وزارة التعليم العالي والبحث العلمي

جامعة الجيلالي بونعامة -خميس مليانة-

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

ميدان التكوين في العلوم الاقتصادية والتجارية وعلوم التسيير

مطبوعة بيداغوجية بعنوان:

دروس وتمارين في مقياس

الاحصاء 3

مقدمة لطلبة : لطلبة السنة الثانية ليسانس

شعبة: العلوم التجارية

من إعداد: د. شيشة نوال

السنة الجامعية: 2022-2023

الملخـص

تجمع هـذه المطبوعـة المـواد التعليميـة الاساسـية المسـتخدمة في تـدريس مقيـاس الإحصـاء 3 لكـل الشـعب في السـنة الثانيـة ليسـانس بالنسـبة للسداسـي الثالـث في كليـة العلـوم الاقتصـادية والتجاريـة وعلـوم التسـيير، جامعـة الجيلالي بونعامة بخميس مليانة. وعلى الرغم من أن المراجع في هذا المجال واسعة النطاق، الا أنه من المناسب تحرير مطبوعة مع التقيد بالمحتويات المحددة وفق المقرر الرسمي لوزارة التعليم العالي والبحث العلمي. وبالتالي فإن هـذه المطبوعة تحدد منهج مقياس الاحصاء 3 وتحاول تبسيطه قدر الإمكان مع إظهار المفاهيم الأساسية حتى يجد الطلبة سهولة في استيعابها والتعمق في مختلف التقنيات الإحصائية الأكثر استخداما.

وتجـدر الاشـارة الى ان تطـوير المفـاهيم المتعلقـة بمقيـاس الاحصـاء 3 تتطلـب معرفـة مسـبقة لمقياسـي الاحصـاء 1 (الاحصاء الوصفي) والاحصاء 2 (الاحتمالات) ، بالإضافة الى الرياضيات ولهذا السبب من الضروري مراعاة المفـاهيم المتعلقـة بالمتوسـط والتبـاين، والمقـاييس الإحصـائية الأخـرى، بالإضـافة الى عـدة معـارف متعلقـة بخصـائص التوزيعات الاحتمالية. وقد حرصنا في اعداد هذه المطبوعة على تبسيط المفاهيم المدعمة بأمثلة تطبيقية بعيدا عن البراهين المعقدة حتى يتمكن الطلبة استيعاب محتويات هذا المقياس.

وتعتبر هـذه المطبوعة ثمرة تجربة عدة سنوات تدريس للإحصاء الوصفي، الاحصاء التطبيقي و النماذج الاحصائية والاحصاء 3 في كلية العلوم الاقتصادية والتجارية وعلوم التسيير لجامعة خميس مليانة، بداية من سنة 2016 الى غاية 2020 ، بحيث تمت المواءمة في مستوى الليسانس سنة 2016، ولقد حاولنا أن نستفيد من هذه التجربة لصياغة محتوى المقياس بما يتماشى مع المقرر الوزاري وطبيعة تخصص الطلبة. لتحقيق هذا الغرض حرصنا على ربط المفـاهيم والقواعـد النظريـة باسـتخداماتها التطبيقيـة، وعملنـا علـى إعطـاء أمثلـة وتمـارين عـن مختلـف المفـاهيم الـتي تتضمنها هذه المطبوعة.

و اذ نأمـل ان تسـاعد هـذه المطبوعـة في إحـداث تقـدم شـامل في المفـاهيم الأساسـية الـتي لهـا تطبيقـات أكـبر في المشكلات العملية لمجالات مختلفة من الإحصاء 3 بالنسبة لمستوى الثانية ليسانس في ميدان العلوم الاقتصادية والتجارية وعلم التسيير.

فهــرس المحتويـات

	الصفحة
الملخص	
فهرس المحتويات	IV - I
قائمة الجداول والاشكال	V
المقدمـة	أ — د
الفصـل التمهيـدي: مفـاهيـم أساسيـة	1
تمهيـد	2
1. الاحصاء الوصفي والاستدلالي	2
2. المجتمع والعينة	3
3. حجم العينة	5
1.3. الحالة الأولى: عندما يكون N غير معروف أو عدد عناصره غير محدودة	6
2.3. الحالة الثانية: عندما يكون حجم المجتمع معروفا N	7
3.3. الحالة الثالثة: عندما تكون البيانات نوعية	7
4. أنواع العينات	8
1.4. العينة العشوائية البسيطة	9
2.4. العينات الطبقية	11
3.4. العينة العشوائية المنتظمة	12
4.4. العينة العنقودية	13
5. التوزيعات الاحتمالية	15
1.5. التوزيع الطبيعي	15
2.5. توزيع مربع كاي Khi Deux	19
3.5. توزيع t-Student	21
4.5. توزيع F	23

خلاصة الفصل 25

الفصل الأول: نظرية توزيع المعاينة 26

تمهيد 27

1. توزيعات المعاينة 27

2. توزيع المعاينة للمتوسط الحسابي 28

1.2. حالة المعاينة بالإرجاع 29

2.2. حالة المعاينة بدون إرجاع 32

3. طبيعة توزيع متوسط توزيع المعاينة ونظرية النهاية المركزية 33

4. توزيع المعاينة للنسبة 37

1.4. طبيعة توزيع المعاينة للنسبة 39

5. توزيع المعاينة للفرق للفروق والمجاميع 40

1.5. توزيع المعاينة للفرق بين متوسطي عينتين مستقلتين 40

2.5. طبيعة توزيع المعاينة للفرق بين متوسطين 42

3.5. توزيع المعاينة للفرق بين نسبتي عينتين 43

6. توزيع المعاينة للتباين 44

1.6. في حالة السحب بالارجاع 45

2.6. في حالة السحب بدون ارجاع 45

3.6. طبيعة توزيع المعاينة للتباين 45

4.6. توزيع المعاينة لنسبة تبايني عينتين 46

خلاصة الفصل 47

الفصل الثاني: نظرية التقدير 48

تمهيد 49

1. الاساس النظري 49

2. مفاهيم أساسية حول المقدرات النقطية 51

1.2. الفضاء المعلمي 51

2.2. قيم المقدر النقطي 52
3.2. المقدر غير المتحيز 54
4.2. مقدرات غير متحيزة لتوزيعات معينة 56
3. التقدير بمجال 56
1.3. مجال الثقة للمتوسط المجتمع 57
1.1.3. تقدير متوسط المجتمع باستخدام التوزيع الطبيعي عندما يكون σ معروف 57
2.1.3. تقدير متوسط المجتمع عندما يكون σ مجهول 59
3.1.3. أمثلة متنوعة لتقدير المتوسط 61
2.3. مجال الثقة للفرق بين متوسطي مجتمعين 63
1.2.3. فترات الثقة للفرق بين متوسطي مجتمعين بتوزيع طبيعي والانحرافين معروفين 63
2.2.3. فترات الثقة للفرق بين متوسطي مجتمعين عنندما تكون الانحرافات غير معروفة 64
3.3. مجال الثقة للنسبة والفرق بين نسبتين 65
1.3.3. مجال الثقة للنسبة 65
2.3.3. مجال الثقة للفرق بين نسبتين (للعينات الكبيرة) 66
4.3. مجال الثقة للتباين 67
5.3. فترات الثقة لنسبة تبايني مجتمعين 68
خلاصة الفصل 70
الفصل الثالث: اختبار الفرضيات 71
تمهيد 72
1. أساسيات اختبار الفرضيات 72
1.1. مناطق الرفض والقبول 74
2.1. أنواع الأخطاء في اختبار الفرضيات 75
3.1. الاختبار من طرفين والاختبار من طرف واحد 80
2. اختبار الفرضيات لمعلمات مجتمع بتوزيع طبيعي 81
1.2. اختبار فرضيات متوسط المجتمع عندما يكون التباين معلوم 81

2.2. اختبار فرضيات متوسط المجتمع عندما يكون التباين مجهول 83

3. اختبار الفرضيات للفرق بين متوسطي مجتمعين 85

1.3. حالة تبايني المجتمعين معلومين 85

2.3. حالة تبايني المجتمعين غير معلومين مع $\sigma_1^2 = \sigma_2^2$ 87

3.3. حالة تبايني المجتمعين غير معلومين مع $\sigma_1^2 \neq \sigma_2^2$ 89

4. اختبار الفرضيات لنسبة المجتمع 91

1.4. اختبار الفرضيات للفرق بين نسبتين 92

5. اختبار الفرضيات لتباين المجتمع 93

1.5. اختبار تساوي تباينين 94

خلاصة الفصل 94

الفصل الرابع: اسئلة التقييم الذاتي وتمارين مقترحة 95

1. أسئلة التقييم الذاتي 96

2. تمارين مقترحة 98

قائمة المراجع 111

الملاحق

قائمة الجداول والأشكال

قائمة الجداول		
الرقم	**العنوان**	**الصفحة**
1–0	أهم الصيغ الرياضية لحساب حجم العينة	7
1–1	توزيع المعاينة لمتوسط العينة	30
1–2	أكثر المقدرات غير المتحيزة شيوعا	55
1–3	أنواع الاخطاء الاحصائية في اختبارات الفروض	77

قائمة الاشكال		
الرقم	**العنوان**	**الصفحة**
1–0	المعاينة والاستدلال	2
2–0	انواع العينات	8
3–0	التمثيل البياني لدالة كثافة التوزيع الطبيعي	16
4–0	تمثيل دالة كثافة الاحتمال لتوزيع x^2 مع 3 و 6 و 9 و 12 درجة من الحرية	20
5–0	دالة الكثافة t–Student ، عند 1، 2، 5 ، و30 درجة حرية	22
1–1	دالة التوزيع الاحتمالي لـ لمتوسط المجتمع ومتوسط العينة ذات الحجم 2 مأخوذة من المجتمع {1 ، 3 ، 6 ، 8}	31
2–1	رسم توضيحي لنظرية النهاية المركزية ($X\sim exp(1)$)	35
1–2	تمثيل بعض فترات الثقة	50
1–3	قيم احتمالات الأخطاء من النوع الأول والثاني	79

المقدمة

لحل مشاكل العالم الحقيقي، لا تكفي المعرفة النظرية فقط بل يجب على المرء أن يلجأ إلى التفكير المنطقي النقدي الذي يمكن على أساس إطار مفاهيمي استنتاج الحلول. يجب أن تمكن عملية التعليم والتعلم بشكل عام كل طالب من التفكير المنطقي النقدي، خاصة بالنسبة للإحصاء.

فالإحصاء هو علم يشمل عدة أساليب علمية لجمع البيانات وتنظيمها وتلخيصها وتحليلها، وكذلك لاستخلاص النتائج واتخاذ القرارات بناءا على نتائج التحليل. فالأساليب الإحصائية ضرورية لكل من الإدارة المالية، وكذلك لإدارة العمليات والمبيعات والتسويق والتحصيل واللوجستيات وإدارة شؤون الموظفين، وغيرها من الأنشطة الأخرى.

وبخلاف القيمة الجوهرية للمعرفة بموضوع معين، هناك خمسة أسباب مهمة على الأقل، بسببها نحتاج إلى دعم الإحصاء في حياتنا اليومية.

- .كأداة عمل: في جميع العلوم ، تساهم الاساليب الإحصائية في التوليف والتمثيل واستخلاص النتائج حول سلوك بيانات معينة
- .حل المشكلات: تعتبر المساهمة التي يوفرها الإحصاء ضرورية للإجابة على عدة أسئلة
- .في البحث النظري: تساعد في إنشاء نظريات تسمح بالتنبؤ بالسلوك في ظل ظروف معينة، خاصة في الظروف التي لا تخضع فيها الأحداث لقوانين مادية أو حتمية
- .استخدام البحث: في كل مجال، يساعد الاحصاء على فهم البيانات المتولدة في البحث النظري أو التطبيقي، حيث يتم إنشاء كمية كبيرة من البيانات، وهو نفس الشيء الذي يتم تحليله من خلال النظرية الإحصائية.
- .الرضا الشخصي: في البداية، يميل الطلبة إلى الاعتقاد بأن عملية جمع البيانات وتحليلها ليست ممتعة للغاية، حيث انهم يعتبرونها عملية معقدة للغاية. لكن بمساعدة التكنولوجيا، يتم جميع البيانات المنهجية والضرورية في النهاية نحصل على النتائج والاستنتاجات.

وكما هو معروف، ينقسم الإحصاء إلى فرعين أساسيين: الإحصاء الوصفي والإحصاء الاستدلالي. الإحصاء الوصفي الذي يتعامل مع جمع المعلومات وتصنيفها وتبسيطها، حيث يتم تلخيص البيانات التي في الجداول والرسوم البيانية التي تصف بشكل مناسب سلوك هذه البيانات، ، وكذلك حساب المقاييس الرقمية التي تسمح بإبراز أهم جوانبها.

الإحصاء الاستدلالي يتعامل مع عمليات التقدير والتحليل واختبارات الفرضيات، للوصول إلى الاستنتاجات التي توفر أساسا علميا مناسبا لاتخاذ القرار، مع مراعاة بيانات العينة التي تم جمعها. بمعنى آخر، يتعامل الإحصاء الاستدلالي مع التحليل وتفسير النتائج والاستنتاجات التي يمكن الوصول إليها من البيانات التي تم الحصول عليها من عينة من أجل توسيع نتائجها لتشمل المجتمع قيد الدراسة.

في هذا الفرع من الإحصاء، المقصود هو الحصول على استنتاجات عامة من مجتمع معين من خلال دراسة عينة تمثيلية مأخوذة منها. مما يعني انه مع القياسات او الاحصائيات التي سنحصل عليها، يمكننا إنشاء قيم المعلمات ثم يمكننا أن نستنتج أن هذه الإحصائيات تدرس مجتمع ما باستخدام البيانات والنتائج التي تم الحصول عليها من عينة. من الأمثلة على هذا النوع نجد تطبيق علاجات جديدة بأدوية جديدة، أو التوقعات التي يمكن أن يقدمها خبراء السوق حول كيفية تأثير الإعلان على قطاعات معينة، ...الخ.

وتجدر الاشارة الى انه في دراسة الإحصاء الاستدلالي من الضروري الإجابة على أسئلة معينة تتطلب دراستها معرفة الاحتمالات والرياضيات. على سبيل المثال: كيف يتم اختيار العينة؟ وكيف يتم الاستدلال؟ وما هي درجة الثقة ؟ للإجابة على السؤال الأول سوف نتطرق أولا الى ادراج المادة الخام للإحصاء والتي تتمثل في مجموعات الأرقام التي تم الحصول عليها عن طريق عد أو قياس عناصر لبعض الظواهر قيد الدراسة.

لهذا السبب يجب أن نولي عناية خاصة للتأكد من أن المعلومات كاملة وصحيحة. بعد ذلك يجب تحديد المعلومات والكمية التي سيكون من الضروري جمعها والتي تمثل أهم مشاكل الإحصائيين، لأنه بناءً على ذلك يتم تحديد موثوقية النتائج. من ناحية أخرى ، وللإجابة على السؤالين الثاني والثالث سنتعرف على تقنيات تساعدنا في تنفيذ الاستنتاجات التي يمكننا من خلالها تحديد درجة الثقة.

لذلك، ستتناول هذه المطبوعة في خمسة فصول أهم أساليب الاستدلال الإحصائي بطريقة واضحة وبسيطة، حتى يتمكن الطلبة من التغلب على الصعوبات التي تنشأ بشكل متكرر.

بعد ادراج مختلف المفاهيم الاساسية في الفصل التمهيدي، سننتقل في الفصل الأول لدراسة نظرية المعاينة، حيث سنناقش توزيعات المعاينة للفرق المتوسط والمتوسط للنسبة والتباين. وسنقوم بتوسيع توزيعات العينات للفروق والمجاميع، مع مراجعة لنظرية النهاية المركزية.

في الفصل الثاني سنتطرق لنظرية التقدير وسنتحدث بإيجاز عن مقدرات النقاط وأهم خصائصها، بالاضافة الى التقدير بمجال وفترات الثقة. حيث سندرج المفاهيم الأساسية حول خصائص المقدر وسنراجع الجزء المنهجي من فترات الثقة لمعلمات المجتمع، للنسبة، والتباين.

في الفصل الثالث والاخير سننتقل الى اختبار الفرضيات مع مراجعة لأهم المفاهيم الأساسية، ومنهجية اختبار الفرضيات. يشمل هطا الفصل ايضا ماهية الفرضية الإحصائية والأخطاء التي نرتكبها عند إجراء الاختبار. كما نلقي نظرة فاحصة على قوة الاختبار. في النهاية، نقوم بمراجعة الجزء المنهجي من اختبارات الفرضيات.

وفي كل فصل سيتم تبسيط المفاهيم بأمثلة قبل ان يتم تجميع مجموعة من الاسئلة ليتمكن الطالب من اجراء تقييم ذاتي لدرجة استيعابه لمختلف المفاهيم في الفصل الخامس. في هذا الفصل يستم ايضا ادراج مجموعة من التمارين حتى يتمكن الطالب من ملاحظة وتطبيق المعرفة التي يكتسبها في حل المشكلات والحالات.

مع هذه الفصول الخمسة نأمل ان نتمكن من إحداث تقدم كامل في المفاهيم الأساسية التي لها تطبيقات أكبر في المشكلات العملية لمجالات مختلفة من الإحصاء الاستدلالي في ميدان العلوم الاقتصادية والتجارية وعلوم التسيير.

الفصل التمهيدي

مفاهيم أساسية

تمهيـد

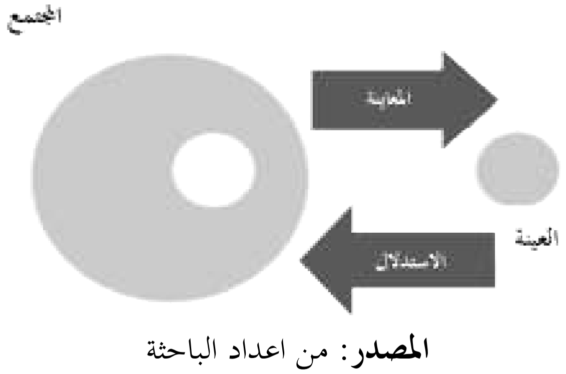
قبل الشروع في دراسة الأساليب المختلفة المدرجة في هذه المطبوعة يستحسن أن يفهم الطالب بعض المفاهيم الاساسية ذات العلاقة حتى تكون له نظرة عامة حول المقياس. وتتمثل هذه المفاهيم أساسا في مفهوم كل من الاحصاء والاحصاء الاستدلالي، المجتمع والعينة، وحجم العينة. كما يحتاج الطالب إلى بعض المفاهيم ايضا المتعلقة بالتوزيعات الاحتمالية التي سيتم استخدامها في هذه المطبوعة.

1. الاحصاء الوصفي والاستدلالي

تزايدت أهمية علم الاحصاء في العامل المعاصر نتيجة للأهمية التي تكتسيها عمليات الاستقصاء التي تقوم بها المؤسسات الإقتصادية ومنظمات الأعمال المعاصرة، وذلك لتحقيق عدة أهداف كمعرفة توجهات المستهلكين، وأذواقهم، تفضيلاتهم،.....إلخ. ويتم جمع البيانات الإحصائية المتعلقة بمشكلة أو ظاهرة معينة وفق أسلوبين، يتمثل الأسلوب الأول في الحصر الشامل والثاني في اسلوب العينة. لكن يعتبر الأسلوب الأول جد مكلف، مما يتطلب مجهودات كبيرة، وفترة زمنية. ومن الأمثلة الموجودة في الواقع العملي على الحصر الشامل هو احصاء السكان الذي تقوم به الدول.

وهنا تتجلي أهمية الاحصاء في امكانية دراسة مجتمع ما من خلال اختيار عينة عشوائية منه، عوضا عن اجراء مسح شامل للمجتمع. و يمكن القول أن n فرد من المجتمع يشكلون عينة مختارة بطريقة عشوائية. وباستخدام الاساليب الاحصائية، وتطبيقها على العينة سنتمكن من التوصل إلى نتيجة تجريبية، تساعد على رسم استدلال احصائي حول معالم المجتمع الكلي. ينقسم الإحصاء إلى فرعين أساسيين هما الإحصاء الوصفي والإحصاء الاستدلالي. حيث يهتم الإحصاء الوصفي بجمع وتنظيم وعرض البيانات في جداول أو رسومات بيانية(جلاطو جيلالي، 2002)، وبالإضافة الى حساب المقاييس التي تسمح بوصف وإبراز أهم جوانب المتغيرات.

الشكل رقم 0-1: المعاينة والاستدلال



المصدر: من اعداد الباحثة

ويستخدم علم الإحصاء لجمع وتنظيم وتفسير البيانات (دومينيك سالفاتور، 2001) التي تنتشر في مختلف المجالات كاستطلاعات الرأي، والاقتصاد، والرياضة، وبيانات الطقس، وجودة المنتج، الخ. لذا نحتاج إلى معارف أساسية في الإحصاء لتقييم كل هذه البيانات و تفسير الظواهر المختلفة.

تعريف

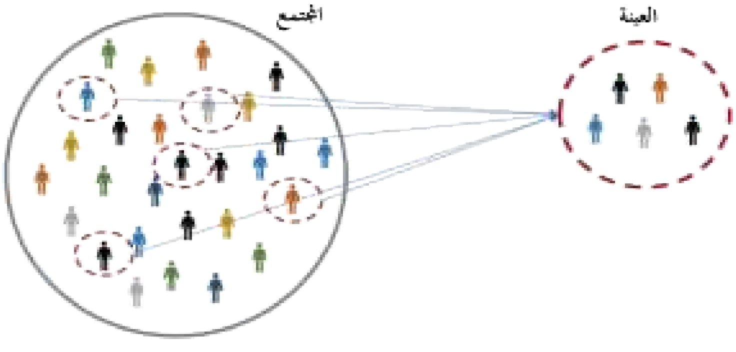
الإحصاء هو فرع من فروع الرياضيات الذي يوفر طرقا لجمع المعلومات وتنظيمها وتحليلها واستخدامها لاستخلاص استنتاجات مختلفة يمكن أن تساعد في حل المشكلات وصنع القرارات.

ويعتبر الإحصاء أساسي للعديد من فروع العلوم من الطب إلى الاقتصاد. لكن قبل كل شيء، من الضروري قراءة البيانات وتفسيرها، ومعالجتها، لاستخلاص النتائج. وتتمثل المادة الخام للإحصاء في مجموعات البيانات التي تم الحصول عليها، ونظرا لأن طبيعة الظواهر التي يمكننا تحليلها تختلف اختلافا كبيرا فمن الضروري تحديد مجموعات البيانات التي نريد تحليلها.

في حين يعتمد الإحصاء الاستدلالي على النتائج التي تم الحصول عليها من العينة لاستخلاص النتائج واتخاذ القرارات بشأن المجتمع قيد الدراسة (معتوق امحمد، 2007). وهنا سنحتاج إلى مزيد من المفاهيم الاساسية للمجتمع والعينة. لتسهيل استيعاب هذه المفاهيم سنقدم بعض الأمثلة على سبيل التوضيح.

2. المجتمع والعينة

يتعامل الاستدلال الإحصائي مع مشكلة الحصول على معلومات حول مجتمع من خلال نتائج تحليل عناصر العينة. وبما أن المجتمع الحقيقي قد يتكون من أشخاص ومؤسسات وما إلى ذلك، فإن البحث الإحصائي يسعى للحصول على معلومات حول خصائص معينة للموضوعات، والظواهر التي يمكن تحليلها. وبالتالي يرتبط كل موضوع في المجتمع بمقاييس معينة (Sedkaoui, 2018).



ومنه يمكننا القول ان المجتمع عبارة عن مجموعة من المفردات (أفراد، أعداد، ...) ذات خصائص مشتركة تدور الدراسة الإحصائية حولها (مجموعة العناصر التي تعتمد عليها الدراسة الإحصائية).

تذكير

عند الحديث عن المجتمع فلا بدمن ذكر عناصره أو ما سميناه بمفردات المجتمع، والتي يمكن وصفها بواحد أو أكثر من خاصية، وبالتالي يطلق على سمة المفردة أي خاصية يمكن من خلالها تصنيف العنصر ودراسته. فمثلا يعبر كل من الجنس، الحالة الاجتماعية، عدد الاخوة، الطول، وما الى ذلك على مجموعة من الخصائص. أما اذا كانت عناصر المجتمع عبارة عن أجهزة الكمبيوتر مثلا، فيمكن أن تكون الخصائص عبارة عن أحد المكونات، سرعة المعالج، وسعة القرص الصلب، وغيرها.

تجدر الاشارة الى أن أحد أهداف الإحصاء الاستدلالي هو الحصول على معلومات من المعلمات. على سبيل المثال:

- .نفترض أن إدارة المؤسسة التي تصنع الأجهزة تريد معرفة متوسط العمر لنموذج معين من الثلاجات. في هذه الحالة يجب مراقبة جميع الثلاجات الشيء الذي سيكلف المؤسسة الكثير. لذلك يتم إجراء استنتاج بخصوص متوسط عمر نموذج الثلاجات المذكور.
- .أيضا لنفترض أن مدير التصنيع في المؤسسة التي تصنع المصابيح الكهربائية يود معرفة متوسط عمر المصابيح التي ينتجها. كما في الحالة السابقة، إذا كان يريد معرفة متوسط عمر هذه المصابيح ، فسيتعين عليه اختبارها جميعا. ولكن ستكون هذه الدراسة شاقة للغاية ومكلفة. لذلك يتم إجراء استنتاج بخصوص متوسط عمر جميع المصابيح.
- .نفس الشيء ينطبق على الانتخابات المقبلة، حيث نود أن نتعرف على اتجاهات المرشحين لرئاسة الجمهورية. عشية الانتخابات يلعب معرفة توجهات الناخبين تجاه مرشح معين دورا مهما للغاية لأنه ليس من الممكن دراسة كل الناخبين.

ومنه نقوم ان كل عنصر من العناصر التي تدخل في تعريف المجتمع يعتبر عنصر أو مفردة من المجتمع. فمثلا إذا كانت مجموعة البيانات مكونة من جميع طلاب الجامعيين كلية العلوم الاقتصادية والتجارية وعلوم التسيير لجامعة خميس مليانة، فسيكون كل طالب عنصرا إحصائيا، بينما سيشكل جميع الطلاب المجتمع.

وغالبا ما تتكون المعلومات المتاحة للدراسة من جزء أو مجموعة فرعية من المجتمع والمعروف بالعينة، وهي مجموعة جزئية مأخوذة من المجتمع تشترك في خاصية أو خصائص معينة ويشترط أن تكون ممثلة للمجتمع. وهذه المجموعة

الجزئية (العينة) تغنينا عن دراسة كل عناصر المجتمع خاصة في حالة المجتمعات الغير محدودة والتي يصعب حصر عناصرها. ومن المثال السابق يمكن أن تتشكل العينة من جميع الطلاب في السنة الثانية علوم التسيير.

على سبيل المثال إذا كان عدد عناصر المجتمع مكون من جميع عملاء متجر تجاري كبير يقومون بإجراء تغييرات أو إرجاع أحد المنتجات، فستكون العينة عبارة عن عدد معين من العملاء يتم اختيارهم بموجب طرق واساليب المعاينة المعروفة.

ونرمز عادة لحجم المجتمع ب N، ولحجم العينة بـ n. ويمكننا استخدام احصائية (احصاءة) العينة (متوسط العينة) للاستدلال حول معلمة المجتمع (متوسط المجتمع μ). ويقصد بمعالم المجتمع مجموعة من الخصائص مثل: المتوسط μ ،الانحراف المعياري σ ،إلخ. فيما يتعلق بالخصائص الاكثر شيوعا فيمكننا تلخيصها في الحجول الموالي:

الخاصية	المعلمة (المجتمع)	الاحصائية (العينة)
المتوسط الحسابي	μ	m_i
التباين	σ^2	σ_m^2
الحجم	N	n

ملاحظة

كل من المعلمة والإحصائية هي مقاييس موجزة، والفرق هو أن أحدهما يستخدم جميع عناصر المجتمع بينما يستخدم الآخر عناصر العينة.

ومن خصائص المجتمع أيضا طبيعة توزيعه الاحتمالي كالتوزيع الطبيعي مثلا. أما الاحصائية فهي متغير عشوائي أنشئ انطلاقا من مشاهدات العينة وصمم لتقدير معلمة محل الاهتمام. وعليه يمكن القول أن المعلمة تعبر عن خاصية من خصائص المجتمع، بينما تمثل الاحصائية خاصية من خصائص العينة.

3. حجم العينة

يحتاج الباحثون والطلاب إلى معرفة الحجم المثالي للعينة لإجراء الدراسة الميدانية للبحث الذي يقومون به. وتجدر الاشارة الى أن عملية تحديد حجم العينة للدراسات تعتبر من بين الصعوبات التي يواجهها الباحثين، وذلك لعدم الإلمام بالصيغ الرياضية المستخدمة في الاختيار.

وبالرغم من صعوبة تحديد حجم العينة بشكل دقيق الا انه توجد مجموعة من التقنيات التي تساعد الى حد ما في تحديد حجم العينة. ولكن عند تحديد الحجم المناسب للعينة، يجب أن يؤخذ في عين الاعتبار ما يلي:

- التمثيل: يجب أن يكون لكل عنصر من عناصر المجتمع نفس فرصة الظهور لتكوين العينة.
- كافية وصالحة: يجب أن يكون خطأ العينة هو الحد الأدنى الممكن فيما يتعلق بالمجتمع.
- المصداقية: يجب الحصول على حجم العينة من خلال بعض العمليات الحسابية التي تتفادى حدوث الخطأ.

يمكننا أن نثبت أن حساب حجم العينة هو أحد الجوانب الرئيسية في المرحلة الأولية لأي بحث علمي، لأنه بهذا يمكن تحديد درجة المصداقية التي يمكننا تخصيصها لنتائج البحث. علاوة على ذلك، من خلال اختيار حجم عينة جيد وتقنية مناسبة لأخذ العينات، فاننا نقوم بشكل ضمني بجمع المعلومات التي تلبي الخصائص المذكورة أعلاه للعينة: التمثيل، الصلاحية والموثوقة.

كما أن حل مشكلة الحجم ليس بالبساطة التي قد يتخيلها البعض، فهذا يتطلب معرفة مسبقة بالموضوعات الإحصائية التي سنقوم بدراستها لاحقا، كفترات الثقة واختبار الفرضيات، بالإضافة إلى أهداف الدراسة وخصائص المجتمع المعني.

في هذا القسم نعرض الصيغ الأساسية لتحديد حجم العينة[1]، مع التأكيد على أن المشكلة لن تحل بهذه الصيغ، لأنه يجب أن تحتوي أي معادلة لحساب حجم العينة على ثلاثة عوامل:

أ. نسبة الثقة التي تريد تعميم بيانات العينة بها على المجتمع

ب. نسبة الخطأ المسموح بها لقبول التعميم

ج. مستوى التباين أو الاحتمال الذي تحدث به ظاهرة الدراسة، سيتم الإشارة إلى القيم بـ p

1.3. الحالة الأولى: عندما يكون N غير معروف أو عدد عناصره غير محدودة

عندما لا نعرف حجم المجتمع ، يمكن حساب حجم العينة لتقدير المتوسط باستخدام الصيغة التالية:

$$n \geq \frac{p(1-p)Z_{1-\alpha}^2}{\varepsilon^2}$$

[1] هناك عدة مواقع الكترونية تقوم بحساب حجم العينة الأمثل نذكر منها:
http://www.raosoft.com/samplesize.html ؛ /https://www.surveymonkey.com/mp/sample-size-calculator

حيث: n هو حجم العينة؛ $Z_{1-\alpha}$ هي القيمة الجدولية للتوزيع الطبيعي لاحتمال مركزي قدره $1-\alpha$ ؛ ε^2 هو الخطأ المسموح به في أخذ العينات؛ و p هو التباين الإيجابي (الموثوقية).

2.3. الحالة الثانية: عندما يكون حجم المجتمع معروفا N

عندما نعرف حجم المجتمع، يتم حساب حجم العينة لتقدير المتوسط باستخدام الصيغة التالية:

$$n \geq \frac{Np(1-p)Z_{1-\alpha}^2}{(N-1)\varepsilon^2 + p(1-p)Z_{1-\alpha}^2}$$

حيث: n هو حجم العينة ؛ N هو حجم السكان ؛ $Z_{1-\alpha}$ هي القيمة الجدولية للتوزيع الطبيعي لاحتمال مركزي قدره $1-\alpha$ ؛ ε^2 هو الخطأ المسموح به في أخذ العينات ، p هو التباين الإيجابي. يتم الحصول على هذه الصيغة من:

$$\bar{x} - Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{x} + Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

الحالة الثالثة: عندما تكون البيانات نوعية

في الظواهر الاجتماعية التي يتم فيها استخدام المقياس الاسمي للتحقق من غياب أو وجود الخاصية (او الخصائص) المراد دراستها، يتم حساب حجم العينة لتقدير النسبة باستخدام الصيغة الموالية:

$$n \geq \frac{p^*}{1 + p^*/N}$$

حيث: n هو حجم العينة؛ N هو حجم المجتمع؛ و $p^* = \frac{p(1-p)}{(se)^2}$ خطأ معياري و p درجة الموثوقية.

كما ان هناك عدد من الصيغ الرياضية والتي تساعد على حساب حجم العينة. والتي يمكن تلخيصها في الجدول الموالي:

الجدول رقم 0-1: أهم الصيغ الرياضية لحساب حجم العينة

المعادلة	الصيغة
معادلة ريتشارد جيجر	$\frac{\left[\frac{z}{d}\right]^2 * (0.50)}{1+\frac{1}{N}\left[\left[\frac{z}{d}\right]^2 * (0.50)^2 - 1\right]}$

معادلة ستيفن ثامبسون	$\frac{N*P(1-P)}{[[(N-1)*(d^2/z^2)]+(1-P)]}$
معادلة روبرت ماسون	$\frac{M}{[(S^2*(M-1))/P(1-P)]+1}$
معادلة هيربرت اركن	$\frac{P(1-P)}{(SE/t)+[P(1-P)/N]}$

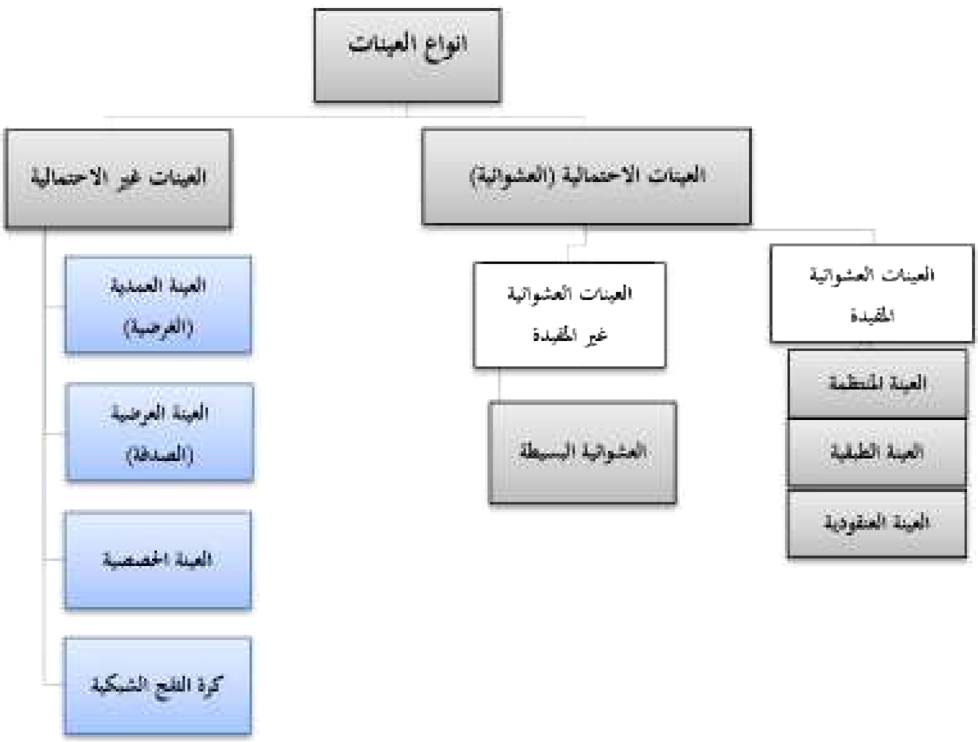
المصدر: من اعداد الباحثة

حيث p تمثل الصفة المتوفرة في بعض مفردات المجتمع، Z تمثل القيمة المعيارية للتوزيع الطبيعي عند فترة ثقة 95% وهي تساوي 1.96 (انظر جدول التوزيع الطبيعي في قائمة الملاحق)، SE تمثل تباين المجتمع، d تمثل الخطأ المسموح به، وعادة ما يتم اختياره 0.05، و يمثل حجم مجتمع الدراسة.

4. أنواع العينات

تنقسم العينات الى نوعين: عينات عشوائية (احتمالية) وعينات غير عشوائية (غير احتمالية)، يمكن تلخيصها في الشكل الموالي:

الشكل رقم 0-2: انواع العينات



المصدر: من اعداد الباحثة

يتمثل الاختلاف الأساسي بين هذه الأنواع في أنه في العينات الاحتمالية، يمكن قياس المخاطر التي تفترضها عملية المعاينة[1]، بينما لا يكون ذلك ممكنا في العينات غير الاحتمالية.

كما أن العشوائية في عملية المعاينة تعني أن لكل عناصر العينة نفس فرصة الظهور أو الاختيار(رويسات عبد الناصر، 2006)، كما لا نكون على علم مسبق بالنتائج التي سيتم التوصل اليها.

ويتوفر أسلوب المعاينة العشوائية على العديد من المزايا نلخص اهمها في:

- **حذف التحيز**: هذا الأسلوب يؤكد أن عناصر العينة تم اختيارها بدون تحيز، وعلى الرغم من أن الاختيار العشوائي لا يضمن بأن تكون العينة ممثلة للمجتمع، إلا أنه يحذف مخاطر الاختيار المتحيز.
- **تحديد الثقة**: أسلوب العينات العشوائية يضع الأساس الإحصائي لتحديد الثقة المقترنة بالاستنتاج الإحصائي ،والذي لا يمكن تنفيذه إذا تم اختيار عناصر العينة بطريقة أخرى بغير الطريقة العشوائية.
- **التحكم في خطأ المعاينة** : يسمح هذا الاسلوب بالتحكم في خطأ المعاينة من خلال اختيار حجم العينة، ولكن مع الطرق غير العشوائية فإنه لا يمكن أن يتحقق مستوى مقبول من خطأ المعاينة.

وتتمثل أساليب المعاينة العشوائية في:

1.4. العينة العشوائية البسيطة

حيث تؤخذ العينة بشكل يعطي لأي عنصر من عناصر المجتمع نفس الفرصة لأن يكون ضمن العينة ،وتستخدم في حالة المجتمعات المتجانسة والمحدودة (Labatte، 2012)، ويتم اختيار أفراد العينة بكتابة أرقام أفراد المجتمع على بطاقات متشابهة وخلطها بشكل جيد ثم تتم عملية السحب عن طريق القرعة، أو باستخدام جداول الأرقام العشوائية أو باستخدام الحاسب الآلي (تطبيق Excel مثلا).

ملاحظة

العينة العشوائية البسيطة هي اختيار عناصر العينة عشوائيا بحيث أن كل عناصر المجتمع لها نفس الاحتمال في أن تكون منتقاة في العينة. ونميز بين نوعين من المعاينة العشوائية باختلاف طريقة السحب: المعاينة بالإرجاع (عينة غير نفادية) والمعاينة بدون إرجاع (عينة نفادية).

- **العينة الغير نفادية**: عندما يكون السحب بالإرجاع حيث يمكن أن تظهر المفردة أكثر من مرة في العينة، وتسمى هذه العينة غير نفادية لان تكرار العملية لا يؤدي إلى تقليص عدد المفردات في المجتمع. وفي هذه الحالة نستخدم القانون التالي لإيجاد عدد العينات الممكنة: N^n

[1] المعاينة هي الطريقة أو العملية التي تسمح بانتقاء مجموعة فرعية من المجتمع.، وسوف نتعمق في نظرية المعاينة في الفصل الأول لهذه المطبوعة.

- **العينة النفادية**: تسمى المعاينة بدون ارجاع معاينة نفادية حيث لا يمكن أن تظهر المفردة في العينة أكثر من مرة واحدة. وفي هذه الحالة نستخدم القانون التالي لإيجاد عدد العينات الممكنة:

$$C_N^n = \frac{N!}{(N-n)!\,n!}$$

إلا أنه في بعض الحالات التي يكون فيها حجم المجتمع غير منته، يصعب تكوين العينات العشوائية بالطريقة السابقة، لذلك يلجأ الباحثون إلى استخدام ما يعرف بالأرقام العشوائية أو الجداول العشوائية (أنظر الملحق).

مثال 0–1

من مجتمع متكون من طلبة السنة الثانية علوم التسيير، اخترنا عشوائيً عينة من 10 طلبة للتحليل ظاهرة معينة. لاحترام العشوائي ، يمكننا اجراء عملية المعاينة بطرف مختلفة ، والأكثر شيوعً هو تخصيص رقم مختلف لكل طالب، ثم بمساعدة جدول الأرقام العشوائية أو برنامج Excel، نختار 10 أرقام عشوائية ونشرع في إجراء المقابلات مع الطلبة المختارين.

لنفرض أننا أحصينا جميع الطلاب في المجتمع، وكانت النتيجة 350 طالب، ثم نقوم بإعطاء ارقام لكل طالب من 0 ، 1 ، 2 ، حتى 350. باستخدام جداول الأرقام العشوائية أو برنامج Excel. تم إنشاء 10 أرقام بين 0 و350، لنفترض أن النتائج شملت الطلبة اصحاب الأرقام التالية:

45، 78، 92، 184، 197، 236، 248، 269، 275، و 291

أي أننا اخترنا الطلاب العشرة باستخدام أسلوب المعاينة العشوائية البسيطة.

على الرغم من أن تعريف العينة العشوائية يبدو بسيط، الا انه يمكننا أن نلاحظ أنه يتضمن عدة مفاهيم من نظرية الاحتمالات. بهذا المعنى فإن المتغيرات X_1، X2، ...، X_n تشكل عينة عشوائية عندما تكون[1]:

- المتغيرات X_1، X2، ...، X_n مستقلة.
- المتغيرات X_1، X2، ...، X_n لها نفس التوزيع.

[1] في الإحصاء تعرف المتغيرات العشوائية X_1، X2، ...، X_n التي تشكل عينة عشوائية في شكل مبسط على انها متغيرات مستقلة وموزعة بشكل متماثل.

2.4. العينات الطبقية

عندما يكون لدينا مجموعة يمكن تقسيمها إلى عدة مجموعات فرعية نسميها طبقات، وفقًا لخصائص معينة يجب، فإننا نفكر في نوع طبقي من العينات. وتستخدم هذه الطريقة في الحالات التي تكون فيها نتائج البحث تعتمد على تحليل المتغيرات المفسرة كالعمر، الجنس، ... الخ.

ففي هذه الحالة يتم تقسيم المجتمع إلى مجموعات جزئية اعتمادا على هذه الخصائص وتسمى بالطبقات، ثم باستخدام طريقة العينة العشوائية يتم اختيار عينة جزئية يتناسب حجمها مع حجم الطبقة، وتشكل مجموعات العينات الجزئية المختارة ما يعرف بالعينة الطبقية. يوصى بهذا النوع من العينات عندما تريد أن يكون لديك ممثلين عن كل مجموعة فرعية في العينة.

مثال 0-2

نفترض أننا بصدد تحديد عينة بحجم 2% من إجمالي عدد طلاب كلية العلوم الاقتصادية والتجارية وعلوم التسيير لجامعة خميس مليانة ، والتي تضم 12500 طالبا. يجب أن تستوفي العينة شرط وجود ممثل واحد على الأقل من كل من الأقسام التي يتم تشملها هذه الكلية كما يلي: العلوم الاقتصادية (4200)، علوم التسيير (3250)، العلوم التجارية (850)، العلوم المالية والمحاسبة (1700)، بالاضافة الى طلبة الماستر لكل الأقسام (2500).

من خلال بيانات هذا المثال، يمكننا إجراء المعاينة باستخدام اسلوب العينات الطبقية، حيث يبلغ حجم العينة 250 طالب (2% من 12500). يتم الحصول على حجم العينة لكل طبقة بهذه الطريقة:

العلوم الاقتصادية	$\frac{4200}{12500} \approx 0.336 \ \rightarrow n_1 = 0.336 * 250 = 84$
علوم التسيير	$\frac{3250}{12500} \approx 0.26 \ \rightarrow n_2 = 0.26 * 250 = 65$
العلوم التجارية	$\frac{850}{12500} \approx 0.068 \ \rightarrow n_3 = 0.068 * 250 = 17$
العلوم المالية والمحاسبة	$\frac{1700}{12500} \approx 0.136 \ \rightarrow n_4 = 0.136 * 250 = 34$
طلبة الماستر لكل الأقسام	$\frac{2500}{12500} \approx 0.200 \ \rightarrow n_5 = 0.200 * 250 = 50$

ومنه فحجم العينة يساوي: $n = n_1 + n_2 + n_3 + n_4 + n_5 =$

$$84 + 65 + 17 + 34 + 50 = 250$$
$$n = 250$$

3.4. العينة العشوائية المنتظمة

يتم اختيارها من خلال تحديد مجتمع الدراسة ووضع أفراده في قائمة بشكل عشوائي وإعطاء كل منهم رقما، ثم يتم تحديد قاعدة الاختيار وفق قسمة حجم المجتمع على حجم العينة من أجل الحصول على طول الفترة. وبعد ذلك يتم انتقاء أحد الأرقام عشوائيا من بين الأرقام التي تساوي أو أقل من طول الفترة، ليتم اعتباره كعنصر أول من مفردات العينة ويشرع في إضافة طول الفترة لها للحصول على العنصر الثاني، وهكذا إلى غاية الوصول إلى العدد الممثل لحجم العينة المطلوب (Sedkaoui, 2018).

بطريقة اخرى يمكننا القول أنه يتم سحب العينة المنتظمة من عناصر مجتمع يكون مرتبا بشكل تصاعدي أو تنازلي، ويتم اختيار عناصر هذه العينة بعد تحديد:

- . فترة السحب أو الزيادة المنتظمة (j) والتي تمثل المسافة الفاصلة بين كل اختيار وآخر، ويمكن حسابها بهذه الصيغة:

$$j = \frac{N}{n}$$

حيث N تمثل حجم المجتمع الكلي و n تمثل حجم العينة المراد تكوينها

- .نقطة البداية، والتي تشير إلى النقطة التي تنطلق أو تبدء بها العينة، وهي قيمة نختارها عشوائية حيث تكون محصورة بين 1 وطول فترة السحب.

وبهذه الطريقة نكون قد حصلنا على متتالية حسابية حدها الأول هو الرقم المختار عشوائيا في البداية، أساسها طول الفترة، وعدد حدودها هو حجم العينة .

مثال 0-3

نريد اختيار عينة عشوائية حجمها 10 وفق طريقة العينة العشوائية المنتظمة من مجتمع يضم 350 طالب. باعتبار أن القائمة مرتبة عشوائيا، فنحسب طول الفترة كما يلي:

طول الفترة = 10/350 = 35

يتم اختيار الرقم الأول عشوائيا على أن يكون أقل من أو يساوي العدد 35. ونفرض أن العدد الأول من الصف هو 12. نشرع في إضافة 35 في كل مرة ليتم الحصول على العينة التالية والتي نفرض أنها تشكل الطلبة ذوي الارقام: 12، 41، 73، 90، 123، 150، 181، 212، 240، 275

4.4. العينة العنقودية

هذا النوع من العينات مشابه للعينة الطبقية، حيث يتم تقسيم المجتمع إلى مجموعات فرعية (طبقات) ، ولكن على عكس الطبقية، فإنه لا يتطلب ممثل لكل طبقة في العينة، لأنه في المقام الأول نختار عينة من الطبقات كما انه يتم اختيار عينة من كل منها لتشكيل العينة المطلوبة.

يستخدم هذا الاسلوب في مجموعات كبيرة للغاية، وعلى عكس المذكورة سابقا، فإنه لا يتطلب إطار لأخذ العينات ، كما انه يتم استخدام هذا الاسلوب عندما تكون الوحدات متفرقة جغرافيا على نطاق واسع. أما اختيار عناصر العينة المراد دراستها فإنه يتم بنفس الطريقة المنتهجة في أسلوب العينة العشوائية البسيطة في حالة المجموعة الواحدة، أو العينة العشوائية الطبقية في حالة مجموعتين أو أكثر.

مثال 0–4

لنفترض أننا نريد إجراء مسح لمستخدمي مترو الجزائر العاصمة (حوالي 100 ألف مستخدم يوميا حسب الاحصائيات المتوفرة). نظرا لأن مجتمع الدراسة كبير جدا، يمكننا تقسيمه إلى طبقات. على سبيل المثال، محطات مترو الانفاق. بعد ذلك، نختار عينة من المحطات وننتقل إلى إجراء مسح للمستخدمين في المحطات المختارة (يمكن أن يكون ذلك بأخذ عينات منهجية). هذا الشكل من أخذ العينات يقلل بشكل كبير من التكلفة لأنه ليس من الضروري الترقيم المسبق لوحدات المجتمع.

وبالرغم من اهمية الأساليب العشوائية للمعاينة، فإننا نجد ان بعض البحوث تتجه الى اختيار أساليب غير عشوائية خاصة في البحوث النوعية (Sedkaoui, 2018)، حيث يسعون إلى التركيز على جوانب أخرى في عينات الدراسة، أهمها:

- .غزارة البيانات والمعلومات المتوفرة عند أفراد العينة
- .قربهم من الأحداث والموضوعات المعنية بالدراسة
- .استعداد افراد العينة لإعطاء المعلومات

فأساليب المعاينة غير العشوائية تقوم على مبدأ عدم تحكم الباحث في اختيار عناصر العينة، وعدم معرفة عناصر المجتمع، وهذا ما يؤدي إلى عدم تساوي الفرصة لعناصر المجتمع لتكون ضمن العينة، وبالتالي عدم تمثيل المجتمع، وتعميم النتائج المتوصل إليها من خلال نتائج تحليل العينة.

والمعاينة غير الاحتمالية تتضمن مجموعة من الأساليب موضحة في الشكل السابق وتتمثل في:

– **العينة العرضية أو عينة الصدفة:** حيث لا يخضع اختيار الباحث لعناصر العينة لأي اعتبار وإنما يكون على سبيل المصادفة، كأن يشرع في توزيع الاستبيان على مجموعة من الطلبة وذلك بتسليم الاستمارة للطلبة الذين يصادفهم أثناء زياراته الميدانية .

– **أسلوب المعاينة العمدية:** وهي عينة يتم اختيار عناصرها بشكل مقصود ومستهدف لتوفر بعض الخصائص في هذه العناصر بما يخدم أهداف الدراسة، ويلجأ الباحث عادة إلى هذا الاسلوب من المعاينة عند توفر البيانات اللازمة للدراسة لدى فئة محددة من المجتمع الأصلي للدراسة. فمثلا اذا اراد الباحث دراسة سلوك المستهلكين لمنتج معين فعلى الباحث هنا الاخذ بعين الاعتبار المستهلكين الحقيقيين لهذا المنتج وتجنب جمع البيانات من أشخاص لا يستهلكون هذا المنتج.

– **أسلوب المعاينة الحصصية:** سميت بالحصصية لأن مجتمع البحث يقسم إلى فئات طبقا لخصائصها الاساسية، وتمثل كل فئة بنسبة وجودها، ويتجلى الفرق بين هذا الاسلوب واسلوب المعاينة الطبقية في كون أن في العينة الطبقية عملية اختيار عناصر العينة يتم بشكل عشوائي فضلا على عدم محدودية عناصر المجتمع في حالة المعاينة الحصصية .

– **كرة الثلج (الشبكية):** أخذ عينات كرة الثلج هي عملية المعاينة غير الاحتمالية التي يبدأ فيها الباحث بمجموعة صغيرة من الأفراد المعروفين ويوسع العينة عن طريق سؤال هؤلاء المشاركين الأوليين لتحديد الآخرين الذين يجب أن يشاركوا في الدراسة. وبعبارة أخرى، تبدأ العينة صغيرة ولكن "كرات الثلج" في عينة تكبر خلال مسار البحث. اسلوب المعاينة من كرة الثلج هو أسلوب شائع خاصة بالنسبة للباحثين الراغبين في العمل مع مجموعة يصعب تحديدها أو تحديد موقعها.

وعلى العموم عند اللجوء إلى أساليب المعاينة غير العشوائية لا بد من توخي الحذر عند تفسير النتائج. ولا شك في أن نتائج الدراسة المتوصل إليها وفق طرق العينات العشوائية تكون أكثر مصداقية في مختلف الدراسات الإحصائية لخلوها من عنصر التحيز.

5. التوزيعات الاحتمالية

سيسهل مقياس الإحصاء 03 على الطالب تعميم النتائج انطلاقا من العينة المدروسة، ويستطيع في نفس الوقت أن يختبر صحة الفرضيات حول المجتمع المدروس. لذا فقبل تناول توزيعات المعاينة والانتقال إلى نظرية التقدير والتطرق الى كيفية اختبار الفرضيات، نحتاج إلى التعرف او تذكر بعض التوزيعات الاحتمالية المستمرة الشهيرة. وسنلخص فما يلي بعض التوزيعات الاحتمالية الشهيرة، والكثيرة الاستخدام في الدراسات حتى يتمكن الطالب من فهم استخدامها في هذا المقياس.

1.5. التوزيع الطبيعي

يعتبر هذا التوزيع من أشهر التوزيعات الاحتمالية مع وجود عدد كبير من الظواهر العشوائية المستمرة التي تتبع هذا التوزيع (رويسات عبد الناصر، 2006). ويمكن تعريف دالة الكثافة الاحتمالية لهذا التوزيع كما يلي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad x \in]-\infty, +\infty[$$

حيث أن μ ، σ يمثلان على التوالي المتوسط والانحراف المعياري لهذا التوزيع،

ونكتب: $X \sim N(\mu, \sigma^2)$

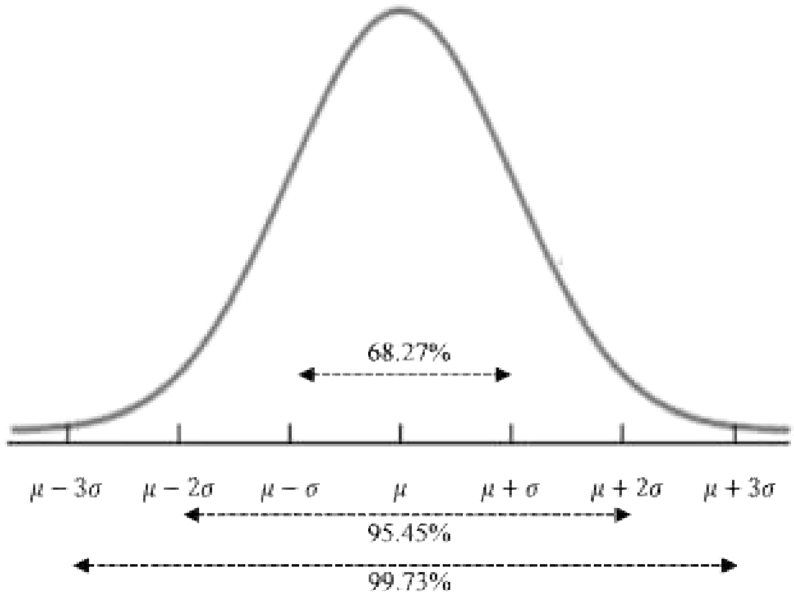
تذكير

المميزات العددية لهذا المتغير العشوائي هي:

$$E(X) = \mu$$
$$V(X) = \sigma^2$$

ويمكن تمثيل الشكل البياني لهذا التوزيع كما يلي:

الشكل رقم 0–3: التمثيل البياني لدالة كثافة التوزيع الطبيعي



ونلاحظ ان شكل التوزيع يأخذ شكل الجرس، كما انه متناظر، ويكون محور تناظره هو المتوسط μ. وتجدر الاشارة الى ان:

- .مساحة هذا المنحنى تساوي الواحد لأنها دالة كثافة احتمالية، كما ان نقاط الانعطاف الأولى لهذه الدالة تساوي $\mu \mp \sigma$ ؛
- .وتبلغ المساحة الممتدة بين هاتين النقطتين 68.27%، أما النقطتين $\mu \mp 2\sigma$ فإن المساحة بينهما تساوي 95.45%؛
- .في حين تساوي المساحة بين $\mu \mp 3\sigma$ مقدار 99.73%.

مثال 0–5

احسب احتمال أن يأخذ المتغير العشوائي X قيمة أقل أو تساوي a.

الحل

هنا نحسب التكامل التالي:

$$P(X \le a) = \Phi(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

حيث: $\Phi(a)$ تمثل دالة التوزيع للمتغير العشوائي X

لحساب هذا التكامل نحتاج الى وقت كثير، من أجل ذلك قام الاحصائيون بوضع جداول تحتوي على هذه الاحتمالات (الموضحة في الملاحق)، لذا سنقوم بما يلي:

نضع : $Z = \frac{X-\mu}{\sigma}$

هذا يعني أننا سنتحصل على متغيرة عشوائية جديدة Z تتبع التوزيع الطبيعي المعياري $Z \sim N(0,1)$

مثال 0-6

لتكن المتغيرة العشوائية X التي تتبع التوزيع الطبيعي ، حيث: $X \sim N(9,16)$

احسب الاحتمال التالي: $P(X \leq 14.04)$

لحساب الاحتمال سنقوم بتحويل X من التوزيع الطبيعي إلى التوزيع الطبيعي المعياري ، كما يلي:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-9}{4}$$

حيث أن: $\sigma = \sqrt{16} = 4$ و $\mu = 9$

$$P(X \leq 14.04) = P\left(\frac{X-9}{4} \leq \frac{14.04-9}{4}\right)$$
$$= P(Z \leq 1.26)$$
$$= \Phi(1.26)$$

بعد حساب الاحتمال نقوم الان لاستخدام جدول التوزيع الطبيعي المعياري من خلال البحث عن القيمة 1.26، حيث تكتب هذه القيمة كما يلي:

$$0.06 + 1.2 = 1.26$$

وتود الكتابة السابقة لكون جدول التوزيع الطبيعي المعياري يتكون من أسطر وأعمدة، العمود الأول يحتوي على رقم واحد بعد الفاصلة، والسطر الأول يحتوي على رقمين بعد الفاصلة.

في مثالنا هذا سنبحث في العمود الأول على القيمة 1.2 وفي السطر نبحث عن القيمة 0.06

بعد ذلك نقوم بتحديد نقطة التقاطع بين القيميتين والتي ستعطي لنا الاحتمال المساوي للقيمة: 0.8962 كما يتضح من الشكل الموالي:

X	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319

وهذا يعني ان:

$$P(X \leq 14.04) = 0.8962$$

ملاحظة

لا يمكن استخدام جدول التوزيع الطبيعي المعياري في حالة حساب احتمال أن يكون Z أكبر من قيمة موجبة، أو أن يكون أكبر أو اقل من قيمة سالبة إلا بعد الاعتماد على العلاقات التالية:

لنفرض أن a و b عددين موجبين، لحساب الاحتمالات التالية نستعمل العلاقات الموالية والتي تم الوصول إلى هذه النتائج باستخدام خاصية التناظر التي يتميز بها التوزيع الطبيعي:

$$\mathrm{P}(\mathrm{Z} > \mathrm{a}) = 1 - \mathrm{P}(\mathrm{Z} \leq \mathrm{a}) = 1 - \Phi(\mathrm{a})$$

$$\mathrm{P}(\mathrm{Z} \leq -\mathrm{a}) = 1 - \mathrm{P}(\mathrm{Z} \leq \mathrm{a}) = 1 - \Phi(\mathrm{a})$$

$$\mathrm{P}(\mathrm{Z} > -\mathrm{a}) = \mathrm{P}(\mathrm{Z} \leq \mathrm{a}) = \Phi(\mathrm{a})$$

$$\mathrm{P}(\mathrm{a} \leq \mathrm{Z} \leq \mathrm{b}) = \mathrm{P}(\mathrm{Z} \leq \mathrm{b}) - \mathrm{P}(\mathrm{Z} \leq \mathrm{a}) = \Phi(\mathrm{b}) - \Phi(\mathrm{a})$$

$$\mathrm{P}(-\mathrm{a} \leq \mathrm{Z} \leq \mathrm{a}) = 2.\Phi(\mathrm{a}) - 1$$

مثال 0–7

لتكن X المتغيرة العشوائية التي تتبع التوزيع الطبيعي كما يلي:

$$X \sim N(145,144)$$

أحسب الاحتمالات التالية:

$$P(X \geq 124)\ ,\ P(X \leq 114)\ ,\ P(X > 179)\ ,\ P(X \leq 149)\ ,\ P(130 \leq X \leq 164)$$

الحل

لدينا: $\mu = 145$ و $\sigma = \sqrt{144} = 12$

نضع:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-145}{12}$$

$$P(X \leq 149) = P\left(\frac{X-145}{12} \leq \frac{149-145}{12}\right)$$
$$= P(Z \leq 0.33)$$
$$= \Phi(0.33)$$

نبحث عن هذه القيمة في جدول التوزيع الطبيعي بنفس الطريقة السابقة ونجد أن الاحتمال يساوي:

$$P(X \leq 149) = 0.6293$$

$$\mathrm{P}(\mathrm{X} > 179) = 1 - \mathrm{P}(\mathrm{X} \leq 179) = 1 - P(Z \leq \frac{179-145}{12})$$

$$\mathrm{P}(\mathrm{X} > 179) = 1 - \Phi(2.83) = 1 - 0.9977 = 0.0023$$

$$P(X \leq 114) = P\left(Z \leq \frac{114-145}{12}\right) = \Phi(-2.58)$$

$$P(X \leq 114) = 1 - \Phi(2.58) = 1 - 0.9951 = 0.005$$

$$P(X \geq 124) = P\left(Z \geq \frac{124-145}{12}\right) = P(Z \geq -1.75)$$

$$P(X \geq 124) = P(Z \leq 1.75) = \Phi(1.75) = 0.9599$$

$$P(130 \leq X \leq 164) = P\left(\frac{130-145}{12} \leq Z \leq \frac{164-145}{12}\right)$$

$$P(130 \leq X \leq 164) = P(-1.25 \leq Z \leq 1.58) = \Phi(1.58) - \Phi(1.25)$$
$$= 0.9429 - 0.8944 = 0.0485$$

2.5. توزيع مربع كاي Khi Deux

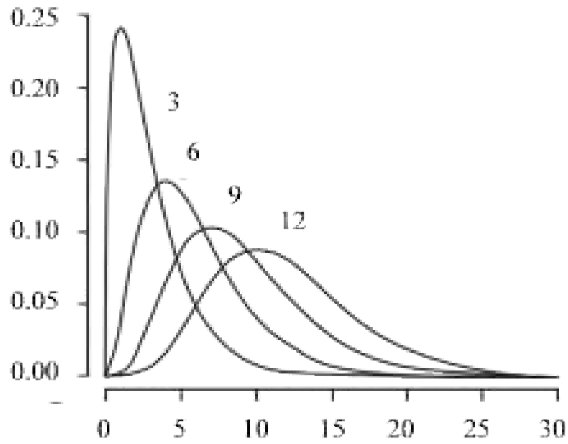
يتم تحديد دالة الكثافة لمتغير عشوائي مستمر مع توزيع مربع كاي (Khi Deux) والمعلمة n من خلال:

$$f(x,v) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\Gamma\dfrac{v}{2}}\left[x^{\frac{v}{2}-1}\, e^{-\frac{x}{2}}\right] & x \geq 0 \\ 0 & x < 0 \end{cases}$$

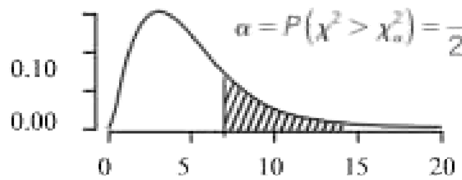
تعرف المعلمة n بدرجات الحرية ، وترتبط ارتباطا وثيقا بحجم العينة. يتم تمثيل هذا التوزيع بـ χ^2 (مربع كاي).

يوضح الشكل الموالي بعض الرسوم البيانية لمربع كاي لقيم مختلفة لـ n.

الشكل رقم 0–4: تمثيل دالة كثافة الاحتمال لتوزيع χ^2 مع 3 و 6 و 9 و 12 درجة من الحرية



ويمكن استخدام جدول توزيع مربع كاي لحساب قيم التوزيع لبعض الاحتمالات نعتمد على العرض الموضح في الشكل الموالي:



$$\alpha = P\left(\chi^2 > \chi^2_\alpha\right) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_{\chi^2_\alpha}^{\infty} \left(\chi^2\right)^{(\nu/2)-1} e^{-\chi^2/2} d\chi$$

n	α										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.94	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19

يوضح الجدول قيم توزيع مربع كاي التي تساوي بها المنطقة اليمنى أسفل المنحنى. بمعنى إذا كان المتغير العشوائي X له توزيع مربع كاي مع n درجة من الحرية ، إذن:

$$P(X_n > k) = \alpha = 1 - P(X_n \leq k) = 1 - F_{X_{n(k)}}$$

يشير إلى احتمال أن تكون X (مع درجات n من الحرية) أكبر من القيمة k و F دالة التوزيع التراكمي.

ويمكن استخدام الجدول على النحو التالي: يظهر العمود الأول درجات حرية المتغير. بعد ذلك يتم تشكيل أزواج من الأعمدة التي توضح قيمة الاحتمال في الأعلى.

مثال 0-8

أوجد القيمة المقابلة للاحتمال المشار إليه لتوزيع مربع كاي.

$$P(X_8 > k) = 0.99$$

الحل

الاحتمال ذو طرف أيمن لذلك يتم البحث عن القيمة عند $\alpha = 0.99$ في الجدول. ثم في عمود درجات الحرية يوجد 8 ويكون تقاطع الصف والعمود هو القيمة المطلوبة.

n	α										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.94	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59

3.5. توزيع[1] t–Student

اذا كان: $Z \sim N(1,0)$ و $Y \sim \mathcal{X}2(n)$

فإن المتغير العشوائية X والناتج من العلاقة التالية: $X = \frac{Z}{\sqrt{Y/n}} \sim t(n)$

يتبع توزيع آخر هو توزيع (t–Student) بدرجة حرية n، وتأخذ دالة كثافة هذا التوزيع الشكل التالي:

$$f(x) = \frac{1}{\sqrt{\pi . n}} . \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} . \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

حيث $x \in \mathcal{R}$

[1] تم نشر توزيع الاحتمالات هذا لأول مرة في عام 1908 من قبل الأيرلندي W. S. Gosset في ذلك الوقت كان يعمل في مصنع أيرلندي رفض نشر الأوراق البحثية ، لذلك نشر Gosset عمله. بالاسم المستعار "Student". لهذا السبب ، تم تسمية هذا التوزيع باسم t–Student.

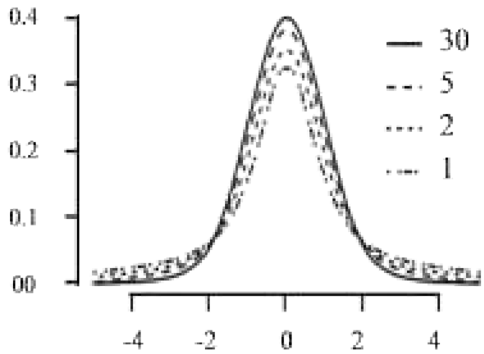
تذكير

المميزات العددية لـ t-student:

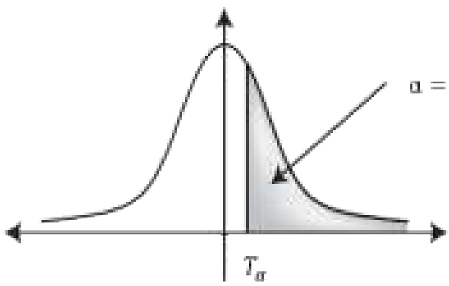
$$E(X) = 0$$
$$V(X) = \frac{n}{n-2} \; ; n > 2$$

وبنفس الطريقة نقوم فيماي يلي بتوضيح التمثيل البياني لهذا التوزيع:

الشكل رقم 0-5: دالة الكثافة t-Student ، عند 1، 2، 5 ، و30 درجة حرية



يستخدم جدول توزيع $t\text{-}Student$ لحساب قيم المتغير لاحتمالات معينة، ولديها العرض التقديمي الموضح في الشكل الموالي:



كما يتضح ، تيظهر الجدول قيم توزيع t-Student للاحتمالات المختلفة، والتي يُشار إليها بـ α. بمعنوإذا كان المتغير العشوائي X له توزيع t-Student مع n درجة من الحرية ، فإن:

$$P(X_n > k) = \alpha = 1 - P(X_n \leq k) = 1 - F_{X_{n(k)}}$$

والتي تشير إلى احتمال أن X مع درجات n حرية الحرية أكبر من قيمة k

ويتشكل جدول توزيع Student على النحو التالي: تظهر درجات الحرية للمتغير في العمود الأول، وتظهر قيم الاحتمالات في الصف الأول بينما تقاطعات درجات الحرية والاحتمالات تظهر قيم المتغير k التي تتوافق مع:

$$P(X_n > k) = \alpha \; o \; F_{X_{n(k)}} = 1 - \alpha$$

مثال 0-9

أوجد القيمة المقابلة للاحتمال المشار إليه لتوزيع Student.

$$P(X_7 > k) = 0.99$$

الحل

إن قراءة جدول توزيع ستودنت تشابه إلى حد كبير توزيع كي 4. ، حيث نعتمد على درجة الحرية وقيمة الاحتمال، فمثلا عند درجة حرية 7 و احتمال قدره 0.99 لدينا القيمة هي: 2.998

نبحث في الجدول عن قيمة $\alpha = 0.99$. بعد ذلك، في عمود درجات الحرية نبحث عن 7 ويكون تقاطع الصف والعمود هو القيمة المطلوبة.

n\α	0,6	0,7	0,8	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
1	0,325	0,727	1,376	3,078	6,314	12,70	31,82	63,65	318,2	636,5
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,32	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,21	12,92
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587

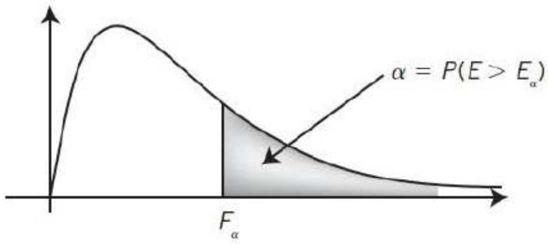
4.5. توزيع F

المتغير العشوائي المستمر X له توزيع احتمالي F عندما يمكن تمثيل دالة الكثافة الخاصة به كما يلي:

يتم تحديد دالة الكثافة لمتغير عشوائي مستمر مع التوزيع F والمعلمات v_1 و v_2 من خلال:

$$f(x) = \frac{\Gamma(\frac{v_1 + v_2}{2})(\frac{v_1}{v_2})^{\frac{v_1}{2}}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma(\frac{v_2}{2})} \, .x^{\frac{v_1}{2}-1}\left[1 + \frac{v_1}{v_2}x\right]^{-\frac{v_1+v_2}{2}} \quad ;x > 0$$

تعرف المعلمتان v_1 و v_2 بدرجات حرية توزيع F للبسط والمقام على التوالي. يتم استخدام هذا النوع من التوزيع بشكل متكرر في الإحصاء عند العمل مع النسبة بين الفروق. يمكن توضيح كيفية استخدام جدال التوزيع F من خلال ما يلي:



لاستخدام الدالة التراكمية يتم استخدام تكملة التوزيع:

$$F(f_\alpha) = P(f < f_\alpha) = 1 - P(f > f_\alpha) = 1 - \alpha$$

قيم توزيع F للاحتمالات المختلفة إلى اليمين، والتي يتم الإشارة إليها بواسطة α ؛ على عكس التوزيعين الآخرين لكل قيمة من قيمة α هناك صفحتان من قيم المتغير بناء على درجات الحرية من 1 إلى 30 ثم للقيم التي تم تخطيها. أي ، إذا كان المتغير العشوائي X له توزيع F مع v_1 درجة من الحرية في البسط و v_2 درجة من الحرية في المقام:

$$P(X(v_1, v_2) > k) = \alpha = 1 - P(X(v_1, v_2) \leq k) = 1 - F_{X(v_1,v_2)}(k)$$

تشير إلى احتمالية أن X (بدرجة v_1 و v_2 من الحرية في البسط والمقام على التوالي) أكبر من القيمة k ، مع F دالة التوزيع التراكمي من X.

يتم استخدام الجدول على النحو التالي: في العمود الأول تظهر درجات حرية المقام وفي الصف الأول ، قيمة الاحتمال الذي تتوافق معه الجداول بينما يظهر الصف الثاني درجات حرية البسط.

مثال 0–10

أوجد قيمة k ، بحيث يكون: $P(X(5,7) > k) = 0.05$، حيث X لها توزيع F.

الحل

نظرا لأن الاحتمال ذو الطرف الأيمن يتم البحث عن قيمة $\alpha = 0.005$ في الجدال ثم في عمود درجات الحرية للبسط ، يوجد 5 ، ثم يوجد التقاطع مع الصف المقابل لدرجات الحرية للمقام 7، وقيمة التقاطع هي القيمة المطلوبة. كما يظهر ذلك من خلال الجدول الموالي، فان القيمة المذلوبة تساوي 3.972:

v_2 \ v_1	1	2	3	4	$\boxed{5}$	6	8	10	12	18	24	30	50	60	120
1	161,446	199,499	215,707	224,583	230,160	233,988	238,884	241,882	243,905	247,324	249,052	250,096	251,774	252,196	253,254
2	18,513	19,000	19,164	19,247	19,296	19,329	19,371	19,396	19,412	19,440	19,454	19,463	19,476	19,479	19,487
3	10,128	9,552	9,277	9,117	9,013	8,941	8,845	8,785	8,745	8,675	8,638	8,617	8,581	8,572	8,549
4	7,709	6,944	6,591	6,388	6,256	6,163	6,041	5,964	5,912	5,821	5,774	5,746	5,699	5,688	5,658
5	6,608	5,786	5,409	5,192	5,050	4,950	4,818	4,735	4,678	4,579	4,527	4,496	4,444	4,431	4,398
6	5,987	5,143	4,757	4,534	4,387	4,284	4,147	4,060	4,000	3,896	3,841	3,808	3,754	3,740	3,705
$\boxed{7}$	5,591	4,737	4,347	4,120	3,972	3,866	3,726	3,637	3,575	3,467	3,410	3,376	3,319	3,304	3,267
8	5,318	4,459	4,066	3,838	3,688	3,581	3,438	3,347	3,284	3,173	3,115	3,079	3,020	3,005	2,967
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,137	3,073	2,960	2,900	2,864	2,803	2,787	2,748

خلاصة الفصل

بعد ان تطرقنا الى مختلف المفاهيم الاساسية والضرورية بالاضافة الى اهم التوزيعات الاحتمالية التي سنيتند عليها في ما تبقى من هذه االمطبوعة، سننتقل الان الى ادراج اهم نظريات وطرق الاستدلال الاحصائي والتي تعتمدها العديد من الابحاث والدراسات في مختلف المجالات.

الفصـل الأول

نظريـة توزيـع المعـاينـة

تمهيد

تنتشر في مجتمعاتنا المعاصرة عمليات الاستقصاء، ففي عالم الأعمال تقوم المؤسسات عن طريق مصالح التسويق ومصالح البحوث والتطوير بإجراء استقصاءات للاطلاع على توجهات المستهلكين، وفي وسائل الإعلام لا يمر يوم دون أن يعلن عن نتائج استقصاء أجرته مجلة أو جامعة حول مواضيع سياسية أو اجتماعية متعددة، منها الاستقصاءات المثيرة للجدل حول الآراء السياسية للمواطنين أثناء الحملات الانتخابية. فما هي الأسس النظرية الرياضية التي تستند عليها الاستقصاءات المختلفة ؟ أو كيف يمكن الاستدلال من خلال بيانات عينة على خصائص المجتمع الذي أخذت منه؟ الإجابة على هذه الأسئلة وغيرها تتطلب فهم العلاقات الرياضية بين الخصائص المختلفة للمجتمع مثل المتوسط، التباين وغيرها، والخصائص المناظرة لها في العينة وهو ما سنتناوله في هذا الفصل.

1. توزيعات المعاينة

عندما يكون من المستحيل أو غير العملي تحليل المجموعة الكاملة من الملاحظات التي يتألف منها المجتمع، ولكننا نريد استخلاص استنتاجات حول بعض مقاييس المجتمع، فغالبا ما يتم استخدام أخذ عينات عشوائية بسيطة. لقد رأينا من قبل أن هذا النوع من أخذ العينات يتميز بأن أي عينة بالحجم n من لها نفس احتمالية أن يتم اختيارها مثل أي عينة أخرى من نفس الحجم. وبعبارة أخرى فإن أخذ العينات العشوائية يلغي أي مشكلة يتم فيها المبالغة في تقدير أو التقليل من بعض خصائص المجتمع (Jarkko، 2014). وبالتالي يجب أن تكون الملاحظات التي يتم إجراؤها مستقلة وعشوائية.

من خلال المفاهيم الموضحة في الفصل السابق يمكننا القول أن مصطلح المجتمع يقصد به القياسات أو القيم وليس الأفراد أو الأشياء التي تم قياسها، كما أن المجتمع قد يكون محدودا أو غير محدود، أما العينة فهي عادة تكون محدودة، ونرمز عادة لحجم المجتمع ب N، ولحجم العينة ب n.

ملاحظة

هناك فرضيتان تتكرران في عدد من العلاقات الرياضية التي سنراها لاحقا، هما فرضية أن قيم مفردات العينة مستقلة والمجتمع لانهائي. يتحقق شرط الاستقلال إذا كانت المعاينة غير نفادية، وإذا كانت كذلك، يمكن اعتبار المجتمع مجتمعا غير محدود.

في مشاكل الاستدلال، عادة ما نهتم بتقدير معلمة μ من المجتمع (على سبيل المثال، متوسط المجتمع) من خلال X_1، X_2، ...، X_n من القيم. لهذا نستخدم الاحصائية m_i (على سبيل المثال متوسط العينة) وبناءا على القيمة التي تم الحصول عليها لـ m_i من عينة معينة سنتخذ القرارات التي تتطلبها المشكلة.

وقد اعتبرت الاحصائية m_i متغير عشوائي لأنها تعتمد على العينة المسحوبة (Jarkko، 2014)؛ وستعطي عينات مختلفة قيما مختلفة لـ m_i. ومنه فتوزيع المعاينة للإحصائية هو التوزيع الاحتمالي لهذه الإحصائية المحسوب لكل العينات الممكنة والتي حجمها n والمأخوذة من المجتمع الإحصائي المدروس مهما كان حجمه وطريقة السحب (بالإرجاع او بدون ارجاع).

فيما يلي بعض توزيعات العينات الأكثر شيوعا المستخدمة في دراسة الإحصاء الاستدلالي.

2. توزيع المعاينة للمتوسط الحسابي

إذا أخذنا عينات متكررة من مجتمع ما وقمنا بقياس متوسط كل عينة، فسنجد أن معظم هذه المتوسطات تختلف عن بعضها البعض، ويسمى التوزيع الاحتمالي لمتوسطات العينات **بتوزيع المعاينة للمتوسط**. ولهذا التوزيع متوسط يرمز له بالرمز: m_i وانحراف معياري أو خطأ معياري يرمز له بـ: σ_m.

ملاحظة

توزيع المعاينة للمتوسطات الحسابية هي عبارة عن التوزيع التكراري للمتوسطات الحسابية لعدد كبير من العينات العشوائية المتساوية الحجم ومن مجتمع إحصائي واحد. كما أن المتوسط الحسابي لتوزيع معاينة إحصائيات العينة يتفق تماما مع المعلمة التي أخذت منم هذه العينات.

لتحديد العلاقة بين المتوسط الحسابي للمجتمع والمتوسط الحسابي لتوزيع المعاينة في حالة عينة غير نفاذية (السحب بالإرجاع) والعينة النفادية (السحب بدون ارجاع) نعتد على بيانات المثال التالي:

مثال 1-1

ليكن المجتمع التالي والذي يشمل العناصر التالية: {1 ، 3 ، 6 ، 8}

تمثل هذه القيم مجموعة القيم الخاصة بخاصية اهتمام المجتمع قيد الدراسة. لذلك بالنسبة لهذه الفئة من المجتمع، يمكن حساب المتوسط والتباين كما يلي:

$$E(X) = \mu = \frac{1}{4}(1+3+6+8) = 4.5$$
$$V(X) = \sigma^2$$
$$= \frac{1}{4}[(1-4.5)^2 + (3-4.5)^2 + (6-4.5)^2 + (8-4.5)^2] = 7.25$$

نفترض أننا سنستخرج من هذه المجموعة عينة حجمها $n = 2$، والإحصائية التي سنحسبها هي متوسط العينة.

1.2. حالة المعاينة بالإرجاع

بعض نماذج الاحتمالات التي يمكن الحصول عليها هي: {1،1} و {1،3} و {6،8} ، والتي تكون قيم متوسطاتها هي 1 و 2 و 7 على التوالي.

يمكننا أن نرى أن هناك تنوع في قيم الإحصائية وبالتالي سيكون من المثير للاهتمام التعمق أكثر في هذا التنوع. لمعرفة متوسط العينة سيتعين علينا معرفة جميع القيم الممكنة للعينة ذات الحجم 2. في هذا المثال ونظرا لأنه لدينا 4 عناصر فقط من المجتمع ، فإن الحصول على جميع العينات الممكنة ذات الحجم 2 ليس بالأمر الصعب.

ونظرا لأن السحب يتم من خلال الاستبدال، فلدينا 4 احتمالات في كل من السحوبات الممكنة. ومن ثم فإن العدد الإجمالي للعينات العشوائية البسيطة هو 4 × 4 = 16. من ناحية أخرى، في كل سحب يكون لكل عنصر من المجتمع فرصة متساوية في الظهور نظرا لوجود 4 عناصر فلكل عنصر احتمال 4/1.

وفي الاخير، ونظرا لأن السحوبات مستقلة فحتى نحصل على احتمال وجود زوج من العناصر الذي ينتمي إلى العينة نقوم بضرب الاحتمالات:

$$P(A \cap B) = P(A)P(B)$$ (عندما يكون A و B مستقلين)

يتم ضمان الاستقلالية عن طريق استبدال كل عنصر.

في الجدول أدناه نقوم بإدراج جميع العينات الممكنة مع الاحتمالات الخاصة بكل منها، ولكل منها نقدم قيمة متوسط العينة.

بالتعمق في هذا الجدول، يمكننا أن نرى أن قيم المتوسطات الحسابية المحتملة هي: 1، 2، 3، 3.5، 4.5، 5.5، 6، 7، 8 ويمكننا إنشاء دالة التوزيع الاحتمالي الخاصة بها، مع ملاحظة على سبيل المثال أنه يمكن الحصول على القيمة 2 من عينتين: (1،3) أو (3،1).

تنويه

إن أحد الأسباب الرئيسية لدراسة التوزيعات الاحتمالية هو القدرة على تحديد وتقييم خصائص توزيعات الإحصاءات قيد الدراسة ، لأنها توفر الأسس التي تمكن من إجراء استنتاجات أفضل حول معلمات التوزيع. من وجهة نظر عملية يمثل توزيع العينات نموذجا نظريا للرسم البياني للترددات النسبية أو المطلقة التي يتم الحصول عليها بالقيم المقابلة. وذلك لأن شكل التوزيع النظري لأخذ العينات للإحصائية يعتمد على توزيع متغيرات العينة.

ونظرا لأن هذه العينات تتوافق مع أحداث متنافية ، فإن احتمال الحصول على عينة متوسط يساوي 2 هو:

$$\begin{aligned} p(\bar{X}=2) &= P(\{1,3\} \cup \{3,1\}) \\ &= P(\{1,3\} + P\{3,1\}) \\ &= \frac{1}{16} + \frac{1}{16} = \frac{2}{16} \end{aligned}$$

الجدول رقم 1-1: توزيع المعاينة لمتوسط العينة

العينات الممكنة	المتوسط	الاحتمال	العينات الممكنة	المتوسط	الاحتمال
1,1	(1+1)/2=1	(1/4)* (1/4)=1/16	6,1	3.5	(1/4)* (1/4)=1/16
1,3	(1+3)/2=2	(1/4)* (1/4)=1/16	6,3	4.5	(1/4)* (1/4)=1/16
1,6	3.5	(1/4)* (1/4)=1/16	6,6	6	(1/4)* (1/4)=1/16
1,8	4.5	(1/4)* (1/4)=1/16	6,8	7	(1/4)* (1/4)=1/16
3,1	2	(1/4)* (1/4)=1/16	8,1	4.5	(1/4)* (1/4)=1/16
3,3	3	(1/4)* (1/4)=1/16	8,3	5.5	(1/4)* (1/4)=1/16
3,6	4.5	(1/4)* (1/4)=1/16	8,6	7	(1/4)* (1/4)=1/16
3,8	5.5	(1/4)* (1/4)=1/16	8,8	8	(1/4)* (1/4)=1/16

بنفس المنطق نحصل على دالة توزيع الاحتمالات للمتوسطات التالية:

$\bar{x}$	1	2	3	3.5	4.5	5.5	6	7	8
$P(\bar{X}=\bar{x})$	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16

وبالتالي:

$$E(m_i) = 1*\frac{1}{16} + 2*\frac{2}{16} + 3*\frac{1}{16} + 3.5*\frac{2}{16} + 4.5*\frac{5}{16} + 5.5*\frac{2}{16} + 6*\frac{1}{16} + 7*\frac{2}{16} + 8*\frac{1}{16} = 4.5 = \mu$$

$$\sigma_m^2 = (1-4.5)^2*\frac{1}{16} + (2-4.5)^2*\frac{2}{16} + (3-4.5)^2*\frac{1}{16} + (3.5-4.5)^2*\frac{2}{16} + (4.5-4.5)^2*\frac{5}{16} + (5.5-4.5)^2*\frac{2}{16} + (6-4.5)^2*\frac{1}{16} + (7-4.5)^2*\frac{2}{16} + (8-4.5)^2*\frac{1}{16}$$

$$= 3.625 = \frac{7.25}{2} = \frac{\sigma^2}{2} = \frac{\sigma^2}{n}$$

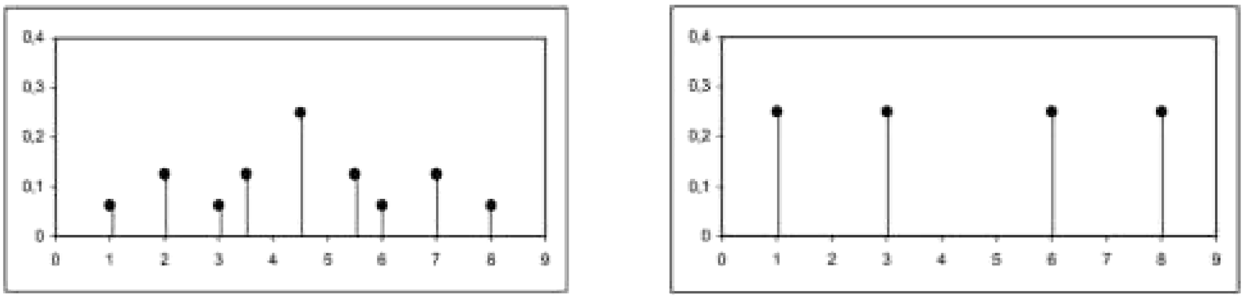
من خلال هذا المثال يمكننا أن نلاحظ أن:

$m_i = \mu$ و $\sigma_m^2 = \frac{\sigma^2}{2}$ حيث 2 هو حجم العينة.

تخبرنا هذه النتائج أن متوسط توزيع المعاينة يساوي متوسط المجتمع وأن تباينه يساوي تباين المجتمع مقسوما على حجم العينة.

الشكل الموالي يوضح الرسوم البيانية للتوزيع الاحتمالي لمتوسط المجتمع في الجزء (أ) و لمتوسط العينة في الجزء (ب).

الشكل رقم 1–1: دالة التوزيع الاحتمالي لـ لمتوسط المجتمع ومتوسط العينة ذات الحجم 2 مأخوذة من المجتمع {1 ، 3 ، 6 ، 8}



ب أ

يمكننا أن نلاحظ أن متوسط كلاهما هو 4.5 وأن توزيع المعاينة للمتوسط لديه تشتت أقل حول هذا

2.2. حالة المعاينة بدون إرجاع

بالاعتماد على بيانات المثال السابق وفي حالة السحب بدون إرجاع، فات عدد العينات الممكنة مكن حسابها من خلال:

$$C_N^n = C_4^2 = 6$$

والعينات الممكن سحبها هي: (1،3)، (1،6)، (1،8)، (3،6)، (3،8)، (6،8) بمتوسطات: 2، 3.5، 4.5، 4.5، 5.5، و7 على التوالي.

وبالتالي:

$$E(m) = 2 * \frac{1}{6} + 3.5 * \frac{1}{6} + 4.5 * \frac{2}{6} + 5.5 * \frac{1}{6} + 7 * \frac{1}{6} = 4.5 = \mu$$

ومنه فالقيمة المتوقعة لمتوسط توزيع المعاينة في حالة المعاينة بدون ارجاع هي:

$$m_i = \mu$$

أما بالنسبة للتباين فهي:

$$\sigma_m^2 = (2-4.5)^2 * \frac{1}{6} + (3.5-4.5)^2 * \frac{1}{6} + (4.5-4.5)^2 * \frac{2}{6} + (5.5-4.5)^2 * \frac{1}{6} + (7-4.5)^2 * \frac{1}{6} = 2.41$$

أو بطريقة اخرى:

$$\sigma^2 = E(X^2) - E(X)^2 = \frac{4 + 12.25 + 20.25 + 20.25 + 30.25 + 49}{6} - 4.5^2 = 2.41$$

المقارنة بين تباين متوسط العينة و تباين المجتمع نجد:

$$2.41 = \frac{7.25}{2}\left(\frac{4-2}{4-1}\right) = \sigma_m^2 = \frac{\sigma^2}{2}\left(\frac{N-n}{N-1}\right)$$

مما سبق يمكن كتابة النظريتين التالتين:

نظرية 1

إذا كان متغير عشوائي يمثل مجتمع ما و m_i متغير عشوائي يمثل متوسط عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لمتوسط العينة $E(m)$ تكتب كما يلي: (في حالة السحب بالإرجاع وبدون ارجاع): $E(m) = m_i = \mu$

نظرية 2

إذا كان متغير عشوائي يمثل مجتمع ما و m_i متغير عشوائي مثل متوسط عينة مسحوبة من ذات المجتمع بالإرجاع، فإن تباين m_i (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي:

في حالة السحب بالإرجاع: $\sigma_m^2 = \frac{\sigma^2}{2}$ حيث n حجم العينة.

في حالة السحب بدون ارجاع:

$$\sigma_m^2 = \frac{\sigma^2}{2}\left(\frac{N-n}{N-1}\right)$$

حيث: $\left(\frac{N-n}{N-1}\right)$ يمثل معامل الارجاع

وعادة يتم استخدام معامل الارجاع عند حساب التباين أو الانحراف المعياري في حالة السحب بدون ارجاع (معاينة نفادية) أو اذا كان حجم المجتمع صغير أو منته، حيث تكون N صغيرة مقارنة بـ n ($\frac{n}{N} \geq 0.05$). وكلما كان حجم المجتمع كبيرا أو غير منته تقترب هذه النسبة من الواحد.

مثال 1-2

قمنا بسحب عينة عشوائية من مجموعة كبيرة من الكتب متوسطها 50 وتبيانها 336. اذا علمت ان حجم العينة المسحوبة هو 12، احسب كل من: المتوسط الحسابي للعينة، التباين، والانحراف المعياري للعينة.

الحل

المجتمع غير محدود ومنه:

المتوسط الحسابي: $m_i = \mu = 50$

التباين: $\sigma_m^2 = \frac{\sigma^2}{2} = \frac{36}{2} = 18$

الانحراف المعياري: $\sigma_m = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{2}} = 4.24$

3. طبيعة توزيع متوسط توزيع المعاينة ونظرية النهاية المركزية

يعتمد شكل ونوع التوزيع الاحتمالي لمتوسطات العينات العشوائية على توزيع المجتمع الذي اخذت منه هذه العينات العشوائية. أي أنه في حالة ما إذا كان المجتمع موزعا توزيعا طبيعيا بمتوسط حساب μ وتباين σ^2 ، فإن متوسط العينة المسحوبة منه يتبع أيضا التوزيع الطبيعي بمتوسط m وتباين σ^2/n

ونكتب: $m \sim N(\mu, \sigma^2/n)$

ولكن بشكل عام بالنسبة لأي توزيع تكون المشكلة معقدة بعض الشيء، ومع ذلك هناك طريقة مقاربة تستخدم لتحديد احتمالات متوسط المتغيرات من عينة عشوائية من أي توزيع، والشرط الوحيد هو أن يكون لها متوسط وتباينات محدودة. بعد ذلك نقوم بإضفاء الطابع الرسمي على هذه النتيجة المقاربة.

وبالتالي ففي حالة عدم معرفة طبيعة توزيع المجتمع الذي أخذت منه العينات العشوائية، فإن توزيع النهائي عينة عشوائية سيؤول إلى التوزيع الطبيعي كلما ارتفع حجم العينة n. وهو ما تشير إليه نظرية النهاية المركزية (The central limit theorem).

نظرية 3: نظرية النهاية المركزية

لنفترض أن X_1، X_2، ...، X_n هي متغيرات عينة عشوائية من التوزيع ذي القيمة المتوسطة μ والتباين المحدود σ^2 .

إذا كان المجتمع ليس بالضرورة طبيعيا فإن المتغيرة المعيارية لـ m أي: $Z = \frac{m-\mu}{\sigma_m} = \frac{m-\mu}{\sigma/\sqrt{n}}$ لها توزيع طبيعي معياري عندما يكون n كبيرا ($n \geq 30$) ونكتب:

$$Z \sim N(0,1)$$

التفسير العملي لنظرية النهاية المركزية يمكن توضيحه على النحو التالي:

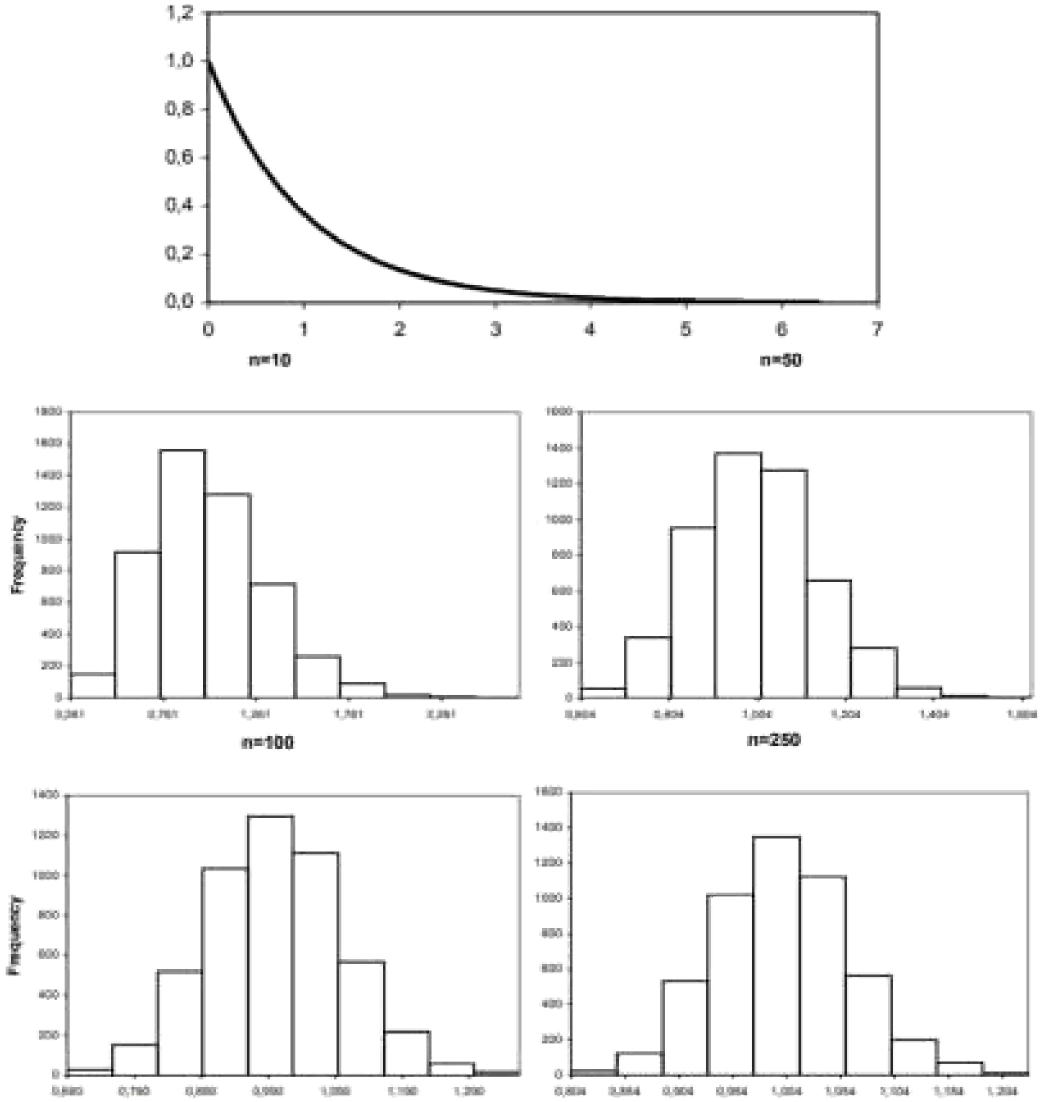
بالنسبة للعينات الكبيرة من أي مجتمع، يمكننا تقريب توزيع المعاينة لـ X عن طريق التوزيع الطبيعي بنفس المتوسط والتباين الذي يساوي تباين المجتمع مقسوم على حجم العينة.

يعتمد الحجم الذي يجب أن تكون عليه العينة للحصول على تقدير تقريبي جيد على خصائص توزيع المجتمع. إذا كان توزيع المجتمع لا ينحرف كثيرا عن التوزيع الطبيعي، فسيكون التقريب جيد، حتى بالنسبة لأحجام العينات الصغيرة.

يوضح الشكل الموالي هذه النظرية للتوزيع الأسي، أي لمجتمع مُوزع وفق عدد أسي مع المعلمة $\lambda = 1$.

الشكل رقم 1-2: رسم توضيحي لنظرية النهاية المركزية ($X \sim exp(1)$)

التوزيع الأسي بمتوسط 1



يمثل الرسم البياني العلوي توزيع المجتمع في حين تمثل الرسومات البيانية المتبقية توزيع المعاينة لـ X على 5000 عينة من الأحجام 10 ، 50، 100،و 250. وهكذا يمكننا أن نلاحظ أنه على الرغم من اختلاف عدد مفردات المجتمع فإن توزيع X يصبح أقرب بشكل متزايد إلى التوزيع الطبيعي مع زيادة n.

*على أي نوع من توزيع المعاينة تنطبق نظرية النهاية المركزية؟

في صياغة هذه النظرية يمكن ملاحظة أنها تنطبق على متوسط المتغيرات وأيضا على مجموع المتغيرات. إذن السؤال الذي يطرح نفسه: هل هناك أي قيود لنكون قادرين على تطبيقها؟

في صياغة نظرية النهاية المركزية يمكننا أن نرى أنه من المطلوب فقط أن تكون القيمة المتوقعة والتباين في توزيع المتغير العشوائي محددا وأن تكون محدودة.

بهذا المعنى في صياغة نظرية النهاية المركزية نتحدث عن حجم عينة لانهائي، ولكن من الناحية العملية ما هو حجم العينة الذي يمكن أن يعطي تقريبا جيدا؟ نحن نعلم أن حجم العينة اللانهائي لا يمكن الحصول عليه نظريا أو حتى عمليا، ولكن قد ثبت أنه من خلال عينات كبيرة الحجم (($n \geq 30$)، فإن تطبيق هذه النظرية يعطي تقديرات تقريبية جيدة، ولهذا السبب يتم تطبيقها على عينات ذات أحجام أكبر من أو تساوي 30.

مثال 1-3

يتم تصنيع نوع معين من المسامير بقطر 10 مم وانحراف معياري قدره 1 مم.

- ما هو احتمال أن يكون متوسط قطر العينة العشوائية المكونة من 400 مسمار أقل من أو يساوي 10.05 مم؟

- ما هو الحد الأدنى لحجم العينة الذي ينبغي اختياره إذا كان المنتج يرغب أن يختلف متوسط قطر عينة المسمار بمقدار 0.10 مم على الأكثر من القطر الحقيقي مع احتمال أكبر من أو يساوي 0.95؟

الحل

ليكن X_1 و X_2 و ... و X_{400} هي المتغيرات العشوائية التي تمثل الأقطار بالمليمترات الخاصة بـ 400 مسار

لدينا: $\mu = 10$ ، $\sigma = 1$ ، $n = 400$

- يتم تمثيل الاحتمال المطلوب من خلال:

$$P(m \leq 10.05)$$

نظرا لعدم معرفة توزيع المتغيرات العشوائية لا يمكننا معرفة توزيع m ، لذلك لا يمكن حساب الاحتمال المطلوب. ومع ذلك فإن متوسط وتباين المجتمع محددين. علاوة على ذلك فان حجم العينة كبير. ومنه يمكننا تقريب الاحتمال بتطبيق نظرية النهاية المركزية على المتغير: $Z = \frac{m-\mu}{\sigma/\sqrt{n}}$

ونلاحظ أن Z سيقترب من التوزيع الطبيعي المعياري لأحجام العينات الكبيرة:

$$P(m \leq 10.05) = P\left(\frac{m-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{10.05-10}{\frac{1}{\sqrt{400}}}\right) = P(Z \leq 1) \cong 0.8413$$

- نفترض أن X_1 و Xn هما المتغيران العشوائيان اللذان يمثلان الأقطار بالمليمترات الخاصة بـ n مسمار.

$$P(|m-\mu| \leq 0.10) \geq 0.95$$

نظرا لأن توزيع البيانات غير معروف يمكن استخدام النظرية:

$$P(|m-\mu| \leq 0.10) = P\left(|Z| \leq \frac{0.10}{\frac{1}{\sqrt{n}}}\right) = P\left(|Z| \leq 0.10\sqrt{n}\right) \geq 0.95$$

من جدول التوزيع الطبيعي لدينا:

$$P(|Z| \leq 1.96) = 0.95$$

$$0.10\sqrt{n} \geq 1.96$$

$$n \geq \left(\frac{1.96}{0.10}\right)^2 = 384.16 \quad ; \quad n \geq 385$$

4. توزيع المعاينة للنسبة

في بعض الاحيان يجب تقدير او اتخاذ قرار يتعلق بنسبة عناصر أو مفردات المجتمع الاحصائي التي تحمل صفة معينة. على سبيل المثال نسبة الطلبة المصابين بفيروس كورونا في كلية العلوم الاقتصادية والتجارية وعلوم التسيير بجامعة خميس مليانة، او أو نسبة المتحصلين على شهادة الدكتوراه العاطلين عن العمل، أو نسبة التلف في الانتاج لسلعة ما...إلخ.

إن إحصائية العينة التي عادة ما تستخدم لإعطاء معلومات عن نسبة مفردات المجتمع الاحصائي والتي تحمل صفة معينة تعرف **بالنسبة** (علي عبد السالم العماري، علي حسين العجيلي، 2000).

ونقصد بالنسبة في المجتمع : $p = \frac{N_a}{N}$ ، حيث ان N تمثل حجم المجتمع، و N_a تمثل عدد مفردات المجتمع التي تتحقق فيها الصفة.

اما النسبة في العينة فيمكن تمثيلها كما يلي: $p' = \frac{n_a}{n}$ ، حيث يمثل n حجم العينة، و n_a عدد مفردات العينة التي تتحقق فيها نفس الصفة.

ومنه يمكن كتابة النظرية الموالية والتي متوسط، تباين، و طبيعة توزيع الإحصائية p'

نظرية 4

لتكن X متغيرة عشوائية تمثل مجتمع ما غير محدود وموزع طبيعيا حيث p نسبة المفردات في المجتمع ذات صفة معينة، ولتكن $'$ متغيرة عشوائية تمثل نسبة المفردات ذات الصفة المذكورة في العينة المسحوبة من ذات المجتمع. حيث يتميز التوزيع النظري لمعاينة النسبة بما يلي:

توقع (متوسط توزيع المعاينة للنسبة): $E(p') = \mu_{p'} = p$

خطأ معاينة (الانحراف المعياري لتوزيع المعاينة للنسبة): $\sigma_{p'} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$

ونكتب: $p' \approx N(p, \sigma_{p'}^2)$ (في حالة $n \geq 30$ ، في حالة مجتمع طبيعي)

إذا كانت المعاينة نفادية، أو حجم المجتمع محدود، أي: $n/N > 0.05$ ، فإن:

$$\sigma_{p'} = \sqrt{\frac{pq}{n}}(\sqrt{\frac{N-n}{N-1}})$$

مثال 1-4

في منطقة ما كان 18% من المراهقين يتم اعتقالهم من قبل الشرطة لأسباب تتعلق بجنوح الأحداث. تم اختيار عينة عشوائية من 100 مراهق من هذه المنطقة. ما هي القيمة والتباين المتوقعان لنسبة المراهقين الذين تم اعتقالهم من الشرطة لعينات عشوائية بحجم 100؟

الحل

لتكن X_1 و X_2 و X_{100} هي المتغيرات العشوائية التي تمثل ما إذا كان المراهق تم اعتقاله من طرف الشرطة. حسب ظروف المشكلة، فإن المتغيرات لها توزيع برنولي (Bernoulli)[1]، حيث أن قيم المتغير هي صفر للحالة التي لم يكن لدى المراهق مشاكل مع الشرطة وواحد للحالة المعاكسة.

لدينا: $X = X_1 + \cdots + X_{100}$

النسبة هي:

$$p' = \frac{X}{100}$$

[1] في نظرية الاحتمالات يسمي قانون برنولي (الذي سمي على اسم عالم الرياضيات السويسري جاك برنولي) قانون الاحتمالات لمتغير عشوائي منفصل يأخذ القيمة 1 مع الاحتمال p و 0 مع الاحتمال $q = 1 - p$.

من النظرية رقم 04 يتضح ان:

$$E(p') = P = 0.18$$
$$\sigma_{p'}^2 = \frac{P(1-P)}{n} = 0.001476$$

1.4. طبيعة توزيع المعاينة للنسبة

عموما يمكن القول بأنه كلما زاد حجم العينة كلما اقترب توزيع المعاينة للنسبة، من التوزيع الطبيعي حسب نظرية النهاية المركزية.

ونكتب: $Z = \frac{p'-p}{\sigma_{p'}} \sim N(0,1)$

لكي يكون التقريب باستخدام التوزيع الطبيعي مقبول يجب أن يتحقق ما يلي:

$$\begin{cases} n \geq 30 \\ np > 5 \\ nq > 5 \end{cases}$$

مثال 1–5

عينة عشوائية حجمها 40 طالب قمنا بإدراج علاماتهم في مقياس الاحصاء 3، وكانت النتائج كما يلي:

1	2.5	3	5	8.5	10.5	8.5	2.5	3.5	7.5
17	15.5	12	14	10.5	15	11.5	18.5	10.5	17.5
3	12	2.5	5.5	7.5	6	14	6	5	11
3.5	11	14	9.5	10	10	12.5	8.5	1.5	9

– احسب نسبة الطلبة الحاصلين على المعدل في العينة؟

– ما هو توزيع نسبة الطلبة الحاصلين على المعدل في العينة من مجموع الطلبة البالغ 500؟

الحل

– نسبة الطلبة الحاصلين على علامة اكبر او تساوي 10:

لدينا 12 طالب تحصلوا على علامة اكبر او تساوي 10:

$$p' = \frac{n_a}{n} = \frac{19}{40} = 0.475$$

– توزيع نسبة الطلبة الحاصلين على المعدل في العينة:

القيمة المتوقعة في العينة: $E(p') = \mu_{p'} = p = 0.475$

لحساب تباين توزيع المعاينة للنسبة (خطأ المعاينة) نحتاج الى اختبار النسبة: $\frac{n}{N}=\frac{40}{500}=0.08$ وهي أكبر من 0.05 وبالتالي نستخدم العلاقة لحساب تباين توزيع المعاينة للنسبة:

$$\sigma_{p'}=\frac{pq}{n}\frac{N-n}{N-1}=\frac{0.475(1-0.475)}{40}\frac{500-40}{500-1}=0.0057$$

ونكتب: $p'\sim N(0.475,0.0057)$

5. توزيع المعاينة للفرق للفروق والمجاميع

في عدة حالات يتم التركيز على دراسة مجتمعين وليس مجتمع واحد فقط، لذلك لا بد من دراسة الفرق أو المجموع بين احصائيتين (المتوسط، النسبة، الخ) لعينتين مسحوبتين من مجتمعين مختلفين. فإذا كانت لدينا عينة عشوائية حجمها n_1 من مجتمع احصائي ما ذو متوسط حسابي μ_1 ، وانحراف معياري σ_1 ، وعينة عشوائية حجمها n_2 من مجتمع احصائي ما ذو متوسط حسابي μ_2 ، وانحراف معياري σ_2 ، نحسب لكل عينة مسحوبة من المجتمع الأول الاحصائية S_1، ونحسب نفس الاحصائية في كل عينة مسحوبة من المجتمع الثاني S_2.

إن الفرق S_1-S_2 يشكل بدوره متغيرة عشوائية ذو المتوسط والتباين:

$$\mu_{S_1-S_2}=\mu_{S_1}-\mu_{S_2} \text{ ،}\quad \sigma^2_{S_1-S_2}=\sigma^2_{S_1}+\sigma^2_{S_2}$$

إذا أردنا حساب مجموع الاحصائيتين بدال من الفرق بينهما فإن:

$$\mu_{S_1+S_2}=\mu_{S_1}+\mu_{S_2} \text{ ،}\quad \sigma^2_{S_1+S_2}=\sigma^2_{S_1}+\sigma^2_{S_2}$$

وذلك بشرط استقلال العينات ، مما يعني استقلال المتغيرات.

1.5. توزيع المعاينة للفرق بين متوسطي عينتين مستقلتين

إذا كانت الاحصائية S_X و الاحصائية S_Y تمثلان متوسطي عينتين مسحوبتين من مجتمعين مستقلين طبيعيين ذوي التباينين: σ_X^2 و σ_Y^2 على التوالي، فإن المتوسط الحسابي لتوزيع المعاينة للفرق بين متوسطي عينتين هو الفرق بين متوسطي المجتمعين ونكتب: $\mu_{\bar{X}-\bar{Y}}=\mu_{\bar{X}}-\mu_{\bar{Y}}=\mu_X-\mu_Y$

وتباين توزيع المعاينة للفرق بين متوسطي عينتين يكتب كما يلي: $\sigma^2_{X-Y}=\frac{\sigma_X^2}{n_1}+\frac{\sigma_Y^2}{n_2}$

مثال 1–6

لكن لدينا المجتمعين التاليين المشكلين من العناصر التالية:

المجتمع الأول: 2، 3، 4

المجتمع الثاني: 1، 2

تحقق من أن:

$$\mu_{1-2} = \mu_1 - \mu_2 \text{ ،} \quad \sigma_{1-2}^2 = \sigma_1^2 + \sigma_2^2$$

$$\mu_{1+2} = \mu_1 + \mu_2 \text{ ،} \quad \sigma_{1+2}^2 = \sigma_1^2 + \sigma_2^2$$

الحل

المجتمع الاول	**المجتمع الثاني**
المتوسط: $\mu_1 = \frac{2+3+4}{3} = 3$	المتوسط: $\mu_2 = \frac{1+2}{2} = 1.5$
التباين: $\sigma_1^2 = \left(\frac{\sum X^2}{N}\right) - \mu_1$	التباين: $\sigma_2^2 = \left(\frac{\sum Y^2}{N}\right) - \mu_2$
$= \frac{4+9+16}{3} - 9 = 0.67$	$= \frac{1+4}{2} - 2.25 = 0.25$

الفرق بين المجتمعين: 1،2، 3، 0، 1، 2.

حساب متوسط هذه العناصر: $\mu_{1-2} = \frac{1+2+3+0+1+2}{6} = 1.5$

حساب تباين هذه العناصر: $\sigma_{1-2}^2 = \frac{1+4+9+0+1+4}{6} - (1.5)^2 = 0.92$

التحقق من العلاقات: $\mu_{1-2} = \mu_1 - \mu_2 = 3 - 1. = 1.5$

$$\sigma_{1-2}^2 = \sigma_1^2 + \sigma_2^2 = 0.67 + 0.25 = 0.92$$

مجموع المجتمعين: 3،4، 5، 4، 5، 6.

حساب متوسط هذه العناصر: $\mu_{1+2} = \frac{3+4+5+4+5+6}{6} = 4.5$

حساب تباين هذه العناصر: $\sigma_{1-2}^2 = \frac{9+16+25+16+25+36}{6} - (4.5)^2 = 0.92$

التحقق من العلاقات: $\mu_{1+2} = \mu_1 + \mu_2 = 3 + 1.5 = 4.5$

$$\sigma_{1+2}^2 = \sigma_1^2 + \sigma_2^2 = 0.67 + 0.25 = 0.92$$

2.5. طبيعة توزيع المعاينة للفرق بين متوسطين

إذاكانت العينتين المسحوبتين من مجتمعين طبيعيين فإن الفرق بين متوسطي العينتين أو مجموعهما يتبع من التوزيع الطبيعي (محمد صبحي، عدنان محمد، 2004).

نظرية 5

في حالة $n_1 \geq 30$ و n_2 ، إن توزيع المعاينة للفرق بين متوسطين يتوزع تقريبا وفق التوزيع الطبيعي، ونكتب:

$$(\bar{X}-\bar{Y}) \sim \mathrm{N}(\mu_X-\mu_Y, \sigma^2_{\bar{X}-\bar{Y}})$$

وعليه فإن توزيع المتغير المعياري: $(\bar{X}-\bar{Y})$ يقترب من التوزيع الطبيعي

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_{\bar{X}}-\mu_{\bar{Y}})}{\sqrt{\frac{\sigma^2_{\bar{X}}}{n_1}+\frac{\sigma^2_{\bar{Y}}}{n_2}}} \sim N(0,1)$$

وعليه بالعودة للمثال السابق :وعليه فإن توزيع المعاينة للفرق بين متوسطين يكتب بالشكل:

$$(\bar{X}-\bar{Y}) \sim \mathrm{N}(1.5, 0.92)$$

كما ان توزيع المعاينة للمجموع بين متوسطين يكتب بالشكل: $(\bar{X}+\bar{Y}) \sim \mathrm{N}(4.5, 0.92)$

** تجدر الاشارة الى انه اذاكان المجتمعين مجهولي التباين فإننا نميّز بين حالتين:

الحالة الأولى: ذاكان حجم العينتين كبير $(30 \geq)$ فإن توزيع المعاينة للفرق بين متوسطي عينتين يتبع التوزيع الطبيعي بمتوسط $\mu_X-\mu_Y$ وبتباين $\sigma^2_{\bar{X}-\bar{Y}}=\frac{S^2_{\bar{X}}}{n_1}+\frac{S^2_{\bar{Y}}}{n_2}$

ويكون التوزيع الاحتمالي كماي يلي:

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_{\bar{X}}-\mu_{\bar{Y}})}{\sqrt{\frac{S^2_{\bar{X}}}{n_1}+\frac{S^2_{\bar{Y}}}{n_2}}} \sim N(0,1)$$

الحالة الثانية: اذاكان حجم احدى العينتين أوكالاهما صغير (أقل من 30) والمجتمعين موزعين توزيعا طبيعيا فإن توزيع المعاينة للفرق بين متوسطي عينتين يتبع التوزيع الطبيعي بمتوسط $\mu_X-\mu_Y$ وتباين $\sigma^2_{\bar{X}-\bar{Y}}=\frac{S^2}{n_1}+\frac{S^2}{n_2}$

ويكون التوزيع ويكون التوزيع الاحتمالي كما هو مبين في العلاقة التالية:

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_{\bar{X}}-\mu_{\bar{Y}})}{\sqrt{\frac{S^2}{n_1}+\frac{S^2}{n_2}}}\sim N(0,1)$$

حيث أن: S^2 يعرف بالتباين المشترك لتبايني العينتين $S_{\bar{X}}^2$ و $S_{\bar{Y}}^2$ ويمكن حسابه كما يلي:

$$S^2=\frac{(n_1-1)S_{\bar{X}}^2+(n_2-1)S_{\bar{Y}}^2}{n_1+n_2-2}$$

3.5. توزيع المعاينة للفرق بين نسبتي عينتين

من المثير للاهتمام أحيانا مقارنة نسب عينتين، على سبيل المثال نسب مبيعات عنصرين ونسب العناصر الجيدة التي تنتجها آلتان ، وهكذا.

نظرية 6

لنفترض أن X_1 و X_2 و X_{n1} و Y_1 و Y_2 و Y_{n2} عينتان عشوائيتان مستقلتان مأخوذتان من مجتمعين متوسطهما على التوالي: p_1 و p_2 ، والنسبيتين: $p_1'=\bar{X}$ و $p_2'=\bar{Y}$ على التوالي، اذن:

$$E(p_1'-p_2')=p_1-p_2 \quad \text{و} \quad \sigma^2_{p_1'-p_2'}=\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}$$

توزيع المعاينة للفرق بين نسبتي عينتين حجمهما كبير (أكبر أو يساوي من 30) ومسحوبتين من مجتمعين مستقلين، فإن توزيع المعاينة للفرق بين نسبتي عينتين يتبع التوزيع الطبيعي تقريبا، ونكتب:

$$(p_1'-p_2')\sim N(\mu_{p_X-p_Y},\sigma^2_{p_X-p_Y})$$

وعليه فإن توزيع المتغير المعياري للفرق بين نسبتي عينين يقترب من التوزيع الطبيعي المعياري ونكتب:

$$Z=\frac{(p_1'-p_2')-(\mu_{p_X'}-\mu_{p_Y'})}{\sqrt{\sigma^2_{p_X'-p_Y'}}}\sim N(0,1)$$

مثال 1–7

اذا كانت نسبة الاصابة بفيروس كورونا في كلية العلوم الاقتصادية والتجارية وعلوم التسيير لجامعة خميس مليانة هي 35% لدى الاناث و 25% لدى الذكور. اذا قمنا بسحب عينتين عشوائيتين من المجتمعين حجم كل منهما 100. أوجد احتمال أن الفرق بين نسبة الاصابة بين الاناث ونسبة الاصابة بين الذكور أكبر من 20%.

الحل

$$n_1 = n_2 = 100 > 30$$

لدينا ايضا:

$$n_1 p_1 (100)0.35 = 35 > 5$$
$$n_1 q_1 (100)0.65 = 65 > 5$$

و

$$n_2 p_2 (100)0.25 = 25 > 5$$
$$n_2 q_2 (100)0.75 = 75 > 5$$

وعليه فإن توزيع المعاينة للفرق بين نسبتي العينتين يقترب من التوزيع الطبيعي بمتوسط:

$$E(p'_1 - p'_2) = p_1 - p_2 = 0.35 - 0.25 = 0.1$$

وانحراف معياري قدره:

$$\sigma_{p'_1 - p'_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0.35 * 0.65}{100} + \frac{0.25 * 0.75}{100}} = 0.062$$

احتمال أن يكون الفرق بين نسبة الاصابة بين الاناث ونسبة الاصابة بين الذكور أكبر من 20% :

$$P(p'_1 - p'_2) > 0.2$$
$$P\left(\frac{(p'_1 - p'_2) - (\mu_{p'_1} - \mu_{p'_2})}{\sigma^2_{p'_1 - p'_2}} > \frac{0.2 - 0.1}{0.062}\right)$$
$$= P(Z > 1.62) = 1 - P(Z \leq 1.62)$$
$$= 1 - 0{,}9474 = 0.0526$$

6. توزيع المعاينة للتباين

يمكن الحصول على توزيع المعاينة للتباين من خلال سحب كل العينات الممكنة ذات الحجم n وحساب كل التباينات الممكنة لكل العينات الممكنة، وبالتالي نحصل على مجتمع جديد يعرف بتوزيع المعاينة للتباين، يتبع توزيعا يعرف بتوزيع مربع كاي.

بالاعتماد على معطيات المثال 1-1 أين لدينا مجتمع مكون من العناصر التالية: {1 ، 3 ، 6 ، 8}، سنقوم بحساب في حالة السحب بالارجاع وبدون ارجاع كل من: تباين المجتمع σ^2، أحسب تباين كل عينة S_i^2، القيمة المتوقعة لعناصر التباينات الجديدة.

لدينا: متوسط المجتمع: $\mu = 4.5$ ، تباين المجتمع: $\sigma^2 = 7.25$

1.6. في حالة السحب بالارجاع

التباينات الممكنة S_i^2:

0	1	6.25	12.25	1	0	2.25	6.25
6.25	2.25	0	1	12.25	6.25	1	0

$$E(S^2) = \frac{\sum S_i^2}{n} = \frac{58}{16} = 3.625$$

$$E(S^2) = \sigma^2(n - 1/n) \qquad 3.625 = 7.25(2 - 1/2)$$

2.6. في حالة السحب بدون ارجاع

التباينات الممكنة S_i^2:

1	6.25	12.25	2.25	6.25	1

$$E(S^2) = \frac{\sum S_i^2}{n} = \frac{29}{6} = 4.83 \qquad E(S^2) = \sigma^2(n - 1/n)(\frac{N}{N-1})$$

$$4.83 = 7.25(2 - 1/2)(\frac{4}{4-1})$$

نظرية 7

اذا كان X المتغير العشوائي من مجتمع ما و S^2 متغيرة عشوائية تمثل تباين عينة مسحوبة من ذات المجتمع فإن القيمة المتوقعة لتباين العينة يعطى كالتالي:

السحب بالارجاع: $E(S^2) = \sigma^2(n - 1/n)$

السحب بدون ارجاع: $E(S^2) = \sigma^2(n - 1/n)(\frac{N}{N-1})$

تجدر الاشارة إلى أنه كلما زاد حجم المجتمع إن قيمة النسبة $\frac{N}{N-1}$ ستؤول إلى الواحد. ($\frac{N}{N-1} \approx 1$). كما أنه كلما كان حجم العينة كبيرا ($n \geq 30$) فإن: $E(S^2) \approx \sigma^2$

3.6. طبيعة توزيع المعاينة للتباين

مما سبق نجد أن: $$E\left(S^2 \frac{n}{n-1}\right) = \sigma^2$$

ونقول عن $S^2 \frac{n}{n-1}$ أنه مقدر غير منحرف لـ σ^2 ويرمز له بالرمز $\hat{S}^2$ حيث:

$$\hat{S}^2 = S^2 \frac{n}{n-1}$$

نظرية 8

انطلاقا من عينة عشوائية حجمها n من المشاهدات المستقلة مسحوبة من مجتمع طبيعي بتباين σ^2 فان:

$$\frac{nS^2}{\sigma^2}=\frac{(n-1)\hat{S}^2}{\sigma^2}\sim x_{n-1}^2$$

مثال 1-8

نفرض ان علامات 30 طالب لمقياس معين موزعة طبيعيا كما يلي: $X\sim N(25,10)$

أحسب احتمال أن يكون تباين علامات الطلاب في العينة أقل من 10.

الحل

احتمال أن يكون تباين نقاط الطالب أقل من 10:

$$P(S^2<10)=P(nS^2/\sigma^2<(n-1)\hat{S}^2/\sigma^2)$$
$$=P\left(x_{n-1}^2<\frac{(29)10}{10}\right)=P(x_{29}^2<18)=1-P(x_{29}^2\geq 18)$$

من جدول كاي مربع نستخرج قيمة الاحتمال المقابل لدرجة الحرية وعليه فإن:

$$P(S^2<10)=P(x_{29}^2<18)=1-P(x_{29}^2\geq 18)=0.1$$

4.6. توزيع المعاينة لنسبة تبايني عينتين

اذا كان لدينا $X\sim x_{v_1}^2$ و $Y\sim x_{v_2}^2$ وكان X و Y مستقلين فان: $Z=\frac{(X/n_1)}{(Y/n_2)}$ يتبع توزيع F (Fisher) بدرجتي حرية v_1 و v_2 حيث: $v_1=n_1-1$ و $v_2=n_2-1$

ونكتب: $Z\sim F_{v_1v_2}$

ومنه نستنتج أن توزيع المعاينة لنسبة تبايني عينتين مستقلتين يتبع توزيع فيشر، ويمكن صياغة النظرية التالية:

نظرية 9

ن نسبة تباين عينتين مستقلتين حجمهما على الترتيب n_1 و n_2 مسحوبتين من مجتمعين طبيعيين تبايناهما على التوالي: σ_1^2 و σ_2^2 يمكن كتابتها على الشكل التالي:

$$F=\frac{\left[\frac{S_1^2n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}}=\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}\sim F_{v_1v_2}$$

مثال 1–9 (مأخوذ من بوعبد الله صالح، 2006)

عينتين حجمهما 8 و10 مسحوبتين من مجتمعين طبيعيين تبايناهما على التوالي 20 و36. ما احتمال أن يكون تباين الأولى أكبر من ضعف تباين الثانية؟

الحل

$$P(S_1^2 > 2S_2^2) = P(\frac{S_1^2}{S_2^2} > 2) =$$

$$P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}}\right) = P\left(\frac{S_1^2\left(\frac{n_1}{n_1-1}\right)\frac{1}{\sigma_1^2}}{S_2^2\left(\frac{n_2}{n_2-1}\right)\frac{1}{\sigma_2^2}} > 2\frac{\left(\frac{8}{7}\right)\frac{1}{20}}{\left(\frac{10}{9}\right)\frac{1}{36}}\right) =$$

$$= P(F_{7,9} > 3.7)$$

من الجدول نجد $0.05 > P(F_{7,9} > 3.7) > 0.01$

وفي الحقيقة $P(F_{7,9} > 3.7) = 0.036$

خلاصة الفصل

درسنا من خلال هذا الفصل مجموعة من النظريات للعلاقة الرياضية بين احصائيات العينة والمعالم المناظرة لها في المجتمع مثل المتوسط، التباين، النسبة...كما تطرقنا ايضا الى العلاقة بين شكل توزيع المجتمع وشكل التوزيع الاحتمالي لإحصائيات العينة. تظهر هذه العلاقات كتوصيف لخصائص العينة ومعالمها ولكنها تستخدم أكثر لتقدير خصائص ومعالم المجتمع محل الدراسة، وهذا ما سنتعرف عليه في هذا الفصل الموالي مع التطرق الى كيفية تفسير وتحليل ودمج البيانات لتحديد فترات الثقة التي تسمح بتقدير معلمات المجتمع.

الفصـل الثاني

نظريـة التقديـر

تمهيد

رأينا سابقا أن متوسط العينة هو مقدر جيد لمتوسط المجتمع. لكننا رأينا أيضا أن هناك تباين في قيم هذا المتوسط، أي أن كل عينة تؤدي إلى قيمة مختلفة للمقدر. وبالتالي من المنطقي أن نفترض أن الاستنتاج الذي تم إجراؤه باستخدام قيمة النقطة ليس هو الأنسب. لهذا السبب، من الأفضل الإشارة إلى نطاق من القيم يتم فيه تقدير موقع المعلمة قيد الدراسة بدرجة معينة من الثقة. لذا سنقوم بدراسة نظرية التقدير في هذا الفصل والتي يمكن اعتبارها الأساس النظري لتطوير الاستدلال الإحصائي.

1. الاساس النظري

يؤدي هذا إلى دراسة فترات الثقة، ولهذا نفترض أن لدينا مجموعة سكانية تكون معاملتها ، θ، غير معروفة ، وأنه في ظل ظروف معينة، نجد أن$\theta \epsilon \left(\hat{\theta}_i, \hat{\theta}_s\right)$ ، حيث ترتبط $\hat{\theta}_i$ و$\hat{\theta}_S$ بقيمة الإحصائية $\hat{Q}$ لتحقيق معين لعينة عشوائية وتعرف باسم الحد الأدنى والأعلى. وبما أن حدي المجال $\hat{\theta}_i$ و$\hat{\theta}_S$ تعتمد على المعاينة، فإنها تمثل قيما معينة للمتغيرات العشوائية $\hat{Q}_i$ و $\hat{Q}_S$.

بناءا على المتغيرات العشوائية السابقة، من الممكن حساب احتمال وجود المعلمة θ في المجال المحدد ، من اجل $1-\alpha$ مع $\alpha \in (0,1)$، لدينا:

$$P\left(\hat{Q}_i < \theta < \hat{Q}_s\right) = 1 - \alpha$$

المجال الذي يقع فيه المعلمة θ $(\hat{\theta}_i < \theta < \hat{\theta}_s)$ يعرف بفترة الثقة او بمجال الثقة، بينما تمثل القيمة $1-\alpha$ درجة الثقة. اما حدود هذا المجال $\hat{\theta}_i$ و $\hat{\theta}_s$ فتمثل حدود الثقة.

ولكن كيف يمكن أن نفسر فترة الثقة؟

لنفترض مجتمع ما بمعلمة θ ، حيث نقوم باختيار عينة عشوائية X_1، X_2، ...، X_n ، والتي من خلالها نحدد الإحصائية $\hat{Q}$ ، المقابلة للمعلمة θ.

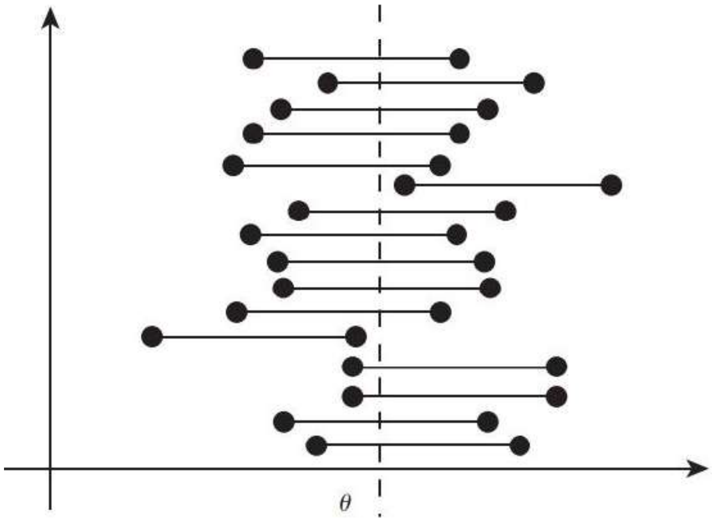
من ناحية أخرى نحدد المتغيرات العشوائية $\hat{Q}_i$ و $\hat{Q}_S$ بحيث يكون لكل: x_1، x_2، ...، x_n للعينة العشوائية X_1، $X2$، ...، X_n ، نحصل على مجال ثقة $\left[\hat{\theta}_i, \hat{\theta}_s\right]$، حيث $\hat{\theta}_i$ و$\hat{\theta}_S$ هي قيم المتغيرات $\hat{Q}_i$ و $\hat{Q}_S$، ومنه:

$$P\left(\hat{Q}_i < \theta < \hat{Q}_s\right) = 1 - \alpha$$

نشير إلى أن $(1-\alpha)$ 100% من الفترات التي تم تحديدها $\left[\hat{\theta}_i, \hat{\theta}_s\right]$ لكل x_1، x_2، ...، x_n للعينة العشوائية تشمل المعلمة.

ويمكن تمثيل ذلك بيانيا من خلال الشكل الموالي:

الشكل رقم 2–1: تمثيل بعض فترات الثقة



تمثل مقاطع الخط في الشكل 2–1 طول المجالات لكل عينة، بينما يعمل الخط العمودي المنقط كمرجع لإظهار فترات الثقة (المجالات) التي تشمل المعلمة.

ملاحظة

في هذه المرحلة، يجدر توضيح خطأ التفسير الذي يحدث بشكل متكرر في فترات الثقة. لذا نفرض مجال الثقة $\left[\hat{\theta}_i, \hat{\theta}_s\right]$ الذي تم الحصول عليه من x_1، x_2، ...، x_n لعينة عشوائية للمعلمة θ . يقال في كثير من الأحيان أن:

$$P\left(\hat{\theta}_i < \theta < \hat{\theta}_s\right) = 1-\alpha$$

وهذا خطأ لأنه في $P\left(\hat{\theta}_i < \theta < \hat{\theta}_s\right)$ ليس لدينا متغيرات عشوائية لذلك لا توجد احتمالات لحساب المعلمة (في هذه الحالة يكون الاحتمال صفر أو واحد).

مثال 2–1

تبين من عينة عشوائية تتكون من 20 مفردة أن متوسط المدة بالساعات هو 750، وبناءً على هذه القيمة، فإننا يمكن ان نقدر المعلمة μ مع احتمال $1-\alpha$ (محدد مسبقا) في الفترة (740 ,760).

ولكن لا ينبغي تفسير ذلك على أنه: $P(740 < \mu < 760) = 1-\alpha$

ومنه نقول انه نظرا لعدم وجود متغيرات عشوائية لا يمكننا حساب الاحتمالات.

في كثير من الحالات قد نحتاج الى الاجابة على عدة اسئلة ككيفية معرفة المعلمة التي يجب تقديرها؟ كيفية تحديد قيمة لمقدر المعلمة؟ وما الى ذلك. ويمكن إعطاء الإجابات بناءً على أساسيات الاستدلال الإحصائي والتي سنناقشها في الأقسام التالية.

2. مفاهيم أساسية حول المقدرات النقطية

بعد مناقشة الاسس العامة فيما يتعلق بفترات الثقة، سننتقل فيما يلي الى الأقسام المنهجية التي ستساعدنا في الحصول على فترات الثقة لمعلمات مجتمع بتوزيع طبيعي. بمعنى سنتعرف على كيفية تحديد فترات الثقة للمتوسط ولمتوسط الفروق، التباين والنسبة.

وتتمثل المبادئ الأساسية للاستدلال الإحصائي في إنشاء طرق للتوصل إلى استنتاجات حول المجتمع، وتنقسم هذه الطرق إلى مجموعتين (Sedkaoui, 2018): الطرق الكلاسيكية وطريقة بايز (Bayes).

ملاحظة

في الطرق الكلاسيكية (طريقة المعقولية العظمى، طريقة العزوم، طريقة المربعات الصغرى) يتم الاستدلال من خلال نتائج المعاينات العشوائية، اما في الثانية فيتم تنفيذه بناءا على المعرفة السابقة حول توزيع المعلمات غير المعروفة (في طريقة بايز تعتبر المعلمات متغيرات عشوائية). وقد حققت الطرق البايزية ازدهارا كبيرا في مجالات الإحصاء المختلفة لكن دراستها خارجة عن أهداف هذه المطبوعة.

1.2. الفضاء المعلمي

حيث نبدأ دراسة الاستدلال الإحصائي بتحديد مجموعة القيم الممكنة للمعلمة، ولتكن X_1، ...، X_n عينة عشوائية بدالة الكثافة $f(x,\theta)$ حيث يكون شكل الدالة معروف ولكن المعلمة θ غير معروفة، نوكل ما نعرفه هو انها تنتمي إلى مساحة المعلمة التي يشار إليها بالرمز Ω.

أمثلة

1. المتغيرات العشوائية لها دالة كثافة أسية بالمعلمة β،

$$\begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & , \quad x \geq 0 \quad , \quad \beta > 0 \\ 0 & \text{لكل قيمة أخرى لـ } x \end{cases}$$

في هذه الحالة ، تكون المعلمة $\theta = \beta$ ومجموعة قيم المعلمة هي Ω = (0، ∞)

2. المتغيرات لها دالة كثافة طبيعية والمعلمات μ و σ^2

$$f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

حيث: $\mu \epsilon R$ و $\sigma > 0$

في هذه الحالة يكون فضاء المعلمة هو: $\Omega = \mathrm{R} \times \mathrm{R}^{+}$

3. المتغيرات لها دالة كثافة برنولي والمعلمة p

$$\begin{cases} p & x=1 \\ 0 & x=0 \end{cases}$$

حيث: $p\epsilon[0,1]$

في هذه الحالة يكون فضاء المعلمة هو: $\Omega = [0,1]$

من الأمثلة السابقة يمكننا أن نلاحظ أنه في الاستدلال الإحصائي عندما نتحدث عن المعلمة نشير إلى مجموعة من الكثافات (معتوق امحمد، 2007)، بمعنى يجب علينا دراسة كيفية الحصول على معلومات حول المعلمة لدالة الكثافة.

2.2. قيم المقدر النقطي

لنفترض أننا نريد إجراء استنتاج بشأن متوسط علامات جميع الطلاب الذين يدرسون موضوع حساب التفاضل والتكامل، حيث نقوم بتحليل عينة عشوائية متكونة من 10 من الطلاب المتحصلين على العلامات التالي:

3، 7، 2، 8، 6، 9، 9، 4، 8، 6

باستخدام البيانات المذكورة أعلاه نحسب قيمة الإحصائية m

$$m=\frac{1}{10}(3+7+2+8+6+9+9+4+8+6)=6.2$$

بناءًا على القيمة المحسوبة لإحصائية m ، من الممكن استنتاج المعلمة μ لعلامات الطلبة، بمعنى يمكن إجراء تقدير نقطي لمتوسط المعلمة فيما يتعلق بمؤهلات الطلبة. في هذه الحالة يتم تقدير المعلمة μ.

وفيما يتعلق بالمقدر النقطي m يتم تعريفه بشكل عام على النحو التالي:

تعريف

ليكن المجتمع ذو المتوسط θ و $X_1, \dots X_n$ عينة عشوائية من ذات المجتمع، مع $\hat{Q} = U(X_1, \dots X_n)$، الإحصائية المقابلة لـ θ ، في جزء تقدير $\hat{Q}$ يسمى مقدر θ ، بينما تسمى القيمة $\hat{\theta}$ لـ $\hat{Q}$ التي تم الحصول عليها من العينة العشوائية بالمقدر النقطي لـ θ.

مع هذا يمكن استنتاج أنه إذا كانت θ هي المعلمة مع فضاء Ω، فيجب تضمين أي مقدر دقيق لـ θ في Ω. أي ، إذا كان $U(X_1, \dots X_n)$ هو مقدر المعلمة فسيكون لتوزيعها Ω كمجال.

استطعنا أن نلاحظ أن المقدر $\hat{Q}$ المقابل للمعلمة θ هو إحصائية، لذلك يمكن الإجابة جزئيا على السؤال حول المقدر الذي يجب استخدامه. لنفترض أننا نرغب في الحصول على مقدر لـ θ. نبدأ بمراجعة العلاقات بين المعلمة والقيم التي نعرفها في الإحصاء الوصفي مثل المتوسط، التباين ،الانحراف المعياري، إلخ.

مثال 2-2

المتغيرات العشوائية لها دالة كثافة أسية بالمعلمة β،

$$\begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & , \quad x \geq 0 \quad , \quad \beta > 0 \\ 0 & \text{لكل قيمة أخرى لـ } x \end{cases}$$

لنفترض أن X_1، X_2، ...، X_6 هي عينة عشوائية لتقدير β والتي تم من خلالها اختيار القيم التالية: 3.2، 4.5 ، 2.3 ، 1.9 ، 3.5 ، و 2.8 .

استخدم هذه القيم لتقدير المعلمة $\beta = \theta$.

الحل

من التوزيع الأسي نعرف ان المعلمة $\beta = E(X)$ و من ناحية أخرى ، يمكن تقدير القيمة المتوقعة بالمتوسط الحسابي بحيث:

$$\hat{\beta} = \frac{1}{6}(3.2 + 4.5 + 2.3 + 1.9 + 3.5 + 2.8) = 3.033$$

بناءا على ذلك لدينا المقدر النقطي لـ

كما ان فضاء المعلمة للتوزيع الأسي هي: $\Omega = (0, \infty)$ و 3.033 تنتمي الى Ω.

3.2. المقدر غير المتحيز

لا يمكن توقع القيمة المحددة لتقدير المعلمة لأنها تعتمد أيضا على الإحصائيات المستخدمة. على سبيل المثال إذا كان عدد الطلاب لديهم متوسط علامة قدره 6.5 في مقياس الاحصاء 3، واذا اخذنا عينة عشوائية تتكون من ثلاث طلاب X_1 ، X_2 ، و X_3 بالقيم التالية: 3 ، 6 ، و 6 يتم اعتبارها لتقدير المعلمة نتج عن ذلك:

$$m = \frac{1}{3}(3 + 6 + 6) = 5$$

أي أن متوسط الإحصائية يختلف عن متوسط المعلمة بمقدار 1.5. بينما وسيط الإحصائية هو 6 ويختلف عن المعلمة بمقدار 0.5 فقط. بهذه الطريقة يقدر وسيط الإحصائية المعلمة بشكل أفضل. ولكن ماذا سيحدث إذا كانت العلامات لتقدير المعلمة في الحالة الثانية لعينة عشوائية تشمل القيم:4 ، 4 ، 10؟

سيكون لديها:

$$m = \frac{1}{3}(4 + 4 + 10) = 6$$

وبالتالي يختلف هنا متوسط الإحصائية عن المعلمة بمقدار 0.5 وحدة، بينما يختلف الوسيط (الذي يساوي 4) ، بمقدار 2.5 وحدة.

أي انه في الحالة الثانية فإن المتوسط يقدر المعلمة بشكل أفضل. لذلك تثار الأسئلة حول: هل هناك مقدر نقطي أفضل من غيره؟ ما هي الخصائص التي يجب أن يفي بها أفضل المقدر؟ وغيرها من الاسئلة.

لذلك يجب أن تفي المقدرات الجيدة بخصائص معينة تتمثل في:

- المقدر الذي تنتمي قيمه دائما إلى فضاء المعلمة.
- تكون القيمة المتوقعة لتوزيع المقدر مساوية للمعلمة.
- الكفاءة: تتعلق كفاءة مقدر ما بمقدار التباين لتوزيع المعاينة للإحصائية، فاذا كان لمقدرين (إحصائيتين) نفس المتوسط نقول عن المقدر ذو توزيع المعاينة الأقل تباينا أنه الأكثر كفاءة.
- التقارب: كما ان نقول عن مقدر أنه متقارب إذا كان يؤول إلى قيمة المعلمة المقدرة عندما يؤول حجم العينة إلى ما لا نهاية.

ومنه يمكن تعريف المقدر غير المتحيز كما يلي:

تعريف

المقدر $\widehat{Q} = U(X_1, \dots X_n)$ لـ $g(\theta)$ للمعلمة θ يسمى المقدر غير المتحيز لـ $g(\theta)$ إذا كان:

$$E_p(\widehat{Q}) = g(\theta)$$

في الحالة المغايرة $(E_p(\widehat{Q}) \neq g(\theta))$ يُعرف بالمقدر المتحيز لـ $g(\theta)$

مثال 2–3

لتكن X_1، X_2، ...، X_5 عينة عشوائية . أي من مقدرات μ التالية غير متحيزة:

$$T_1 = \frac{X_1+X_2+X_5}{3} \quad , \quad T_2 = \frac{X_1+X_2+\cdots+X_5}{10} \quad , \quad T_3 = \frac{X_1+X_2+X_3-X_4+X_5}{3}$$

الحل

للتحقق من المقدرات غير المتحيزة سيتم استخدام التعريف والخصائص الخطية للقيمة المتوقعة.

$$E(T_1) = E\left[\frac{X_1+X_2+X_5}{3}\right] = \frac{1}{3}E[X_1+X_2+X_5]$$

$$= \frac{1}{3}[E(X_1)+E(X_2)+E(X_3)] = \frac{1}{3}(3\mu) = \mu$$

$$E(T_2) = E\left[\frac{X_1+X_2+\cdots+X_5}{10}\right] = \frac{1}{10}E[X_1+X_2+\cdots+X_5]$$

$$= \frac{1}{10}[E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5)] \quad = \frac{1}{10}(5\mu) = \frac{1}{2}\mu \neq \mu$$

$$E(T_3) = E\left[\frac{X_1+X_2+X_3-X_4+X_5}{3}\right] = \frac{1}{3}E[X_1+X_2+X_3-X_4+X_5]$$

$$= \frac{1}{3}[E(X_1)+E(X_2)+E(X_3)-E(X_4)+E(X_5)]$$

$$= \frac{1}{3}(\mu+\mu+\mu-\mu+\mu) = \frac{1}{3}(3\mu) = \mu$$

أكثر المقدرات غير المتحيزة شيوعا يمكن تلخيصها في الجدول الموالي:

الجدول رقم 2–1: أكثر المقدرات غير المتحيزة شيوعا

	المعلمة	**الاحصائية غير المتحيزة**
المتوسط	μ	m
الفرق بين متوسطين	$\mu_x - \mu_y$	$m_x - m_y$
التباين	σ^2	S_{n-1}^2
النسبة	p	p'
الفرق بين نسبيتين	$p_x - p_y$	$p'_x - p'_y$

2.4. مقدرات غير متحيزة لتوزيعات معينة

في الأمثلة السابقة ، للتحقق مما إذا كان المقدر غير متحيز أم لا ، لا يتطلب ذلك معرفة توزيع المتغير لأنه كان من الضروري فقط اللجوء إلى خطية عامل القيمة المتوقعة، لكننا بحاجة إلى معرفة توزيع المتغير لحساب قيمته المتوقعة.

مثال 2–4

لنفترض أن X_1، X_2، ...، X_n عينة عشوائية من مجتمع معين حيث:

$$\begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & , \quad x \geq 0 \quad , \quad \beta > 0 \\ 0 & \text{لكل قيمة أخرى لـ } x \end{cases}$$

أثبت أن $T = \bar{X}$ هو مقدر غير متحيز لـ β.

الحل

حتى نثبت ان $T = \bar{X}$ هو مقدر غير متحيز لـ β، سنقوم بحساب قيمته المتوقعة.

رأينا في الاقسام السابقة لهذه المطبوعة ان:

$$E(\bar{X}) = \mu$$

بما انه توزيع أسي فان: $E(X) = \beta$

وبهذا نستنتج:

لتكن T مقدرا للمعلمة $g(\theta)$، الدالة التي تمثل الفرق بين القيمة المتوقعة للمقدر والمعلمة تسمى التحيز:

$$S(\theta) = E(T) - g(\theta)$$

3. التقدير بمجال

نادرا ما يتساوى تقدير النقطة والمعلمة المراد تقديرها، لذلك يتم بدلا من ذلك تحديد فترة أو مجال. ويشير التقدير بفترة إلى مدى يبين مجالا من القيم التي يحتمل أن تقع القيمة الفعلية المجهولة ضمنها (طالب محمد عوض، 2000) ، أي تقدير معلمة المجتمع بنقطتين يحددان مجال لقيمة المعلمة.

وبشكل عام لتحديد فترة الثقة لاي معلمة كانت، لا بد من ايجاد حدين (حد أدنى: Lower limit، و حد أقصى: Upper limit)، والصيغة الرياضية لتقدير حدود الثقة هي:

$$P(L \le X \le U) = 1 - \alpha$$

L : الحد الادنى لمجال الثقة

U : الحد الاقصى لمجال الثقة

$1 - \alpha$: مستوى الثقة

α: احتمال الخطأ ويسمى أيضا مستوى المعنوية

فمثلا اذا كانت لدينا علامات الطلبة في مقياس الاحصاء 3 محصورة بين 8 و13 بمستوى معنوية 5%، فان:

$$1 - \alpha = 1 - 0.05 = 0.95$$

أي ان مستوى الثقة يساوي 95%.

حدود مجال الثقة هي: الحد الادنى: 8 والحد الاقصى: 13 ، مجال الثقة (أو فترة الثقة) هو: [8,13]

1.3. مجال الثقة للمتوسط المجتمع

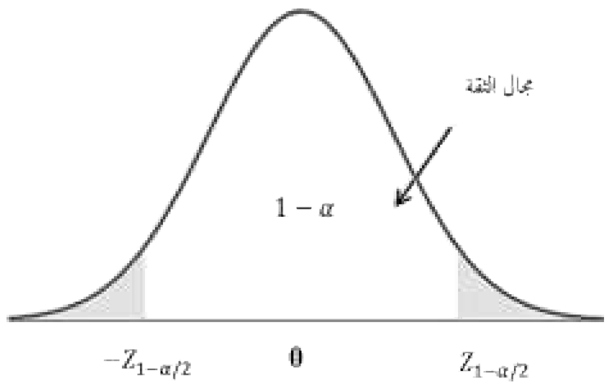
يقدر متوسط المجتمع μ من خلال الاحصائية المقابلة في العينة m، ومن أجل انشاء مجال الثقة حول متوسط المجتمع نميز بين حالتين:

1.1.3. تقدير متوسط المجتمع باستخدام التوزيع الطبيعي عندما يكون σ معروف

لتقدير مجال الثقة لمتوسط المجتمع المجهول μ، نسحب عينة عشوائية حجمها n. لنفرض ان المجتمع الذي سحبت منه هذه العينة يتبع التوزيع الطبيعي وتباينه σ^2 معلوم. أو في حالة العينات الكبيرة $(n \ge 30)$، حسب نظرية النهاية المركزية فإن يتبع التوزيع الطبيعي، ونكتب: $m \sim N(\mu, \sigma_m)$ أي أن:

$$\frac{m-\mu}{\sigma_m} \sim N(0,1)$$

ومنه يمكننا ان نمثل هذا التوزيع كما في الشكل الموالي:



يتصح من الشكل ان: $P\left(-Z_{1-\frac{\alpha}{2}} \leq Z \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$

وبما أن: $Z = \frac{m-\mu}{\sigma_m} = \frac{m-\mu}{\sigma/\sqrt{n}}$

ومنه فان:

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{m-\mu}{\sigma/\sqrt{n}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(-Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n} \leq m-\mu \leq Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\right) = 1-\alpha$$

وعليه يمكن تكوين فترة الثقة لمتوسط المجتمع كما يلي:

$$P\left(m - Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n} \leq \mu \leq m + Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\right)$$

ويسمى المقدار $ME = Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}$ بهامش الخطأ (Margin error)

ومنه يمكننا كتابة مجال ثقة لمتوسط المجتمع على النحو التالي:

$$P(m - ME \leq \mu \leq m + ME) = 1-\alpha$$

ملاحظة

في حالة تباين المجتمع مجهول وحجم العينة $n \leq 30$، سنستخدم الانحراف المعياري للعينة كتقدير للانحراف المعياري للمجتمع σ، وستكون فترة الثقة لمتوسط المجتمع μ كالتالي:

$$m + Z_{1-\frac{\alpha}{2}} S/\sqrt{n-1} \quad \text{أو} \quad m \pm Z_{1-\frac{\alpha}{2}} S/\sqrt{n}$$

كما أنه في حالة المعاينة نفادية، أو المجتمع محدود ($n<0.05N$) فإن مجال الثقة يكتب بالشكل التالي:

$$m \pm Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n} \cdot \sqrt{\frac{N-n}{N-1}}$$

مثال 2-5

ليكن مجتمع يتبع التوزيع الطبيعي انحرافه المعياري $\sigma = 2.5$، ومتوسطه μ. قدر فترة الثقة لمتوسط المجتمع انطلاقا من عينة عشوائية مسحوبة من نفس المجتمع حجمها $n = 20$ ومجموع مفرداتها 2200 عند مستوى المعنوية 5%.

الحل

بما ان العينة مسحوبة من مجتمع طبيعي تباينه معلوم فنتخدم التوزيع الطبيعي:

$$m \pm Z_{1-\frac{\alpha}{2}} {}^{\sigma}\!/\!_{\sqrt{n}}$$

$$m = \frac{\sum m_i}{n} = \frac{2200}{20} = 110$$

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$

ومنه فمجال الثقة هو: $\left[110 - (1.96)\left(\frac{2.5}{\sqrt{20}}\right); 110 + (1.96)\left(\frac{2.5}{\sqrt{20}}\right)\right]$

2.1.3. تقدير متوسط المجتمع عندما يكون σ مجهول

في الحقيقة عندما يكون متوسط المجتمع μ مجهولا فإن تباين المجتمع σ^2 أيضا يكون مجهولا، وعندما يكون المجتمع موزعا توزيعا طبيعيا و لكن σ غير معلومة وحجم العينة n صغير فإنه لا يمكننا استخدام التوزيع الطبيعي لتحديد مجال الثقة لمتوسط المجتمع. في هذه الحالة يمكننا استخدام توزيع ستيودنت Student.

وعند استخدام توزيع Student لتقدير مجال الثقة لـ μ، يتم تعويض معلمة المجتمع σ المجهولة بمقدرها الانحراف المعياري للعينة $\hat{S}$. ومن العلاقات السابقة للحصول على المتغير المعياري يمكن كتابة:

$$T = \frac{m-\mu}{\hat{S}/\sqrt{n}} \sim t_{n-1}$$

وعليه فإن: $P\left(-t_{(\frac{\alpha}{2},n-1)} \leq T \leq t_{(\frac{\alpha}{2},n-1)}\right) = 1-\alpha$

بالتعويض نجد: $P\left(-t_{(\frac{\alpha}{2},n-1)} \leq \frac{m-\mu}{\hat{S}/\sqrt{n}} \leq t_{(\frac{\alpha}{2},n-1)}\right) = 1-\alpha$

وبالتالي: $P\left(m - t_{(\frac{\alpha}{2},n-1)} \hat{S}/\sqrt{n} \leq \mu \leq m + t_{(\frac{\alpha}{2},n-1)} \hat{S}/\sqrt{n}\right) = 1-\alpha$

وعليه فمجال الثقة لمتوسط المجتمع في حالة العينات الصغيرة يكتب كما يلي:

$$P\left(m - t_{(\frac{\alpha}{2},n-1)} \hat{S}/\sqrt{n} \leq \mu \leq m + t_{(\frac{\alpha}{2},n-1)} \hat{S}/\sqrt{n}\right)$$

مثال 2–6

تدعي إحدى المؤسسات المصنعة لآلات تحضير المشروبات الغازية أن منتجاتها تنتج ما معدله 240 مل في 99.9% من الحالات. يقرر المشتري فحص إحدى الآلات لذلك يأخذ عينة عشوائية من 20 مشروبا غازيا، يحصل منها على القياسات التالية:

243	250	240	248	245	250	238	246	252	247
246	240	250	249	248	240	245	247	238	248

حدد ما إذا كانت ادعاءات هذه المؤسسة صحيحة.

الحل

لتحديد مجال الثقة نحتاج إلى حساب المتوسط غير المتحيز وتباين البيانات:

$$m = 245.5$$
$$S_{n-1}^2 = 18.37$$
$$S_{n-1} = 4.29$$

وبما ان حجم العينة هو 20 وتباين المجتمع غير معروف فسنستخدم توزيع Student.

لهذا ،نقوم بحساب القيمة $t_{\frac{\alpha}{2}}$ مع درجة الحرية تساوي $n - 1 = 20 - 1 = 19$

كما ان: $1 - \alpha = 0.999$

وبالتالي: $\alpha = 0.001$ و $\frac{\alpha}{2} = \frac{0.001}{2} = 0.0005$

من جدول توزيع t–Student: $t_{0.0005,19} = 3.883$

وبالتالي: $245.5 - 3.883\, {}^{4.29}/_{\sqrt{20}} \leq \mu \leq 245.5 + 3.883\, {}^{4.29}/_{\sqrt{20}}$

ومنه فمجال الثقة لمتوسط المجتمع هو: $241.77 \leq \mu \leq 249.22$

ملاحظة

كما يتضح يقتصر استخدام العلاقة السابقة على جداول توزيع **t-Student**، وبشكل عام عندما يكون حجم العينة $(n \leq 30)$. لذلك ، فإن السؤال الذي يطرح نفسه ، ماذا نفعل عندما تكون الانحراف المعياري مجهول وحجم العينة أكبر من 30؟

عندما تكون العينة كبيرة يمكن استخدام النتيجة التي تقول: يقترب توزيع **t-Student** من التوزيع الطبيعي عندما تكون درجات الحرية كبيرة. أي عندما يكون حجم العينة كبيرا.

3.2.3. أمثلة متنوعة لتقدير المتوسط

لقد رأينا حالات مختلفة لإجراء تقدير للمتوسط بفواصل زمنية. يبقى تحليل الأمثلة حيث يتم إجراء تعديلات معينة على التعبيرات التي تتدخل في كل حالة من الحالات الثلاث للمتوسط.

نفترض أن θ معلمة و $\hat{\theta}$ مقدرها. يشير خطأ التقدير إلى الفرق الذي يمكن أن يوجد بين θ و $\hat{\theta}$ بغض النظر عما إذا كانت $\hat{\theta}$ أكبر أو أقل من المعلمة. لذلك ، إذا أشرنا إلى الخطأ بواسطة ε، فسيكون لدينا:

$$\varepsilon = \left|\theta - \hat{\theta}\right|$$

من التعريف أعلاه والصيغ لتقدير مجال الثقة، من الممكن الحصول على خطأ التقدير في كل حالة. على سبيل المثال ، عندما يُعرف الانحراف المعياري:

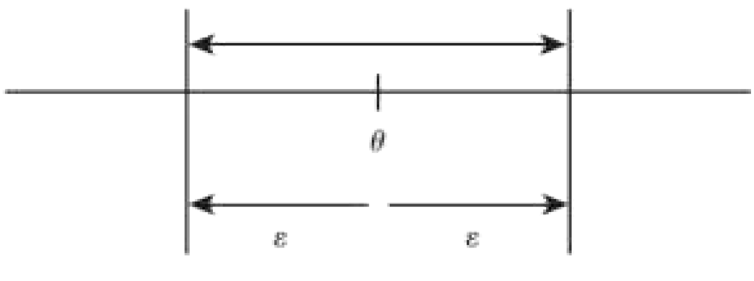
$$m - Z_{1-\frac{\alpha}{2}} \sigma_0/\sqrt{n} \leq \mu \leq m + Z_{1-\frac{\alpha}{2}} \sigma_0/\sqrt{n}$$

بطرح متوسط العينة يصبح: $$Z_{1-\frac{\alpha}{2}} \sigma_0/\sqrt{n} \leq \mu - m \leq Z_{1-\frac{\alpha}{2}} \sigma_0/\sqrt{n}$$

من خلال تحديد القيمة المطلقة لدينا: $$\varepsilon = |\mu - m| < Z_{1-\frac{\alpha}{2}} \sigma_0/\sqrt{n}$$

ويمكن تمثيل ذلك بيانيا كما يلي:

مجال الثقة حيث يمكن تحديد موقع مقدر θ



طول يساوي حجم الخطأ إلى يسار ويمين المعلمة

ولكن كيف نجد حجم العينة المناسب؟

بمجرد معرفة حجم خطأ التقدير وتباين المجتمع ودرجة الثقة أو قيمة $Z_{1-\frac{\alpha}{2}}$ ، يمكن حساب الحد الأدنى لحجم العينة الذي يلبي خطأ التقدير.

$$n \geq \left(\frac{Z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2 \sigma_0 \text{ o } n = \left[\left(\frac{Z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2 \sigma_0\right] + 1$$

الحالة 01: الانحراف المعياري معروف

$$|\mu - m| \le Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}} \quad ; \quad \varepsilon = Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}} \ \text{o} \ n \ge \left(\frac{Z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2 \sigma_0$$

الحالة 02: الانحراف المعياري غير معروف

$$|\mu - m| \le t_{\frac{\alpha}{2}} S_{n-1}/_{\sqrt{n}} \quad ; \quad \varepsilon = t_{\frac{\alpha}{2}} S_{n-1}/_{\sqrt{n}}$$

الحالة 02: الانحراف المعياري غير معروف، عينات

$$|\mu - m| \le Z_{1-\frac{\alpha}{2}} S_{n-1}/_{\sqrt{n}} \quad ; \quad n \ge \left(\frac{Z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2 S_{n-1}^2$$

غالبا ما تستخدم أخطاء التقدير لحساب الحد الأدنى لحجم العينة الضروري لانحراف متوسط العينة عن متوسط المجتمع بمقدار أقل من أو يساوي ε . تتوافق هذه الصيغ مع تلك التي تم الحصول عليها الفصل السابق مع:

$$\varepsilon = |\mu - m|$$

اذا اردنا حساب حجم العينة عن طريق خطأ التقدير، في كل من الحالة الثانية والثالثة، توجد مشكلة معرفة قيمة S_{n-1}. لهذا فحجم العينة مطلوب. في الحالة الثالثة، يمكن حل المشكلة عن طريق أخذ تقدير S_{n-1} مع حجم معين ، ولكن لا يمكن استخدام الحالة الثانية لأنها تتطلب أيضا حجم العينة لحساب القيمة $t_{\frac{\alpha}{2}}$.

اذن كيف نحسب حجم العينة عندما يُعطى طول الفترة؟ من تعريف خطأ التقدير يمكن ان نلاحظ أن طول فترة الثقة يساوي: طول المجال = 2ε

مكن التحقق من ذلك بسهولة نظرا لأن طول الفترة يساوي الحد الأعلى مطروحا منه الحد الأدنى.

$$m - Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}} \le \mu \le m + Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}}$$

يتم طرح الحدود ونحصل على:

$$m + Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}} - \left(m - Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}}\right) = 2Z_{1-\frac{\alpha}{2}} \sigma_0/_{\sqrt{n}} = 2\varepsilon$$

2.3. مجال الثقة للفرق بين متوسطي مجتمعين

على سبيل المثال لمقارنة متوسط الإنتاجية لخطي إنتاج أو عمليتين قد نفكر في أخذ عينة من كل مجموعة ودراسة متوسط الفرق بينهما. وعند مقارنة مجموعتين من مجتمعين هناك عدة حالات ولكننا سنتطرق هنا الى حالتين فقط.وقبل إنشاء فترات الثقة للفرق بين المتوسطين، يجب التأكيد على أن الهدف الأساسي لهذا هو مقارنة متوسطي مجتمعين.

1.2.3. فترات الثقة للفرق بين متوسطي مجتمعين بتوزيع طبيعي والانحرافين معروفين

بنفس الطريقة كما في حالة المجتمع بالتوزيع الطبيعي، تبدأ مقارنة الفرق بين متوسطي مجتمعين عندما تُعرف تباينات المجتمعين ، كما هو موضح فيما يلي:

$$(m_1 - m_2) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (m_1 - m_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال 2–7

تم مقارنة قوة الشد لنوعين من الخيوط اللولبية، حيث يتم اختبار 12 قطعة من كل نوع من الحبال في ظل ظروف مماثلة وتم الحصول على النتائج المبينة في الجدول الموالي (بالكيلوجرام):

النوع	1	2	3	4	5	6	7	8	9	10	11	12
الاول	68	70	72	69	71	72	70	69	75	69	70	71
الثاني	75	73	73	68	68	67	69	75	74	68	73	74

ذا كانت μ_1 و μ_2 هي متوسطي قوة الشد و$\sigma_1^2 = 5$ و $\sigma_2^2 = 10$، الانحرافين المعياريين لقوة الشد من النوع الأول والثاني على التوالي:

- قدر بفترة ثقة 90% $\mu_1 - \mu_2$، بافتراض أن التوزيع طبيعي وأن العينات مستقلة.

- عند مستوى ثقة 90% ، هل يمكن القول أن النوع الثاني لديه قوة شد أعلى من النوع الأول؟

الحل

$$m_2 = 71.42 \quad \text{و} \quad m_1 = 70.5$$

الان نقدر فترة الثقة كما يلي:

$$(70.5-71.42)-1.645\sqrt{\frac{5}{12}+\frac{10}{12}}\leq\mu_1-\mu_2\leq(70.5-71.42)-1.645\sqrt{\frac{5}{12}+\frac{10}{12}}$$

$$-2.76\leq\mu_1-\mu_2\leq 0.92$$

أي أن ألفرق بين متوسطي قوة الشد من النوع الأول والنوع الثاني يقع في الفترة (-2.76، 0.92) مع احتمال 0.90.

وبما أن مجال الثقة يشمل قيم سالبة وقيم موجبةمع مستوى ثقة 90% ، فلا يمكن أن نصمن أن الخيط من النوع الثاني لديه قوة شد أكبر من النوع الأول، حيث أنه مع النتيجة التي تم الحصول عليها استنتج أن كلا متوسطي المقاومة متساوية بمستوى ثقة 90%.

2.2.3. فترات الثقة للفرق بين متوسطي مجتمعين عنندما تكون الانحرافات غير معروفة

هنا نميز بين حالتين:

الحالة الاولى: حالة تساوي التباينين

في هذه الحالة نقدر الفرق بين متوسطي مجتمعين كما يلي:

$$(m_1-m_2)-t_{\frac{\alpha}{2}}(S_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq\mu_1-\mu_2\leq(m_1-m_2)+(m_1-m_2)+t_{\frac{\alpha}{2}}(S_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

حيث: (S_p) تمثل التقدير الشائع للانحراف المعياري للمجتمع وتحسب كما يلي:

$$(S_p)=\sqrt{\frac{(n_1-1)\hat{S}_1^2+(n_2-1)\hat{S}_2^2}{n_1+n_2-2}}$$

درجة الحرية هناا تساوي: (n_1+n_2-2)

$\hat{S}_1^2$ و $\hat{S}_1^2$: التباينات غير المتحيزة للعينتين

الحالة الثانية: حالة التباينين غير متساويين

هنا لتقدير الفرق بمجال ثقة نستخدم العلاقة التالية:

$$(m_1-m_2)-t_{\frac{\alpha}{2}}\sqrt{\frac{S_1}{n_1}+\frac{S_2}{n_2}}\leq\mu_1-\mu_2\leq(m_1-m_2)+t_{\frac{\alpha}{2}}\sqrt{\frac{S_1}{n_1}+\frac{S_2}{n_2}}$$

مع درجة الحرية المعتمدة لاستخراج القيمة الجدولية من جدول توزيع t-Student هي:

$$v = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\left(\frac{S_1^2}{n_1}\right)^2\left(\frac{1}{n_1-1}\right) + \left(\frac{S_2^2}{n_2}\right)^2\left(\frac{1}{n_2-1}\right)}$$

3.3. مجال الثقة للنسبة والفرق بين نسبتين

1.3.3. مجال الثقة للنسبة

في الفصل السابق تعرفنا على مفهوم النسبة وقمنا بحساب متوسطها وتبيانها وانحرافها المعياري. وقد تمت الإشارة إلى النسبة بـ: p' والمعرفة كما يلي: $p' = p$

حيث: $E(p') = \mu_{p'} = p$ و $\sigma_{p'}^2 = \frac{p(1-p)}{n}$

لتقدير النسبة يمكننا استخدام نظرية النهاية المركزية، فإذا كانت p' هي نسبة النجاحات في إجراء عينة عشوائية بالحجم $n > 30$ ، فيمكن ايجاد فترة ثقة لـ: $(1-\alpha)100\%$ لمتوسط النسبة في المجتمع كما يلي:

$$p' - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq p' + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

ونوضح من خلال المثال التالي استخدام هذه العلاقة:

مثال 2–8

في عينة عشوائية تشمل 100 عميل محتمل، صرح 70 منهم انهم يفضلون المنتوج A. اوجد فترة ثقة عند مستوى 95 % لنسبة جميع العملاء المحتملين الذين يفضلون نفس المنتوج.

الحل

بالنسبة لمجال الثقة للنسبة سنقوم أولا بحساب قيمة نسبة الأشخاص الذين يفضلون هذا المنتوج:

$p' = \frac{70}{100} = 0.7$ $\qquad$ $q' = 1 - p' = 1 - 0.7 = 0.3$

مستوى الثقة في هذه الحالة هو 95% ، وبالتالي: $1 - \alpha = 0.95$

وبالعودة لجدول التوزيع الطبيعي المعياري لدينا: $Z_{1-\frac{\alpha}{2}} = 1.96$

ومنه:

$$0.70 - 1.96\sqrt{\frac{0.70.0.30}{100}} \leq p \leq 0.70 + 1.96\sqrt{\frac{0.70.0.30}{100}}$$

$$0.6102 \leq p \leq 0.7898$$

وقول أن نسبة العملاء الذين يفضلون المنتوج A تتراوح بين 61.02 و 78.98% بمستوى ثقة 95%.

2.3.3. مجال الثقة للفرق بين نسبتين (للعينات الكبيرة)

أما فيما يخص الفرق بين نسبتين فغالبا ما تستخدم النسب لدراسة التفضيلات بين منتجين أو التحسين بين عمليتين إنتاجيتين، وغيرها. في جميع الحالات التي يتم فيها تقسيم المجتمع إلى فئتين يمكن استخدام الفرق في النسب للعينات الكبيرة (انظر نظرية النهاية المركزية).

إذا كانت p_1' و p_2' نسبتي عينتين عشوائيتين حجمهما على التوالي: n_1 و n_2 ($n_1 > 30$ و $n_2 > 30$) ومستوى الثقة $(1-\alpha)100\%$، فانه تقدير الفرق بين p_1 و p_2 $(p_1 - p_2)$ يكون كما يلي:

$$(p_1' - p_2') - Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}} \leq p_1 - p_2 \leq (p_1' - p_2') + Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}}$$

مثال 2–9

تم إجراء اختبار سريري لمعرفة ما إذا كان تلقيحا معينا يؤثر على حدوث المرض. تم الاحتفاظ بعينة من 1000 فأر في بيئة خاضعة للرقابة لمدة عام واحد، تم تلقيح 500 منها. في المجموعة التي لم يطبق عليها التلقيح ، كانت هناك 120 حالة من هذا المرض ، بينما في المجموعة التي تم تلقيحها، أصيبت 90 حالة منها. إذا كان p_1 يمثل احتمال حدوث المرض في الفئران غير المعالجة و p_2 هو احتمال الإصابة بعد التلقيح، قدر بفترة ثقة 95% لكل من p_1 و p_2، وحدد ما إذا كانت نسبة حدوث المرض أعلى في الفئران غير المعالجة.

الحل

نظرا لأن p_1 هو احتمال حدوث المرض في الفئران غير المعالجة، وبما أنه من بين 500 فأر غير معالج أصيب 120 أي:

$$n_1 = 500 \quad و \quad q_1' = 0.76 \quad و \quad p_1' = \frac{120}{500} = 0.24$$

بنفس الطريقة بالنسبة للفئران التي تم تلقيحها لدينا:

$$n_2 = 500 \quad و \quad q_2' = 0.82 \quad و \quad p_2' = \frac{90}{500} = 0.18$$

من جدول التوزيع الطبيعي لدينا القيمة الجدولية تساوي *1.96* ومنه نجد مجال الثاقة كما يلي:

$$(p_1' - p_2') - Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}} \leq p_1 - p_2 \leq (p_1' - p_2') + Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}}$$

$$(0.24-0.18)-1.96\sqrt{\frac{0.24*0.76}{500}+\frac{0.18*0.82}{500}}\leq p_1-p_2$$

$$\leq(0.24-0.18)+1.96\sqrt{\frac{0.24*0.76}{500}+\frac{0.18*0.82}{500}}$$

$$0.0096\leq p_1-p_2\leq 0.1104$$

4.3. مجال الثقة للتباين

نفرض أن المجتمع يتبع التوزيع الطبيعي، وبما أن المتغيرة $\frac{S_2}{\sigma^2}$ تتبع توزيع مربع كاي درجة حرية $(n-1)$ فلتقدير تباين المجتمع بمجال ثقة نستخدم الخاصية التالية:

$$\frac{(n-1)\hat{S}^2}{\sigma^2}\sim x^2(n-1)$$

وبالتالي لتقدير التباين بفترة ثقة سنعتمد على $\frac{(n-1)\hat{S}^2}{\sigma^2}$.

فإذا كانت X_1، X_2، ... X_n عينة عشوائية حجمها n، مأخوذة من مجتمع يتبع التوزيع الطبيعي بمتوسط μ وتباين σ^2، مع S^2 قيمة تباين العينة، فان مجال الثقة يمكن ايجاده بواسطة العلاقة:

$$\frac{(n-1)\hat{S}^2}{x^2_{\alpha/2}}\leq\sigma^2\leq\frac{(n-1)\hat{S}^2}{x^2_{1-\alpha/2}}$$

ومنه نستنتج مجال الثقة للانحراف المعياري كما يلي:

$$\frac{\sqrt{n}S}{\chi_{0.975}}\leq\ \sigma\ \leq\frac{\sqrt{n}S}{\chi_{0.025}}$$

أو

$$\frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}}\leq\ \sigma\ \leq\frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$$

مثال 2-10

قام عالم الأنثروبولوجيا بقياس العرض (بالسنتيمتر) لعينة عشوائية من تسع جماجم من أفراد قبيلة معينة وكانت النتائج كما يلي:

13.3 ، 14.2 ، 13.5 ، 16.7 ، 11.1 ، 13.1 ، 13.0 ، 12.2 ، 13.0

قدر مجال الثقة عند مستوى ثقة 95% للتباين.

الحل

نحسب أولا التباين غير المتحيز للبيانات ونجد:

$$\hat{S}^2 = 2.33$$

كما أن: $1 - \alpha = 0.95$ وبالتالي: $\alpha/_2 = 0.025$ و $1 - \alpha/_2 = 0.975$

من جدول توزيع مربع كاي مع درجة حرية $n - 1 = 9 - 1 = 8$ ، نحصل على:

$$x^2_{\alpha/_2, n-1} = x^2_{0.025,8} = 17.5345$$

و

$$x^2_{1-\alpha/_2, n-1} = x^2_{0.975,8} = 2.1797$$

ومنه فمجال الثقة محصور بين:

$$\frac{(9-1)2.33}{17.5345} \le \sigma^2 \le \frac{(9-1)33}{2.1797}$$

$$1.06 \le \sigma^2 \le 8.55$$

أي ، مع احتمال 0.95 فان تباين عرض الجماجم تتراوح بين 1.06 و 8.55 سم.

5.3. فترات الثقة لنسبة تبايني مجتمعين

عندما يكون لدينا مجموعتان طبيعيتان بالمعلمات التالية: (μ_1, σ_1^2) و (μ_2, σ_2^2) على التوالي، فيمكننا، كما رأينا سابقا، مقارنة الفرق بين متوسطيهما. الآن سوف ندرس فترات الثقة لنسبة تبايني مجتمعين. أي أننا سنحدد السكان الأكثر تباينًا ، ولهذا سوف نلجأ إلى نتيجة توزيع المعاينة التي تحدد:

$$\frac{\hat{S}_1^2\ \sigma_2^2}{\hat{S}_2^2\ \sigma_1^2}$$

التي لها توزيع F مع: $v_1 = n_1 - 1$ و $v_2 = n_2 - 1$، درجات الحرية في البسط والمقام، على التوالي. الصيغة التي سنستخدمها هنا لفترات الثقة للنسبة بين التباينين.

إذا كانت $\hat{S}_1^2$ و $\hat{S}_2^2$ تبايني غير متحيزة لعينتين عشوائيتين مستقلتين بأحجام n_1و n_2 من المجتمعين يتبعان توزيع طبيعي بالمعلمات : (μ_1, σ_1^2) و (μ_2, σ_2^2)، فإن مجال الثقة $(1 - \alpha)$ 100% لنسبة الفروق:

$$\sigma_1^2/_{\sigma_2^2}$$

تعطى من خلال:

$$(\frac{\hat{S}_1^2}{\hat{S}_2^2})\frac{1}{f\alpha_{/2}(v_1,v_2)} \le \frac{\sigma_1^2}{\sigma_2^2} \le (\frac{\hat{S}_1^2}{\hat{S}_2^2})f\alpha_{/2}(v_1,v_2)$$

حيث $f\alpha_{/2}(v_1,v_2)$هي قيمة توزيع F ، مع و $v_1 = n_1 - 1$ و $v_2 = n_2 - 1$ درجات الحرية للبسط وللمقام على التوالي. وسوف نقوم بتطبيق ذلك من خلال المثال التالي:

مثال 2–11

لتكن بيانات التجربتين التاليتين:

التجربة الاولى: 15.8 ، 12.7 ، 13.2 ، 16.9 ، 10.6 ، 18.8 ، 11.1 ، 14.3 ، 17.0 ، 12.5.

التجربة الثانية: 24.9 ، 23.6 ، 19.8 ، 22.1 ، 20.4 ، 21.6 ، 21.8 ، 22.5.

اذا فرضنا أن: و $\sigma_1^2 = \sigma_2^2$ ، قدر بمجال ثقة نسبة التباينين وحدد ما إذا كان الافتراض صحيحا بمستوى ثقة 90%.

الحل

نحيب تباينات العينتين نجد: $\hat{S}_1^2 = 7.50$ ، $n_1 = 10$ ، $\hat{S}_2^2 = 2.68$ ، $n_2 = 8$

ومن جدول توزيع F نستخرج القيم التالثة:

$$f_{0.05}(9,7) = 3.677 \quad ، \quad f_{0.05}(7,9) = 3.293$$

ةمنه فان مجالالثقة لنسبة التباينين هو:

$$\left(\frac{7.50}{2.68}\right)\frac{1}{3.677} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{7.50}{2.68}\right)3.293$$

$$0.76 \le \frac{\sigma_1^2}{\sigma_2^2} \le 9.22$$

من مجال الثقة للنسبة بين التباينن يمكن ان نلاحز أن القيمة 1 يشملها مجال الثقة. لذلك ، مع مستوى ثقة 90% ، فإن الافتراض بأن $\sigma_1^2 = \sigma_2^2$ له ما يبرره، بما أن:

$$\frac{\sigma_1^2}{\sigma_2^2} = 1 \in [0.76, 9.22]$$

وإذا ضربنا طرفي المساواة في σ_2^2 نحصل على $\sigma_1^2 = \sigma_2^2$.

خلاصة الفصل

في هذا الفصل تناولنا كيفية تقدير معالم المجتمع من خلال بيانات العينة وبعض خصائص المقدر الجيد. ولكن في كثير من الاحيان يحتاج الباحث في مرحلة ما من بحثه إلى اختبار فرضية أو أكثر بخصوص المجتمع المدروس. كاختبار فرضية بخصوص معدل الدخل في منطقة معينة، اختبار فرضية نسبة شفاء لدواء معين، وغيرها. ويمكن للباحث اختبار فرضياته من خلال صياغة فرضية عن المجتمع المدروس (أو المجتمعات المدروسة) ومن ثم محاولة الحصول على دليل إحصائي ينفي أو يثبت هذه الفرضية وذلك من خلال بيانات عينة (أو أكثر) عشوائية بسيطة. تخص الفرضية أحد معالم المجتمع كالمتوسط، النسبة أو التباين، ونعتمد في إثباتها أو رفضها على خصائص إحصائية معاينة مختارة. لذا سننتقل في الفصل الموالي الى كيفية اختبار فرضيات موضوعة حول معالم مجتمع أو أكثر.

الفصـل الثالث

اختبـار الفرضيـات

تمهيد

في الفصل السابق قمنا بمراجعة تقدير المعلمات باستخدام التقدير النقطي والتقدير بمجال. وسوف ننتقل الان لنوع آخر من طرق الاستدلال الاحصائي لزيادة المعرفة حول معلمات توزيع المجتمع قيد الدراسة، والتي تتبلور في وضع افتراضات أو تخمينات حول المعلمات التي يتم تحليلها. لذا سوف نتناول في هذا الفصل كيفية استنتاج سلوك خصائص المجتمع من خلال تطبيق طرق اختبار الفرضيات التي تساعد في عملية صنع القرار. وسنتعرف من خلال محتويات الفصل الى مفهوم اختبار الفرضيات، وأهميتها، بالاضافة الى أنواع الأخطاء في اختبار الفرضيات. من خلال هذا الفصل سندرس المفاهيم الأساسية لنظرية اختبار الفرضيات، ثم نتعامل مع الجزء المنهجي لاختبار الفرضيات، مما يساعد في اتخاذ القرار في المشكلات المتعلقة بالمجتمعات التي يصعب أو يستحيل تحليلها بالكامل. بطريقة مشابهة لما تم القيام به مع فترات الثقة، سنقوم بتحليل الجزء المنهجي على المجتمعات ذات التوزيع الطبيعي. أي أننا سنبدأ من العينة العشوائية بناءا على الصيغ التي سنقوم ببنائها وسنحصل على قواعد القرار لاختبارات الفرضيات المطلوبة.

1. أساسيات اختبار الفرضيات

نفترض أن لدينا مجموعة (مجتمع) قيد الدراسة تتمثل في علامات طلبة كلية العلوم الاقتصادية والتجارية وعلوم التسيير لجامعة خميس مليانة، والتي لا نعرف معلماتها ولكن نود أن نعرف ما إذا كان متوسط علامات الطلبة أكبر من 7.0

لتحليل صحة الفرضية نلاحظ اننا لا نعرف متوسط المجتمع والذي يمكننا من خلاله تخمين أنه أكبر من 7.0، بالرغم من اننا نعرف توزيع المجتمع. لتسهيل الامر، نختار عينة عشوائية من علامات الطلبة حجمها 15 طالب: والمتمثل في: 8 ، 4 ، 6 ، 7 ، 6 ، 9 ، 5 ، 7 ، 8 ، 10 ، 7 ، 3 ، 9 ، 8 ، 4.

نقوم بحساب متوسط هذه العلامات للحصول على تقدير نقطي للمعلمة، كما يلي:

$$\bar{x} = \frac{\sum x_i}{n} = 6.7$$

انطلاقا من النتيجة السابقة هل يمكن القول ان الفرضية غير صحيحة؟

لنفترض أننا اخترنا اخرى من نفس الحجم وتشمل العلامات التالية: 6 ، 10 ، 7 ، 5 ، 7 ، 9 ، 10 ، 8 ، 7 ، 9 ، 7 ، 8 ، 4 ، 6 ، 8.

في هذه الحالة متوسط العلامات يساوي 7.4، في هذه الحالة هل يمكن القول بأن الفرضية صحيحة ؟

تكمن إحدى المشكلات الرئيسية عند بدء دراسة اختبار الفرضيات في استيعاب أن عدم الامتثال أو الوفاء بالفرضية من خلال بيانات العينة لا يكفي بالنسبة لنا للقول إحصائيا أنه خاطئ أو صحيح. وبالتالي في العينة الاولى ، لا يمكن القول إحصائيا أن عينة عشوائية بحجم 15 والفرضية $\mu > 7$ خاطئة فقط لان متوسط العينة اقل (6.7). وبالطريقة نفسها لا يمكننا أن نؤكد مع العينة الثانية صحة الفرضية فقط لان المتوسط اكبر من 7.

تعريف

سوف نسمي الفرضية الإحصائية أي تأكيد يشير إلى معلمات مجتمع واحد أو أكثر.

يتكون اختبار الفرضية الإحصائية من البحث عن أدلة لتقرير ما إذا كنا سنرفض البيان الذي تم الإدلاء به أم لا. فمثلا بالنسبة للمصابيح الكهربائية يمكننا أن نفرض ان متوسط عمرها يتجاوز 750 ساعة، لنفترض الآن أنه تم اختيار عينة من هذه المصابيح واتضح أن متوسط عمرها كان 730 ساع. فهل ستكون هذه نتيجة دليل كافي للإشارة إلى أن الفرضية غير صحيحة؟

في اختبار الفرضيات، لا يمكن معرفة الحقيقة فيما يتعلق بالقرار الذي تم اتخاذه برفض الفرضية المقدمة أو قبولها إلا من خلال دراسة المجتمع بالكامل. لذلك يجب أن نعرف أن عدم رفض فرضية بناءا على عينة يشير إلى أنه من البيانات التي تم الحصول عليها لا يوجد دليل للقيام بذلك.

عند صياغة فرضية حول حدث معين وإجراء اختبار الفرضية (الرفض أو القبول) من المنطقي أن نعرف على أي أساس سيتم رفض الفرضية او قبولها(جلاطو، 2002).

كما نلاحظ أنه عند إنشاء فرضية سيكون هناك دائما فرضية أخرى تعارضها، بحيث يتم إعطاء الفرضيات المصاغة اسم الفرضيات الصفرية (فرضية العدم) والفرضية البديلة، والتي سنرمز لها بالرمزين H_0 و H_1 على التوالي. وهكذا تبدأ إحدى أولى المشكلات في دراسة اختبار الفرضيات والتي تتمثل في كيفية تحديد الفرضيات الصفرية والبديلة؟ للإجابة على هذا السؤال نحتاج إلى مزيد من المفاهيم.

1.1. مناطق الرفض والقبول

لنفترض أننا نواجه مشكلة متوسط عمر المصابيح حيث يكون لدى المجتمع سلوك يمكن وصفه من خلال دالة الكثافة $f\ (x\ ,\mu)$ والمعلمة μ لها فضاء معلمة $\Omega=[0,\infty]$ (μ متوسط عمر المصابيح والذي لا يمكن أن يكون سالب). بعد ذلك يمكننا تحديد اختبار الفرضيتين التاليتين:

$$H_0: \mu \leq 750$$
$$H_1: \mu > 750$$

أي أن فضاء المعلمة $\Omega=[0,\infty]$ مقسمة إلى منطقتين المنطقة المقابلة للمعلمة في الفرضية الصفرية والتي نرمز لها بالرمز ω و المنطقة المقابلة للمعلمة في الفرضية البديلة والتي نرمز لها بالرمز $\Omega-\omega$. بهذه الطريقة يمكننا إنشاء اختبار الفرضيتين السابقتين بطريقة أكثر عمومية وبما يعادل:

$$H_0: \mu \in \omega$$
$$H_1: \mu \in \Omega-\omega$$

حتى الآن لم نتحدث بعد عن المشكلة التي ستكون ذات أهمية عملية والتي تتمثل في: ما يجب القيام به عندما يكون لدينا بيانات فقط لتحديد الفرضيات الصحيحة أو ما الذي سنفهمه من خلال اختبار الفرضية.

لنفترض أن لدينا مشكلة اختبار الفرضية لنصف عمر المصابيح وأن كل مصباح له عمر موصوف بواسطة متغير عشوائي بدالة الكثافة $f\ (x\ ,\mu)$. من ناحية أخرى لدينا عينة عشوائية من هذه المتغيرات $x_1, x_2, \dots, x_n$ يرمز لها بالمتجه (الشعاع) X.

هنا سنحتاج إلى العمل ليس مع العينة العشوائية ولكن مع تحقيقها من خلال تمثيل R كمجموعة من جميع تحققات X: $\mathrm{R}=\{x|x\}$

لذلك في مشكلة عملية يمكننا تقسيم المجموعة R واتخاذ قرار بشأن صحة الفرضية الصفرية بناءا على نتائج الملاحظات.

تعريف

يسمى اختبار الفرضية H_0 مقابل $1\mathrm{H}$ تقسم R إلى مجموعتين، والتي نشير إليها بواسطة R_a و R_r ومنطقة القبول أو منطقة الرفض أو المنطقة الحرجة، على التوالي.

نلاحظ من خلال التعريف السابق أنه من الممكن إنشاء عملية يمكننا من خلالها اتخاذ قرار من عينة عشوائية فيما إذا كانت H_0 صحيحة أم لا. بهذا المعنى كيف نحدد ما إذا كان يجب رفض فرضية العدم أم لا؟ بالنسبة لاختبار الفرضية للمعلمة θ، بشكل عام:

$$H_0: \theta \in \omega$$
$$H_1: \theta \in \Omega - \omega$$

سيتم إعطاء قاعدة القرار بناءا على تحقق x على النحو التالي:

- نرفض H_0 إذا كانت x تنتمي الى R_r (منطقة الرفض أو المنطقة الحرجة).
- نقبل H_0 إذا كانت x تنتمي الى R_a (منطقة القبول)

بالنسبة للحالة الخاصة لمتوسط عمر المصابيح الكهربائية، لدينا:

$$H_0: \mu \leq 750$$
$$H_1: \mu > 750$$

إذن يمكننا اختيار القيمة $\mu_0 \in \Omega$ ، وليس فقط $\mu_0 = 750$، بحيث يتم إعطاء قسم R بواسطة:

$$R_a = \{x/x \in R \quad , \qquad \bar{x} \leq \mu_0\}$$
$$R_r = \{x/x \in R \quad , \qquad \bar{x} > \mu_0\}$$

لاحظ أنه لكل قيمة مختارة من $\mu_0 \in \Omega$ لدينا اختبار فرضية أو قسم من R. تسمى القيمة التي تقسم مناطق الرفض ومناطق القبول "بالقيمة الحرجة".

على سبيل المثال يمكننا اعتبار القيمة الحرجة $\bar{x} = 760$ ساعة، والتي يتم من خلالها إنشاء مناطق القبول ($\bar{x} \leq 760$) والرفض أو المنطقة الحرجة ($\bar{x} > 760$). اي انه في هذه الحالة تم اختيار الرقم 760 كمثال توضيحي لماهية منطقتي الرفض والقبول. ولكن أي من الاختبارات سيكون جيد؟

2.1. أنواع الأخطاء في اختبار الفرضيات

في القسم الفرعي السابق تعاملنا مع تعريف اختبار الفرضيات، وبناءا على قاعدة القرار ، قررنا رفض الفرضية الصفرية أم لا. لكننا نعلم أنه في الإحصاء، النتائج تعتمد دائما على الظروف العشوائية لتغير الظاهرة قيد الدراسة (Sedkaoui, 2018). لهذا السبب عند اتخاذ قرار بشأن صحة الفرضية الصفرية، من المحتمل أن نرتكب أحد الخطأين اللذين سنحاول أن يكونا صغيرين قدر الإمكان.

ملاحظة

هناك نوعين من الاخطاء: خطأ من النوع الأول عندما يتم رفض الفرضية الصفرية، على الرغم من أنها صحيحة بالفعل، والخطأ من النوع الثاني هو عندما تقبل الفرضية الصفرية، على الرغم من أنها خاطئة.

بالنظر إلى تعريف أهم خطأين في رفض أو قبول فرضية العدم، فإن السؤال الذي يطرح نفسه هو احتمال ارتكاب أخطاء من النوع الأول أو النوع الثاني؟

الإجابة على هذا السؤال تتطلب اعتبار الاختبار الذي يقلل احتمالات كلا النوعين من الأخطاء. وبالمثل فإن أفضل اختبار (إن وجد) سيكون هو الاختبار الذي يقلل احتمالات كلا الخطأين فيما يتعلق بجميع الاختبارات الممكنة الأخرى.

يبدو أن العثور على اختبار جيد هو مهمة بسيطة، ولكن مع ذلك عندما يتم تقليل احتمال حدوث أحد الأخطاء إلى الحد الأدنى، يزداد احتمال النوع الآخر من الخطأ. في الواقع عد الإجابة على الأسئلة أعلاه إحدى المشكلات الرئيسية في نظرية اختبار الفرضيات وتتطلب بعض الوقت وفهم أفضل للمشكلة. لهذا سنبدأ بإدراج ما يلي:

- احتمال (الخطأ من النوع الاول باستخدام $\mathrm{R}_r|\mathrm{H}_0$) =احتمال ارتكاب الخطأ من النوع الأول مع صحة الفرضية الصفرية.

وبالمثل لدينا:

- احتمال (الخطأ من النوع الثاني باستخدام $\mathrm{R}_a|\mathrm{H}_1$) =احتمال ارتكاب الخطأ من النوع الثاني مع صحة الفرضية البديلة.

نلاحظ أن حساب احتمالات الخطأ من النوع الثاني يمكن أيضا حسابه في منطقة الرفض عن طريق المكمل.

احتمال (الخطأ من النوع الثاني باستخدام $\mathrm{R}_a|\mathrm{H}_1$) = $1-$ احتمال (الخطأ من الثاني الاول باستخدام $\mathrm{R}_r|\mathrm{H}_1$)

وفيما يلي جدول يلخص علاقة القرار بطبيعة الفرضية:

الجدول رقم 3–1: أنواع الاخطاء الاحصائية في اختبارات الفروض

القرار	نوع الفرضية	
	$\mathbf{H_0}$ **صحيحة**	$\mathbf{H_1}$ **صحيحة**
قبول $\mathbf{H_0}$	قرار صحيح	خطأ من النوع الثاني
رفض $\mathbf{H_0}$	خطأ من النوع الاول	قرار صحيح

وللتعرف على الاخطاء من النوع الاول والثاني سنعرض المثالين التاليين:

مثال 3–1

لنفترض أنه في حالة العمر الافتراضي للمصابيح الكهربائية، فإن هذه المصابيح لها انحراف معياري يساوي 50 ساعة، إذا تم أخذ عينة من 49 مصباح مع الفرضيات:

$$H_0: \mu \leq 750$$
$$H_1: \mu > 750$$

حدد في منطقة الرفض لنصف عمر أكبر من 760 ساعة ، ثم احسب:

- احتمال ارتكاب خطأ من النوع الأول للحالة $\mu = 740$
- احتمال ارتكاب الخطأ من النوع الثاني للحالة التي يكون فيها $\mu = 755$.

الحل

لحساب الاحتمالات نعرف أن منطقة الرفض تعطى بواسطة $\bar{x} > 760$، من ناحية أخرى حجم العينة هو 49، لذلك يمكننا استخدام نظرية النهاية المركزية:

$$P(\bar{x} > 760|\mu \leq 750) = P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{760 - \mu}{50/\sqrt{49}}|\mu = 740) = \mathrm{P}(\mathrm{Z} > \frac{760 - 740}{50/\sqrt{49}}) = P(Z > 2.8) = 0.0026$$

بنفس الطريقة، لحساب احتمال الخطأ من النوع الثاني، نستخدم نظرية النهاية المركزية وحقيقة أن R_a معطى بواسطة $\bar{x} \leq 760$.

$$P(\bar{x} \leq 760|\mu > 750) = P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{760 - \mu}{50/\sqrt{49}}|\mu = 755) = \mathrm{P}(\mathrm{Z} > \frac{760 - 755}{50/\sqrt{49}}) = P(Z > 0.70) = 0.7580$$

يمكن أيضا حسابه كما يلي:

$$1-P(\bar{x}>760|\mu>750)=1-P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}>\frac{760-\mu}{50/\sqrt{49}}|\mu=755)$$

$$=1-\mathrm{P}(\mathrm{Z}>\frac{760-755}{50/\sqrt{49}})=0.7580$$

مثال 3–2

فترض أنه في المثال السابق، تم اعتبار منطقة الرفض لنصف عمر أكبر من 752 ساعة. احسب:

- احتمال ارتكاب خطأ من النوع الأول حيث $\mu=740$
- احتمال ارتكاب الخطأ من النوع الثاني التي يكون فيها $\mu=755$

الحل

نفس الطريقة نحصل على:

$$P(\bar{x}>752|\mu\leq 750)=P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}>\frac{752-\mu}{50/\sqrt{49}}|\mu=740)$$

$$=\mathrm{P}\left(\mathrm{Z}>\frac{760-740}{50/\sqrt{49}}\right)=\mathrm{P}(\mathrm{Z}>1.68)=0.0465$$

بنفس الطريقة، بواسطة $\bar{x}>752$.

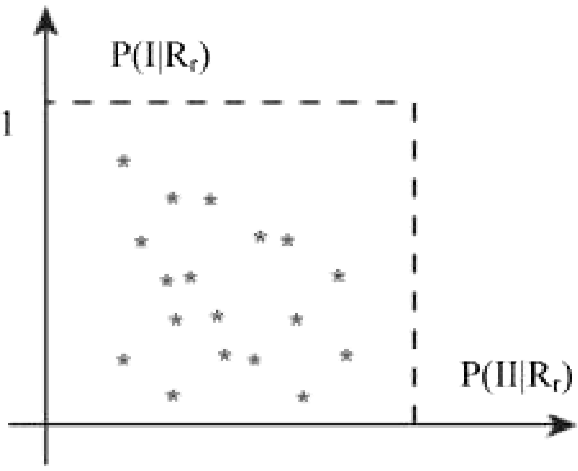
$$P(\bar{x}\leq 752|\mu>750)=P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}>\frac{752-\mu}{50/\sqrt{49}}|\mu=755)=\mathrm{P}(\mathrm{Z}>\frac{752-755}{50/\sqrt{49}})$$

$$=P(Z>-0.42)=0.3372$$

مقارنة:

مقارنة الاختبارين$\bar{x}>752$ و $\bar{x}>760$، اختبار القسم $\bar{x}>752$ و اختبار القسم $\bar{x}\leq 752$ أفضل. ونظرا لأن احتمال الخطأ من النوع الأول صغير (حوالي 5%)، في حين انخفض احتمال الخطأ من النوع الثاني بشكل كبير مقارنة بالقسم المقابل. كما يظهر من الشكل التالي:

الشكل رقم 3–1: قيم احتمالات الأخطاء من النوع الأول والثاني

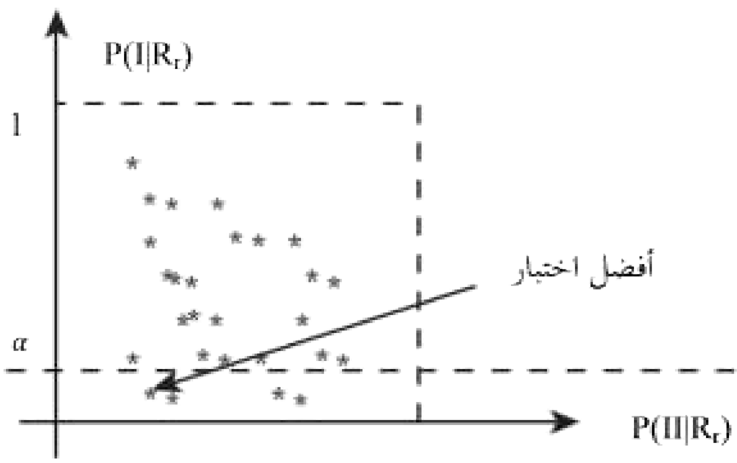


من المثال وتعريفات اختبار الفرضيات وتقسيم المجموعة R، فمشكلة العثور على اختبار جيد يمكن أن تعمل كتحديد قسم جيد للمجموعة Rمع تعريف واحد لمنطقة R_r. وبالتالي يتوافق كل اختبار مع زوج من الاحتمالات $\mathrm{P}(\mathrm{I}|\mathrm{R}_r)$ و $\mathrm{P}(\mathrm{II}|\mathrm{R}_r)$.

ويكون أفضل اختبار هو الذي تكون منطقة رفضه R_r^* (بحيث يكون زوج الاحتمالات $\mathrm{P}(\mathrm{I}|R_r^*)$ ، $\mathrm{P}(\mathrm{II}|R_r^*)$) هو الأقرب إلى أصل الإحداثيات، ومنه يمكن اعطاء التعريف التالي:

تعريف

يعرف اختبار α باختبار R_r الذي يحقق: $\mathrm{P}(\mathrm{I}|\mathrm{R}_r) \leq \alpha$ ($\alpha \in [0,1]$)، بالإضافة إلى ذلك، إذا كان اختبارα يحتوي على الحد الأدنى من احتمال الخطأ من النوع الثاني، فسيتم اعتباره الاختبار الأقوى. اي نختار الخط الذي يحتوي على أقل قيمة لاحتمال الخطأ من النوع الثاني.



3.1. الاختبار من طرفين والاختبار من طرف واحد

تجدر الاشارة بأن الفرضية البديلة هي التي تحدد نوع الاختبار، حيث نميز بين نوعين من الاختبارات، اختبار من طرفين، واختبار من طرف واحد.

في حالة اختبار من طرفين (Two Tails) يتم صياغة فرضية العدم (الصفرية) كما يلي:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

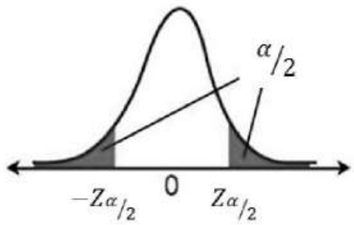
فمثلا اذا اردنا اجراء الاختبار على متوسط أجور الاساتذة في جامعة خميس مليانة، والذي يقدر بـ: 40000 دينار جزائري، حسب مسؤول قسم المحاسبة. وللتأكد من صحة ذلك نقوم باختيار عينة عشوائية وحساب متوسط الاجور في العينة، وفي هذه الحالة يتم صياغة الفرضية الصفرية التالية:

$$H_0: \mu = 40000$$

وبالتالي تكون الفرضية البديلة هي:

$$H_1: \mu \neq 40000$$

ويمكن تمثيل ذلك بيانيا كما يلي:



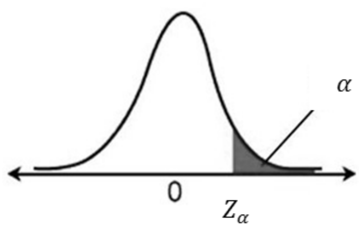
وتسمى المنطقة المحصورة بين: $-Z\alpha/_2$ و $Z\alpha/_2$ منطقة القبول $-Z\alpha/_2 < Z < Z\alpha/_2$ بينما المناطق المظللة (بالأحمر) تسمى منطقتي الرفض.

في حالة اختبار من جانب واحد (One Tail) تكون جهة الرفض إما من اليمين أو من اليسار.

فاختبار من طرف واحد من الجهة اليمنى، واستنادا على المثال السابق تصاغ الفرضية البديلة كما يلي:

$$H_1: \mu > 40000$$

ونمثل ذلك بيانيا كما يلي:

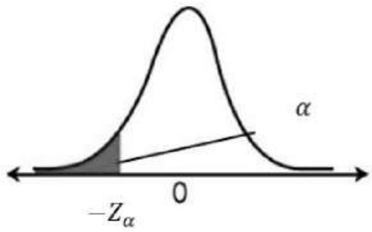


ومنطقة القبول هنا هي: $Z < Z_\alpha$

اما اختبار من طرف واحد من الجهة اليسرى، فمن المثال السابق، الفرضية البديلة تصاغ على النحو التالي:

$$H_1: \mu < 40000$$

بيانيا تمثل على الشكل التالي:



حيث أن منطقة القبول في هذه الحالة هي: $Z > -Z_\alpha$

2. اختبار الفرضيات لمعلمات مجتمع بتوزيع طبيعي

كما في حالة فترات الثقة، سنقوم بتحليل معلمات لتوزيع الطبيعي والمتوسط والتباين.

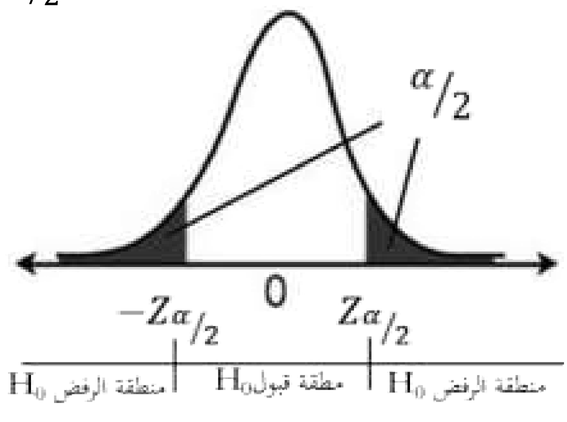
1.2. اختبار فرضيات متوسط المجتمع عندما يكون التباين معلوم

في حالة مجتمع يتبع التوزيع الطبيعي ومعلوم التباين، فان احصائية الاختبار Z تقترب من التوزيع الطبيعي المعياري.

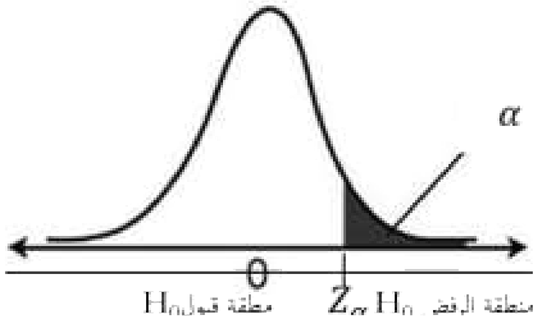
$$Z_{cal} = \frac{m - \mu}{\sigma_m} \sim N(0,1)$$

ويتم تحديد القيم الحرجة على أساس α، وتحديد مناطق الرفض والقبول كما يلي:

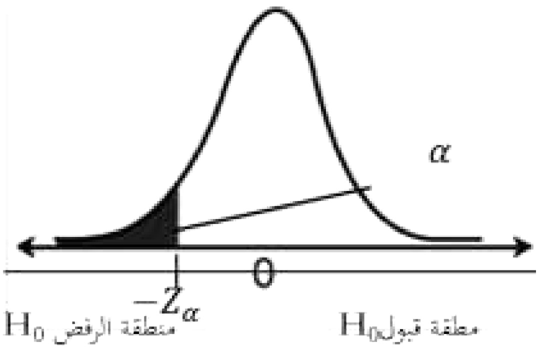
إذا كان الاختبار من طرفين: نقبل فرض العدم إذا تحققت المعادلة التالية: $-Z\alpha_{/2} < Z_{Cal} < Z\alpha_{/2}$



أما اذا كان من طرف واحد (الجهة اليمنى) فنقبل الفرضية الصفرية اذا كانت: $Z_{Cal} < Z_\alpha$ (ونرفض الفرضية الصفرية في حالة $Z_{Cal} > Z_\alpha$).



أما اذا كان من طرف واحد (الجهة اليسرى) فنقبل الفرضية الصفرية اذا كانت: $Z_{Cal} < Z_{\alpha}$ (ونرفض الفرضية الصفرية في حالة $Z_{Cal} > Z_{\alpha}$).



مثال 3-3

تنتج آلة أجزاء معدنية أسطوانية الشكل. يتم أخذ عينة من تسع قطع أقطارها 9.8 و 9.5 و 9.8 و 11.5 و 9.0 و 10.4 و 9.8 و 10.1 و 11.2 سم. افترض أن الأقطار موزعة بشكل طبيعي بتباين 0.64. إذا أكدت المؤسسة المصنعة أن متوسط القطر هو 10 سم. اختبر صحة هذا الادعاء اذا علمت ان $\alpha = 0.01$.

الحل

يطلب اختبار فرضية بأن متوسط القطع المعدنية يساوي 10 سم، أي ان الفرضية الصفرية هي:

$$H_0 : \mu = 10$$

وتكون الفرضية البديلة كما يلي:

$$H_1 : \mu \neq 10$$

بما أن العينة مسحوبة من مجتمع طبيعي تباينه معلوم، إذا توزيع المعاينة المستخدم هو توزيع Z.

نقوم بحساب احصائية الاختبار Z_{cal} على النحو التالي: $Z_{cal} = \frac{m-\mu_0}{\sigma/\sqrt{n}}$

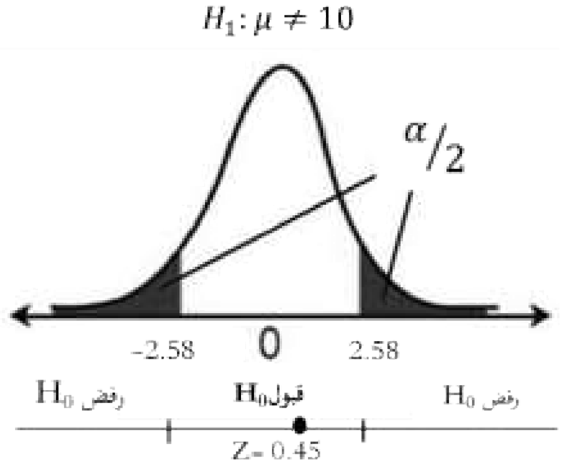
$\sigma = \sqrt{0.64} = 0.80$ ، $n = 9$ ، $m = 10.12$ ، $\mu_0 = 10$

وبالتالي
$$Z_{cal} = \frac{10.12-10}{0.8/\sqrt{9}} = 0.45$$

ومن ثم نحدد المناطق الحرجة: $\alpha = 0.01$

ومنه: $Z_{1-\alpha/2} = Z_{0.995} = \pm 2.58$

ويمكن تمثيل منطقة الرفض، ومنطقة القبول بيانيا كما يلي:



بما أن القيمة المحسوبة وقعت في منطقة القبول اذن نقبل الفرضية الصفرية ، أي أن ادعاء المؤسسة صحيح ولا يوجد هناك فرق معنوي بين المتوسط الحقيقي والمتوسط المدعى عند مستوى معنوية 1%.

2.2. اختبار فرضيات متوسط المجتمع عندما يكون التباين مجهول

عندما يكون تباين المجتمع مجهول وحجم العينات صغير ($n < 30$) نستبدل الاحصائية Z بالإحصائية t.

$$T_{cal} = \frac{m-\mu_0}{S/\sqrt{n}} \sim t_{(n-1,\alpha/2)}$$

ويتم تحديد القيم الحرجة على أساسα ، وتحديد مناطق الرفض والقبول كما يلي:

إذا كان الاختبار من طرفين: نقبل فرض العدم إذا تحققت المعادلة التالية:

$$-t_{(n-1,\alpha/2)} < T_{cal} < t_{(n-1,\alpha/2)}$$

نرفض فرض العدم إذا تحققت إحدى المعادلتين:
$$\begin{cases} T_{cal} > t_{(n-1)} \\ T_{cal} < t_{(n-1)} \end{cases}$$

إذا كان الاختبار من طرف واحد الجهة اليمنى: نقبل فرض العدم إذا تحققت المعادلة: $T_{cal} < t_{(n-1)}$

إذا كان الاختبار من طرف واحد الجهة اليسرى: نقبل فرض العدم إذا تحققت المعادلة: $T_{cal} > t_{(n-1)}$

مثال 3-4

ادعى أحد المسؤولين في شركة النقل بالشكك الحديدية أن معدل فترة الانتظار في كل شباك لا تزيد عن 15 دقائق، وبهدف اختبار هذا الادعاء تم دراسة عينة مكونة من 10 شبابيك لخدمة الزبائن على مستوى احدى محطات القطار فوجد أن معدل الانتظار 16.5 دقيقة بانحراف معياري 4 دقائق. على فرض أن مدة الانتظار لكل الشبابيك تتبع التوزيع الطبيعي اختبر ادعاء المؤسسة عند مستوى معنوية 0.05.

الحل

$$\mu_0 = 15 \text{ ،} \quad m = 16.5 \quad \text{،} \quad n = 10 \quad \text{،} \quad S = 4$$

ويمكن صياغة فرضيات الدراسة كما يلي:

$$H_0\colon \mu = 15$$
$$H_1\colon \mu > 15$$

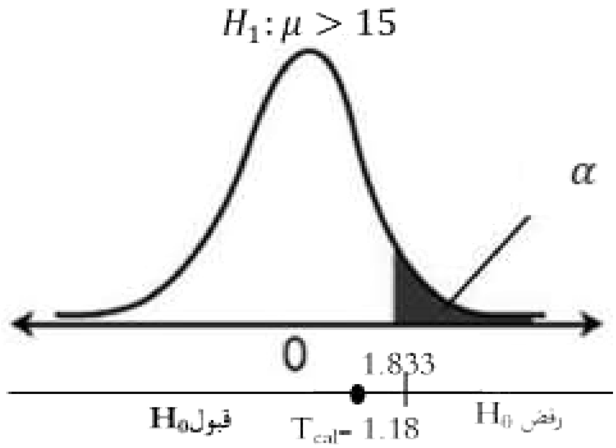
بما أن العينة مسحوبة من مجتمع طبيعي تباينه مجهول، وحجم العينة صغير اذن نستخدم توزيع t، و نحسب قيمة الاحصائية T كما يلي:

$$T_{cal} = \frac{m-\mu_0}{S/\sqrt{n}} = \frac{16.5-15}{4/\sqrt{10}} = 1.18$$

وبعد ذلك نستخرج القيمة الجدولية من جدول توزيع Student: $\alpha = 0.05$

$$t_{(n-1,\alpha)} = t_{(9,0.05)} = 1.833$$

ويمكن تمثيل منطقة الرفض، ومنطقة القبول بيانيا كما يلي:



بما أن القيمة المحسوبة وقعت في منطقة القبول اذن نقبل الفرضية الصفرية، ومنه فمتوسط الانتظار لا يتعدى 15 دقيقة عند مستوى معنوية 5%.

3. اختبار الفرضيات للفرق بين متوسطي مجتمعين

إن مشكلة اختبار الفرضية الخاصة بالفرق بين متوسطي مجتمعين لها نفس الاساس في تطبيقها مثل فترات الثقة. أي أنه سيكون لدينا نفس الحالات لمقارنة متوسط الفرق: عندما يكون التباينين معروفين، عندما يكون التباينين غير معروفين ولكن من المعروف أنهما متساويين، عندما يكون التباينين غير معروفين ولكنهما غير متساويين.

يمكن تحديد المشكلة على النحو التالي: ليكن هناك مجموعتان مستقلتان بتوزيع طبيعي: $N\ (\mu_1, \sigma_1^2)$ و $N\ (\mu_2, \sigma_2^2)$ والتي تمثل سلوك ظاهرتين مهمتين نرغب في مقارنتها من خلال تساوي متوسطيهما من عدمه، أي نقارن: $H_0: \mu_1 = \mu_2$ مع $H_1: \mu_1 \neq \mu_2$

او ان يكون متوسط احدهما اكبر او اصغر من الاخر، كما يلي:

$$H_0: \mu_1 \leq \mu_2 \text{ مع } H_0: \mu_1 > \mu_2$$

أو

$$H_0: \mu_1 \geq \mu_2 \text{ مع } H_0: \mu_1 < \mu_2$$

يمكننا تعميمها بشكل أكبر اعتمادا على الادعاء الذي نريد إثباته. ومن الشائع استخدام هذه الاختبارات لتقييم نتائج البحث في الزراعة والطب والتعليم وعلم النفس وعلم الاجتماع والعديد من مجالات العلوم الأخرى. ونميز بين ثلاث حالات نوضحها فيما يلي:

1.3. حالة تبايني المجتمعين معلومين

يمكن استنتاج هذه الحالة بسهولة من توزيعات المعاينة نظرا لأن الفرق بين متوسطي عينتين من مجتمعين موزعين طبيعيا لهما توزيع طبيعي بمتوسط: $\mu_1 - \mu_2$ ، وتباين: $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ (σ_1^2 و σ_2^2 معيوفين). ويتم حساب إحصائية الاختبار كما يلي:

$$Z_{cal} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

مثال 3-5

يدعي مصنع للخيوط اللولبية من النوع A و B أن متوسط مقاومة الشد للنوع الأول يفوق متوسط بمقاومة الشد من النوع الثاني بـ 3 وحدات. للتحقق من ادعائه اختبر تم اختبار وبشكل مستقل 100 قطعة من كل نوع في ظل ظروف مماثلة، وتم الحصول على النتائج التالية: النوع A لديه متوسط قوة 88، بينما النوع B لديه متوسط قوة 83. بافتراض أن قوة الشد موزعة طبيعيا، مع $X_A \sim N\ (\mu_A, 25)$ و $X_B \sim N\ (\mu_B, 81)$. اختبر احصائيا عند مستوى معنوية 0.05 هذا الادعاء، ثم احسب قوة الاختبار من اجل: $\mu_A - \mu_B = 2$.

الحل

سوف نقوم بمقارنة فيما اذا كانت قوة شد النوعين A و B أكثر من 3 او لا. يعني نقوم بصياغة الفرضيتين التاليتين:
$$\begin{cases} H_0: \mu_A - \mu_B = 3 \\ H_1: \mu_A - \mu_B \neq 3 \end{cases}$$

باتباع الخطوات المنهجية لإجراء الفحص وتحديد اختبار عند مستوى معنوية 0.05، سيكون لدينا:

- بما أن حجمي العينتين أكبر من 30 وتباينا المجتمعين معلومين، إذن فالمجتمعان يتبعان توزيع طبيعي وبالتالي نستخرج القيمة الجدولية كما يلي:

بما ان : $\alpha = 0.05$ ومنه $1 - \alpha/2 = 1 - 0.05/2 = 1 - 0.025 = 0.975$

$$Z_{Tab} = Z_{1-\alpha/2} = Z_{0.97\ 5} = \pm 1.96$$

من جدول التوزيع الطبيعي فان القيمة الجدولية هي: ± 1.96 (من الطرفين)

لدينا المعطيات التالية:

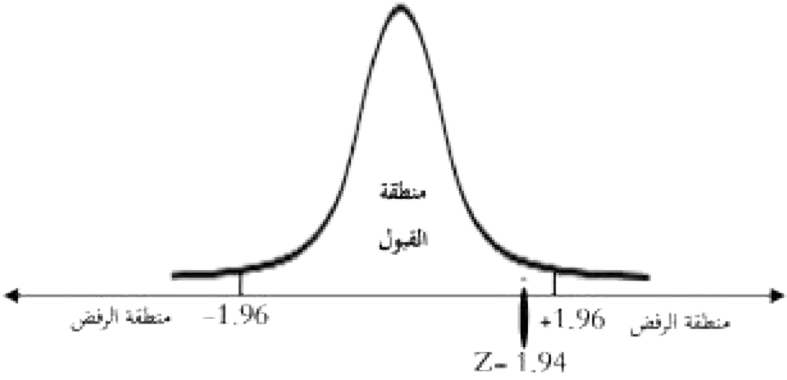
$$m_A = 88 \quad ، \quad m_B = 83 \quad ، \quad \sigma_A^2 = 25 \quad ، \quad \sigma_B^2 = 81 \quad ، \quad n_A = n_B = 100$$

ومنه: $m_A - m_B = 88 - 83 = 5 \quad \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{25}{100} + \frac{81}{100}} = 1.03$

اذا كان: $\mu_1 - \mu_2 = 3$ فان:

$$Z_{cal} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5-3}{1.03} = 1.94$$

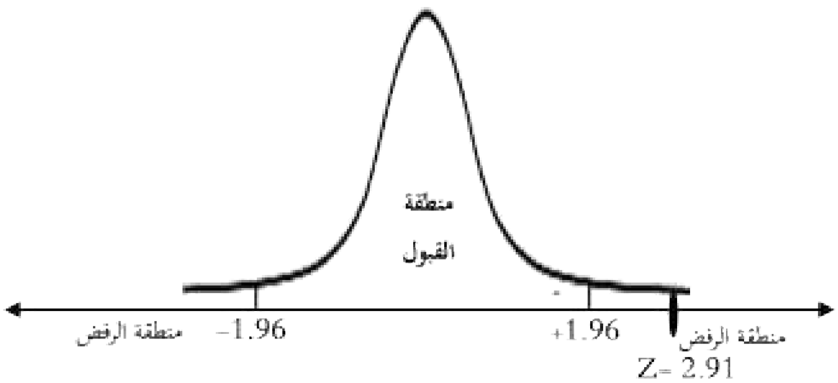
اذن هنا نقبل الفرضية الصفرية ونقول ان ادعاء المؤسسة صحيح لان القيمة المحسوبة (1.94) تقع في منطقة القبول وهذا ما يمكن توضيحه من خلال الشكل الموالي:



إذا كان متوسط الفرق الحقيقي يساوي 2 : $\mu_1 - \mu_2 = 2$ فان:

$$Z_{cal}=\frac{(m_1-m_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{5-2}{1.03}=2.91$$

ومنه نلاحظ ان القيمة المحسوبة اكبر من القيمة الجدولية وتقع في منطقة الرفض، وبالتالي نرفض هنا الفرضية الصفرية ونقول ان الفرق لا يساوي 2، كما يتضح من الشكل:



2.3. حالة تبايني المجتمعين غير معلومين مع $\sigma_1^2=\sigma_2^2$

تُعرف مشكلة مقارنة الوسائل في حالة تساوي التباينات بمشكلة Behrens-Fisher والتي تطورت بين عامي 1935 و 1939. في هذه الحالة يكون لإحصائية الاختبار توزيع *t-Student* مع n_1+n_2-2 درجة حرية. وتعطى الاحصائية بالعلاقة التالية:

$$T_{cal}=\frac{(m_1-m_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{n_1+n_2-2}$$

مثال 3-6

تتم مقارنة قوة الشد لنوعين من الخيوط اللولبية. يتم اختبار 12 قطعة من كل نوع بشكل مستقل في ظل ظروف مماثلة، والنتائج موضحة في الجدول الموالي:

النوع	1	2	3	3	4	5	6	7	8	10	11	12
A	78	76	80	79	78	80	82	81	79	83	80	82
B	83	80	82	83	81	80	79	80	82	78	79	81

بافتراض ان قوة الشد موزعة طبيعيا، مع تباينات غير معروفة ولكنها متساوية، نريد معرفة ما إذا كان من الممكن أن تستنتج إحصائيا أن متوسط قوة الشد للنوع الأول يساوي متوسط قوة الشد للنوع الثاني. حدد اختبار الفرضية

المناسب لهذه المشكلة، ثم اختبر إذا كان الاستنتاج صحيحا عند مستوى معنوية يساوي 0.05، ثم احسب قوة الاختبار لـ $\mu_1 - \mu_2 = -2$

الحل

سوف نقوم بمقارنة فيما اذا كانت قوة شد من النوع A أقل من قوة الشد للنوع B. يعني نقوم بصياغة الفرضيتين التاليتين:

$$\begin{cases} H_0: \mu_A - \mu_B = 0 \\ H_1: \mu_A - \mu_B \neq 0 \end{cases}$$

بما أن حجمي العينتين أقل من 30 وتباينا المجتمعين غير معلومين، إذن فالمجتمعان يتبعان توزيع Student وبالتالي نستخرج القيمة الجدولية كما يلي:

بما ان : $\alpha = 0.05$ ومنه $\alpha/2 = 0.05/2 = 0.025$

ومن جدول توزيع Student نجد: $T_{Tab} = T_{n_1+n_2-2,\alpha/2} = T_{22,0.025} = 2.074$

لدينا المعطيات التالية:

$m_A = 79.8333$ ، $m_B = 80.6667$ ، $S_A^2 = 3.9697$ ، $\sigma_B^2 = 2.6061$ ،

$$n_A = n_B = 12$$

ومنه اذا كان: $\mu_1 - \mu_2 = 0$ فان

$$T_{cal} = \frac{(79.8333 - 80.6667) - 0}{\sqrt{\frac{(12-1)3.9697 + (12-1)2.6061}{12+12-2} . \left(\frac{1}{12} + \frac{1}{12}\right)}} = \frac{-0.8337}{0.55} = 1.52$$

اذن نلاحظ ان القيمة المحسوبة (1.52) اقل من القيمة الجدولية (2.074) و تقع في منطقة القبول ومنه نقبل الفرضية الصفرية ونقول ان ادعاء المؤسسة صحيح.

إذا كان متوسط الفرق الحقيقي يساوي -2 : $\mu_1 - \mu_2 = -2$ فان:

$$T_{cal} = \frac{(79.8333 - 80.6667) - (-2)}{\sqrt{\frac{(12-1)3.9697 + (12-1)2.6061}{12+12-2} . \left(\frac{1}{12} + \frac{1}{12}\right)}} = \frac{1.1666}{0.55} = 2.12$$

في هذه الحالة نلاحظ ان القيمة المحسوبة اكبر من القيمة الجدولية وتقع في منطقة الرفض، وبالتالي نرفض الفرضية الصفرية ونقول ان الفرق لا يساوي -2.

3.3. حالة تبايني المجتمعين غير معلومين مع $\sigma_1^2 \neq \sigma_2^2$

إذا أخذت عينة عشوائية حجمه n_1 (أصغر من 30) من مجتمع متوسطه μ_1 وتباينه σ_1^2 ، وأخذت عينة عشوائية ثانية ومستقلة عن الاولى حجمها n_2 (أصغر من 30) من مجتمع متوسطه μ_2 وتباينه σ_2^2 ، حيث أن تبايني كلا المجتمعين مجهولين، فان احصائية الاختبار للفرضية الصفرية $H_0: \mu_1 - \mu_2 = 0$ تحسب كماي يلي:

$$T_{cal} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$$

حيث:

$$v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$

مثال 3–7

قارنت إحدى الدول استهلاك الوقود لنوعين من سيارات الديزل المجهزة بشكل متماثل. تم استخدام 12 سيارة من النوع الاول و10 من النوع الثاني في اختبارات سرعة ثابتة تبلغ 90كم/ساعة، حيث اشارت النتائج الى ان متوسط الاستهلاك للنوع الاول هو 16 كم/ لتر بانحراف معياري قدره 1.0كم/لتر ، بينما كانت نتائج متوسط الاستهلاك بالنسبة لسيارات النوع الثاني 11كم/لتر بانحراف معياري قدره 1.8كم/لتر. مع هذه القيم أكد الشخص الذي يقوم بالتجربة أن سيارات النوع الاول في المتوسط تتجاوز سيارات النوع الثاني بمقدار 4كم/لتر. اذا فرضنا أن العائد لكل لتر لكل نوع من السيارات يتم توزيعه بشكل طبيعي مع تباينات مختلفة. اختبر صحة الادعاء عند مستوى معنوية 0.10، ثم اوجد حدد الاختبار لمتوسط فرق يبلغ 3 كم/لتر.

الحل

يُطلب اختبار فرضية للفرق بين المتوسطين مع إثبات أن متوسط الكيلومترات لكل لتر لسيارات النوع الاول يتجاوز تلك التي تقطعها سيارات النوع الثاني بمقدار 4كم/لتر، وبالتالي المطلوب اختبار الفرضيتين التاليتين:

$$\begin{cases} H_0: \mu_1 - \mu_2 = 4 \\ H_1: \mu_1 - \mu_2 \neq 4 \end{cases}$$

لدينا المعطيات التالية:

$m_1 = 16$ ، $m_2 = 11$ ، $S_1^2 = 1$ ، $\sigma_2^2 = 3.24$ ، $n_1 = 12$ ، $n_2 = 10$

نحسب v كما يلي:

$$v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{({S_1^2}/{n_1})^2}{n_1 - 1} + \frac{({S_2^2}/{n_2})^2}{n_2 - 1}} = \frac{(\frac{1}{12} + \frac{3.24}{10})^2}{\frac{({1}/{12})^2}{12 - 1} + \frac{(\frac{3.24}{10})^2}{10 - 1}} = 13.4946 \approx 13$$

بما ان : $\alpha = 0.10$ ومنه $\alpha/2 = 0.10/2 = 0.05$

من جدول توزيع Student نجد: $T_{Tab} = T_{,\alpha/2} = T_{13,0.05} = 1.771$

ومنه فان:

$$T_{cal} = \frac{(16 - 11) - (4)}{\sqrt{\frac{1}{12} + \frac{3.24}{10}}} = \frac{1}{0.638} = 1.567$$

اذن نلاحظ ان القيمة المحسوبة تقع في منطقة القبول وهي أقل من القيمة الجدولية، ومنه نقبل الفرضية الصفرية ونقول أن سيارات النوع الاول تتجاوز سيارات النوع الثاني بمقدار 4 كم/لتر.

إذا كان متوسط الفرق الحقيقي يساوي 3 : $\mu_1 - \mu_2 = 3$ فان:

$$T_{cal} = \frac{(16 - 11) - (3)}{\sqrt{\frac{1}{12} + \frac{3.24}{10}}} = \frac{2}{0.638} = 3.13$$

في هذه الحالة نلاحظ ان القيمة المحسوبة اكبر من القيمة الجدولية وتقع في منطقة الرفض، وبالتالي نرفض الفرضية الصفرية ونقول ان الفرق لا يساوي 3، كما يتضح من الشكل:

ملاحظة

الحالة الثانية: تباينا المجتمعين مجهولين وحجمي العينتين كبير فان الاحصائية تعطى بالعلاقة التالية:

$$Z_{cal} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

4. اختبار الفرضيات لنسبة المجتمع

إذا سحبت عينة عشوائية حجمها n من مجتمع يتبع توزيع ذو الحدين (عدنان بن ماجد البري واخرون، 1997) وكان حجم العينة كبيرا، ونريد اختبار فرضية تساوي نسبة ظاهرة معينة في المجتمع لقيمة ما، فإن الفرضيات الاحصائية يمكن صياغتها على النحو التالي:

نوع الاختبار	صياغة الفرضية	القرار
من الطرفين	$\begin{cases} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{cases}$	قبول H_0 اذا كانت Z محصورة بين $Z_{\alpha/2}$ و $-Z_{\alpha/2}$
الجهة اليمنى	$\begin{cases} H_0 : P = P_0 \\ H_1 : P > P_0 \end{cases}$	قبول H_0 اذا كانت Z أقل من $Z_{\alpha/2}$
الجهة اليسرى	$\begin{cases} H_0 : P = P_0 \\ H_1 : P < P_0 \end{cases}$	$-Z_{\alpha/2}$ قبول H_0 اذا كانت Z أكبر من $-Z_{\alpha/2}$

حيث أن إحصائية الاختبار تعطى بالعلاقة التالية: $$Z_{cal} = \frac{'-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

مثال 3–8

صرح مدير عام قناة تلفزيونية أن نسبة الجمهور الذي يشاهد برنامج معين ليلة الخميس تزيد عن 40%. تم اختيار عينة من 100 مشاهد تمت مقابلتهم، ووجد أن 45 منهم شاهدوا فعلا البرنامج. عند مستوى معنوية 5%، اختبر ما إذا كان التصريح صحيحا.

الحل

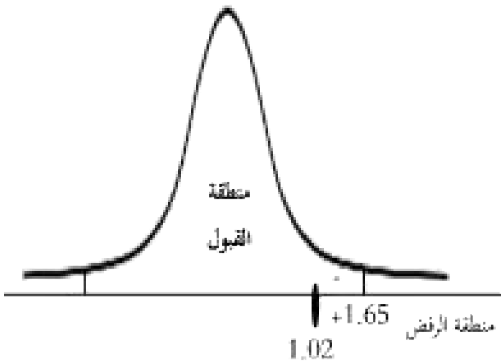
سيتم اختبار فرضية لنسبة المشاهدين الذين يشاهدون البرنامج ليلة الخميس، والطي قال عنه مدير القناة أنه أكبر من 0.40، وبالتالي سيكون علينا اختبار الفرضيتين التاليتين: $\begin{cases} H_0 : P = P_0 \\ H_1 : P > P_0 \end{cases}$

لدينا المعطيات التالية: $p'_a = 45$ ، $n = 100$ ، $P_0 = 0.40$

ومنه: $$1 - p_0 = 1 - 0.4 = 0.6 \quad و \quad p' = \frac{{}'_{a}}{n} = \frac{45}{100} = 0.45$$

من جدول التوزيع الطبيعي نجد: $$Z_{Tab} = Z_{1-\alpha} = Z_{0.95} = +1.65$$

ومنه فان: $$Z_{cal} = \frac{0.45-0.4}{\sqrt{\frac{0.4(0.6)}{100}}} = \frac{0.05}{0.049} = 1.02$$



ونلاحظ ان القيمة المحسوبة تقع في منطقة القبول ومنه نقبل الفرضية الصفرية ونقول أن نسبة الجمهور الذي يشاهد برنامج معين ليلة الخميس تزيد عن 40%.

1.4. اختبار الفرضيات للفرق بين نسبتين

اذا كانت x و y عينتين عشوائيتين مسحوبتين من مجتمعين طبيعيين، حيث تمثل p'_x نسبة صفة معينة في العينة x و p'_x نسبة نفس الصفة في العينة y. ذا كان حجم العينتين كبير (n_1 و n_2 أكبر من 30) فإن احصائية اختبار الفرق بين نسبتي مجتمعين p_x و p_y تقترب من التوزيع الطبيعي المعياري ونكتب:

$$Z_{cal} = \frac{(p'_x - p'_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_1} + \frac{p_y(1-p_y)}{n_1}}}$$

ملاحظة

تجدر الاشارة إلى أنه في حالة ما اذا كانت p_x و p_y متساويتين، فان p'_{ax} و p'_{ay}سيكون كلاهما تقدير بقيمة واحدة لنفس المعلمة، وهنا نستخدم متوسط مرجح للنسبتين $\bar{p}$ حيث: $\bar{p} = \frac{p'_{ax}+p'_{ay}}{n_1+n_2} = \frac{n_1p'_x+n_2p'_y}{n_1+n_2}$

اي ان احصائية الاختبار تعطى بالعلاقة: $Z_{cal} = \frac{p'_x - p'_y}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1}+\frac{\bar{p}(1-\bar{p})}{n_2}}}$

مثال 3-9

توزع إحدى شركات تصنيع السجائر نوعين x و y. ويريد مدير المبيعات معرفة ما إذا كانت إحداهما متفوقة على الأخرى، حيث أجرى استبيانين مستقلين من أجل ذلك. وكانت النتيجة أن 56 من بين 200 مدخن يفضلون النوع x ، بينما 29 من أصل 150 مدخن يفضلون النوع الاخر. هل يمكن للمدير أن يستنتج أن السجائر من النوع x تتفوق على السجائر من النوع y؟ (مستوى الدلالة 0.05).

الحل

سنقوم باختبار الفرضية التالتين لمعرفة ما اذا كانت نوع من السجائر يتفوق على الاخر: $\begin{cases} H_0: P_x = P_y \\ H_1: P_x > P_y \end{cases}$

لدينا المعطيات التالية: $n_1 = 200$ ، $n_2 = 150$

ومنه: $p'_x = \frac{56}{200} = 0.28$ و $p'_y = \frac{29}{150} = 0.193$

$$\bar{p} = \frac{p'_{ax} + p'_{ay}}{n_1 + n_2} = \frac{56 + 29}{200 + 150} = 0.2429$$

من جدول التوزيع الطبيعي نجد: $Z_{Tab} = Z_{1-\alpha/2} = Z_{1-0.025} = +1.96$ ومنه فان:

$$Z_{cal} = \frac{0.28 - 0.193}{\sqrt{\frac{0.2429(1-\ 0.2429)}{200} + \frac{0.2429(1-\ 0.2429)}{150}}} = \frac{0.097}{0.04532} = 2.14$$

ونلاحظ ان القيمة المحسوبة تقع خارج منطقة القبول وهي اكبر من القيمة الجدولية ومنه نرفض الفرضية الصفرية ونقول أن اجد نوعي السجائر متفوق على الاخر.

5. اختبار الفرضيـات لتبايـن المجتمـع

الامر هنا يتعلق باختبار فيما اذا كان تبايـن المجتمـع يساوي قيمة معينة، وبالرجوع للفصل الاول (توزيع المعاينة) نعلم أن الاحصائية $\frac{(n-1)\hat{S}^2}{\sigma^2}$ تتبع توزيع مربع كاي χ^2 ومنه يمكن ان نكتب: $\chi^2_{cal} = \frac{(n-1)\hat{S}^2}{\sigma_0^2} \sim \chi^2_{n-1}$

مثال 3–10

تنتج آلة أجزاء معدنية أسطوانية الشكل، نأخذ عينة من تسع قطع أقطارها 9.8 ، 9.5 ، 9.8 ، 11.5 ، 9.0 ، 10.4 ، 9.8 ، 10.1 ، 11.2 ملم على التوالي. نفرض أن أقطار الأجزاء موزعة بشكل طبيعي. إذا أكدت الشركة المصنعة للقطع المذكورة أن تباين قطرها يقل عن 1 مم^2. اختبر اذلك عند مستوى 0.01.

الحل

نريد إثبات أن تباين الأجزاء المعدنية أقل من 1 مم^2 ، اي اختبار الفرضيتين التاليتين: $\begin{cases} H_0: \sigma^2 = 1 \\ H_1: \sigma^2 < 1 \end{cases}$

$\sigma_0^2 = 1$ ، $\alpha = 0.01$ ، $n - 1 = 9 - 1 = 8$ ، $\hat{S}^2 = 0.637$

$$\chi^2_{cal} = \frac{(n-1)\hat{S}^2}{\sigma_0^2} = \frac{9(0.637)}{1} = 5.733$$

$$\alpha = 0.01 \quad \chi^2_{v,\alpha} = \chi^2_{8,0.01} = 1.647$$

منه نرفض الفرضية الصفرية أن تباين قطرها لا يقل عن 1 مم^2

1.5. اختبار تساوي تباينين

يمكننا ان نلاحظ أن طريقة مقارنة تباينين من مجتمعين موزعين توزيعا طبيعيا هي عن طريق إحصائية الاختبار $\hat{S}^2_{n_1-1}/\hat{S}^2_{n_2-1}$ ، باستخدام التوزيع:

$$F_{cal} = \frac{\chi^2_{n_1-1}}{\chi^2_{n_2-1}} = \left(\frac{\hat{S}^2_{n_1-1}}{\hat{S}^2_{n_2-1}}\right)\frac{1}{\sigma_1^2/\sigma_2^2}$$

مثال 3–11

تتم مقارنة قوة شد نوعين من الخيوط اللولبية. تم اختبار 12 قطعة من كل نوع في ظل ظروف مماثلة، والنتائج مبينة في جدول المثال رقم 3–6. نريد اختبار ما إذا كان الافتراض الوارد في المثال 3–6 حول تساوي التباينين $(\sigma_1^2 = \sigma_2^2)$ صحيح عند مستوى معنوية 0.05.

الحل

نختبر الفرضيتين:
$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

$\frac{\sigma_1^2}{\sigma_2^2} = 1$ ، $\hat{S}_2^2 = 2.61$ ، $\hat{S}_1^2 = 3.97$ ، $n_1 - 1 = n_2 - 1 = 12 - 1 = 11$

بما ان : $\alpha = 0.05$ ودرجات الحرية $v_1 = v_2 = 9$ وبالرجوع لجدول F نجد:

$$F_{\alpha/2,v_1,v_1} = F_{0.05,11,11} = 3.474$$

$$F_{1-\alpha/2,v_1,v_1} = \frac{1}{F_{0.05,11,11}} = \frac{1}{3.474} = 0.2879$$

$$F_{cal} = \left(\frac{3.97}{2.61}\right)\frac{1}{1} = 1.52$$

نلاحظ أن: $1.52 \in [0.2879,3.474]$ وبالتالي نقبل الفرضية الصفرية القائلة ان التباينين متساويين.

خلاصة الفصل

تعتبر عملية اختبار الفرضيات الاحصائية أحد أهم المواضيع التي يشملها الاستدلال الاحصائي والذي يساعد في عملية اتخاذ القرار بشأن القيمة المختبرة أو المعلنة لمعلمة المجتمع من خلال تحليل الفرضياتوالادعاءات حولها انطلاقا من الاثباتات المقدمة من العينة.

الفصـل الرابع

اسئـلة التـقييم الذاتي وتمـارين مقترحـة

1. أسئلة التقييم الذاتي

1. ما هو الفرق بين المعلمة والإحصائية؟

2. أجب عن الأسئلة التالية واشرح إجابتك.

- هل المقاييس النزعة المركزية تمثل مدى انتشار مركز البيانات؟
- هل يمثل الانحراف قيمة تشير إلى مدى التقارب بين قيم العينة أو انتشارها؟
- هل الانحراف المعياري يشير إلى شكل توزيع العينة؟
- من أجل الوصف الإحصائي لسلوك ملاحظات العينة، هل يكفي حساب جميع مقاييسها المركزية؟

3. إذا كان عدد مفردات المجتمع 1.000 فهل حجم عينة مكونة من 400 عنصر سيكون له نتائج جيدة؟

4. بعد مراجعة المفاهيم الاساسية في هذا الفصل ما هي المنهجية المثلى لتقديم تقرير إحصائي عن سلوك عينة ما؟

5. في المواقف التالية وضح كيف ستقوم بتشكيل عينة حجمها n مع الأخذ في الاعتبار واحدة على الأقل من الأساليب التي تمت مراجعتها في هذا الفصل.

- عملاء بنك الفلاحة والتنمية الريفية فرع خميس مليانة
- جميع أصحاب الحسابات المصرفية في جميع أنحاء الوطن.
- لخط إنتاج المشروبات الغازية من نوع حمود بوعلام.
- لإنتاج سيارات Renault في مدينة وهران.
- لنزلاء فندق في مدينة عنابة.
- عن مسار الدينار العام الماضي.
- لجميع مشجعي كرة القدم في العاصمة.

6. لنفترض أن Z هو متغير نموذج التوزيع الطبيعي المعياري، احسب (بدون استخدام الجداول أو الآلة الحاسبة) كل من:

$$P(Z > 7.23)\ ;\ \ P(Z \leq -3\pi)\ ;\ \ P(-8 < Z < 8)\ ;\ \ P(0 < Z < 12.79)$$

7. هل صحيح أن توزيع مجموع ومتوسط التوزيعات الطبيعية هي توزيع طبيعي آخر؟

8. كيف تختلف نظريات حساب احتمال مجموع القواعد المعيارية عن نظرية النهاية المركزية؟

9. ما هو حجم العينة الموصي به لاستخدام نظرية النهاية المركزية ؟

10. إذا عرفنا توزيع متغير عشوائي، فلماذا من الضروري استخدام نظرية النهاية المركزية لحساب احتمالات مجموع أو متوسط عينة عشوائية من هذه المتغيرات؟

11. هل صحيح أن الإحصائية التي تأتي من عينة عشوائية من المتغيرات يجب أن تحتوي دائما على نفس معاملات هذه المتغيرات؟

12. هل توزيع أخذ العينات للمقدر هو التوزيع على جميع العينات الممكنة من نفس الحجم n؟

13. عادة ما يسمى الانحراف المعياري لتوزيع العينات لأي إحصائية بالخطأ المعياري؟

14. هل صحيح أن التقدير النقطي دائمًا أفضل من التقدير بمجال؟

15. ما هو المقدر المتحيز؟

16. ما هو حجم العينة الموصي به لاستخدام نظرية النهاية المركزية؟

17. ما هو الفرق بين المقدر والإحصائية؟

18. هل يمكن تطبيق الصيغ التي تم تحديدها لفترات الثقة للمتوسط على أي توزيع؟

19. لماذا من الضروري أن تكون العينات مستقلة في فترات الثقة لنسبة التباين؟

20. ماذا يعني مستوى الثقة؟

21. في أي حالة من فترات الثقة للفرق بين المتوسطات يجب أن تكون الملاحظات غير مستقلة؟

22. ما هو المقصود باختبار الفرضيات؟

23. ما هي منطقة الرفض ومنطقة عدم الرفض؟

24. ما هي فلسفة تحديد اختبار الفرضية؟

25. كيف يتم تحديد اختبار الفرضيات؟

26. ما هو اختبار الفرضية من طرف واحد؟

27. ما هو اختبار الفرضية من الطرفقين؟

28. هل صحيح أن الصيغ التي تم تحديدها لاختبارات فرضية المتوسط يمكن تطبيقها على أي توزيع؟

29. لماذا من الضروري أن تكون العينات مستقلة في الاختبارات الفرضية لنسبة التباينات؟

30. ما هو الخطأ من النوع الأول وكيف يتم حساب احتمال هذا النوع من الخطأ؟

31. ما هو الخطأ من النوع الثاني وكيف يتم حساب احتمال حدوث هذا النوع من الخطأ؟

32. ماذا نقصد بمستوى المعنوية؟

33. ما هي إحصائية الاختبار؟

34. ما هي المنطقة الحرجة؟

35. ما هي الخطوات التي يجب اتباعها لاختبار الفرضية الإحصائية؟

36. إحصائيا هل يصح القول بأن الفرضية الصفرية مقبولة؟

2. تمارين مقترحة

التمرين 01: ترغب إدارة إحدى شركات التصنيع في حساب تكاليف الإصلاح السنوية لآلة معينة والتي تجري دراسة من أجلها ، حيث اظهرت النتائج أن تكاليف الإصلاح السنوية تتبع التوزيع الطبيعي بمتوسط 400.000$ وانحراف معياري قدره 50.000$.

- أحسب احتمال أن تتراوح تكاليف الإصلاح لهذا العام بين 300000 و 500000 $
- ما هي أقل تكلفة ميزانية للإصلاح السنوي للآلات في 10% من الحالات؟

التمرين 02: تدفع شركة خدمات عامة لموظفيها متوسط أجر قدره 10 دولارات للساعة بانحراف معياري قدره دولار واحد. إذا كان للأجور توزيع طبيعي.

- ما هي نسبة العاملين الذين يتقاضون رواتب تتراوح بين 9 و 11 دولار في الساعة؟
- كم تمثل نسبة 5% من الموظفين الأعلى أجرا؟

التمرين 03: يتبع عمر أنواع معينة من بطاريات السيارات التوزيع الطبيعي بمتوسط 1200 يوم وانحراف معياري 100 يوم.

- من بين 3000 بطارية المعروضة للبيع، كم عدد البطاريات التي ستستمر لأكثر من 1300 يوم؟
- إلى متى يجب ضمان البطاريات إذا كانت الشركة المصنعة تريد استبدال 10% فقط من البطاريات المباعة؟

التمرين 04: أجرت سلطات العاصمة دراسة عن عمليات السطو التي ستعرض لها سكان المنطقة، وقد اظهرت النتائج ان أن 15% منهم قد تعرضوا للسرقة.

- تم اختيار عينة عشوائية من 100 مواطن من العاصمة، ما هو احتمال تعرض ما بين 10 و 20% منهم للسرقة؟

- صرح ممثل عن السلطات أن نسبة السرقات في عينة ما أقل من 17% مع احتمال أكبر من 0.95. ما هو الحد الأدنى لحجم العينة حتى يكون بيانه صحيح إحصائيا؟

التمرين 05: لاختيار موظفيه، يستخدم المسؤول التنفيذي اختبارا بمتوسط درجة 140 نقطة وانحراف معياري قدره 10. اذا فرضنا ان توزيع الدرجات يتبع التوزيع الطبيعي وأن الحد الأدنى من الدرجة 130 يسمح لمقدم الطلب بأن يؤخذ طلبه بعين الاعتبار. ما هو احتمال أن يتم بموجب هذا المعيار النظر في مقدم الطلب الموالي؟

التمرين 06: أرادت مؤسسة وضع برنامج سلامة العمل للتقليل من الوقت الضائع بسبب حوادث العمل بغرض تسويقه. وقامت احدى المؤسسات المهتمة بمقارنة الوقت الضائع قبل وبعد تنفيذ البرنامج وقد قررت هذه المؤسسة اقتناء البرنامج اذا كان احتمال تقليل 14 ساعة على الأقل في المتوسط أكبر من 0.80، خلال 10 اسابيع. لنفترض أن المؤسسة تعلم أن الوقت الضائع الحالي في الأسبوع له توزيع طبيعي $N(108, 12^2)$ بينما التوزيع الأسبوعي للوقت الضائع مع تنفيذ البرنامج يتبع توزيع طبيعي $N(91, 14^2)$. لبيع هذا البرنامج يؤكد مدير التسويق في المؤسسة للمؤسسات المهتمة أن برنامجه يقلل من متوسط الوقت الضائع بأكثر من 14 ساعة في الأسبوع في أكثر من 90% من الحالات. ما هو الحد الأدنى لحجم الأسابيع التي يجب أن يختارها المدير لإثبات أن ادعائه إحصائيا؟

التمرين 07: من المعروف أن الدواء المستخدم في علاج مرض معين كان فعالا خلال ثلاثة أيام في 75% من الحالات التي تم استخدامه فيها. عند تقييم فعالية دواء جديد لعلاج نفس المرض، تم إعطاؤه لعشرة مرضى. باعتبار أن الدواء الجديد فعال على الأقل مثل الأول، ما هو احتمال ان يتعافى ثلاثة مرضى على الأكثر من هذه العينة؟

التمرين 08: من المعتقد أن 16% من الأسر في جنوب البلاد لديها إجمالي الدخل المصنف في المستوى الاقتصادي المرتفع. ويعتقد أن هذه النسبة في شمال البلاد تبلغ 11%. إذا كانت هذه الأرقام صحيحة:

- احسب احتمال أن تكون نسبة عينة الأسر في الجنوب أكبر من تلك الموجودة في الشمال بنسبة 4% كحد أقصى، إذا تم اختيار عينات من 500 و 625 أسرة.

– أوجد الحجم الأدنى للعينة الذي يجب مراعاته بحيث يكون الاحتمال في السؤال السابق أقل من 0.20.

– حل السؤالين السابقين لميزة أقل من 2% مع شرح الفروق.

التمرين 09: تهتم شركة المحاسبة بتقدير نسبة الحسابات التي يوجد بها تناقض بين الميزانيات العمومية المبلغ عنها بين العملاء والبنوك. كم عدد الحسابات التي يجب اختيارها بحيث يكون هناك احتمال 90% أن نسبة العينة تختلف بما لا يزيد عن 0.02 وحدة من النسبة الحقيقية ، وهي 35%؟

التمرين 10: مصنع يقوم بتصنيع المصابيح الكهربائية التي يتبع عمرها التوزيع الطبيعي بمتوسط 780 ساعة وانحراف معياري قدره 50 ساعة.

– لمواجهة شكاوى الزبائن، سيقوم المصنع باجراء دراسة جودة إذا كان في عينة عشوائية مكونة من 25 مصباح احتمال أن يكون متوسط عمر المصابيح يتراوح ما بين 750 و 800 أقل من 0.95. هل سيتعين على المصنع اجراء هذه الدراسة؟

– إذا ادعت مجموعة الزبائن أنهم يستطيعون إثبات إحصائيا أن أقل من 1% من المصابيح تدوم في المتوسط لأكثر من 790 ساعة، فما هو الحد الأدنى لحجم العينة الذي يجب عليهم اختياره حتى يتمكنوا من تأكيد ذلك؟

التمرين 11: لنفترض أن X_1 و X_2 و X_3 و X_4 عينة عشوائية منتقاة من مجتمع بمتوسط μ وانحراف معياري σ. لتكن المقدرات التالية لـ μ:

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{6},\ T_2 = \frac{X_1 + X_2 + X_3 + 2X_4}{5},\ T_3 = \frac{X_1 + 2X_2 + 3X_3 + 4X_4}{10}$$

التمرين 12: لنفترض أن X_1 و X_2عينة عشوائية مع $\mu = \theta$ وانحراف معياري $\sigma = 1$. لتكن المقدرات المقدرات الثلاثة التالية لـμ .

$$T_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2,\ T_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2,\ T_3 = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

– بيّن أن T_i هو مقدر غير متحيز لـ μ من اجل: $i = 1,2,3$

التمرين 13: يتم إعداد آلة للمشروبات الغازية بحيث يتم توزيع كمية السائل التي يتم صرفها هو توزيع طبيعي مع انحراف معياري يساوي 0.15 ديسيلتر. قدر عند مستوى ثقة 95% متوسط العدد الحقيقي للمشروبات الغازية التي تقدمهاالآلة إذا كان متوسط العينة العشوائية المكونة من 36 هو 2.25 ديسيلتر.

التمرين 14: تنتج آلة معينة أجزاء معدنية أسطوانية الشكل. يتم أخذ عينة من القطع التي يبلغ قطرها 10 ، 12 ، 11 ، 11.5 ، 9 ، 9.8 ، 10.4 ، 9.8 ، 10 ،و 9.8 مل. نفترض أن الأقطار موزعة بشكل طبيعي. بمستوى ثقة 99%:

- أوجد مجال ثقة لمتوسط قطر جميع الأجزاء (نفترض أن الانحراف المعياري يساوي 1).
- حدد الحجم الأدنى للعينة التي يجب اختيارها بحيث يكون خطأ الأقطار أقل من ربع ملليمتر.
- إذا كان الحد الأدنى لفترة الثقة 9.75 مل، فما هو الحد الأعلى ومستوى الثقة؟
- اوجد مجال ثقة لمتوسط قطر جميع الأجزاء في هذا الجهاز ، إذا فرضنا ان الانحراف المعياري غير معروف.

التمرين 15: تدعي غرفة التجارة في العاصمة أنه وفقا لدراساتها الاقتصادية، فإن متوسط مبلغ المال الذي ينفقه الأشخاص الذين يحضرون المؤتمرات يوميا، والذي يشمل وجبات الطعام والإقامة والترفيه، أقل من 950 دينار. لاختبار هذا الادعاء، اختار مشرف الغرفة 16 شخصا يحضرون المؤتمرات وسألهم عن مقدار الأموال التي ينفقونها يوميا، وكانت اجاباتهم كما يلي: 940، 875، 863، 948، 942، 989، 835، 874، 868، 852، 958، 884، 1034، 1046، 955، 963.

نفترض أن المبلغ الذي يتم إنفاقه في اليوم يتم يتبع التوزيع الطبيعي. يقرر المشرف أنه إذا كان الحد الأعلى لفاصل الثقة للمتوسط أقل من 960 دينار ، فإن بيانها يكون صالحا. بثقة 95%:

- أوجد مجال الثقة لمتوسط جميع النفقات اليومية ، افترض أن الانحراف المعياري للبيان الصحيح يساوي 45 دينار.
- قم بتحديد الحد الأدنى لحجم العينة الذي يجب اختياره بحيث يكون تقدير المتوسط ضمنمجال بطول 40 دينار.
- إذا كان الحد الأدنى لفترة الثقة 900 دينار، فما هو الحد الأعلى ومستوى الثقة؟
- اوجد مجال الثقة لمتوسط جميع المصاريف اليومية، وإذا لم الانحراف المعياري غير معروف ، فهل البيان صحيح؟

التمرين 16: تم تسجيل القياسات التالية لساعات التجفيف لعلامة تجارية للطلاء:

$$3.4, 2.5, 4.8, 2.9, 3.6, 2.6, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8$$

نفترض أن القياسات تمثل عينة عشوائية. قدر بفترة ثقة الانحراف المعياري عند مستوى 90%.17. تحتاج مؤسسة منتوج معين يتم تصنيعه بواسطة موزعين مختلفين A و B. لتقدير الفرق في مدة المنتوج بين هذين الموزعين، تم إجراء تجربة باستخدام 12 عنصرا من كل موزع. والنتائج تضير الى ان المتوسط والانحراف المعياري في A هما 70.5 و 1.88 على التوالي، بينما بالنسبة للعينة B فان المتوسطيساوي 73.4 والانحراف المعياري 3.12. حدد بفترة ثقة عند مستوى 95% لـ $\frac{\sigma_A^2}{B}$. وما هو الموزع يقوم بتصنيع منتج أكثر تجانسا. نفس السؤال عند مستوى 80%.

التمرين 17: تم جمع البيانات التالية في تجربة مصممة للتحقق مما إذا كان هناك اختلاف منهجي في الأوزان التي تم الحصول عليها بمقياسين. أي أنه يتم وزن كل جسم على كلا المقياسين. بافتراض أن الاختلافات في الأوزان لكلا المقياسين تتبع التوزيع الطبيعي.

المقياس 1	11.23	14.36	8.33	10.5	23.42	9.15	13.47	60.47	12.40	*19.38*
المقياس 2	11.27	14.41	8.35	10.52	23.41	9.17	13.52	60.46	12.45	19.35

عند ميتوى ثقة 92%.

- اوجد فترة ثقة للفرق في المتوسطات وقم اختبارها إذا علمت انه لا توجد اختلافات بينهما.
- إذا كان الحد الأدنى لمجال الثقة يساوي 0.4-، ما هو الحد الأعلى، ثم اوجد قيمة $1-\alpha$ ؟

التمرين 18: يدعي مدير العلامة التجارية A للسجائر أنه يتفوق على منافسه، العلامة التجارية B ، بنسبة 11% على الأقل مع احتمال 0.95. للتحقق من صحة البيان إحصائيا، يقوم المدير بمسح مجموعتين من المدخنين بشكل مستقل. في المجموعة 1 كان السؤال: هل تفضل العلامة التجارية A ؟، بينما في المجموعة 2: هل تفضل العلامة التجارية B؟ في المجموعة 1 المكونة من 200 شخص، أجاب 41 بنعم ، بينما في المجموعة 2، أجاب 18 من 150 بنعم. عند مستوى ثقة 95% احصل على:

- لوجد مجال الثقة للفرق الحقيقي في النسب. هل تأكيد المدير مبرر؟
- الحد الأدنى مجال الثقة للفرق: $p_A - p_B$ يساوي 0.02 ، فأوجد الحد الأعلى وقيمة $1-\alpha$.

التمـرين 19: إذا كشـفت دراسـة أوليـة أن 56 مـن بـين 200 مـدخن يفضـلون السـجائر مـن النـوع A ، أوجدحجم العينـة حـتى يكـون تفضـيل النـوع A علـى انهـا تقـع في حـدود 0.05 مـن الفعليـة للمـدخنين الـذين يفضلون هذا النوع، صحيح (*بنسبة* 99%)

التمرين 20: في دراسة نوعين من الجرائم التي يرتكبهاالاشخاص المحتجزون في بعض مؤسسات اعادة التاهيل، خلال فترة 10 سنوات ، تم الحصول على البيانات التالية:

نوع الجريمة	**حجم العينة**	**العدد**
A	200	40
B	220	35

- أوجد مجال الثقة عند مستوى 95% للاختلاف في نسب نوع الجريمة. وضح ما إذا كـان مـن الممكـن التأكد عند 95% أن نسبة الجريمة من النوع A أكبر من النسبة B. اشرح إجابتك.

- اذا كان الحد الأعلى لمجال الثقة للفرقبين $p_A - p_B$ هو 0.1، فأوجد الحد الأدنى وكذا $1-\alpha$.

التمرين 21: في تجربة ، تمت مقارنة الوقود لنوعين من المركبات. تم استخدام 12 سيارات من النوع الاول و 10 من النوع الثاني، في اختبارات سرعة ثابتة تبلغ 90 كم/ساعة. إذا حصلنا على متوسط 12.5 كم/لتر مع انحراف معياري 2.0 كم/لتر بالنسبة لسيارات النوع الاول و متوسط 14.2 كم/لتر وانحراف معياري قدره 1.8 كم/لتر للنوع الثاني. وبافتراض المسافة المقطوعةلكل نوع تتبع التوزيع الطبيعي، مع مستوى ثقة قدره95%.

- أوجد مجال ثقة للفرق في متوسط الأميال لكل لتر من النوعين بافتراض: التباينين هما التوالي 5، 4.

- تحديد الحجم الأدنى للعينة التي يجب اختيارها بحيث يكون الخطأ في تقدير الفرق في متوسط السيارات أقل من 1 كم.

- إذا كـان الحـد الأدنى مـن مجـال الثقـة يسـاوي 23، مـا المقـدار الـذي يتوافـق مـع الحـد الأعلـى ، مـع تحديـد $1-\alpha$.

التمرين 22: يقال إن النظام الغذائي الجديد يقلل من وزن الشخص بمعدل 4.5 كج في أسبوعين. تم تسجيل أوزان سبع نساء اتبعت هذا النظام الغذائي قبل هذه الفترة وبعدها. افترض أن توزيع الفرق لكل شخص طبيعي تقريبا.

الوزن / المرأة	1	2	3	4	5	6	7
قبل	58.5	60.3	61.7	69.0	64.0	62.6	56.7
بعد	60.0	57.9	58.1	62.1	58.5	59.9	54.4

مع مستوى ثقة قدره96%:

- أوجد فترة الثقة للفرق بين متوسطي الأوزان، ثم قم باختبار مدى فاعلية النظام الغذائي.
- إذا كان الحد الأعلى لمجال الثقة هو 5 ، ما هو الحد الأدنى وبأي نسبة $1-\alpha$ يمكننا الحصول عليه؟

التمرين 23: أفاد مركز أبحاث الطب الرياضي في ماي 2020 ، فيما يتعلق بالاختلافات في معدلات استهلاك الأكسجين للذكور الجامعيين المدربين بطريقتين مختلفتين ومستقلتين، أحداهما تستخدم تدريب متقطع والآخرى مستمر بمدة متساوية. فيبين الجدول الموالي النتائج التي تم تسجيلها، والمعبر عنها بالملل لكل كج/دقيقة من عينتين عشوائيتين مستقلتين.

التدريب المتقطع	التدريب المستمر
$n_i = 27$	$n_j = 22$
$m_i = 39.63$	$m_j = 43.71$
$S_i = 9.68$	$S_j = 4.87$

- بافتراض ان المجتمعين موزعين طبيعيا، حدد بفترة ثقة عند مستوى 99% *الفرق بين متوسطي كلا الطريقتين.*

التمرين 24: تعتقداحدى المؤسسات المنتجة للبطاريات أنها تدوم 40 ساعة في المتوسط ، مع انحراف معياري قدره ساعة واحدة. فإذا كانت 09 من هذه البطاريات لها عمر 40.5 و 38 و 38.5 و 41 و 38.6 و 40.5 و 37.9 و 39.1 و 39 ساعة،وبافتراض أن عدد ساعات عمر البطارية موزعة بشكل طبيعي ومستوى الثقة هو 95%.

- قدر بمجال ثقة متوسط عمر البطارية، بافتراض أن : $\sigma = 1$ (ساعة واحدة)، وبين ما إذا كان افتراض الشركة الصانعة (بان المتوسط يساوي 40) صالحا في ظل هذه الظروف.
- إذا كان الحد الأعلى لفترة الثقة 40 ، فما هو الحد الأدنى ومستوى الثقة؟

التمرين 25: في عملية كيميائية، تتم مقارنة محفزين للتحقق من تأثيرهما على نتيجة تفاعل العملية. تم تحضير عينة من 12 عملية باستخدام المحفز 1 وواحدة من 10 مع المحفز 2. في الحالة الأولى، تم الحصول على متوسط قدره 84.5 مع انحراف معياري 3، بينما كان المتوسط للعملية الثانية هو 81 بالانحراف معياري قدرره 5.5. بافتراض ان المجتمع تبيع التوزيع الطبيعي.

- قدر بمجال عند مستوى ثقة 99% نسبة التبايني كلا المحفزات.
- بناءا على النتيجة السابقة، حدد مجال الثقة عن مستوى 90% للفرق بين متوسطي المحفزين في تفاعل العملية. – هل يمكننا أن تستنتج عن مستوى 90% أن هناك فرقا حقيقيا بين المتوسطين؟

التمرين 26: في مدينتي خميس مليانة وعين الدفلى تم إجراء دراسة حول تكلفة المعيشة لتقدير متوسط تكلفة الطعام للعائلات المكونة من 04 أفراد. تم اختيار عينات عشوائية مستقلة من 21 عائلة من كل من هاتين المدينتين وكانت النتائج:

$$\sum_{i=1}^{21} X_i^2 = 1.103.192500 \quad ، \quad \sum_{i=1}^{21} X_i = 139.150$$
$$\sum_{i=1}^{21} Y_i^2 = 1.114.252.400 \quad ، \quad \sum_{i=1}^{21} Y_i = 139.720$$

- إذا كانت تكلفة المعيشة في كلتا المدينتين موزعة طبيعيا، قدر بفترة ثقة عند مستوى 90% $\frac{\sigma_X^2}{\sigma_Y^2}$
- هل يبدو أن التباين في كلتا المدينتين هو نفسه؟ اشرح اجابتك.

التمرين 27: تم تصميم الأدوات الدقيقة التي تستخدمها Naftal للمراقبة وتوفير قياس يعتبر صحيحا فقط في المتوسط. اشتكى الزبائن الذين يصلون إلى محة الوقود من أنهم يتلقون في المتوسط كمية أقل من البنزين مما هو محدد على العداد. لذا تقوم المؤسسة بإرسال مراقب لفحص مضخة البنزين. في مضخة محطة الوقود، يقوم هذا المراقب بإجراء 10 قياسات على أباريق سعة 20 لترا: 20.5 و 19.99 و 20.0 و 20.3 و 19.90 و 20.05 و 19.79 و 19.85 و 19.95 و 20.15. إذا افترضت أن التوزيع طبيعي في قياسات 20 لتر، تحقق إحصائيا من ادعاء الزبائن.

- حدد الاختبار المناسب لهذه المشكلة.
- حدد عند مستوى دلالة 0.05، إذا كان ادعاء الزبائن صحيحا.

التمرين 28: خلص طبيب بيولوجي من خلال دراسة دواء معين إلى أن متوسط الوقت الذي يستغرقه فعالية الدواء لدى المرضى أقل من 40 دقيقة. لاختبار استنتاجه، يختار الطبيب عينة عشوائية من 15 مريض، وواظهرت النتائج ان المتوسط سيساوي 35 دقيقة بتباين قدره 170 دقيقة. من الدراسات السابقة ، من المعروف أن التوزيع طبيعي مع تباين 100 دقيقة. يرى الطبيب أن كلاهما لهما فرق كبير؛ لذلك، قرر أنه للحصول على قدر أكبر من اليقين الإحصائي، يجب عليه أولا إجراء اختبار للتباين. حدد اختبار الفرضية المناسب للتحقق من أهمية افتراض التباين عند 0.05.

التمرين 29: تدعي احدى المؤسسات المصنعة للمشروبات الغازية أنها تنتج ما متوسطه 250 مل ، ولكن بسبب شكاوى المستهلكين حول آلة معينة، قررت هذه المؤسسة مراقبة ذلك من خلال استعمال ال آلة 20 مرة والحصول على متوسط قدره 247 مل مع انحراف معياري قدره 10.5 مل. ادرس عند مستوى معنوية 0.10 ان كان ادعاء المؤسسة المصنعة صحيحا.

التمرين 30: بدأت إحدى الشركات التي تستخدم عبوات زجاجية سعة 2 لتر في الشكوى من المنتوج ، لأن مدير المؤسسة المنتجة لهذه العبوات يدعي أن متوسط الزجاجات أقل من 4 مم مع انحراف معياري أكبر من 0.07 مم. من أجل تقصي الأمر ارسل المدير مهندس مراقبة الجودة لاختبار ادعائه. يختار المهندس عينة من 25 زجاجة بشكل عشوائي ويقيس سمكها ويجد ان المتوسط والانحراف امعياري يقدران بـ 3.9 مم و 0.09 مم على التوالي. نفترض أن التوزيع طبيعي. هل ادعاء المدير صحيح عند مستوى معنوية 0.05.

التمرين 31: نفترض أن فعالية الأسمنت متغير عشوائي بتوزيع طبيعي $N\sim(\mu, 14400)$، تم تطوير نوع جديد من الأسمنت ، وحسب المقاول فان فعالية هذا النوع الجديد تساوي 5000 كجم/سم^2 ولهذا الغرض يريد اختبار الفرضية التالية:

$$H_0: \mu = 5000$$
$$H_1: \mu \neq 5000$$

للتحقق من اعتقاده ، يقوم بتقييم عينة بحجم 50 تحضيرا من نفس النوع، ويقرر أنه سيرفض استخدام هذا النوع في مشاريعه اذا كان متوسط المحاولات يساوب 4970. نفترض أنه من المعروف من التجربة أن الانحراف المعياري للمجتمع فيما يتعبق بفعالية الأسمنت لا يتغير. احسب احتمال ارتكاب خطأ من النوع الأول عندما يكون متوسط الفعالية الحقيقي يساوي 5010.

التمرين 32: يدعي مدير شركة انتاج مواد الغسيل أن مزيل البقع الجديد فعال في أقل من 70% من وقت استخدامه. للتحقق من هذا الادعاء، سيتم تطبيق هذا المنتج على 12 يقعة تم اختيارها عشوائيا. إذا تمت إزالة أقل من 10 بقع، فلن يتم رفض الفرضية الصفرية القائلة بأن $p = 0.7$.

- أوجد قيمة الخطا من النوع الاول، لنفرض ان $p = 0.8$
- عبر على احتمال الخطأ من النوع الثاني.

التمرين 33: لاحظ باحث يقوم بتحضير أطروحة الدكتوراه حول حوادث العمل أن الحوادث في المصانع لها توزيع طبيعي، هو لا يملك معلومات حول المعلمات، لكن وفقا لدراسته فان متوسط عدد الحوادث أقل من 13. لتحليل ادعائه إحصائيا قام بأخذ عينات عشوائية من 30 مصنعا، ووجد ان المتوسط الحسابي يساوي 10 بانحراف معياري قدره 9.25.

- حدد اختبار الفرضية المناسب لهذه المشكلة.
- اختبر عند مستوى معنوية 0.05 إذا كان ادعاء الباحث صحيحا إحصائيا، بفرض ان الانحراف المعياري يساوي 8.5.

التمرين 34: أسفرت عينة عشوائية من نفقات التشغيل الشهرية لشركة ما على مدى 26 شهر عن متوسط قدره 5774 دينار بانحراف معياري قدره 800 دينار. نفرض ان المجتمع موزع طبيعيا، بانجراف معياري قدره 750 دينار. اذا تأكد مسؤولو الشركة من أن متوسط نفقات التشغيل أعلى من 5400.

- حدد اختبار الفرضية المناسب لهذه المشكلة.
- اختبر الفرضية عند مستوى 0.05.

التمرين 35: عادة ما يتم إنتاج معدن معين بواسطة عملية تقليدية. تم تطوير عملية جديدة يتم فيها إضافة سبيكة إلى إنتاج المعدن. لكل معدن يتم اختيار 10 عينات بشكل عشوائي ويتعرض كل منها للضغط حتى ينكسر. يوضح الجدول الموالي ضغوط كسر العينات بالكيلوغرام لكل سم^2.

عملية تقليدية	438	462	478	442	453	445	433	480	448	462
عملية جديدة	457	473	443	372	458	474	419	479	474	449

- إذا فرضنا أن أخذ العينات قد تم في توزيعين عاديين ومستقلين. اختبر عند 0.10 دلالة إذا كانت تبايني المجتمعين متساويين.

– يدعي المصنعون أن السبيكة الجديدة تزيد عن متوسط مقاومة كسر المعادن التي تنتجها العمليتان بـ 2 كجم/سم^2. حدد اختبار الفرضية المناسب لاختبار ادعاء المؤسسة المصنعة عند مستوى 10%. استخدم نتيجة السؤال السابق.

التمرين 36: بعد عدة سنوات من تحليل نتائج حساب التفاضل والتكامل ، ثبت أنهما مرتبطين. لذلك يؤكد عميد الكلية أن علامات الطلبة في حساب التفاضل أعلى في المتوسط من درجات حساب التكامل بما يتراوح ما بين 1 و 2 نقطة. لاختبار هذا الادعاء، يتم اختيار عينة عشوائية من 10 طلبة ويتم تسجيل علاماتهم الموضحة في الجدول الموالي:

الطالب	1	2	3	4	5	6	7	8	9	10
حساب التفاضل	7	8	6	5	7	9	8	6	8	7
حساب التكامل	5	6	3	2	6	10	7	12	6	6

– نفترض أن الفرق بين العلامات موزع طبيعيا، اختبر إحصائيا ما إذا كانت ادعاءات العميد صحيحة.

– حدد اختبار الفرضية المناسب لهذه المشكلة، ثم قم باختبار الفرضيات مستوى 0.10.

التمرين 37: تدعي الشركة المصنعة للبطاريات أن نسبة العناصر المعيبة التي ينتجونها أقل من 5%. لاختبار ذلك تم أخذ عينة عشوائية من 250 بطارية ووجد ان هناك 8 بطاريات معينة.

– حدد اختبار الفرضية المناسب لهذه المشكلة.

– عند مستوى معنوية 10%، اختبر فيما إذا كانت ادعاءات الشركة لها ما يبررها.

التمرين 38: أخذ مصنع عينة من 640 عنصرا من المنتوج المصنع ووجد أن 15% من العناصر التي تنتجهاالالة A بها عيب طفيف، بينما في عينة مستقلة أخرى مكونة من 760 عنصر تم انتاجها بواسطة الالة B،كان هناك 8% فقط من هذا النوع من العيوب. يدعي المصنعون أن نسبة العناصر المعيبة التي تنتجها الآلة Aأكبر من تلك التي تنتجها الآلة B.

– اقترح اختبار الفرضيات المناسب للمشكل، ثم اختبر صحتها عندرعند مستوى معنوية 5.

التمرين 39: يهتم مصنع للمشروبات الغازية بمقارنة علامتين تجاريتين لعصير الليمون (A و B). لأنهم يشتبهون في تفضيل العلامة التجارية B على العلامة التجارية A ، يتم اختيار عينتين مستقلتين من 200 شخص تمت

مقابلتهم، منهم 116 يفضلون العلامة التجارية B ومن 150 آخرين، 78 يفضلون العلامة التجارية A. قم بصياغة اختبار مناسب للافتراض واختبر عند مستوى معنوية 10% إذا كانت شكوك ألمصنع في محلها.

التمرين 40: في عملية كيميائية تتم مقارنة محفزين للتحقق من تأثيرهما على نتيجة تفاعل العملية. تم تحضير عينة من 12 مرة باستخدام محفز العلامة التجارية L و12 من العلامة التجارية M. والنتائج موضحة في الجدول الموالي:

0.97	0.68	0.61	0.65	0.67	0.76	0.87	0.78	0.76	0.68	0.92	0.89	**L**
0.78	0.56	0.72	0.56	0.62	0.56	0.70	0.68	0.64	0.62	0.79	0.95	**M**

نفرض ان التوزيع طبيعي في تأثيرات المحفزات L و M. يؤكد المسؤولون عن العملية أن تبايني المجتمعين لكلا المحفزين متساوية. اختبر عند مستوى 0.05 إذا كان التباينين متساويين حقا.

قـائمة المـراجع

المراجع باللغة العربية

- جلاطو جيلالي، **الإحصاء مع تمارين ومسائل محلولة**، ديوان المطبوعات الجامعية، 2002.
- دومينيك سالفاتور، **الاحصاء والاقتصاد القياسي**، الطبعة الخامسة العربية، الدار الدولية للاستثمارات الثقافية، القاهرة، مصر، 2001.
- معتوق امحمد، **الاحصاء الرياضي والنماذج الاحصائية**، ديوان المطبوعات الجامعية، الجزائر، 2007.
- رويسات عبد الناصر، **الاحصاء الوصفي ومدخل الاحتمالات "دروس وتمارين"**، ديوان المطبوعات الجامعية، وهران، 2006.
- علي عبد السالم العماري، علي حسين العجيلي، **الاحصاء والاحتمالات: النظرية والتطبيق**، دار النشر ELGA، فاليتا، مالطا، 2000، ص 429.
- محمد صبحي أبو صالح، عدنان محمد عوض، **مقدمة في الإحصاء مبادئ وتحليل باستخدام SPSS**، دار المسيرة ،عمان، الأردن ،2004، ص 192.
- بوعبد الله صالح، **محاضرات الإحصاء الرياضي**، لطلبة كلية العلوم الاقتصادية، جامعة المسيلة، 2005-2006، ص 10.
- طالب محمد عوض، **مقدمة في الاقتصاد القياسي**، منشورات الجامعة الاردنية، عمان، الاردن، 2000، ص 131
- عدنان بن ماجد البري، محمود محمد هندي، أنور أحمد عبد هللا، **مبادئ الاحصاء والاحتمالات**، الطبعة الثالثة، جامعة الملك سعود، الرياض، السعودية، 1997، ص 362.

المراجع باللغة الاجنبية

- Jarkko. I.(2014). Basics of Statistics, course held in University of Tampere, Finland.
- Labatte, J-M, (2012), Biostatistiques, université d’Angers, disponible sur : https://math.univ-angers.fr/~labatte/enseignement%20UFR/MTVPS.html, Consulté le 09-03-2022
- Sedkaoui, S. (2018), Data Analytics and Big Data, ISTE-Wiley, London

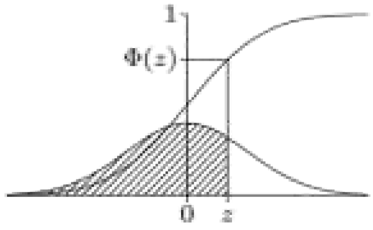
..

..

الملاحق

Loi normale $\mathcal{N}(0,1)$

1° *Fonction de répartition de la loi Normale.* — La fonction de répartition Φ de la loi Normale $\mathcal{N}(0,1)$ est définie par $\Phi(z) = \int_{-\infty}^{z} \mathrm{e}^{-u^2/2}\,\mathrm{d}u/\sqrt{2\pi}$, $z \in \mathbb{R}$. Pour tout $z \in \mathbb{R}$, on a $\Phi(z) = 1 - \Phi(-z)$.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

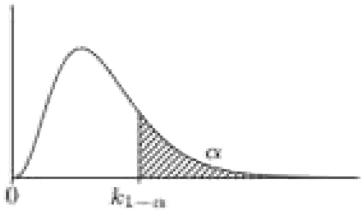
Exemples. — $\Phi(0{,}25) \approx 0{,}5987$, $\Phi(-0{,}32) = 1 - \Phi(0{,}32) \approx 1 - 0{,}6255 = 0{,}3745$.

Lois de χ^2

Si X est une variable aléatoire suivant la loi du χ^2, ou de Pearson, à ν degrés de liberté, la table donne, pour α fixé, la valeur $k_{1-\alpha}$ telle que

$$\mathbb{P}\{X \geqslant k_{1-\alpha}\} = \alpha.$$

Ainsi, $k_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi du χ^2 à ν degrés de liberté.



ν \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	18,4668
5	0,5543	0,8312	1,1455	1,6103	9,2364	11,0705	12,8325	15,0863	20,5150
6	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,4494	16,8119	22,4577
7	1,2390	1,6899	2,1673	2,8331	12,0170	14,0671	16,0128	18,4753	24,3219
8	1,6465	2,1797	2,7326	3,4895	13,3616	15,5073	17,5345	20,0902	26,1245
9	2,0879	2,7004	3,3251	4,1682	14,6837	16,9190	19,0228	21,6660	27,8772
10	2,5582	3,2470	3,9403	4,8652	15,9872	18,3070	20,4832	23,2093	29,5883
11	3,0535	3,8157	4,5748	5,5778	17,2750	19,6751	21,9200	24,7250	31,2641
12	3,5706	4,4038	5,2260	6,3038	18,5493	21,0261	23,3367	26,2170	32,9095
13	4,1069	5,0088	5,8919	7,0415	19,8119	22,3620	24,7356	27,6883	34,5282
14	4,6604	5,6287	6,5706	7,7895	21,0641	23,6848	26,1189	29,1412	36,1233
15	5,2293	6,2621	7,2609	8,5468	22,3071	24,9958	27,4884	30,5779	37,6973
16	5,8122	6,9077	7,9616	9,3122	23,5418	26,2962	28,8454	31,9999	39,2524
17	6,4078	7,5642	8,6718	10,0852	24,7690	27,5871	30,1910	33,4087	40,7902
18	7,0149	8,2307	9,3905	10,8649	25,9894	28,8693	31,5264	34,8053	42,3124
19	7,6327	8,9065	10,1170	11,6509	27,2036	30,1435	32,8523	36,1909	43,8202
20	8,2604	9,5908	10,8508	12,4426	28,4120	31,4104	34,1696	37,5662	45,3147
21	8,8972	10,2829	11,5913	13,2396	29,6151	32,6706	35,4789	38,9322	46,7970
22	9,5425	10,9823	12,3380	14,0415	30,8133	33,9244	36,7807	40,2894	48,2679
23	10,1957	11,6886	13,0905	14,8480	32,0069	35,1725	38,0756	41,6384	49,7282
24	10,8564	12,4012	13,8484	15,6587	33,1962	36,4150	39,3641	42,9798	51,1786
25	11,5240	13,1197	14,6114	16,4734	34,3816	37,6525	40,6465	44,3141	52,6197
26	12,1981	13,8439	15,3792	17,2919	35,5632	38,8851	41,9232	45,6417	54,0520
27	12,8785	14,5734	16,1514	18,1139	36,7412	40,1133	43,1945	46,9629	55,4760
28	13,5647	15,3079	16,9279	18,9392	37,9159	41,3371	44,4608	48,2782	56,8923
29	14,2565	16,0471	17,7084	19,7677	39,0875	42,5570	45,7223	49,5879	58,3012
30	14,9535	16,7908	18,4927	20,5992	40,2560	43,7730	46,9792	50,8922	59,7031

Lorsque le degré de liberté ν est tel que $\nu > 30$, la variable aléatoire

$$Z = \sqrt{2X} - \sqrt{2\nu - 1}$$

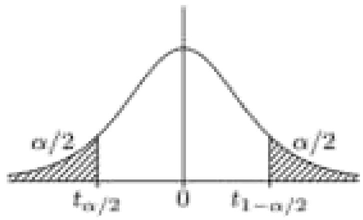
suit approximativement la loi normale centrée réduite.

Lois de Student

Si T est une variable aléatoire suivant la loi de Student à ν degrés de liberté, la table donne, pour α fixé, la valeur $t_{1-\alpha/2}$ telle que

$$\mathbb{P}\{|T| \geqslant t_{1-\alpha/2}\} = \alpha.$$

Ainsi, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à ν degrés de liberté.



ν \ α	0,900	0,500	0,300	0,200	0,100	0,050	0,020	0,010	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6193
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,1303	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,1286	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208
14	0,1280	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1405
15	0,1278	0,6912	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	4,0728
16	0,1277	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0150
17	0,1276	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9216
19	0,1274	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8834
20	0,1273	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8495
21	0,1272	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7921
23	0,1271	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,7454
25	0,1269	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7066
27	0,1268	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6896
28	0,1268	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6594
30	0,1267	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
40	0,1265	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	3,5510
60	0,1262	0,6786	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	3,4602
80	0,1261	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4163
120	0,1259	0,6765	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	3,3735
∞	0,1257	0,6745	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758	3,2905

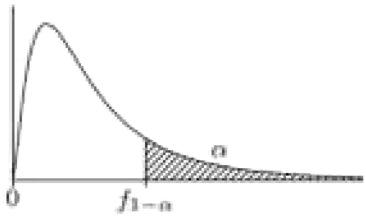
Lorsque $\nu = \infty$, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$.

Lois de Fisher–Snedecor ($\alpha = 0,05$)

Si F est une variable aléatoire suivant la loi de Fisher–Snedecor à (ν_1, ν_2) degrés de liberté, la table donne la valeur $f_{1-\alpha}$ telle que

$$\mathbb{P}\{F \geqslant f_{1-\alpha}\} = \alpha = 0{,}05.$$

Ainsi, $f_{1-\alpha}$ est le quantile d'ordre $1 - \alpha = 0{,}95$ de la loi de Fisher–Snedecor à (ν_1, ν_2) degrés de liberté.



$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	15	20	30	∞
1	161	200	216	225	230	234	239	242	246	248	250	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,94	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,85	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	2,72	2,65	2,57	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	2,62	2,54	2,47	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,20	2,12	2,04	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,15	2,07	1,98	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25	2,11	2,03	1,94	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,22	2,07	1,99	1,90	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,19	2,04	1,96	1,87	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	1,92	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03	1,87	1,78	1,69	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,84	1,75	1,65	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,95	1,79	1,70	1,60	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,93	1,77	1,68	1,57	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,67	1,57	1,46	1,00