الوحدة الرابعة معالجة البيانات

1. مقدمة

تعد جميع النتائج المخبرية معرضة للأخطاء وأن القيم المقاسة تختلف دائماً عن القيمة الحقيقية فإذا أعيد قياس قيمة معينة باستعمال نفس طريقة العمل فيلاحظ وجود اختلاف مهما كان صغيراً بين هذه القيم. ويمكن تقليل هذا الاختلاف بين القياسات بإتباع طرق تحليل متعددة ولكن لا يمكن التخلص من هذا الاختلاف تماماً. وعليه فإن تكرار القياسات لنفس القيم المقاسة ضرورة لابد منها حتى يمكن الحصول على نتائج دقيقة.

2. تصنيف الأخطاء Classification of errors

ويمكن تحديد مصادر الأخطاء في التحليلات الكيميائية بنوعين أساسين هما : الأخطاء المحددة determinate errors والأخطاء غير المحددة indeterminate errors غالباً ما تسمى بالأخطاء العشوائية وهي تتبع قوانين الاحتمالات . وتعرف الأخطاء غير المحددة على أنها تلك الأخطاء التي لا يمكن تحديدها، وتعريفها، وليس لها قيمة مقاسه ، وأما الأخطاء المحددة فهي التي لها قيم محددة يمكن قياسها.

وتحتوي الأخطاء المحددة على الأخطاء الشخصية مثل الإهمال، أو عدم القدرة على تميز الألوان، وكذلك تحتوي على الأخطاء الآلية والأخطاء الناتجة عن طريقة التحليل المختارة. ويمكن تصحيح مثل هذه الأخطاء بإجراء تحاليل لعينات قياسية، أو إجراء اختبارات المقارنة واستخدام أحجام مختلفة من العينة.

1.2 الأخطاء المحددة Determinate errors

وهي الأخطاء التي يمكن تعيين وتحديد مصادرها وتسمى أحياناً بالأخطاء النظامية Systematic errors ويمكن تصنيف هذا النوع من الأخطاء على النحو التالي:

أ- أخطاء آلية Instrumental errors

وهي الأخطاء المقترنة بالآلة وأسبابها :

1- عدم التأكد من قراءة القياس

2 - استعمال أوزان وأدوات غير معايرة

3- استعمال مواد كيميائية مجهولة النقاوة

4- الاستعمال الخاطئ لبعض الأدوات المخبرية

5- الجهاز المستعمل للقياس

ب- أخطاء ناتجة من طبيعة طريقة التحليل methodic errors

وهي ناتجة عن استعمال طريقة غير مناسبة وأسبابها :

-ارتفاع فى قابلية ذوبان الراسب

-التفاعلات غير التامة وغير القياسية

-تلوث الراسب

-تحلل الراسب أثناء عملية الغسل او الحرق

ج- أخطاء تشغيلية Operative errors

وهي ناتجة عن المحلل الكيميائي وأسبابها:

-التحليل في أوان غير مغطاة ودخول مواد غريبة داخل النموذج

-كثرة غسل الراسب وذوبانه

-عدم استعمال المجففات المناسبة

-التجفيف والحرق غير الكاملين

-وزن الجفنة (البوتقة) قبل تبريدها

-فقدان جزء من الراسب أثناء عمليات الترشيح والغسل والحرق والوزن

د- أخطاء شخصية personal errors

وهي مقترنة بالشخص وهو المحلل الكيميائي وأسبابها :

-عدم مقدرة المحلل على التمييز بين الألوان المختلفة

-قلة الإدراك الحسي للمحلل

-التميز فى اختيار نتيجة دون أخري.

2.2 تأثير الأخطاء المحددة على نتيجة التحليل:

تعد الأخطاء المحددة أخطاء ثابتة Constant errors وأخطاء نسبية Proportional errors كلاهما يعتمدان على حجم وكمية النموذج قيد التحليل.

ومن الصعب معرفة الأخطاء المحددة ولكن معايرة الأجهزة والأدوات المستعملة تقلل من مقدار هذه الأخطاء. وفيما يلي بعض الطرق التي يمكن الاستعانة بها لمعرفة هذه الأخطاء وتحديد مقدارها:

-تحليل نماذج قياسية

-إجراء التحليل بطرق مختلفة

-استعمال تحليلات ضابطة

-استعمال تحليلات لنماذج مختلفة الأوزان.

إذا تفحصنا مجموعة من المشاهدات ، أو البيانات الإحصائية عن ظاهرة من الظواهر فإننا نجد أن هذه البيانات تميل إلى التمركز حول قيمة معينة، وهذا الميل نحو تلك القيمة هو ما يسمي بالنزعة المركزية لهذه البيانات وتسمى تلك القيمة بالقيمة المتوسطة للبيانات. وهناك أكثر من قيمة متوسطة تتجه نحوها البيانات ومن هذه القيم يمكننا أن نعدد : الوسط الحسابي، المنوال، الوسيط وقد تتساوى هذه القيم فى بعض التوزيعات، وقد تختلف ، ولكن لكل من هذه القيم طريقة خاصة لحسابها، وتعريفها.

3. الأرقام ذات الدلالة Significant Figures

كما نعلم فإن الكيمياء التحليلية الكمية يتم فيها الحصول على بيانات ونتائج رقمية متعددة .إذ إن نتيجة كل تجربة يتم إجراؤها في المخبر تتضمن عدة عمليات منها على سبيل المثال تحديد الوزن أو الحجم أو التركيز أو النسبة المئوية ، وكل قيمة من تلك القيم تنتج عن متوسط ثلاثة قراءات على الأقل .لذلك من البديهي أن نتعلم الطرق العلمية المتعلقة بتصدير البيانات ومعالجتها

- جميع الأرقام الصحيحة (1-9) تعتبر ذات دلالة ، وعليه يجب عدها من ضمن الأرقام المعنوية .فمثلاً القيمة 592 تحتوي على 3 أرقام معنوية.

- القيم التي تحتوي على أصفار ، وتشمل عدة أنواع:

أ .الأصفار البينية ، أي التي تقع بين الأرقام الصحيحة ، جميعها ذات دلالة .فمثلاً القيمة 400105 تحتوي على 6 أرقام معنوية.

ب .إذا لم تكن هناك علامة عشرية في القيمة المراد تحديد عدد الأرقام المعنوية فيها ، تعتبر الأصفار على يمين ويسار الأرقام الصحيحة غير ذات دلالة .فمثلاً القيمة 320000 تحتوي على رقمين ذات دلالة
فقط ، وأيضاً القيمة 00274 تحتوي على 3 أرقام ذات دلالة.

ت .إذا احتوت القيمة الم ا رد معرفة عدد الأرقام ذات الدلالة فيها على علامة عشرية ، فإن الأصفار التي تقع على يمين الرقم الصحيح تعتبر ذات دلالة ، بينما الأصفار التي تقع بين العلامة العشرية والرقم
الصحيح ليست ذات دلالة .فمثلاً القيمة 0.002601000 تتكون من 7أرقام ذات دلالة ، بينما القيمة 0.0023 تتكون من رقمين ذات دلالة فقط.

4. العمليات الحسابية باستخدام مفهوم الأرقام ذات الدلالة

قبل الخوض في القوانين التي تحكم العمليات الحسابية المختلفة من المفيد أن نتكلم قليلاً عن التقريب .من المعلوم بداهة أن الرقم إذا تجاوز النصف فإنه يقرب بإضافة واحد إلى ما قبله ، ولا خلاف في ذلك .لكن ماذا لو كان الرقم الذي يحتاج إلى تقريب يساوي النصف تماماً؟ مثلاً نرغب في تقريب القيمة
14.75 لتحتوي على رقم عشري واحد ، فهل يتم حذف ال 5 أم إضافة 1 إلى ال7 ؟ في هذه الحال ننظر إلى الرقم الذي يسبق ال 5 فإن كان فردياً فإن ال 5 تقرب ، وان كان زوجياً فإن ال 5 لا تقرب .وعليه فتقريب القيمة 14.75 إلى رقم عشري واحد تعطينا . 14.8 وبالمثل فإن تقريب القيمة 14.85 إلى رقم عشري واحد تعطينا أيضاً 14.8 ، تطبيقاً للقاعدة المذكورة . وقد نشأت هذه القاعدة من الإحصاء حيث أنه ليس من الصواب تقريب ال 5 دائماً لأن ذلك سيؤدي إلى خطأ إيجابي بالضرورة ، بينما هناك احتمال% 50 أن يكون ما قبل ال 5 فردياً و % 50 أن يكون ما قبلها زوجياً ، وعليه فإن ال 5 تقرب في 50 %من ال مرا ت بينما لا تقرب في ال % 50 الأخرى ، وهذا منطقي.
والآن لننظر في كيفية إجراء العمليات الحسابية باستخدام الأرقام ذات الدلالة:

الجمع والطرح

تتبع عمليتا الجمع والطرح نفس الطريقة ، ولهذا فمن الممكن التعامل معهما كعملية حسابية واحدة .وتتلخص فكرة هذا النوع من الحسابات بالنظر إلى القيمة التي تتضمن أعلى درجة من عدم التأكد ، فمثلاً لننظر إلى المسألة التالية علماً بأن جميع القيم هي قيم تجريبية:

أوجد حاصل جمع التالي (34.403 = 4.5 + 3.68 + 1.223 + 25 : من الحاسبة)

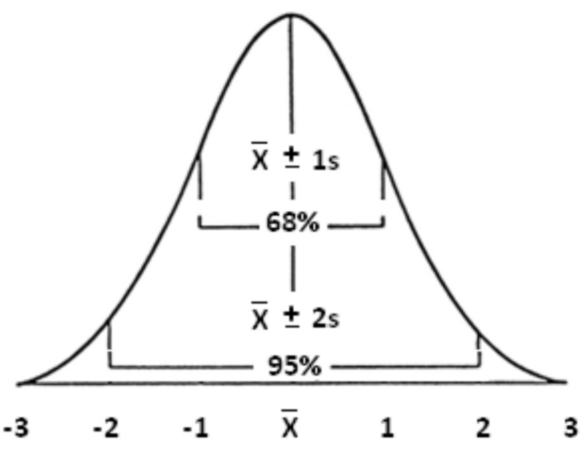
ولتحليل المسألة نقول بأن عدم التأكد في القيمة 25 هو 1 ± (أي في العدد الصحيح)وعدم التأكد في القيمة 1.223 هو ±0.001(أي جزء من ألف)بينما عدم التأكد في القيمة 3.68 هو(0.01 ± أي جزء من مائة)، وأخيرا فإن عدم التأكد في

القيمة 4.5 هو(0.1 ± أي جزء من عشرة .)وبديهي أن عدم التأكد في العدد الصحيح يعتبر أعلى قيمة لعدم التأكد في القيم المذكورة ، وعليه فالجواب يجب أن يحتوي على عدم تأكد في العدد الصحيح ، ولا يمكن أن يحتوي على عدم تأكد في الجزء من عشرة أو المائة.

5. الانحراف المعياري

من أهم طرق التعبير عن الدقة مفهوم الانحراف المعياري ، الذي يعطينا دلالة واضحة جداً عن مدى اتساع النتائج التجريبية التي نحصل عليها بتكرارتجربة ما .ويعبر الانحر اف المعياري تحديداً عن عدم التأكد(وبالتالي الدقة المطلقة)المصاحبة للمتوسط التجريبي .وعليه فإن حساب الانحراف المعياري للنتائج يعتبر من أهم الممارسات التي يقوم بها الكيميائي عند معالجة النتائج والبيانات التي حصل عليها ، ذلك لأنه يلخص أكثر أنواع عدم التأكد المصاحبة لنتائج تجربة ما ، بما ينعكس على قيمة الجواب كما يجب أن يكون.

وعند تحليل عدد كبير من العينات فإن توزيع النتائج حول المتوسط يأخذ شكل الجرس التقليدي ، أو Gaussian distributionكما هو معروف في علم الإحصاء ، والشكل أدناه يبين ذلك:



فإذا كان $\bar{x}$ هو متوسط النتائج و s هو الإنحراف المعياري فإن النتائج من– الناحية الإحصائية – تتوزع حول المتوسط بحيث تكون حوالي % 68 من النتائج واقعة بين قيمة المتوسط $\pm 1s$ ، بينما تقع حوالي % 95 من النتائج بين قيمة المتوسط $\pm 2s$.وكمثال على ما سبق لو فرضنا أننا نرغب في معرفة توزيع أوز ا ن عدد 100 طالب حول متوسط أو زا نهم البالغ 65 كغ ، واذا كان الإنحراف المعياري(ستأتي طريقة حسابه بعد قليل)يبلغ 3± كغ فإننا نعلم من البيانات الإحصائية للتوزيع الطبيعي أن:

- من بين الطلبة ال 100 من المتوقع وجود حوالي 68 طالباً تقع أوز ا نهم بين (3 - 65) و (3 + 65 كغ)، أي بين 68 - 62 كغ.
- أيضاً ، من بين الطلبة ال 100 من المتوقع وجود حوالي 95 طالبأتقع أو زانهم بين 2*3 - 65 و 3 * 2 + 65 كغ ، أي بين71 – 59 كغ
- .أيضاً يمكننا الاستدلال إحصائياً على أن أكثر من % 99 من النتائج تقع بين قيمة المتوسط $\pm 3s$ ، وغير ذلك..

1.5. أنواع الانحر ا ف المعياري

هناك أربعة أنواع من أنواع الانحرا ف المعياري التي سنتعرض لدراستها في هذا المساق .ولكل منها طبيعة معينة وشروط استخدام ومدلول ، إضافة إلى ظروف استخدام معينة.

أولاً :الانحرا ف المعياري لمجتمع إحصائي

وهو نوع من أنواع الانحر ا ف المعياري يتضمن استخدام كافة أفرا د العينة(أو كل العينة)في التحليل ، وليس عينة محدودة منها ، وهو غير مستخدم في الكيمياء التحليلية .فمثلاً إذا رغبنا في تحليل عينة دوائية معينة فإنه يجب لحساب الانحر ا ف المعياري الكلي أن يتم تحليل كامل العينة ، واذا أردنا – على سبيل المثال تحليل الكالسيوم في دم مريض فإنه لا بد من تحليل -كامل دم المريض !!، وهو أمر غير ممكن .أضف إلى ذلك أن هذا النوع من الانحر اف المعياري يشترط وجود ما يزيد على عشرين عينة ، بينما في الكيمياء التحليلية نستخدم من 5 - 3 عينات مكررة .والعلاقة الرياضية التي تستخدم للحصول على الانح ا رف المعياري لمجتمع إحصائي هي:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

حيث أن x_i هي قيمة النتيجة i ،μ هو المتوسط الحقيقي ، و N تعبر عن عدد العينات ، بينما σ هي الانحر ا ف المعياري الكلي.

ثانياً :الانحر ا ف المعياري لعينة إحصائية

وهذا النوع من الانحر ا ف المعياري يعالج عدد قليل من العينات بدلاً من المجتمع الإحصائي كاملاً ، وهو بذلك مناسب للاستخدام في الكيمياء التحليلية .والعلاقة الرياضية التي تستخدم لحسابه هي:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

حيث x_i هي قيمة النتيجة i ، و $\bar{x}$ هو المتوسط التجريبي ، بينما N هي عدد التجارب ، و s هو الانحراف المعياري للعينة . كما يطلق على القيمة $(N{-}1)$ مصطلح درجات التحرر$(degrees\ of\ freedom)$

ثالثاً :الانحر ا ف المعياري المشترك

هذا النوع من الانحرا ف المعياري **يستخدم في حالة وجود أكثر من مجموعة من النتائج ، كل مجموعة منها عبارة عن نتائج لنفس العينة المكررة** ، مع العلم أن عدد النتائج في كل مجموعة ليس شرطاً أن يكون متساوياً مع عدد النتائج في المجموعات الأخرى .ومن الجدير بالذكر أننا في هذا المساق سنعالج النتائج لعينة مكررة عبر مجموعتين فقط لا أكثر .ومن الممكن اشتقاق القانون المستخدم لحساب s_p مباشرة من قانون الانحرا ف المعياري لعينة ، وذلك بإضافة المصطلحات المناسبة في البسط والمقام ، كما هو أدناه:

$$s_p = \sqrt{\frac{\sum_{i=1}^{N1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{N2}(x_{i2} - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

حيث s_p الانحرا ف المعياري المشترك ، $\bar{x}_1$ متوسط نتائج المجموعة الأولى، $\bar{x}_2$ متوسط نتائج المجموعة الثانية ، N_1 و N_2هما عدد نتائج المجموعة الأولى والثانية ، بالترتيب.

رابعاً :الانحر اف المعياري للفروق

وهو نوع من الانحر اف المعياري المستخدم في حساب الانحر ا ف المعياري لمجموعتين من النتائج التي يتم الحصول عليها من عينات متعددة بطريقتين مختلفتين ، ويتميز بأن المجموعتين في هذه الحالة يجب أن تتكونان من نفس العدد من النتائج ، حيث يتم تحليل كل عينة على حدة باستخدام الطريقة
الأولى ثم الثانية ، وبذلك نحصل على نتيجة واحدة(الفرق)لكل منهما .ومن الممكن التعبير عن الانحراف المعياري للفروق بالعلاقة التالية:

$$s_d = \sqrt{\frac{\sum_{i=1}^{N}(D_i - \bar{D})^2}{N - 1}}$$

حيث D_i الفرق بين نتيجة التحليل لكل عينة على حدة(باستخدام الطريقة الأولى والثانية)،$\bar{D}$ هو متوسط مجموع الفروق ، بينما N عدد العينات و s_d الانحر اف المعياري للفروق

6. الاختبارا ت ذات الدلالة($Tests\ of\ significance$)

في كثير من الأحيان ، نتيجة للخبر ا ت المتر ا كمة أو رغبة في تجاوز بعض سلبيات طريقة ما ، أو محاولة لتحسين النتائج التي يمكن الحصول عليها ، فإننا قد نجرب استخدام طريقة جديدة بدلاً من الطريقة القياسية .لكن يجب أن يخضع ذلك لمعايير صارمة إذ من البديهي أن تكون الطريقة الجديدة مكافئة للطريقة القياسية في الأداء .لكن ما هي المعايير التي على أساسها يمكننا القول بأن طريقة ما مكافئة للطريقة القياسية؟ هناك معيار ا ن واضحان ، ألا وهما:

- أن تكون الطريقة الجديدة مكافئة للطريقة القياسية في الدقة، وهذا يعني أنه يجب ألا يكون هناك فرق ذو دلالة إحصائية بين ال ($variances$) للطريقتين ، أو بمعنى آخر فإن الانحر ا ف المعياري للطريقتين يجب أن يكون متقارباً .ونستنتج ذلك بتطبيق اختبار. F
- أن تكون الطريقة الجديدة مكافئة للطريقة القياسية في قيمة النتائج ، بمعنى أن تكون النتائج التي نحصل عليها باستخدام الطريقة الجديدة قريبة من نتائج الطريقة القياسية(النتائج الصحيحة أو المقبولة .)ونستنتج ذلك بتطبيق اختبار t

أولاً: اختبار F

يعطينا اختبار F فكرة عن وجود أو عدم وجود فرق إحصائي جوهري بين ال($variances$)للنتائج التي نحصل عليها باستخدام الطريقة القياسية والجديدة ، ويمكن حساب قيمة F من العلاقة:

$$F = \frac{s_1^2}{s_2^2} > 1$$

قيمة F يجب أن تكون أكبر من 1 ، وعليه فال $variance$ الأكبر يوضع في البسط بغض النظر ما إذا كان للطريقة الجديدة أو القياسية .بعد ذلك نقارن قيمة F المحسوبة من العلاقة أعلاه مع قيمتها المجدولة عند مستوى الثقة المطلوب.

Critical values of $F = s_1^2/s_2^2$ at 95% confidence level

Degrees of freedom for s_2	Degrees of freedom for s_1									
	2	3	4	5	6	7	8	9	10	12
2	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4
3	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.79	8.74
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.58
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91
11	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85	2.79
12	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69
13	3.81	3.41	3.18	3.02	2.92	2.83	2.77	2.71	2.67	2.60
14	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48

ونعتبر أنه ليس هناك فرق إحصائي جوهري بين ال($variances$) للطريقتين إذا كانت قيمة F المحسوبة أقل من قيمتها المجدولة:

$Fcalc < Ftab$

والا فإن الطريقة الجديدة لا يمكن استخدامها لأنها تفتقر إلى الدقة.

ثانياً :اختبار t

أشرنا أعلاه أنه يمكن تطبيق اختبار t لمعرفة إذا ما كان هناك أي فرق إحصائي جوهري بين **النتائج** التي حصلنا عليها باستخدام طريقة جديدة وتلك التي حصلنا عليها باستخدام الطريقة القياسية .وبنفس المعالجة التي ذكرناها عند حديثنا عن اختبار F ، فإننا نقارن بين قيمة t المحسوبة بقيمتها المجدولة عند درجات التحرر المستخدمة ومستوى الثقة المطلوب .فإذا كان:

$t_{calculated} < t_{tabulated}$

$$\pm t = (\bar{x} - \mu)\frac{\sqrt{N}}{s}$$

فإنه في هذه الحالة لا يكون هناك فرق إحصائي جوهري بين نتائج الطريقة الجديدة والقياسية ، بمعنى أنه يمكن استخدام الطريقة الجديدة بدلاً من القياسية (بشرط نجاح اختبار .) F لكن كيف يمكن الحصول على $t_{calculated}$ ؟

من الممكن القول أن هناك ثلاثة أحوال قد تصادفنا عند رغبتنا في حساب $t_{calculated}$ ، ويجب أن نتعلم كيف نحسب قيمة $t_{calculated}$في كل حالة .وهذه الأحوال هي:

1- حساب قيمة $t_{calculated}$ في حالة معرفتنا للقيمة الصحيحة للنتيجة.
2- في حالة عدم معرفتنا للنتيجة الصحيحة ، فإننا نفترض أن نتيجة التحليل بالطريقة القياسية تعطينا القيمة الصحيحة . وعليه فإن حساب في هذه الحالة يتطلب المقارنة بين متوسطي النتائج التي حصلنا عليها باستخدام الطريقتين (الجديدة والقياسية.)
3- أما الحالة الثالثة فتعالج حساب قيمة $tcalculated$ عندما يكون لدينا عينات متعددة ، لا عينة متكررة.

اختبار Q للاستبعاد

في بعض الأحيان نحصل على نتائج متقاربة لتحليل ما ، عدا نتيجة نشك في مصداقيتها ، وهنا لا بد من الإجابة على التساؤل : هل نحتفظ بتلك القيمة الشاذة أم نستبعدها؟ .وللإجابة على هذا التساؤل فإننا نطبق قانون الاستبعاد أو الرفض ، وهو ما يسمى اختبار Q ، وفيه نحسب قيمة Q من العلاقة:

$$Q = \frac{a}{w}$$

حيث أن a هو الفرق الموجب بين القيمة الشاذة(المشكوك فيها)وأقرب قيمة لها ، و w هو الفرق الموجب للقيمة الشاذة وأبعد قيمة عنها ، بينما Q هي معامل إحصائي تعتمد قيمته على مستوى الثقة وعدد النتائج .ويمثل الجدول التالي بعض قيم Q.

مع العلم أن n تمثل عدد النتائج ، وليس عدد درجات التحرر كما هو الحال في اختبار ا ت F و t

Q values at different confidence levels (CL)

n	CL at 90%	CL at 95%	CL at 99%
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568