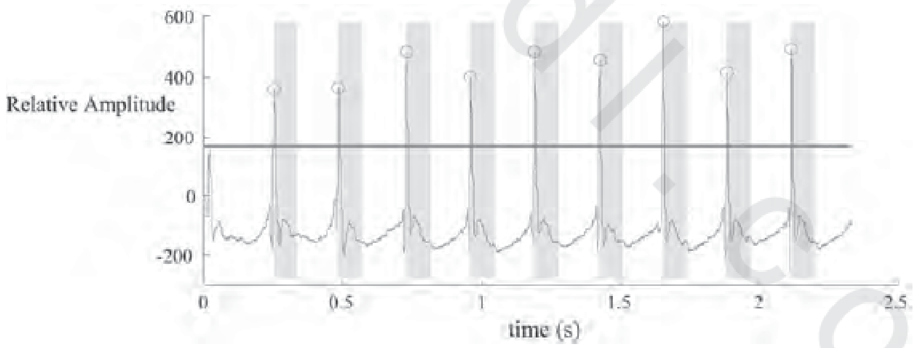


الشكل (١ ,٥). تسجيلات مخطط كهربائي مُقاسة بواسطة إلكترودات موضوعة داخل الأذين الأيسر للقلب. وبالنسبة لكل عملية نَظم هناك تسجيلان للمخطط الكهربائي مأخوذان من موقعين مختلفين في الأذين: الرفرفة الأذينية (AFLUT) والرجفان الأذيني (AF).

إن إحدى الخوارزميات المُستخدمة في معظم الأجهزة لتمييز نَظم القلب الطبيعي عن الرجفان هي خوارزمية المعدل. هذه الخوارزمية هي في الأساس خوارزمية عبور عتبة مطال يحدد الجهاز بواسطتها عدد المرات التي يتجاوز فيها المخطط الكهربائي عتبة

المطال في فترة محددة من الزمن ومن ثم يقدر المعدل من عدد المرات المُكتشفة لعبور العتبة. ويوضح الشكل (٥,٢) كيف تعمل هذه الخوارزمية.

قبل وضع مثل هذه الخوارزمية في أجهزة إزالة الرجفان القابلة للزرع وجب على الباحثين والمطوريّن إثبات ما إذا كان المعدل المُقدَّر من قبل مثل هذا الجهاز مختلفًا فعلاً بين عمليات النَظم الطبيعية والرجفانية (fibrillatory). وكان من المهم أن يكون لديها تداخل بسيط بين عمليات النَظم الطبيعية والرجفانية بحيث يمكن تحديد عتبة المعدل حيث إن المعدلات التي تتجاوز العتبة قد تؤدي بالجهاز إلى إعطاء صدمة. ولكي تعمل هذه الخوارزمية يجب أن يكون هناك اختلاف كبير في المعدلات بين عمليات النَظم الطبيعية والرجفانية وتداخل بسيط من شأنه أن يؤدي إلى صدمات يتم إعطاؤها بشكل غير صحيح والتسبب في آلام شديدة أو خطر تحريض الرجفان. يمكن للتداخل في المعدلات بين عمليات النظم الطبيعية والرجفانية أن ينتج أيضاً عن الجهاز الذي لا يكشف الرجفان بسبب المعدلات المنخفضة.



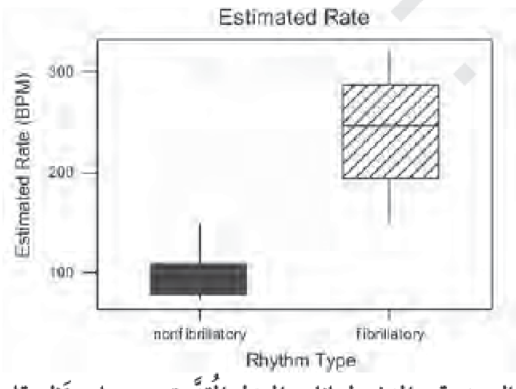
الشكل (٢ , ٥). معدَّل ضربات قلب مُقدَّر من مخطط كهربائي باستخدام خوارزمية عتبة المطال. كلما تجاوز مطال إشارة المخطط الكهربائي (شكل الموجة المستمر) عتبة المطال (الخط الأفقي الرمادي المتصل) خلال فترة زمنية محددة (الأشرطة العمودية المُظللة) يتم الكشف عن "حدث". يتم حساب المعدل عن طريق حساب عدد الأحداث في فترة زمنية معينة.

لتحديد ما إذا كان المعدل خوارزمية جيدة للكشف عن النَظم الرجفاني قد يسجل الباحثون في الواقع إشارات المخطط الكهربائي من مرضى حقيقيين أظهروا نَظمًا طبيعيًّا أو رجفانيًّا وقد يستخدمون خوارزمية المعدل لتقدير المعدل ومن ثم يقارنون معدلات المتوسط للنَظم الطبيعي مقابل معدلات المتوسط للنَظم الرجفاني. على سبيل المثال لنفترض أن الباحثين جمعوا تسجيلات مخطط كهربائي مدتها 15 ثانية للحصول على أمثلة للرجفان ($n_1 = 10$) والنَظم عديم الرجفان ($n_2 = 11$) في 21 مريضًا. نلاحظ أنه تم جمع بيانات الرجفان من أشخاص مختلفين عن البيانات العادية. وهكذا لم يكن لدينا حجب في هذا التصميم التجريبي.

تم تقدير المعدل وذلك باستخدام خوارزمية الجهاز لكل تسجيل مدته 15 ثانية للفرد. يبين الشكل (٥,٣) مخطط الصندوق والمؤشر لكل من مجموعتي البيانات.

ويُعطى الإحصاء الوصفي لمجموعتي البيانات حسب الجدول التالي :

	المخططات الكهربائية عديمة الرجفان (N_1=10)	المخططات الكهربائية الرجفانية (N_1=11)
المتوسط	96.82	239.0
الانحراف المعياري	22.25	55.3



الشكل (٣ , ٥). مخطط الصندوق والمؤشر لبيانات المعدل المُقدَّرة من عينات نَظم قلب عديم الرجفان ونَظم قلب رجفاني.

هناك عدة أشياء يجب ملاحظتها من المخططات والإحصاء الوصفي. أولاً إن حجم العينة صغير وبالتالي فمن المرجح أن التغيُّر لتسجيلات المخطط الكهربائي من النَظم الطبيعي والرجفاني لا يتم تمثيله بشكل مناسب. وعلاوة على ذلك فإن البيانات غير متماثلة ولا تظهر بأنها مُنمذجة بشكل جيد بواسطة التوزيع الطبيعي. وأخيراً يبدو أن التغير في المعدلات للنَظم الرجفاني أكبر بكثير منه للنَظم الطبيعي. وهكذا فإن بعض الافتراضات التي نقوم بها بخصوص تسوية المجتمعات الإحصائية والمساواة بين التباينات ربما تكون ناقصة في تطبيق الاختبار t. ولكن لغرض التوضيح فإننا سوف نجري الاختبار t باستخدام الإحصاء t غير المزدوج.

ولإيجاد الإحصاء T المُقدَّر من بيانات العينة فإننا نستبدل المتوسطات والتباينات وحجم العينة في معادلة الإحصاء T غير المزدوج. وبالنسبة لهذا المثال نجد أن قيمة t المُقدَّرة هي 7.59. وإذا نظرنا إلى جداول t من أجل $[11+10-2]$ درجات حرية نستطيع أن نرفض H_0 عندما $\alpha<0.005$ لأن قيمة T المُقدَّرة تتجاوز إدخالات جدول t عندما $\alpha=0.05\,(t=1.729)$ و $\alpha=0.01\,(t=2.539)$ و $\alpha=0.005\,(t=2.861)$.

وهكذا فإننا نرفض فرضية العدم بثقة مقدارها $(1-0.005)\times100\%$ وندرك أن متوسط معدل ضربات القلب للنَظم الرجفاني أكبر من متوسط معدل ضربات القلب للنَظم الطبيعي. وهكذا ينبغي لخوارزمية المعدل العمل جيداً في التمييز بين النَظم الطبيعي والنَظم الرجفاني. ومع ذلك فقد اختبرنا متوسطات المجتمعات الإحصائية فقط. كما يلاحظ المرء في الشكل (٥،٣) هناك بعض التداخل في العينات الفردية بين النَظم الطبيعي والرجفاني. وهكذا يمكننا أن نتوقع بأن الجهاز سيقوم بأخطاء في إعطاء صدمة بشكل غير صحيح عندما يكون القلب في حالة نَظم طبيعي ولكن متسارع (كما في التمارين) أو قد يفشل الجهاز في إعطاء صدمة عندما يكون القلب مرتجفاً ولكن بمعدل بطيء أو بمطال منخفض.

عند تطبيق الاختبار t فمن المهم أن نلاحظ أنه لا يمكن أبداً إثبات أن متوسطَين متساويان. تستطيع الإحصائيات فقط إظهار أن اختباراً محدداً لا يمكنه إيجاد فرق في متوسطات المجتمعات وعدم إيجاد فرق لا يعني إثبات المساواة. إن فرضية العدم أو عدم الوجود هي أنه لا يوجد فرق في المتوسطات بغض النظر عن ما هو الفرق الحقيقي بين المتوسطَين. إن عدم إيجاد فرق بالبيانات التي تم جمعها والاختبار الإحصائي المناسب لا يعني إثبات أن المتوسطات متساوية. وهكذا فإننا ببساطة نقبل فرضية العدم ولا نقبل بمستوى الثقة أو الأهمية. إننا نحدد مستوى الثقة فقط عندما نرفض فرضية العدم.

(٤,١,١,٥) الاختبار t المزدوج Paired *t* Test

استخدمنا في المثال السابق الاختبار t غير المزدوج لأن مجموعتي البيانات جاءتا من مجتمعين مختلفين غير مترابطين للمرضى. تكمن المشكلة في مثل هذا التصميم التجريبي في أن الاختلافات في مجتمعي المرضى قد تؤدي إلى اختلافات في متوسط معدل ضربات القلب الذي لا علاقة له بنَظم القلب الفعلي وإنما بالاختلافات في حجم أو عمر القلوب بين مجموعتي المرضى أو بعض الاختلافات الأخرى بين مجموعات المرضى. هناك طريقة أفضل لإجراء الدراسة السابقة وهي جمع بيانات القلب الطبيعية والرجفانية من مجتمع واحد من المرضى. وسوف يحتاج فقط أحد أجهزة الرجفان إلى التمييز بين النَظم الطبيعي والنَظم الرجفاني ضمن مريض واحد. يمكننا من خلال حجب أشخاص التخلص من التغير بين الأشخاص في خصائص المخطط الكهربائي التي يمكن أن تعيق خوارزمية المعدل من المجتمعات الإحصائية المنفصلة. قد يكون من المعقول افتراض أن معدلات النَظم الطبيعي والنَظم الرجفاني تختلف داخل المريض أكثر منها عبر المرضى. وبعبارة أخرى نجمع بالنسبة لكل مريض مخططاً كهربائياً خلال عمل القلب الطبيعي وخلال الرجفان. في مثل هذا التصميم التجريبي قد نقارن المتوسطيْن لمجموعتي البيانات باستخدام اختبار t المزدوج.

نستخدم الاختبار t المزدوج عندما يكون هناك ازدواج طبيعي بين كل ملاحظة في مجتمع البيانات X مع كل ملاحظة في مجتمع البيانات Y. على سبيل المثال قد يكون لدينا الفروض التالية التي تستدعي الاختبار t المزدوج:

١- جمع قياسات ضغط الدم من 8 مرضى قبل وبعد ممارسة التمرين.

٢- جمع كل من بيانات الـ MR (الرنين المغناطيسي) والـ CT (التصوير المقطعي المحوسب) من 20 مريض لمقارنة جودة صور الأوعية الدموية.

٣- جمع زمن الحوسبة لخوارزمية معالجة صورة قبل وبعد حدوث تغيير في البرنامج.

في مثل هذه الحالات لم تعد مجموعات البيانات X و Y مستقلة. في المثال الأول أعلاه إن الفسيولوجيا والبيئة البيولوجية التي تؤثر على ضغط الدم قبل التمرين في كل مريض تؤثر أيضاً على ضغط الدم بعد التمرين وذلك لأننا نجمع البيانات قبل وبعد التمرين من نفس الوحدات التجريبية. وهكذا فإن هناك عددًا من المتغيرات التي تؤثر على ضغط الدم لا نستطيع التحكم بها مباشرة ولكن يمكن التحكم بآثارها على ضغط الدم (إلى جانب تأثير التمرين) باستخدام نفس الوحدات التجريبية (الكائنات البشرية في هذه الحالة) لكل مجتمع بيانات. في مثل هذه الحالات تكون الوحدات التجريبية بمثابة تحكم ذاتي بنفسها. وهذا هو عادة التصميم التجريبي المُفضل لمقارنة المتوسطات.

وبالنسبة للاختبار t المزدوج لدينا مرة أخرى فرضية العدم والفروض البديلة كما ذُكر أعلاه للاختبار t غير المزدوج. ومع ذلك فإننا نستخدم في الاختبار t المزدوج الاختبار t على الفرق $W_i = X_i - Y_i$ بين نقاط البيانات المزدوجة من كل مجتمع من المجتمعين الإحصائيين.

نحسب الآن الإحصائية T المزدوجة: (n = عدد الأزواج)

$$T = \frac{\overline{W}}{S_w/\sqrt{n}}$$

حيث $\overline{W}$ هو متوسط الفرق للاختلافات W_i و S_w هو الانحراف المعياري للاختلافات W_i.

وكما هو بالنسبة للاختبار t غير المزدوج لدينا الآن قيمة T المُقدرة التي نستطيع مقارنتها مع قيم t في جدول توزيع t لتحديد ما إذا كانت القيمة T المُقدَّرة تقع ضمن القيم القصوى (أكبر من 95%) للتوزيع t.

ولرفض فرضية العدم عند مستوى أهمية مقدارها α (مستوى ثقة مقدارها $1-\alpha$) يجب أن تكون قيمة T المُقدَّرة أكبر من $t(\alpha, n-1)$ حيث n هو عدد أزواج البيانات.

إذا كانت قيمة T المُقدَّرة تتجاوز $t(\alpha, n-1)$ فإننا نرفض H_0 ونقبل H_1 عند مستوى أهمية يساوي α أو مستوى ثقة مقدارها $1-\alpha$.

مرة أخرى إذا كانت $H_1 : \mu_1 < \mu_2$ أو $H_1 : \mu_1 > \mu_2$ فإننا نجري اختباراً أحادي الجانب حيث يجب أن تكون الإحصائية T أكبر من $t(\alpha, n-1)$ لرفض فرضية العدم عند مستوى مقداره α. إذا كانت فرضية العدم التي نبدأ بها قبل التجربة $H_i : (\mu_1 \neq \mu_2)$ فإننا نجري عندئذ اختباراً ثنائي الجانب ويجب أن تكون الإحصائية T أكبر من $t(\alpha/2, n-1)$ لرفض فرضية العدم عند مستوى أهمية يساوي α.

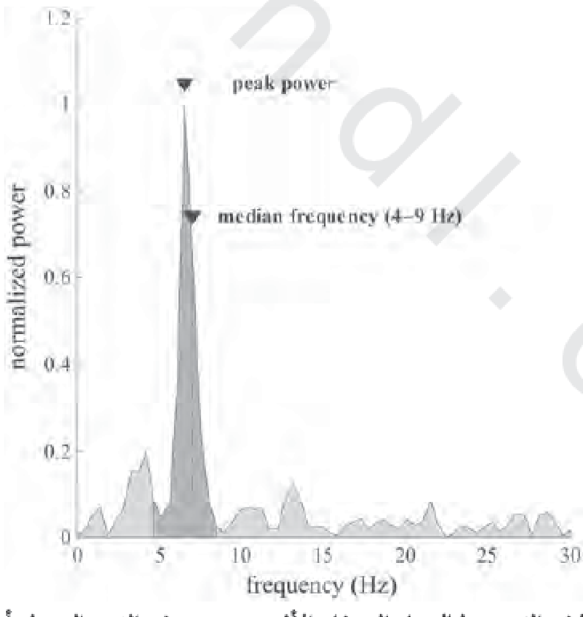
(٥,١,١,٥) مثال لتحدٍّ في الهندسة الطبية الحيوية

Example of a Biomedical Engineering Challenge

فيما يتعلق بالنَظم غير الطبيعي للقلب الذي تم مناقشته سابقاً يمكن استخدام أدوية مضادة لاضطراب النَّظْم لإبطاء النَّظم غير الطبيعي أو إنهائه مثل الرجفان. على سبيل المثال قد يتم استخدام دواء مثل بروكايينامايد (procainamide) لإنهاء الرجفان الأذيني. إن الآلية الدقيقة التي يؤدي بواسطتها الدواء إلى إنهاء النظم غير معروفة بالضبط ولكن يُعتقد أن الدواء يقوم بتغيير فترة الاستعصاء وسرعة التوصيل لخلايا القلب [8]. غالباً ما يستخدم المهندسون الطبيون الحيويون معالجة الإشارة على الإشارات الكهربائية المتولدة بواسطة القلب كوسائل غير مباشرة لدراسة الفسيولوجيا (علم وظائف الأعضاء) الأساسية. ويشكل أكثر تحديداً فقد يستخدم المهندسون التحليل الطيفي أو تحليل فورييه لدراسة التغيرات في المحتوى الطيفي للإشارة الكهربائية

مع مرور الزمن مثل المخطط الكهربائي. قد تخبرنا مثل هذه التغيرات شيئاً عن الفسيولوجيا الكهربائية الأساسية.

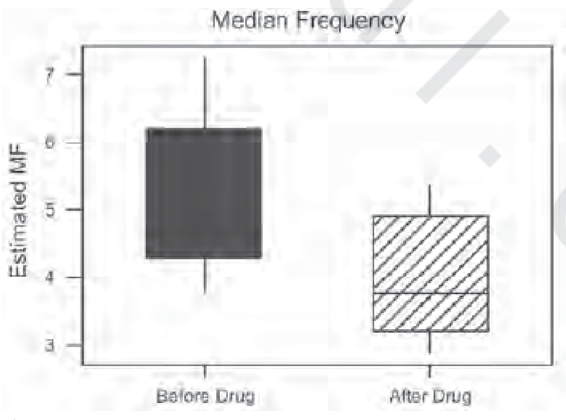
على سبيل المثال فقد تم استخدام التحليل الطيفي لدراسة التغيرات في الطيف الترددي للرجفان الأذيني مع إعطاء الدواء. في إحدى الدراسات [8] كان المهندسون الطبيون الحيويون مهتمين بدراسة التغيرات في التردد الوسط للمخططات الكهربائية الأذينية بعد إعطاء الدواء. يبين الشكل (٥،٤) مثالاً للطيف الترددي للرجفان الأذيني وموقع التردد الوسط الذي هو التردد الذي يقسم قدرة الطيف (المنطقة تحت المنحنى الطيفي بين 4 و 9 هرتز) في النصف. أحد الأسئلة الذي طرحه الباحثون هو ما إذا كان التردد الوسط ينخفض بعد إعطاء دواء مثل البروكاييناميد الذي يُعتقد أنه يقوم بإبطاء النشاط الكهربائي لخلايا القلب.



الشكل (٤ ، ٥). الطيف الترددي لمثال على الرجفان الأذيني. يتم تعريف التردد الوسط بأنه ذلك التردد الذي يقسم المنطقة تحت منحنى القدرة في الحزمة من 4 - 9 هرتز في النصف.

وبالتالي تم إجراء تجربة لتحديد ما إذا كان هناك اختلاف كبير في متوسط التردد الوسط بين الرجفان قبل إعطاء الدواء والرجفان بعد إعطاء الدواء. تم جمع مخططات كهربائية في الأذين الأيمن في 11 مريضًا قبل وبعد إعطاء الدواء. وتم تقييم تسجيلات مدتها خمس عشرة ثانية لطيف التردد وتم تقدير التردد الوسط في حزمة التردد من 4 – 9 هرتز قبل وبعد إعطاء الدواء.

يوضح الشكل (٥,٥) إحصائيات موجزة عن التردد الوسيط قبل وبعد إعطاء الدواء. والسؤال هو ما إذا كان هناك انخفاض ملحوظ في التردد الوسيط بعد إعطاء البروكايينامايد. في هذه الحالة فإن فرضية العدم هي أنه لا يوجد أي تغيير في متوسط التردد الوسط بعد إعطاء الدواء. الفروض البديلة هي أن متوسط التردد الوسط ينخفض بعد إعطاء الدواء. وهكذا فإننا نقارن متوسطين وقد تم جمع مجموعتي البيانات من إحدى مجموعات المرضى. وسوف نحتاج إلى الاختبار t المزدوج لرفض أو قبول فرضية العدم.



الشكل (٥ , ٥). مخطط الصندوق والمؤشر لتردد وسط تم تقديره من عينات للرجفان الأذيني تم تسجيلها قبل وبعد إعطاء الدواء.

لإجراء الاختبار t المزدوج ننشئ عمودًا آخر W_i كما ورد في الجدول التالي. إننا نجـد ما يلي لـ W_i : $\overline{W}=1.262$ و $S_w=0.402$. وإذا استخدمنا هذه التقديرات لـ $\overline{W}$ و S_w في معادلة إحصائية T المزدوجة فإننا نجد أن $T=10.42$ عندما $n=11$ زوج من البيانات. وإذا قارنا قيمة T المُقدَّرة مع التوزيع t نجد أن قيمة T أكبر من إدخال الجدول عندما $t(0.005,11-1)$ ؛ وبالتالي فقد نرفض H_0 عند أهمية أقل من 0.005. وبعبارة أخرى فإننا نرفض فرضية العدم عند مستوى الثقة $[1-0.005]\times 100\%$.

قبل الدواء	بعد الدواء	WI
4.3	2.90	1.4
4.15	2.97	1.18
3.8	3.20	0.60
5.10	3.30	1.80
4.30	3.75	0.55
7.20	5.35	1.85
6.40	5.10	1.30
6.20	4.90	1.30
6.10	4.80	1.30
5.00	3.70	1.30
5.80	4.50	1.30

W_I هو الانخفاض في تردد الوسيط بعد إعطاء البروكايناميد

أخطاء في استخلاص الاستنتاجات من الاختبارات الإحصائية

عندما نجري تحليلاً إحصائياً مثل الاختبار t فقد نكون مخطئين في رفض أو قبول فرضية العدم. وعندما نستخلص استنتاجات من التحليل الإحصائي فإننا نحدد عادة مستوى ثقة لاستنتاجاتنا. وهذا يعني أن هناك دائماً احتمالا إلى حد ما بأن استنتاجاتنا غير صحيحة.

هناك نوعان من الأخطاء التي قد تحدث عند استخلاص استنتاجات من التحليل الإحصائي. ويُشار إلى هذه الأخطاء بالأنواع I وII.

أخطاء النوع I:

١- يُشار إليها أيضاً بأنها الخطأ الموجب الخاطئ.

٢- تحدث عندما نقبل H_1 عندما تكون H_0 صحيحة.

٣- قد تؤدي إلى تشخيص خاطئ للمرض.

أخطاء النوع II:

١- يُشار إليها أيضاً بأنها الخطأ السالب الخاطئ.

٢- تحدث عندما نقبل H_0 عندما تكون H_1 صحيحة.

٣- قد تؤدي إلى عدم التشخيص (غالباً ما تكون أكثر خطورة من خطأ النوع I).

إذا فكرنا في البيئة الطبية فقد يحدث خطأ النوع I عندما يُعطى شخص اختبارًا تشخيصيًّا للكشف عن البكتيريا العقدية والاختبار يشير إلى أن الشخص لديه البكتيريا العقدية عندما لا يكون لدى الشخص هذه البكتيريا في الواقع. إن نتيجة مثل هذا الخطأ تعني أن الشخص ينفق المال على المضادات الحيوية التي لا تخدم أي غرض.

قد يحدث خطأ النوع II عندما يكون بالفعل لدى الشخص نفسه بكتيريا عقدية ولكن الاختبار التشخيصي يعطي نتيجة سلبية ونخلص إلى أن الشخص ليس لديه بكتيريا عقدية. في هذا المثال يكون هذا النوع من الخطأ أكثر خطورة من خطأ النوع I لأن البكتيريا العقدية التي تم تركها دون علاج يمكن أن تؤدي إلى عدد من المضاعفات الخطيرة للجسم.

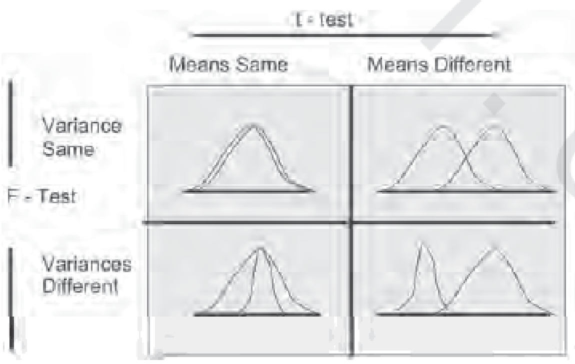
إن قيمة α التي نشير إليها بأنها مستوى الأهمية هي أيضاً الاحتمال لخطأ النوع I. وهكذا كلما كان مستوى الأهمية الذي قد نرفض عنده فرضية العدم أصغر كان خطأ النوع I أصغر وكان احتمال القيام بخطأ النوع I أقل.

نستخدم عادة β للدلالة على خطأ النوع II. وسوف نناقش هذا الخطأ مرة أخرى في نهاية الفصل السابع عندما نناقش اختبارات القدرة.

(٥,٢) مقارنة تباينين
COMPARISON OF TWO VARIANCE

استخدمنا الاختبار t لمقارنة المتوسطين لمجتمعين أو عمليتين. ويمكن أيضاً مقارنة مجتمعين فيما يتعلق بالاختلافات في التباين. وكما تمت المناقشة سابقاً يتم تماماً تمييز المجتمعات المُوزعة بشكل طبيعي بواسطة متوسطاتها وتبايناتها. وهكذا إذا أردنا الاختبار فيما يتعلق بالاختلافات بين مجتمعين طبيعيين فإننا بحاجة فقط إلى مقارنة المتوسطين والتباينين لهذين المجتمعين.

يوضح الشكل (٥,٦) دوال كثافة الاحتمال لمجتمعين طبيعيين (الرسومات السوداء والحمراء). توضح المخططات الأربعة كيف يمكن مقارنة مجتمعين مختلفين موزعين بشكل طبيعي مع بعضهما البعض. تختلف لوحتا اليمين عن لوحتي اليسار في متوسطات المجتمعين الإحصائيين. وتختلف اللوحتان العلويتان عن اللوحتين السفليتين في التباين للمجتمعات الإحصائية.



الشكل (٥,٦). قد يختلف مجتمعان إحصائيان طبيعيان في متوسطاتهما (الصف العلوي) أو تبايناتهما (النصف الأيسر) أو كليهما (الزاوية اليمنى السفلى). يمكن استخدام الاختبارين t وF لاختبار الاختلافات الكبيرة في متوسطات المجتمعات الإحصائية وتباينات المجتمعات على التوالي.

وكما هـو مبيـن في الجـزء العلوي من الرسـومات يتم استخدام الاختبار t لاختبار الاختلافات في المتوسط بين المجتمعين. وكما هو مبين على طول الاتجاه العمودي يتم استخدام الاختبار F لاختبار الاختلافات الكبيرة في تباينات المجتمعات الإحصائية. لاحظ أن المجتمعين الطبيعيين قد يختلفان اختلافاً كبيراً في كل من المتوسط والتباين.

لمقارنة التباينات لمجتمعين نستخدم ما يُشار إليه بأنه الاختبار F. وكما هو بالنسبة للاختبار t يفترض الاختبار F أن البيانات تتألف من عينات عشوائية مستقلة من كل مجتمع من المجتمعين الطبيعيين. إذا كان المجتمعان غير موزعين بشكل طبيعي فقد تكون نتائج الاختبار F لا معنى لها.

كما هو الحال مع الاختبار t يتم استخدام الاختبار F لاختبار الفروض التالية :

١- فرضية العدم: $H_0 : \sigma_1^2 = \sigma_2^2$ و

٢- الفروض البديلة : $H_1 : \sigma_1^2 \geq \sigma_2^2$

حيث σ_1^2 و σ_2^2 هي التباينات للمجتمعين الإحصائيين.

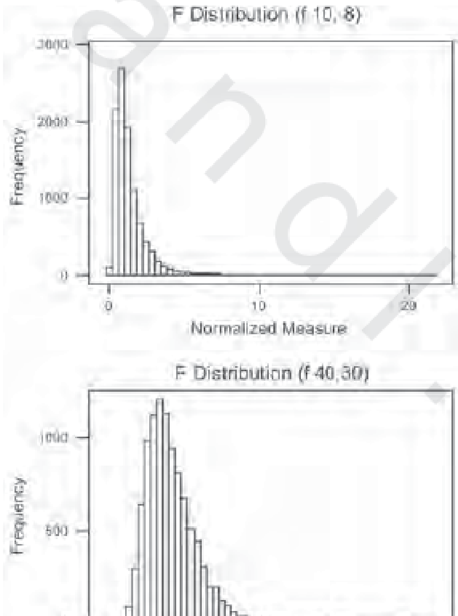
لرفض أو قبول فرضية العدم نحسب الإحصائية F التالية :

$$F = \frac{S_1^2}{S_2^2}$$

حيث s_1^2 و s_2^2 هي تقديرات تباينات العينات لمجتمعين. إن نسبة التباين لمجتمعين طبيعيين هي أيضاً متغير عشوائي يتبع التوزيع F. يوضح الشكل (٥،٧) التوزيع F. وكما هو الحال مع التوزيع t فإن التوزيع F يختلف ببارامترين مثل أحجام العينات لمجتمعين. يبين الجدول (٥،١) جزءًا من الجدول F حيث هناك حاجة إلى درجتي حرية dn و dd لتحديد إحدى قيم F في الجدول F. يتم في هذا الجدول إعطاء إدخالات

الجدول لمستويات أهمية (قيم α) تساوي 0.05 و 0.01. ويتم إعطاء قيم F المرتبطة بمستويات الأهمية 0.05 و 0.01 بخط نظامي وبخط سميك مائل على التوالي. وهكذا بالنسبة لأي درجتي حرية هناك قيمتان لـ F واحدة لمستوى الثقة 95% وواحدة لمستوى الثقة 99%.

لاستخدام الجدول F مع الاختبار F نقوم بتقدير إحصائية F باستخدام تقديرات تباين العينة من كل مجتمع من المجتمعين اللذين نحاول مقارنتهما. لاحظ أنه من أجل استخدام هذا الجدول F ينبغي وضع التباين الأكبر من التباينين في البسط من المعادلة المذكورة سابقاً.



الشكل (٥،٧). المدرج الإحصائي لعينات مأخوذة من توزيعين F مختلفين. في اللوحة العليا تكون درجتا الحرية 10 و 8. في اللوحة السفلى فإن درجتي الحرية هما 40 و 30.

الجدول (١ , ٥). قيم من التوزيع F لمساحات α في الذيل إلى اليمين من $F(dn,dd,\alpha)$.

dn								
11	**10**	**...**	**5**	**4**	**3**	**2**	**1**	***dd***
243, **6082**	242, **6056**		230 **5764**	225, **5625**	216, **5403**	200, **4999**	161, **4052**	**1**
19.40, **99.41**	19.39, **99.40**		19.30, **99.30**	19.25, **99.25**	19.16, **99.17**	19.00, **99.01**	18.5, **98.49**	**2**
8.76, **27.13**	8.78, **27.23**		9.01, **28.24**	9.12, **28.71**	9.28, **29.46**	9.55, **30.81**	10.13, **34.12**	**3**
5.93, **14.45**	5.96, **14.54**		6.26, **15.52**	6.39, **15.98**	6.59, **16.69**	6.94, **18.00**	7.71, **21.20**	**4**
4.70, **9.96**	4.74, **10.05**		5.05, **10.97**	5.19, **11.39**	5.41, **12.06**	5.79, **13.27**	6.61, **16.26**	**5**
								...
2.94, **4.78**	2.97, **4.85**		3.33, **5.64**	3.48, **5.99**	3.71, **6.55**	4.10, **7.56**	4.96, **10.04**	**10**
2.72, **4.22**	2.76, **4.30**		3.11, **5.06**	3.26, **5.41**	3.49, **5.95**	3.88, **6.93**	4.75, **9.33**	**12**
2.51, **3.73**	2.55, **3.80**		2.90, **4.56**	3.06, **4.89**	3.29, **5.42**	3.68, **6.36**	4.54, **8.68**	**15**
								...
2.31, **3.30**	2.35, **3.37**		2.71, **4.10**	2.87, **4.43**	3.10, **4.94**	3.49, **5.85**	4.35, **8.10**	**20**
1.98, **2.62**	2.02, **2.70**		2.40, **3.41**	2.56, **3.72**	2.79, **4.20**	3.18, **5.06**	4.03, **7.17**	**50**
1.88, **2.34**	1.92, **2.51**		2.30, **3.20**	2.46, **3.51**	2.70, **3.98**	3.09, **4.82**	3.94, **6.90**	**100**
1.83, **2.34**	1.87, **2.41**		2.26, **3.11**	2.41, **3.41**	2.65, **3.38**	3.04, **4.71**	3.89, **6.76**	**200**
1.79, **2.24**	1.83, **2.32**		2.21, **3.02**	2.37, **3.32**	2.60, **3.78**	2.99, **4.60**	3.84, **6.64**	∞

dn = درجات الحرية للبسط ؛ *dd* = درجات الحرية للمقام ؛ α = المساحة في ذيل التوزيع إلى يمين $F(dn,dd,\alpha)=0.05$ أو *0.01*.

تابع الجدول (١ , ٥).

dn									
∞	**200**	**100**	**50**	**40**	**30**	**20**	**...**	**14**	**12**
254	254	253	252	251	250	248		245	244
6366	**6352**	**6334**	**6302**	**6286**	**6258**	**6208**		**6142**	**6106**
19.50	19.49	19.49	19.47	19.47	19.46	19.44		19.42	19.41
99.50	**99.49**	**99.49**	**99.48**	**99.48**	**99.47**	**99.45**		**99.43**	**99.42**
8.53	8.54	8.56	8.58	8.60	8.62	8.66		8.71	8.74
26.18	**26.18**	**26.23**	**26.30**	**26.41**	**26.50**	**26.69**		**26.92**	**27.05**
5.63	5.65	5.66	5.70	5.71	5.74	5.80		5.87	5.91
13.46	**13.52**	**13.57**	**13.69**	**13.74**	**13.83**	**14.02**		**14.24**	**14.37**
4.36	4.38	4.40	4.44	4.46	4.50	4.56		4.64	4.68
9.02	**9.07**	**9.13**	**9.24**	**9.29**	**9.38**	**9.55**		**9.77**	**9.89**
2.54	2.56	2.59	2.64	2.67	2.70	2.77		2.86	2.91
3.91	**3.96**	**4.01**	**4.12**	**4.17**	**4.25**	**4.41**		**4.60**	**4.71**
2.30	2.32	2.35	2.40	2.42	2.46	2.54		2.64	2.69
3.36	**3.41**	**3.46**	**3.56**	**3.61**	**3.70**	**3.86**		**4.05**	**4.16**
2.07	2.10	2.12	2.18	2.21	2.25	2.33		2.43	2.48
2.87	**2.92**	**2.97**	**3.07**	**3.12**	**3.20**	**3.36**		**3.56**	**3.67**
1.84	1.87	1.90	1.96	1.99	2.04	2.12		2.23	2.28
2.42	**2.47**	**2.53**	**2.63**	**2.69**	**2.77**	**2.94**		**3.13**	**3.23**
1.44	1.48	1.52	1.60	1.63	1.69	1.78		1.90	1.95
1.68	**1.76**	**1.82**	**1.94**	**2.00**	**2.10**	**2.26**		**2.46**	**2.56**
1.28	1.34	1.39	1.48	1.51	1.57	1.68		1.79	1.85
1.43	**1.51**	**1.59**	**1.73**	**1.79**	**1.89**	**2.06**		**2.26**	**2.36**
1.19	1.26	1.32	1.42	1.45	1.52	1.62		1.74	1.80
1.28	**1.39**	**1.48**	**1.62**	**1.69**	**1.79**	**1.97**		**1.17**	**2.28**
1.00	1.17	1.24	1.35	1.40	1.46	1.57		1.69	1.75
1.00	**1.25**	**1.36**	**1.52**	**1.59**	**1.69**	**1.87**		**2.07**	**2.18**

نقوم الآن بمقارنة الإحصائية F المُقدَّرة مع الإدخالات في الجدول F المرتبطة بدرجات الحرية dn و dd ومستوى الثقة المناسب (يتم تقديم قيم F الـ 95% و 99% فقط في الجدول المُقدَّم). $dn = n_1$ و $dd = n_2$ هي عدد العينات في كل مجتمع مع n_1 هو عدد العينات في المجتمع ذي التباين الموضوع في البسط.

إذا أردنا رفض فرضية العدم المبينة سابقاً يجب أن تكون الإحصائية F المحسوبة أكبر من $F(\alpha, dn, dd)$ في الجدول لرفض H_0 بثقة مقدارها $(1-\alpha)\times 100\%$. درجات الحرية $dn = n_1$ هي القيمة المُستخدمة لتحديد موقع إدخال الجدول في الاتجاه الأفقي (البسط) و $dd = n_2$ هي درجات الحرية المُستخدمة لتحديد موقع إدخال الجدول في الاتجاه العمودي (المقام).

مثال (٢ , ٥)

الاختبار F

	المجتمع الإحصائي A $(N_1 = 9)$	المجتمع الإحصائي B $(N_2 = 9)$
المتوسط	0.026	0.027
التباين	2.0×10^{-5}	7.4×10^{-5}

وبالأخذ في الاعتبار البيانات الواردة في الجدول أعلاه حيث جمعنا 9 عينات من كل مجتمع من المجتمعين A و B يمكننا تقدير الإحصائية F بحيث تكون:

$$F = (7.4E-5)/(2.0E-5) = 3.7$$

وباستخدام الجدول F الذي لديه إدخالات جدول لكل من مستويات الثقة 95% و 99% نجد أن إدخال الجدول بالنسبة إلى α يساوي 0.05 و df 9 لكل من المجتمعين هي:

$$F(0.05, 9, 9) = 3.18$$

تتجاوز القيمة 3.7 المُقدَّرة إدخال الجدول 3.18 ؛ وبالتالي فقد نرفض H_0 بثقة 95% ونقبل أن المجتمع B له تباين أكبر بكثير من المجتمع A. ومع ذلك نلاحظ أن إدخال الجدول من أجل $F(0.01, 9, 9)$ يساوي 5.35 وهو أكبر من القيمة المُقدَّرة لدينا. وهكذا فإننا لا نستطيع رفض فرضية العدم بثقة مقدارها 99%.

(٥,٣) مقارنة ثلاثة متوسطات أو أكثر للمجتمعات الإحصائية

COMPARISON OF THREE OR MORE POPULATION MEANS

ناقشنا حتى الآن التحليل الإحصائي لمقارنة المتوسطات والتباينات (أو الانحرافات المعيارية) لمجتمعين على أساس جمع مجتمعين من البيانات. قد يكون هذان المجتمعان من البيانات مستقلين عن بعضهما بعضاً أو قد يكونان تابعين لبعضهما البعض والتي نشير إليها بالبيانات المزدوجة أو المحجوبة. هناك طريقة أخرى للتفكير في هذا الازدواج وهي تسميتها حجب الوحدة التجريبية وهذا يعني أن كلا المجتمعين من البيانات تم جمعهما من نفس الوحدات التجريبية ولكن تحت ظروف مختلفة. على سبيل المثال قد تكون الوحدات التجريبية كائنات بشرية أو حيوانات أو أجهزة طبية أو خطوط تصنيع أو مستعمرات خلية أو دارات إلكترونية وأكثر من ذلك. إذا كانت الوحدة التجريبية على سبيل المثال كائنات بشرية فإنه يمكن أخذ مجموعتي البيانات قبل وبعد إعطاء الدواء أو قبل وبعد اشتراك الأشخاص في تمرين. وفي مثال آخر قد نقيس نوعين مختلفين من البيانات مثل معدل ضربات القلب وضغط الدم كل منهما من نفس مجتمع الأشخاص. ويُشار إلى هذا بالمقاييس المتكررة وذلك لأننا نأخذ عدة مجموعات من المقاييس من نفس الوحدات التجريبية. في المقاييس المتكررة فإن شيئاً ما في الظروف التجريبية أو الوحدة التجريبية أو القياس الذي يجري جمعه يتباين من إحدى مجموعات المقاييس إلى التي تليها ولكن يتم حجب أخذ العينات من قبل الوحدة التجريبية.

نحتاج في العديد من تطبيقات الهندسة الطبية الحيوية إلى مقارنة المتوسطات من ثلاثة أو أكثر من المجتمعات الإحصائية أو العمليات أو الشروط. في مثل هذه الحالات نستخدم طريقة تحليل إحصائي تُدعى تحليل التباين أو أنوفا (ANOVA). على الرغم من أن الاسم يعني أن المرء يقوم بتحليل التباينات فإن الاستنتاجات التي تنبع من هذا التحليل هي بخصوص الاختلاف الكبير في المتوسطات لثلاثة أو أكثر من المجتمعات الإحصائية أو العمليات.

(٥,٣,١) التجارب أحادية العامل One-Factor Experiments

نبدأ مناقشاتنا لأنوفا (ANOVA) بمناقشة تصميم وتحليل التجارب أحادية العامل. نأخذ العينات في مثل هذه التجارب من ثلاثة مجتمعات إحصائية أو أكثر يتغير فيها عامل واحد من مجتمع إلى آخر.

وكما ورد في الفصل الثاني ينبغي أن يشمل التصميم التجريبي العشوائية والحجب. تضمن العشوائية عدم إدخال انحياز إلى البيانات بسبب ترتيب الآثار في جمع البيانات. وعلاوة على ذلك يساعدنا الحجب على التقليل من تأثير التغير بين الأشخاص على الاختلافات بين المجتمعات الإحصائية

إن بعض تحديات في الهندسة الطبية الحيوية التي قد تتطلب استخدام أنوفا تشمل ما يلي:

١- مقارنة زمن الاهتِراء بين زرعات الورك لمواد مختلفة: في هذا المثال المعالجة = مادة الزرع (على سبيل المثال التيتانيوم والفولاذ وراتنجات البوليمر).

٢- مقارنة متواليات نبضة الـ MR في القدرة على تصوير تلف الأنسجة بعد السكتة الدماغية: في هذا المثال المعالجة = متوالية نبضة الـ MR.

٣- مقارنة قدرة العديد من الأدوية على خفض ضغط الدم المرتفع: في هذا المثال المعالجة = الدواء.

(٥,٣,١,١) مثال لتحدٍّ في الهندسة الطبية الحيوية

Example of Biomedical Engineering Challenge

أحد الأمثلة على مشكلة الهندسة الطبية الحيوية التي قد تستخدم أنوفا لاختبار الفروض هو دراسة آليات الانعكاس (رد الفعل) في المرضى المصابين بالحبل الشوكي [14]. إن أحد اهتمامات الباحثين هو تبعية انثناء الورك (عزم) لحركة الكاحل في المرضى الذين يعانون من إصابات الحبل الشوكي. في هذا المثال هناك بالفعل مجتمعان من العوامل التجريبية التي قد تؤثر على ثني الورك:

١- مجال حركة الكاحل: في هذه الحالة المعالجة = مجال تمديد الكاحل.

٢- سرعة ثني الكاحل: في هذه الحالة المعالجة = سرعة تمديد الكاحل.

ربما نقيّم عاملا واحدًا في الزمن باستخدام أنوفا أحادية العامل أو أحادية الاتجاه. أو ربما نقيّم تأثير كل من العاملين على متوسط ثني الورك باستخدام أنوفا أحادية العامل أو أحادية الاتجاه. وفي كلتا الحالتين لدينا فرضيات العدم والفرضيات البديلة لتأثير كل عامل مثل مجال حركة الكاحل على متوسط المجتمع الإحصائي مثل ثني الورك. إن فرضية العدم هي أنه لا يوجد اختلاف كبير في متوسط ثني الورك عبر حركة الكاحل.

$$H_0 : \mu_1 = \mu_2 = \mu_3 \ldots = \mu_n$$

وتنص الفروض البديلة H_1 على أن متوسطين على الأقل لمجتمعين مثل ثني الورك لحركتين مختلفتين للكاحل يختلفان كثيراً عن بعضها البعض.

إن التصميم الأساسي للتجربة أحادية العامل مُعطى في الجدول (٥,٢). في هذه الحالة هناك k معالجة في عامل واحد. k هو عدد المعالجات y_{ij} هو العينة j الفردية أو نقطة البيانات للمعالجة رقم i $\overline{y}_i$ هو متوسط العينة للمعالجة رقم i و σ_{yi}^2 هو تباين العينة للمعالجة رقم i (مقتبس من [3]).

الجدول (٢ ,٥). تجربة أحادية العامل.

المعالجة	الملاحظات	متوسط العينة	تغيُّر العينة
1	$y_{11}, y_{22}, \dots, y_{1n}$	$\overline{y}_1$	$\sigma^2 y_1$
2	$y_{21}, y_{22}, \dots, y_{2m}$	$\overline{y}_2$	$\sigma^2 y_2$
...			
I	$y_{i1}, y_{i2}, \dots, y_p$	$\overline{y}_k$	$\sigma^2 y_i$
...			
K	$y_{k1}, y_{k2}, \dots, y_{kq}$	$\overline{y}_k$	σy_k

على سبيل المثال قد نكون مهتمين بمقارنة فقدان الوزن لثلاثة أنواع مختلفة من حبوب الحمية. في هذه الحالة $k = 3$. تحت عمود فقدان الوزن سيجد المرء فقدان الوزن لأشخاص فرديين استخدموا واحدة من حبوب الحمية الثلاث. لاحظ في هذا المثال أن عدد العينات يختلف في كل نوع من أنواع الأدوية. وفي تجربة متوازنة ينبغي أن يكون هناك عدد متساو من العينات لكل نوع من أنواع المعالجات. لدينا في العمودين الثالث والرابع متوسط العينة والانحراف المعياري لفقدان الوزن لكل حبة من حبوب الحمية. والسؤال الذي نحاول الإجابة عليه هو ما إذا كان هناك اختلاف كبير في متوسط فقدان الوزن كتابع لنوع حبة الحمية.

حبة الحمية	فقدان الوزن	المتوسط	الانحراف المعياري
بلاسيبو (Placebo)	10, 12, 5, 8, 5, 20	10.0	5.62
الدواء B	30, 5, 12, 20	16.75	10.75
الدواء C	2, 10, 5, 5, 10, 20, 25, 40	14.63	12.92

نقوم بعدد من الافتراضات عند استخدام أنوفا لمقارنة الاختلافات في المتوسطات بين ثلاثة أو أكثر من المجتمعات الإحصائية أو العمليات :

١- يتم تحديد الأشخاص بشكل عشوائي بالنسبة لمعالجة محددة (في هذه الحالة حبوب الحمية).

٢- إن المجتمعات الإحصائية أو العمليات (مثل فقدان الوزن) مُوزَّعة بشكل طبيعي تقريباً.

٣- إن التباين متساوٍ تقريباً عبر مجموعات المعالجة.

لم يتم في هذا المثال المحدد لفقدان الوزن استخدام الحجب. وبعبارة أخرى لم يحصل نفس الشخص في أي حال على أكثر من معالجة واحدة أو حبة حمية واحدة. قد يُحدِث هذا في نهاية الأمر مشكلة ويؤدي إلى استنتاجات خاطئة لأننا لا نأخذ في الاعتبار التغيّر بين الأشخاص كاستجابة لأدوية الحمية. على سبيل المثال قد نفترض أن مقدار فقدان الوزن مرتبط ببداية الوزن. من الممكن أن يكون لجميع الأشخاص المحددين للدواء A بداية أوزان أقل من الأشخاص الذين يتم إعطاؤهم الدواء B. وهكذا قد تكون الاختلافات في فقدان الوزن كبيرة ولكن قد يكون لها علاقة بسيطة بحبوب الحمية الفعلية. وقد يكون الاختلاف في الوزن بسبب الاختلافات في بداية الوزن بدلاً من حبوب الحمية. ونظراً لأن التجربة السابقة لا تستخدم الحجب فإن التغيّر بين الأشخاص الذي لا يتم أخذه في الاعتبار قد يُربِك استنتاجاتنا بخصوص فعالية حبوب الحمية.

والسؤال الذي نحاول معالجته باستخدام أنوفا هو "هل متوسطات المعالجات k متساوية أو هل تم أخذ العينات من معالجات (مجتمعات إحصائية) ذات متوسطات مختلفة؟"

نفترض داخل كل معالجة i (أو مجتمع i) أن التغيّر عبر العينات التي تم ملاحظتها Y_{ij} متأثر بمتوسط المجتمع (المعالجة) μ_i والمتغيرات العشوائية المستقلة e_{ij} المُوزَّعة بشكل

طبيعي مع متوسط يساوي صفر وتباين σ^2. وبعبارة أخرى يمكن التعبير عن العينات Y_{ij} التي تم جمعها لكل تجربة j داخل كل معالجة كما يلي:

$$Y_{ij} = \mu_i + e_{ij}$$

نحاول بواسطة نموذجنا تحديد مقدار التغيُّر في Y بسبب العامل أو المعالجة (مجتمع بمتوسط μ_i) وكم هو هذا المقدار بسبب التأثيرات العشوائية e_{ij} التي لا نستطيع التحكم بها أو التي لم يتم التقاطها في النموذج الذي تم تقديمه أعلاه.

عندما نجري أنوفا لتجربة أحادية العامل يمكننا تنظيم التحليل والنتائج في الجدول التالي (الجدول ٥,٣):

الجدول (٣ ,٥). أنوفا (ANOVA) أحادية العامل بـ k معالجة (من دون حجب).

المصدر	df	SS	MS	F
المعالجة	$k-1$			MS_{treat}/MS_{error}
الخطأ	$N-k$			

N = العدد الكلي للعينات في جميع المعالجات ؛ k = عدد المعالجات ضمن عامل واحد ؛ F = الإحصاء الذي سوف نقارنه مع الجداول F (التوزيع F) إما لرفض أو قبول فرضية العدم ؛ $SS_{treatment}$ = مجموع المربعات بين المعالجات هو مقياس التغير فيما بين متوسطات المعالجات ؛ SS_{error} = مجموع المربعات ضمن المعالجات هو مقياس لمجموع التباينات في جميع المعالجات k ؛ $MS_{treatment} = SS_{treatment}/k-1$ ؛ $MS_{error} = SS_{error}/N-k$.

ومن أجل k معالجة فإن المعادلات المُحددة للعناصر SS هي التالية:

$$SS_{treatment} = \sum_{i=1}^{k} n_i \left(\overline{y}_i - \overline{y}_{grand}\right)^2$$

و

$$SS_{error} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2 = \sum_{i=1}^{k} \sigma_{yi}^2$$

$\overline{Y}_{grand}$ هو متوسط العينة لجميع العينات في جميع العلاجات مجتمعة.

وأخيراً فإن الإحصائية F الأكثر اهتماماً بالنسبة لنا في تقدير رفض أو قبول فرضية العدم هي :

$$F = MS_{treatment} / MS_{error}$$

لاختبار الفروض بمستوى أهمية مقداره α نقارن الآن الإحصائية F المُقدَّرة مع التوزيع F المتوفر في جداول F لإدخال الجدول $F(\alpha; k-1, N-k)$. إذا كانت قيمة F المُقدَّرة أكبر من إدخال الجدول فقد نرفض فرضية العدم بمستوى ثقة مقداره $(1-\alpha)\times 100\%$.

مثال (٣ , ٥)

لدى صانع صمامات القلب ثلاث عمليات مختلفة لإنتاج صمام الوريقة. تم اختيار عينات عشوائية من 50 صمامًا خمس مرات من كل نوع لعملية التصنيع. تم اختبار عيوب آليات الفتح لكل صمام. يتم تلخيص عدد الصمامات التالفة في كل عينة من 50 صمامًا بالجدول التالي :

العملية A	العملية B	العملية C
1	5	3
4	8	1
3	6	1
7	9	4
5	10	0

باستخدام $\alpha = 0.05$ نريد تحديد ما إذا كان متوسط عدد العيوب يختلف بين العمليات (المعالجة). (فرضية العدم H_0 هي أن متوسط عدد العيوب هو نفسه في جميع العمليات).

وللإجابة على السؤال نحتاج إلى إكمال جدول أنوفا التالي :

أنوفا (ANOVA) أحادية الاتجاه: العمليات A و B و C.				
المصدر	df	SS	MS	f
العامل	2			
الخطأ	12			
الإجمالي	14			

نجد أولاً SS_{factor} (= $SS_{treatment}$)

نحن نعلم أن

$$SS_{treatment} = \sum_{i=1}^{k} n_i \left(\overline{y}_i - \overline{y}_{grand}\right)^2$$

نعلم من البيانات المُعطاة أن هناك ثلاثة معالجات A و B و C ؛ وبالتالي $k = 3$. وسوف نحدد $A = 1$ و $B = 2$ و $C = 3$. ونعلم أيضاً أن $n_1 = n_2 = n_3 = 5$. يمكننا تقدير متوسط العينة لكل معالجة أو عملية للحصول على $\overline{y}_1 = 4, \overline{y}_2 = 7.6, \overline{y}_3 = 1.8$ ويمكننا أيضاً استخدام جميع العينات الخمسة عشر لإيجاد $\overline{y}_{grand} = 4.47$.

يمكننا الآن استخدام هذه التقديرات في المعادلة من أجل $SS_{treatment}$:

$$SS_{treatment} = 5(4 - 4.47)^2 + 5(7.6 - 4.47)^2 + 5(1.8 - 4.47)^2 = 85.70$$

ونحل الآن من أجل SS_{error} :

وبالأخذ في الاعتبار أن :

$$SS_{error} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2$$

فإننا بحاجة إلى تقدير المجموع الداخلي لكل معالجة i مكتوبة بواسطة المجموع الخارجي. وهكذا عندما $i = 1$

$$SS_1 = (1-4)^2 + (4-4)^2 + (3-4)^2 + (7-4)^2 + (5-4)^2 = 20;$$

وعندما $i = 2$

$$SS_2 = (5-7.6)^2 + (8-7.6)^2 + (6-7.6)^2 + (9-7.6)^2 + (10-7.6)^2 = 17.2;$$

وعندما $i = 3$

$$SS_3 = (3-1.8)^2 + (1-1.8)^2 + (1-1.8)^2 + (4-1.8)^2 + (0-1.8)^2 = 10.8$$

وبالجمع الآن لجميع المعالجات k حيث $k = 3$ نجد:

$$SS_{error} = SS_1 + SS_2 + SS_3 = 48.0$$

إيجاد متوسط الأخطاء المربعة:

$$MS_{treatment} = SS_{treatment} / DF_{treatment} = 87.5/2 = 43.7$$

$$MS_{error} = SS_{error} / DF_{error} = 48.0/12.0 = 4$$

وأخيراً تُعطى الإحصائية F بواسطة العلاقة التالية:

$$F = MS_{treatment} / MS_{error} = 43.7/4 = 10.92$$

ونكمل الآن جدول أنوفا.

أنوفا (ANOVA) أحادية الاتجاه: العمليات A و B و C.

المصدر	df	SS	MS	f
العامل	2	87.5	43.7	10.92
الخطأ	12	48.0	4.0	

نريد الآن تحديد ما إذا كان متوسط عدد العيوب يختلف فيما بين العمليات A و B و C. إن فرضية العدم هي أن متوسط عدد العيوب لا يختلف فيما بين العمليات. ولرفض هذه الفروض بمستوى ثقة مقداره 95% $\alpha = 0.05$ يجب أن تكون الإحصائية F

المُقدَّرة أكبر من $F(0.05,2,12)=3.83$ الموجودة في جداول التوزيع F. الإحصائية F تساوي 10.92 وهي أكبر من 3.88 ؛ وبالتالي يمكننا رفض فرضية العدم ثقة مقدارها 95% ونخلص إلى أن عدد العيوب يختلف مع اختلاف عملية التصنيع وأن عينات المعالجة A و B و C تم أخذها من مجتمعات ذات متوسطات مختلفة. وفي الواقع يعرض جدول التوزيع F قيمة من أجل $F(0.01,2,12)=6.93$. وهكذا يمكننا كذلك رفض فرضية العدم بمستوى ثقة مقداره 99%.

مثال (٤ , ٥)

يجري تقييم سرعة التقاط الصورة لأربعة أنواع من ماسحات الـ MR. ويلخص الجدول التالي السرعات التي تم قياسها (بالدقائق) لثلاث عينات لكل نوع من الماسحات.

الماسح A	الماسح B	الماسح C	الماسح D
2.0	4.0	3.0	6.0
1.8	4.5	2.5	5.5
2.7	5.5	2.0	3.5

هل يختلف متوسط سرعة التقاط الصورة فيما بين أنواع الماسح الأربعة؟ (استخدم $\alpha=0.01$).

للإجابة على هذا السؤال فقد نعمل من خلال العمليات الحسابية التي أجريناها في المثال (٥,٣). ومع ذلك كما اكتشف المرء في المثال السابق عندما يزداد حجم العينة فإن عدد الحسابات يزداد بسرعة. وهكذا فإن معظم الباحثين يستخدمون حزمة برامج إحصائية لتنفيذ حسابات ANOVA (أنوفا). على سبيل المثال استخدمنا حزمة البرامج

ميني تاب (Minitab Statistical Software, Release 13.32, Minitab, 2000) لتنفيذ ANOVA. تنتج أنوفا الجدول التالي :

أنوفا (ANOVA) للسرعة					
المصدر	df	SS	MS	f	α
العامل	3	19.083	6.361	9.07	0.006
الخطأ	8	5.613	0.702		
الإجمالي	11	24.697			

ينتج عن الإحصائية F لـ 9.07 مساحة مقدارها $\alpha=0.006$ في الذيل الأيمن للتوزيع F. وبما أن $\alpha<0.01$ وهي قيمة α التي كنا نختبر عندها الفروض، فإنه يمكننا رفض فرضية العدم وقبول هذه الفروض البديلة بأن واحداً على الأقل من الماسحات يختلف عن الماسحات المتبقية في متوسط سرعة الالتقاط.

لم نستخدم الحجب في التصميم التجريبي في الأمثلة المذكورة سابقاً. وبعبارة أخرى فإن المجتمعات التي جمعنا منها العينات اختلفت من معالجة إلى معالجة. في بعض التجارب مثل اختبار فقدان الوزن بسبب حبوب الحمية فإنه ليس عملياً أو ممكناً اختبار أكثر من نوع واحد من المعالجات على نفس الوحدة التجريبية.

ومع ذلك عندما يمكن استخدام الحجب ينبغي استخدامه لمقارنة المعالجات للتقليل من التغير بين الأشخاص أو الاختلافات (التي لا يمكن التحكم بها) في النتائج التجريبية. يبين الجدول (٥,٤) التصميم التجريبي لتجربة أحادية العامل تستخدم الحجب. تخضع جميع الوحدات التجريبية في هذا التصميم التجريبي للمعالجة في كل مرة. والنتيجة هي أنه لدينا الآن متوسطات معالجة ومتوسطات حجب. وهكذا عندما نقوم بتحليل أنوفا فقد نقوم باختبار مجتمعين مختلفين من الفرضيات. فرضية العدم

الأولى هي أنه لا يوجد اختلاف كبير في متوسطات المعالجة. وفرضية العدم الثانية هي أنه لا يوجد اختلاف كبير في المتوسطات عبر الأشخاص أو الوحدات التجريبية (التغيّر بين الأشخاص ليس كبيراً).

الجدول (٤ ,٥). عامل واحد بـ k معالجة وحجب

المعالجة (سرعة الكاحل)	رقم الشخص				متوسط المعالجة
	1	2	3	4	
10	5	2	5	6	
20	10	2	8	4	
30	15	8	15	10	
متوسط الحجب					

يمكن في هذا المثال أن تكون المعالجة سرعة الكاحل وقد يكون الحجب المريض بحيث يتم اختبار كل سرعة كاحل على كل شخص مريض. من المهم في مثل هذا التصميم جعل الترتيب الذي يتم بواسطته اختبار سرعات كاحل مختلفة عشوائياً بحيث لا يكون هناك انحياز في ثني الورك بسبب آثار ترتيب سرعة الكاحل أو اهتراء الجهاز.

لاحظ أن المثال سابقاً هو أيضاً تصميم تجريبي متوازن لأن هناك نفس الأعداد من نقاط البيانات في كل خلية من الجدول.

نفترض الآن داخل كل معالجة i (أو مجتمع i) أن التغير عبر العينات التي تم ملاحظتها Y_{ij} يتأثر بمتوسط المجتمع (المعالجة) μ_i وآثار الحجب β_j والمتغيرات العشوائية المستقلة e_{ij} المُوزَّعة بشكل طبيعي بمتوسط مقداره صفر وتباين مقداره σ^2. وبعبارة أخرى يمكن التعبير عن العينات Y_{ij} التي تم جمعها لكل تجربة j داخل كل معالجة i كما يلي:

$$Y_{ij} = \mu_i + \beta_i + e_{ij}$$

وبعبارة أخرى نحاول تحديد مقدار التغير في Y بسبب العامل أو المعالجة (مجتمع بمتوسط μ_i) و كم هو بسبب آثار الحجب β_j والآثار العشوائية e_{ij} التي لا يمكننا التحكم بها أو لم يتم التقاطها في النموذج المُقدَّم سابقاً. نفترض في مثل هذا النموذج أن آثار المعالجة وآثار الحجب هي آثار مُضافة. وهذا ليس افتراضاً جيداً عندما تكون هناك آثار تفاعل بين المعالجة والحجب. تعني آثار التفاعل أن تأثير معالجة معينة قد يعتمد على حجب معين. عندما نعالج تجارب ثنائية العامل في المقطع التالي فسوف نناقش إلى حد كبير آثار التفاعل هذه.

يمكننا تلخيص أنوفا (ANOVA) لعامل واحد مع حجب بجدول أنوفا التالي (بافتراض عدم وجود أي آثار تفاعل) (الجدول ٥,٥) ؛ مقتبسة من [3]).

الجدول رقم (٥ , ٥). أنوفا (ANOVA) أحادية العامل مع حجب.

المصدر	df	SS	MS	F
المعالجة	$k-1$			MS_{treat}/MS_{error}
الحجب	$b-1$			MS_{block}/MS_{error}
الخطأ	$(b-1)(k-1)$			

لاحظ أنه لدينا الآن قيمتين F مُقدَّرتين للأخذ في الاعتبار. تتم مقارنة $F_{\text{treatment}}$ مع إدخال الجدول من أجل $F(\alpha; k-1, (b-1)(k-1))$ ويتم رفض فرضية العدم التي تنص على أنه ليس هناك فرق في متوسط ثني الورك عبر المعالجات إذا كانت قيمة F المُقدَّرة أكبر من إدخال الجدول.

تتم مقارنة F_{block} مع إدخال الجدول من أجل $F(\alpha; b-1, (b-1)(k-1))$ ويتم رفض فرضية العدم التي تنص على أنه ليس هناك فرق في متوسط ثني الورك عبر الأشخاص إذا كانت إحصائية F المُقدَّرة أكبر من إدخال الجدول.

مثال (٥ , ٥)

تجربة أحادية العامل مع حجب.

تقوم شركة HeartSync بتصنيع أربعة أنواع من أجهزة إزالة الرجفان تختلف في قوة الصدمة الكهربائية التي يتم إعطاؤها لحادثة رجفان. تم تقسيم ما مجموعه 280000 مريض إلى أربع مجموعات كل منها 70000 مريض. وتم تخصيص كل مجتمع إلى أحد أجهزة إزالة الرجفان الأربعة وتم تسجيل عدد الصدمات التي فشلت في إزالة الرجفان لمدة أربع سنوات متتالية. وكانت النتائج على النحو التالي:

سنة بعد الزراعة	الجهاز A	الجهاز B	الجهاز C	الجهاز D
1	6	1	9	2
2	8	1	10	2
3	5	3	8	0
4	10	2	11	5

هناك نوعان من الأسئلة التي نتمنى معالجتها بهذه البيانات:

١- باستخدام $\alpha = 0.01$ هل يختلف متوسط عدد الإخفاقات كثيراً كتابع لنوع الجهاز؟

٢- باستخدام $\alpha = 0.01$ هل يختلف متوسط عدد الإخفاقات كثيراً كتابع لعدد السنوات بعد الزرع؟ وبعبارة أخرى هل عدد السنوات بعد الزرع مصدرٌ رئيسيٌّ للتغير بين المجتمعات؟

مرة أخرى بدلاً من تقدير الحسابات يدوياً قد نستخدم البرمجيات الإحصائية مثل Minitab للحصول على النتائج التالية باستخدام أنوفا مع حجب:

أنوفا (ANOVA) لعدد الإخفاقات				
المصدر	df	SS	MS	f
الجهاز	3	173.188	57.729	35.67
السنة	3	20.688	6.896	4.26
الخطأ	9	14.563	1.618	
الإجمالي	15	208.438		

إن الإحصائية F للجهاز $F = 35.67$ أكبر من قيمة F الحرجة $F(0.01,3,9) = 6.99$ المُعطاة في جدول F. وهكذا يمكننا رفض فرضية العدم وقبول هذه الفروض البديلة بأن واحداً على الأقل من الأجهزة يختلف عن الأجهزة المتبقية في متوسط عدد الإخفاقات. الجزء الثاني من الأسئلة يختبر الفروض القائلة بأن متوسط عدد الإخفاقات يختلف اختلافاً كبيراً كدالة لعدد السنوات بعد الزرع (عامل الحجب). والإحصائية F للسنة $F = 4.26$ هو أقل من قيمة F الحرجة $F(0.01,3,9) = 6.99$ المُعطاة في جدول F وبالتالي فإننا نقبل فرضية العدم وهو ما يعني أن معدل الإخفاق لا يختلف بين عدد السنوات بعد الزرع.

(٢, ٣, ٥) التجارب ثنائية العامل Two-Factor Experiments

ناقشنا في التحدي في الهندسة الطبية الحيوية الذي يصف انعكاسات (ردود فعل) ثني الورك اثنين من العوامل التي تؤثر على ثني الورك: سرعة ومجال تمديد الكاحل.

في التجربة ثنائية العامل التي أدت إلى أنوفا ثنائية الاتجاه هناك عاملان يجري تغييرهما A و B حيث إن A لديه a معالجات و B لديه b معالجات وهناك n عينة في كل تركيبة في A و B.

يُقال إن التجربة ثنائية العامل متقاطعة بشكل كامل إذا كانت هناك عينات تم جمعها لكل تركيبة من العوامل A و B. بالإضافة إلى ذلك يُقال إن التجربة متوازنة إذا كان لدينا نفس العدد من العينات لكل تركيبة من العوامل A و B

يوضح الجدول (٥,٦) أدناه تجربة ثنائية العوامل متوازنة ومتقاطعة بشكل كامل. يتم مقاطعة كل معالجة من المعالجتين ضمن العامل A مع كل معالجة من المعالجات الثلاثة ضمن العامل B. لاحظ أيضاً أنه لدينا ثلاث عينات لكل تركيبة من A و B.

الجدول رقم (٦ , ٥). تجربة ثنائية العامل

	العامل B		
العامل A	1	2	3
1	1.2, 1.4, 2.1	2.3, 2.2, 2.6	6.4, 5.8, 3.2
2	3.2, 4.1, 3.6	4.1, 4.3, 4.0	8.2, 7.8, 8.3

السؤال الذي نحاول معالجته بواسطة أنوفا ثنائية العامل هو ما إذا كانت هناك اختلافات في متوسطات المعالجات لكل عامل من العاملين. وبالإضافة إلى ذلك نريد أن نعرف ما إذا كانت هناك آثار تفاعل بحيث يكون هناك اختلافات كبيرة في المتوسطات كدالة لتفاعل الارتباط بين العوامل. وبعبارة أخرى هناك اختلافات كبيرة في متوسطات العينات وبالتالي في متوسطات المجتمعات الإحصائية عندما تحدث تركيبات معينة من العوامل A و B معاً.

وبمجرد جمع البيانات كما هو موضح في الجدول (٥,٦) أعلاه فقد نجري أنوفا ثنائية العامل لاختبار فرضيات العدم الثلاث التالية:

$$H_0 : \mu_{A1} = \mu_{A2} = \mu_{A3} = ... = \mu_{Aa};$$
$$H_0 : \mu_{B1} = \mu_{B2} = \mu_{B3} = ... = \mu_{Bb};$$
$$H_0 : \mu_{A1B1} = \mu_{A1B2} = \mu_{A1B3} = \mu_{A2B1} = \mu_{A2B2} = ... = \mu_{AaBb}$$

إن الفروض البديلة المرافقة لكل فرضية من فرضيات العدم الثلاث هي أن هناك فرقاً كبيراً في اثنين من متوسطات المجتمعات الإحصائية على الأقل لعامل معين أو تركيبة من العوامل.

يمكن تنظيم تحليل ونتائج أنوفا ثنائية العامل كما هو وارد في الجدول (٥,٧) [٣].

الجدول (٥ , ٧). جدول أنوفا (ANOVA) ثنائية العامل (كل منها مع معالجات متعددة).

المصدر	df	SS	MS	F
A	$a-1$			MS_A / MS_{error}
B	$b-1$			MS_B / MS_{error}
AB	$(a-1)(b-1)$			MS_{AB} / MS_{error}
الخطأ	$ab(n-1)$			

إن معادلات SS و MS لكل عامل من العوامل وعوامل التفاعل هي خارج نطاق هذا الكتاب ولكن يمكن إيجادها في [3]. يستخدم الباحثون في الواقع حزمة برامج إحصائية شائعة مثل Minitab أو SPSS أو SAS لتقدير قيم SS و MS (بسبب العبء الحاسوبي) والإشارة ببساطة إلى الإحصائيات F لرفض أو قبول فرضيات العدم.

نلاحظ أن هناك ثلاثة إحصائيات F لاختبار كل فرضية من فرضيات العدم الثلاث التي سبق وصفها. ففي كل حالة نقارن إحصائية F المُقدَّرة مع القيم في الجدول F الذي يمثل التوزيع F. وبشكل أكثر تحديداً فإننا نقارن تقديرات قيم F مع إدخالات الجدول التالي :

لاختبار $H_0 : \mu_{A1} = \mu_{A2} = \mu_{A3} = ... = \mu_{Aa}$ قارن F_A مع $F(\alpha; a-1, ab(-1))$ ؛

لاختبار $H_0 : \mu_{B1} = \mu_{B2} = \mu_{B3} = ... = \mu_{Bb}$ قارن F_B مع $F(\alpha; b-1, ab(-1))$ ؛

لاختبار $H_0 : \mu_{A1B1} = \mu_{A1B2} = \mu_{A1B3} = \mu_{A2B1} = \mu_{A2B2} = ... = \mu_{AaBb}$ قارن F_{AB} مع $F(\alpha; (a-1)(b-1), ab(-1))$.

في كل اختبار من الاختبارات الثلاثة إذا كانت الإحصائية F المُقدَّرة أكبر من إدخالات الجدول للتوزيع F يمكن للمرء رفض فرضية العدم وقبول الفروض البديلة بثقة مقدارها $(1-\alpha)\times 100\%$.

مثال (٦ , ٥)

مثال على تجربة ثنائية العوامل سيتم تقييمها باستخدام أنوفا ثنائية العامل تحدث عندما يبحث المهندسون الطبيون الحيويون في فعالية علاج إعادة التأهيل والعلاج الدوائي في شفاء حركة طرف بعد السكتة الدماغية. تتضمن تفاصيل التصميم التجريبي ما يلي:

١ – العامل T: العلاج المُستخدم (هناك ثلاثة أنواع من العلاجات T1 و T2 و T3)؛

٢– العامل D: الدواء المُستخدم (هناك ثلاثة أنواع من الأدوية D1 و D2 و D3)؛

٣– يتم تحديد 36 مريضًا بشكل عشوائي إلى كل تركيبة من T و D؛

٤– القياس: عدد الأيام لتحقيق معايير الشفاء.

تصميم تجريبي لعاملَين

T3			T2			T1		
D3	D2	D1	D3	D2	D1	D3	D2	D1
7	15	9	16	8	22	13	25	20
10	10	12	19	10	16	12	16	15
9	9	8	11	9	17	22	10	18
9	10	8	21	11	12	10	20	24

الأسئلة التي نحاول معالجتها هي ما إذا كان هناك اختلاف كبير في متوسط أيام الشفاء لأنواع علاج إعادة التأهيل الثلاثة ومتوسط أيام الشفاء لأنواع العلاج الدوائي الثلاثة والاختلافات في متوسط أيام الشفاء للتركيبات التسعة من علاج إعادة التأهيل والعلاج الدوائي (تأثيرات التفاعل).

ينتج تحليل أنوفا الذي يتم إجراؤه باستخدام البرمجيات الإحصائية المعروفة باسم Minitab الجدول التالي الذي يلخص تحليل أنوفا. لاحظ أن هناك ثلاث إحصائيات F مُقدَّرة. يمكننا استخدام قيم F الثلاث لاختبار الفرضيات التالي :

$$H_0 : \mu_{T1} = \mu_{T2} = \mu_{T3};$$
$$H_0 : \mu_{D1} = \mu_{D2} = \mu_{D3};$$
$$H_0 : \mu_{T1D1} = \mu_{T1D2} = \mu_{T1D3} = \mu_{T2D1} = \mu_{T2D2} = \ldots = \mu_{T3D3}$$

أنوفا (ANOVA) ثنائية الاتجاه لأيام الشفاء مقابل T و D.

المصدر	df	SS	MS	F	α
T	2	337.4	168.7	11.4	0.000
D	2	36.2	18.1	1.23	0.309
التفاعل	4	167.8	41.9	2.84	0.043
الخطأ	27	398.3	14.8		

لاحظ أن الإحصائية F المُقدَّرة لعلاج إعادة التأهيل $F = 11.44$ لا تتجاوز القيمة الحرجة عندما $\alpha < 0.05$ (في الواقع α هي أصغر من 0.0001) ؛ وبالتالي فإننا نستنتج أن العينات لعلاجات إعادة التأهيل الثلاثة المختلفة تمثل مجتمعات إحصائية ذات متوسطات مختلفة. وبعبارة أخرى تؤدي علاجات إعادة التأهيل المختلفة إلى أيام شفاء مختلفة. ومع ذلك فإن الإحصائية F المُقدَّرة للعلاج الدوائي لا تتجاوز قيمة F الحرجة في الجدول عندما $\alpha < 0.05$ وبالتالي فإننا نقبل فرضية العدم بأن العلاج الدوائي لا يؤثر تأثيراً كبيراً على أيام الشفاء وأنه لم يتم أخذ العينات من مجموعات ذات متوسطات مختلفة. وعلاوة على ذلك فإن الإحصائية F الثالثة $F = 2.84$ أكبر من قيمة F في الجدول عندما $\alpha < 0.05$ مما يشير إلى أنه قد يكون هناك اختلاف كبير في أيام الشفاء بسبب التفاعل بين علاج إعادة التأهيل والعلاج الدوائي.

(٣, ٣, ٥) إجراء توكي (Tukey) للمقارنات المتعددة

Tukey's Multiple Comparison Procedure

بعد أن أثبتنا أن هناك فارقاً كبيراً في المتوسطات عبر المعالجات ضمن عامل ما فقد نستخدم آخر الاختبارات المخصصة مثل اختبار HSD لتوكي للمعرفة المزدوجة والمقارنات المتعددة (Tukey's HSD multicomparison pair wise test) [3, 9]. تظهر أنوفا ببساطة أن هناك متوسط معالجة واحد على الأقل يختلف عن المتوسطات الأخرى. ومع ذلك فإن أنوفا لا تقدم معلومات محددة عن متوسط (متوسطات) المعالجة الذي يختلف عن متوسط (متوسطات) المعالجة الآخر. ويسمح اختبار HSD لتوكي بمقارنة الفروق الإحصائية في المتوسطات بين جميع أزواج المعالجات. وبالنسبة للتجربة أحادية العامل مع k معالجة فإن هناك $k(k-1)/2$ مقارنة معرفة مزدوجة لاختبارها. والنقطة الهامة التي يجب ملاحظتها هي أنه إذا كانت α هي احتمال خطأ النوع I لمقارنة واحدة فإن احتمال القيام بخطأ النوع I على الأقل للمقارنات المتعددة أكبر بكثير. لذا إذا أردنا ثقة مقدارها $(1-\alpha)\times100\%$ لجميع مقارنات المعرفة المزدوجة الممكنة فإنه يجب أن نبدأ بـ α أصغر بكثير. يسمح إجراء توكي للمقارنات المتعددة بمثل هذه التعديلات في الأهمية عند إجراء مقارنات المعرفة المزدوجة.