الفصل الاول

خطة البحث

1-1 مقدمة:

إن التوزيعات الاحتمالية هي عبارة عن دوال تمثل القيم التي يأخذها المتغير المعرف على فترة معينة في فضاء العينة. تختلف خواص التوزيعات الاحتمالية باختلاف نوع المتغير(منفصل، مستمر) وباختلاف طبيعة البيانات. هذه الخواص سواء أكانت مقاييسا للنزعة المركزية (تقيس قرب البيانات من مركزها) مثل الوسط والوسيط والمنوال، أو كانت مقاييسا للتشتت (تقيس مدى تشتت البيانات أو بعدها عن مركزها) مثل التباين، تلعب الدور الأساسي في المجال التطبيقي.

الالتواء، هو خاصية تتأتى بسبب الاختلاف بين مقاييس للنز عة المركزية، أي أن الوسط والوسيط والمنوال لايتساوون، ينتج عن ذلك الشكل البياني المميز للتوزيع الذي يختلف عن باقي التوزيعات، ويمكن أن يسمى التوزيع بطبيعة التوائه، توزيع متماثل (وهو توزيع تتساوى فيه جميع مقاييس النزعة المركزية) و توزيع ملتو نحو اليمين و توزيع ملتو نحو اليسار.

وعند الحديث عن الالتواء لابد من الحديث عن التفرطح كذلك، لأنهما خاصيتان لشكل التوزيع ومقترنتان ببعضهما. وكما أنه للإلتواء أنواع، للتفرطح أنواع تختلف باختلاف التوزيع المستخدم والبيانات. وفي هذا البحث المراد هو دراسة التفرطح كعامل مؤثر في اختبارات جودة المطابقة الاحصائية.

اختبارات جودة المطابقة الاحصائية هي عبارة عن اختبارات تقيس مدى مطابقة قيم تسمى بالقيم المشاهدة، هذه القيم مأخوذة من أرض الواقع ويتم التحصل عليها بطرق

مختلفة مثل الاستبيانات او العينات العشوائية المختلفة، بنموذج معين أو بنمط معين أو بالقيم المتوقعة لهذا النمط.

هذه الاختبارات عديدة ومتنوعة في الاستخدامات والخواص، من اشهرها واكثرها استخداما اختبار كاي سكوير الذي يستخدم في اختبار عشوائية البيانات، ومن امثلتها ايضا اختبار كلوموقروف سميرنوف الذي سيتم استخدامه والتعرض اليه بشيء من التفصيل في هذا البحث بإذن الله.

تعتبر احصاءات اختبارات جودة المطابقة مؤشرات بتوزيعات احتمالية معروفة، يتم التحصل عليها عادة باستخدام الطرق المستخدمة في اختبارات الفروض.

وكما هو معروف، فان اقتراب العينات الكبيرة من التوزيع الطبيعي (نظرية النهاية المركزية)لا يظهر جليا في العينات الصغيرة، فقد قام عددا من الباحثين وباستخدام اسلوب المحاكاة بتكريس وقتا لدراسة دقة قيم الاحصاءات المتحصل عليها في حالة العينات الصغيرة، اي قام بدراسة كم يجب ان يكون حجم العينة لمختلف النماذج حتى يتم الحصول على دقة يمكن الاعتماد عليها.

من هنا جاءت اهمية هذه الدراسة، إذ أن موضوع دقة الاحصاءات موضوع ذو أهمية لأي باحث يستخدم منطق الاستدلال الاحصائي في بحثه، لان النتائج التي سيتم التوصل اليها ستكون غير حاسمة على الرغم من انفاق العديد من الاموال والحهود في نطاق التحصل على البيانات.

التفرطح، وهو الخاصية تحت الدراسة من الخواص التي لم يتم التعرض لها بشكل كاف عندما يتم تناول خصائص التوزيعات الاحتمالية، إذ نرى انها خاصية لا يجب اهمالها لما تحمله من دلالات، حيث سنرى ان كانت ذات صلة بدقة احصاءات اختبارات جودة المطابقة أم لا...

1-2 مشكلة البحث:

تتمثل مشكلة البحث في الاختلافات الظاهرة بين خصائص التوزيعات الاحتمالية من التواء وتفرطح وغيرها، وكذلك عدم وجود طرق يمكن من خلالها التحكم في هذه الخواص مما قد يؤدي الى عدم الدقة في الاختبار والذي بدوره يؤدي الى عدم الدقة في اتخاذ القرارت التي قد تكون ذات تأثير مهم.

1-3 أهداف البحث:

- تهدف الدراسة الى استخدام المحاكاة لمعرفة خواص التوزيعات الملتوية عموما والتوزيعات الملتوية الى اليمين بشكل خاص

- تهدف هذه الدراسة أيضا الى تسليط الضوء على اختبارات جودة المطابقة ومعرفة العلاقة بين هذه الاختبارات وخاصية التفرطح كخاصية مؤثر ة في دقة احصاءات جودة المطابقة.

- معرفة مدى تأثر قيم اختبار كولموقروف سميرنوف بقيم التفرطح المختلفة.

- استخدام الاختبارات اللامعلمية متمثلة في اختبار Kandal Tau لدراسة معنوية العلاقة بين اختبار كولموقروف سميرنوف وخاصية التفرطح.

4-1 أهمية البحث:

تقل الدراسات التي تتطرق الى دراسة خواص التوزيعات الملتوية بشكل عام والتوزيعات الملتوية الى اليسار بشكل خاص، من هنا تنبع أهمية هذه الدراسة اذ أنها تبحث في خواص أحد هذه التوزيعات وهو توزيع بيتا.

كما يعتبر هذا البحث اضافة للدراسات السابقة للباحثين في دقة اختبارات جودة المطابقة الاحصائية، لأنه يتناول خصائص التوزيعات بشكل عام والتفرطح بشكل خاص وأثره بهذه الاختبارات. يعتبر هذا البحث من قلائل البحوث التي تناولت هذا الموضوع والتي سلطت الضوء على خاصية التفرطح بشكل خاص كخاصية ذات أثر في اختبارات جودة المطابقة في التوزيعات الملتوية.

وأخيرا يعتبر هذا البحث من البحوث الأوائل في دراسة معنوية العلاقة بين اختبارات جودة المطابقة متمثلة في اختبار كلموقروف سميرنوف وخاصية التفرطح باستخدام الاختبارات اللامعلمية.

5-1 منهج البحث:

المنهج المستخدم في هذا البحثمنهج وصفي واستدلالي وذلك أن الدر اسة اعتمدت على بيانات مولدة عشوائيا باستخدام البرامج الاحصائية easyfit حيث تم توليد أربع عينات يحوي كل منها على 100،200،500، 800 مفردة لقياس اثر التفرطح على اختبار جودة المطابقة.

6-1 حدود البحث:

تقتصر هذه الدراسة على ماهية اختبارات جودة المطابقة الاحصائية بشكل عام وأحد امثلتها اختبار كلموقروف سميرنوف المستخدم في هذا البحث واثر خصائص التوزيعات المتمثل في خاصية التفرطح تحت الدراسة وعلاقتها باختبارات جودة

المطابقة حيث تم توليد المفردات عشوائيا وتم اختيار عينات عشوائية باحجام مختلفة وبما أن البيانات مولدة عشو ائيا فليس هنالك قانون لاختيار العينة.

1-7 اجراءات البحث:

ان مجتمع العينة في هذه الدراسة يختص باختبارات جودة المطابقة ويتم التركيز هنا على التفرطح وعلاقته بهذه الاختبارات وشكل العلاقة طردية أم عكسية

لذا فإن مجتمع العينة في هذه الحالة يمثل كل الاثباتات النظرية له ويتم توليد بيانات عشوائية بمتوسط حسابي يتم تغيير القيمة العددية له لدراسة علاقة قيمة معامل التفرطح مع قيمة احصاءة اختبار كلموقروف سميرنوف المستخدم في هذا البحث. كما أن عينة الدراسة أخذت بصورة عشوائية بحيث يكون لكل مفردات المجتمع نفس فرصة الظهور.

1-8 فروض الدراسة :

1. اختبارات جودة المطابقة تتأثر بخواص التوزيعات الملتوية.
2. هنالك علاقة بين اختبار كولموقروف سميرنوف وخاصية التفرطح في التوزيعات الملتوية الي اليمين وهذه العلاقة غير عشوائية.

1-9 الدراسات السابقة:

- بحث بعنوان تأثير حجم العينة عند المقارنة بين اختبارات جودة المطابقة باستخدام المحاكاة، اعداد الباحثة ليمان جودة أحمد، بغداد، 1989.

يجري البحث مقارنة بين مختلف اختبارات جودة المطابقة من حيث الدقة والقوانين وأنسب حالات الاستخدام وتم توليد عينات مختلفة واحجراء الاختبارات عليها.

فرضية الدراسة في هذا البحث هى أن لحجم العينة أثر معنوي في اختبارات جودة المطابقة، وبالفعل تم التوصل الى صحة هذه الفرضية.

- ورقة علمية من اعداد الباحث Lawrence DecarloL عام 1997 تطرق فيها الباحث الى خاصية التفرطح وأثره في التوزيعات، حيث عرف التفرطح بتدبباته المختلفة وتسطحاته الى ثلاثة أنواع. رأى الباحث أن من الاخطاء الشائعة هو الخلط بين التفرطح والتباين عند رؤية شكل التوزيع، وأوصى الباحثين بإجراء دراسات أعمق عن خصائص التوزيع عند استخدام التفرطح في بحوثهم.
- دراسة ماجستير بعنوان العوامل المؤثرة في قوة الاختبار اعداد الباحث أشرف حسن، جامعة السودان للعلوم والتكنلوجيا، 2012. قام الباحث بالكشف عن العوامل المؤثرة في قوة الاختبار الاحصائي، وذلك باستخدام أسلوب المحاكاة.

فرضيات الدراسة هي أن كلا من حجم العينة والانحراف المعياري وخصائص التوزيعات متمثلة في خاصية الالتواء تؤثر في قوة الاختبار، وتم التوصل الى أن هذه الفرضية صحيحة وتم إثباتها.

وعند المقارنة بين الدراسات السابقة والدراسة الحالية، نجد أن هذا البحث تطر ق الى خوص التوزيعات الاحتمالية وركز على خاصية التفرطح، وأثرها على اختبارات جودة المطابقة متمثلة في اختبار كولموقروف.

ولم النوصل الى أن هناك علاقة بين خاصية التفرطح واختبار كولموقروف في التوزيعات الملتوية الى اليمبن.

الفصل الثاني

التوزيعات وخواصها

2-1 التوزيعات الاحتمالية:

يمكن تعريف المتغير العشوائي بأنه متغير يأخذ قيما تحددها نتائج تجربة عشوائية. ذلك أنه كثيرا مايهمنا عند اجراء التجارب نواحي معينة فيها. فمثلا عند القاء زهري نرد قد ينصب اهتمامنا في حاصل ضرب الرقمين على السطحين العلويين، وعند اختيار شخص عشوائيا قد يهمنا قياس طوله أو الدرجة التي حصل عليها في اختبار. . . الخ. وفي كلتا الحالتين هناك كمية رقمية تعتمد على نتيجة التجربة العشوائية ولا يمكن التنبؤ بها مسبقا. وبما أن الكمية الرقمية تختلف باختلاف نتيجة التجربة العشوائية فهي متغير عشوائي. وبما أن قيمتها تعتمد على التنتائج الأولية في فضاء العينة فهي دالة معرفة على النتائج أو النقاط في فضاء العينة. ومن هذا المنطلق يمكن تعريف المتغير العشو ائي بصورة أكثر تحديدا كدالة $x(e)$ تأخذ قيما حقيقية ومعرفة على العناصر في فضاء العينة.

2-1-1 التوزيع الاحتمالي للمتغير العشوائي المنفصل

التوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن يأخذها المتغير، والتي ترتبط باحتمالات النتائج الممكنة في فر اغ العينة، وبمعنى آخر هو التكراري النسبي للقيم التي يمكن أن يأخذها المتغير.

فإذا كان المتغير العشوائي المنفصل X يأخذ القيم، $X:\{x=x_1,x_2,...,x_n\}$، وكان $P(X=x_i)=f(x_i)$ هو احتمال أن المتغير العشوائي يأخذ القيمة x_i، فإنه يمكن تكوين جدول التوزيع الاحتمالي للمتغير العشوائي X، وهو جدول مكون من عمودين، الأول به القيم

الممكنة للمتغير $X:\{x = x_1, x_2, ..., x_n\}$ ، والثاني به القيم الاحتمالية لهذا المتغير $P(X = x_i) = f(x_i)$ ، كما هو موضح في جدول (2−1):

جدول (2−1) التوزيع الاحتمالي للمتغير العشوائي المنفصل

x_i	$f(x_i)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_n	$f(x_n)$
Σ	1

وتسمى الدالة $f(x_i)$ بدالة الاحتمال، ومن خصائص هذه الدالة ما يلي:

$$1\text{-}\quad 0 < f(x_i) < 1$$
$$2\text{-}\quad \sum f(x_i) = 1 \qquad \textbf{...............1}$$

ويحسب الوسط الحسابي للمتغير العشوائي المنفصل كما يلي :

$$\mu = \sum_i x_i \; f(x_i) \qquad \textbf{.............2}$$

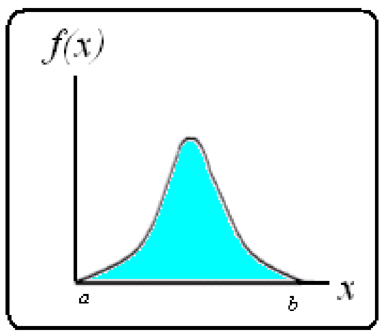
أ− بينما يحسب التباين بتطبيق المعادلة التالية :

$$\sigma^2 = \sum (x_i - \mu)^2 \; f(x_i)$$
$$= \sum x_i^2 \; f(x_i) - \mu^2 \qquad \textbf{............3}$$

2-1-2 التوزيع الاحتمالي للمتغير العشوائي المستمر :

المتغير العشوائي المستمر، هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهائي من القيم الممكنة له داخل مجاله، فإذا كان X متغير عشوائي مستمر، ويقع في المدى ***(a,b)***، أي أن: $\{X = x : a < x < b\}$، فإن للمتغير $\boldsymbol{X}$ عدد لانهائي من القيم تقع بين الحدين الأدنى والأعلى ***(a,b)***، شكل(1-2) يوضح دالة كثافة الاحتمال للمتغير العشوائي المستمر.

الشكل (1-2) دالة كثافة الاحتمال للمتغير العشوائي المستمر



ومن خصائص دالة كثافة الاحتمال $f(x)$ما يلي:

1-الدالة $f(x)$ موجبة داخل المدى ***(a,b)*** أي أن: $f(x) \succ 0$ ، $x \in (a,b)$

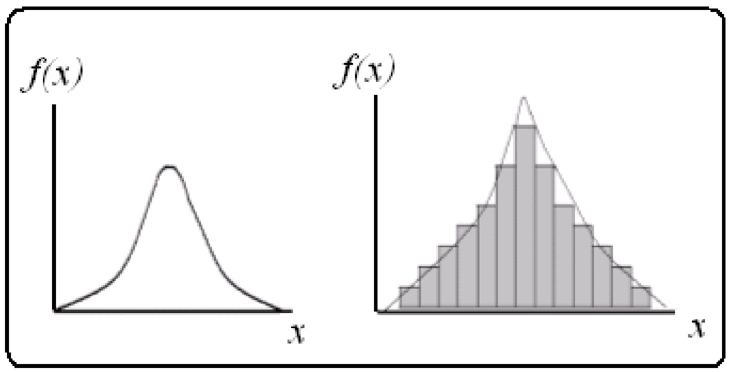
2-التكامل على حدود المتغير من الحد الأدنى $\boldsymbol{a}$ حتى الحد الأعلى $\boldsymbol{b}$ يعبر عن مجموع الاحتمالات الكلية، لذا يساوي الواحد الصحيح، أي أن:

$$\int_{x=a}^{x=b} f(x)\ dx = 1 \qquad \textbf{..............4}$$

و عند تمثيل بيانات المتغير الكمي المستمر في شكل مدرج تكراري نسبي،

نجد أن شكل هذا المدرج هو أقرب وصف لمنحنى التوزيع الاحتمالي للمتغير المستمر، وكلما ضاقت الفترات بين مراكز الفئات، يمكن الحصول على رسم دقيق للمنحنى الخاص بدالة احتمال المتغير المستمر والتي تسمي بدالة كثافة الاحتمال Probability Distribution Function*(p.d.f)*، كما هو مبين بالشكل (2−2):

شكل (2−2) منحنى التوزيع الاحتمالي للمتغير العشوائي المستمر



إذا كانت $f(x)$ هي دالة كثافة الاحتمال للمتغير العشوائي x، $a<x<b$ فإن التوقع الرياضي للدالة $h(x)$ تأخذ الصورة التالية:

$$E(h(x)) = \int^{b} h(x)\ dx \qquad \text{...........5}$$

ومن ثم يمكن كتابة معادلة الوسط والتباين كما يلي.

$$\mu = E(x) = \int_a^b x f(x) dx \qquad \textbf{.........6}$$

$$\sigma^2 = E(x^2) - u^2 \ , \ E(x^2) = \int x^2 f(x) \ dx \qquad \textbf{.......7}$$

2-3 التوزيعات الملتوية وخواصها:

لا تتبع بعض توزيعات البيانات الشكل المتماثل، فقد تضم البيانات قيماً متطرفة فتعمل على امتداد التوزيع من أحد طرفيه وهذا يؤدي إلى التواء المنحنى، وقد تضم البيانات قيماً كثيرة في المنتصف بحيث تظهر قمة منحنى التوزيع مدببة ومرتفعة. وقد يتركز عدد كبير من القيم في المنتصف بحيث يكون التوزيع عريض أو ذو قمة مفلطحة .

علىالرغم من أن التوز يع التكراري يمكن أن يتخذ أي شكل، إلا أنه توجد بعض الأشكال النموذجية التي تناسب معظم التوزيعات التي يقابلها الباحث في المواقف الفعلية، ومن بين هذه التوزيعات: التوزيع الاعتدالي وهو توزيع يشبه الجرس المقلوبوالتوزيع الملتوي التواء ً موجباً والذي تتراكم فيه قيم المتغير حول النهاية الدنيا للتوزيعوالتوزيع الملتوي التواء ً سالباً حيث تتراكم فيه قيم المتغير حول النهاية العليا للتوزيع.

فإذا لم يكن التوزيع اعتدالياً فانه يجب أن لا يكتفي الباحث عند وصف التوزيع بالمتوسط والانحراف المعياريوا ٕنما يحتاج إلى مقياس آخر يعبر عن مدى ابتعاد التوزيع عن الاعتدالية، أي درجة التوائه. ومن المرغوب فيه أيضاً أن يصف التوزيع بمقياس آخر يعبر عن درجة تفلطح أو تدبب التوزيع.

1-3-2 الالتواء Skewness

في حالة عدم تطابق مقاييس النزعة المركزية المنوال والوسيط والوسط الحسابي يعد التوزيع ملتوياً . في حالة التوزيعات المتماثلة يتساوى المتوسط والوسيط والمنوال ، وكلما بعد المنحنى عن التماثل بعدت هذه القيم بعضها عن البعض، ولذلك يمكن استخدام الفرق بين هذه القيم كمقياس للالتواء إلا أن هذا الفرق لا يقيس الالتواء تماماً ، فقد يكون الفرق كبير والالتواء صغيراً ، لأن تشتت قيم البياناتكبيراً ، وقد يكون الفرق صغيراً والالتواء كبيراً لأن تشتت المجموعة صغيراً ، ولذلك يجب أن ينسب هذا الفرق إلى مقياس التشتت المناظر (من نفس نوع مقياس القيمة المتوسطة المستخدم) ويسمى المقياس الناتج بمعامل الالتواء . وبصفة عامة يجب أن يحقق معامل الالتواء الشرطين التاليين:

- أن يساوي صفراً للمنحنيات المتماثلة.
- أن يكون عدداً بحتاً فلا يتوقف على الوحدات التي يقاس بها المتغير .

والالتواء هو درجة عدم التماثل أو الانحراف عن التماثل فإذا كان منحنى التوزيع له طرف على يمين مر كز التوزيع أطول من الطرف الأيسر ، فان التوزيع يسمى ملتوي لليمين أو أن له التواء موجب، وا ٍذا حدث العكس يقال أن التوزيع ملتوي لليسار أو أنه سالب الالتواء.

يعتمد قياس التواءات التوزيعات الإحصائية على معرفة مقاييس النزعة المركزية وهي المتوسط الحسابي، الوسيط، الربيع الأول، الربيع الثاني والمنوال. والمنحنى المتماثل الذي لا يوجد فيه أي التواء تنطبق عليه المقاييس الثلاثة أي أن المتوسط الحسابي = الوسيط =المنوال، أي الفرق بينهم يساوي صفراً .وا ٍذا كان الفرق يختلف عن الصفر كان هذا دليلاً على وجود الالتواء.

وحيث أن المنوال هو أكثر القيم تكراراً فهو يقع تحت قمة المنحنى مباشرة فإذا التوى المنحنى جهة اليمين انتقلت قمته جهة اليمين وانتقلت معه قيمة المنوال إلى اليمين، وكذلك الحال إذا التوى جهة اليسار.

2-3-2 قياس الالتواء بطريقة بيرسون:

سبق الإشارة إلى تساوي قيم المتوسط والوسيط والمنوال في حالة التوزيع المتماثل، وقد تختلف قيمها عن بعض في حالة التوزيع غير المتماثل، بحيث يكون الوسيط في الوسط دائماً بينما يكون المتوسط في جهة القيم المتطرفة (ذيل المنحنى الممتد) حيث أنه يتأثر بها، ويكون المنوال في الجهة الأخرى التي يتركز بها معظم قيم التوزيع ، ولذلك يمكن قياس الالتواء مبدئياً بالصيغة التالية.

الالتواء = المتوسط – المنوال.

ولكن يعيب المقياس السابق للالتواء أنه ليس مقياس يعطي عدد مطلق حيث يأخذ نفس وحدة قياس المتغير، كما أنه لا يأخذ في الاعتبار تشتت التوزيعات المختلفة، فقد تكون قيمة الالتواء واحد لتوزيعين مختلفي التشتت مما يعطي الالتواء معنى مختلف في الحالتين، وللتغلب على تلك العيوب يستخرج مقياس نسبي للالتواء وذلك بقسمة (المتوسط – المنوال) على الانحراف المعياري للتوزيع.

$$\text{الإلتواء} = \frac{\text{المتوسط} - \text{المنول}}{\text{الانحراف المعياري}}$$

ويمكن ايضا استخدام الصيغة :

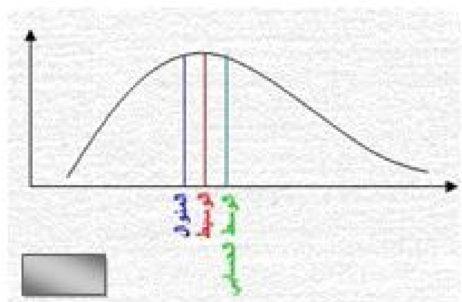
$$\text{الإلتواء} = \frac{3(\text{المتوسط} - \text{المنوال})}{\text{الانحراف المعياري}}$$

حيث ان الوسيط يقع في منتصف المسافة بين المنوال والمتوسط في التوزيعات التكرارية متوسطة الالتواء للمتغيرات المتصلة .

2-3-3 التوزيعات الملتوية نحو اليمين:

وفيها يكون التوزيع غير متماثل بحيث تتراكم معظم التكرارات حول الطرف العلوي للتوزيع وتقل التكرارات كلما اتجهنا نحو الطرف السفلي، ويكون الطرف الأيمن للمنحنى ممتد وله ذيل متجه نحو اليمين، وفي هذه الحالة يكون الوسط الحسابي أكبر من الوسيط، الوسيط أكبر من المنوال كما هو موضح في الشكل (2-3)

شكل (2-3) التوزيعات الملتوية نحو اليمين



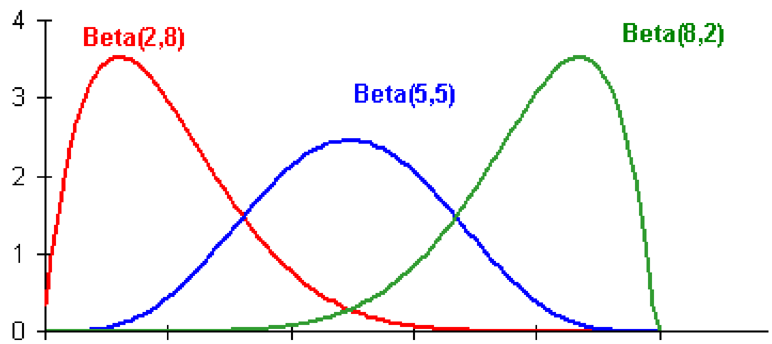
2-3-4 توزيع بيتا:

ومن أمثلة التوزيعات الملتوية لليمين توزيع بيتا وهو توزيع احتمالي مستمر ويعتبر من التوزيعات المرنة التي تمثل البيانات ذات الالتواء الموجب او السالب على السواء .

دالة توزيع بيتا هي دالة مرنة جدا لتقديم النتائج وعرضها في صورة نسب أو احتمالات. معرفة في المدى من صفر الي واحد وتحتوي دالة التوزيع على معلمتان

تعملان سويا على تحديد شكل الدالة. شكل (2−4) يوضح دالة الكثافة الاحتمالية لتوزيع بيتا.

شكل (2−4) توزيع دالة الكثافة الاحتمالية لتوزيع بيتا



تستخدم الدالة ليس فقط لتوضيح الاختلافات الملاحظة بين المشاهدات المختلفة من اراء وغيرها ، بل تستخدم أيضا في طريقة بيز للتقدير.حيث يعتبر توزيع بيتا من التوزيعات المناسبة للاستخدام توزيع قبلي للبيانات. وفيما يلي سنتطرق الي توضيح بعض خواص هذا التوزيع.

دالة بيتا المعرفة في الفترة من(0،1) تأخذ الشكل التالي:

$$Beta(\alpha,\beta):\ prob(x|\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{............8}$$

حيث B هي دالة بيتا:

$$B(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \qquad \text{............9}$$

خصائص دالة بيتا:

المتوسط :

$$E(x) = \mu = \frac{\alpha}{\alpha + \beta} \qquad \textbf{.........10}$$

التباين :

$$V(\mu) = \frac{\beta}{(\alpha + \beta)(\alpha + \beta + 1)} \cdot \mu \qquad \textbf{.......11}$$

$$Variance(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \qquad \textbf{.......12}$$

الالتواء :

$$Skewness(x) = \frac{2(\beta - \alpha)\sqrt{1 + \alpha + \beta}}{\sqrt{\alpha + \beta}(2 + \alpha + \beta)} \qquad \textbf{.......13}$$

التفرطح:

$$Kurtosis(x) = \frac{6[\alpha^3 + \alpha^2(1 - 2\beta) + \beta^2(1 + \beta) - 2\alpha\beta(2 + \beta)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} \qquad \textbf{......14}$$

أحد أهم مزايا توزيع بيتا هوأنه يأخذ العديد من الأشكال وبالتالي فانه يخدم العديد من التطبيقات. كما يجدر بالذكر أن دالة توزيع بيتا ذات التواء نحو اليمين (التواء موجب) وهو سبب اختيار هذا التوزيع في عملية التحليل التي ستأتي لاحقا إنشاء الله.

2-4 خاصية التفرطح

التفرطح هو قياس درجة علو قمة التوزيع بالنسبة للتوزيع الطبيعي وهو يقيس درجة ارتفاع التوزيع والذي عادة ما ينسب إلى التوزيع الاعتدالي، فإذا كان للتوزيع قمة مرتفعة أكبر من التوزيع الاعتدالي يقال أنه مدبب (Leptokurtic)وإ ذا كان التوزيع ذو قمة مسطحة يقال أنه مفلطح (Platykurtic) ،وإ ذا كانت قمة التوزيع متوسطة ليست مدببة وليست مفرطحة يسمى متوسط التفرطح (Mesokurtic) .

وحيث أن ارتفاع قمة التوزيع الطبيعي تساوي 3تقريباً، فان التوزيع يكون مفرطحاً عندما يكون معامل التفرطح أقل من 3، ويكون التوزيع مدبباً عندما يكون معامل التفرطح أكبر من 3.

وخاصية التفرطح ليس لها علاقة بالمتوسط الحسابي للتوزيع فقد يكون هناك أكثر من توزيع لهم نفس المتوسط الحسابي ولكن يختلف شكل المنحنى من مدبب أو مسطح.

تفرطح التوزيع يشير إلى الاستواء أو التدبب في التوزيع بالنسبة لغيره من التوزيعات . فخاصية التفرطح تعد خاصية نسبية.

ويمكن تعريف درجة التفرطح باستخدام العزوم الرابعة حول الوسط:

$$\beta_2 = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{\mu_4}{\sigma^4}, \qquad \text{........15}$$

حيث E هو عامل التوقع، و μ يعني المتوسط، 4μ هي العزوم الرابعة حول الوسط، و σ هو الانحراف المعياري.

يمكن أن نجد في العديد من المراجع أن تفرطح التوزيع الطبيعي يساوي الصفر ، كما يمكن أن تكون قيمة التفرطح سالبة وهذا ليس بالخطأ، حيث تعتمد هذه المراجع مايسمى بال EXCESS KURTOSIS والذي يقترن مفهومه بالتوزيع الطبيعي المساوي 3

نتيجة لذلك فان أقل قيمة يمكن أن يأخذها التفرطح هي −2 ممتدة الى موجب مالا نهاية

ويحسب ال EXCESS KURTOSIS بالصيغة :

$$K = n\,\frac{\sum_{i=1}^{n}(X_i - X_{avg})^4}{\left(\sum_{i=1}^{n}(X_i - X_{avg})^2\right)^2} - 3 \qquad \textbf{..........16}$$

بعض المفاهيم الخاطئة الشائعة فيما يتعلق بدرجة التفرطح:

يمكن الحصول على مزيد من التبصر في درجة التفرطح بالتطرق الى بعض المفاهيم الخاطئة حول هذا الموضوع التي تظهر في عدد من الكتب مثل:

(أ) أن تعرف درجة التفرطح من ناحية الذروة peakedness، بدون أي إشارة إلى أهمية الذيول

(ب) العلاقة بين الذروة وذيول التوزيع يتم وصفه أو توضيحه بشكل غير صحيح.

(ج) الوصف والرسوم التوضيحية لدرجة التفرطح لا يميز بينه وبين والتباين.

وبما أن الهدف الأساسي من الإحصاءات هو تنظيم وتلخيص البيانات، وكما يتم استخدام المتوسط والتباين كمقاييس موجزة للتوزيع الاحتمالي والتغيرات التي تطرأ عليه فبالمثل الانحراف ودرجة التفرطح تقدم معلومات موجزة حول شكل التوزيع. على الرغم من وجود قيود على درجة التفرطح. شكل(2−5) يوضح التغيرات التي

تطرأ على التوزيع الاحتمالي، يصاحبها تغيرات تطرأ على تفرطح التفرطح التوزيع وا إنحرافه المعياري.

شكل(2−5) تغيرات تطرأ على التوزيع الاحتمالي، تصاحبها تغيرات تطرأ على تفرطح التوزيع وانحرافه المعياري

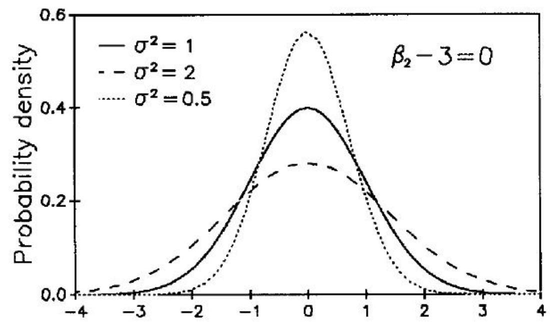
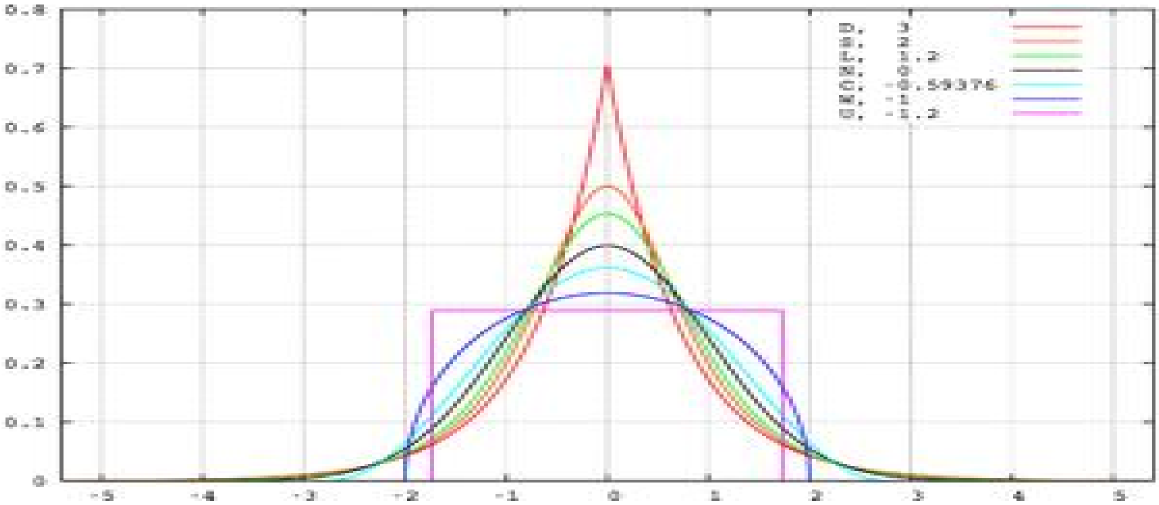


Figure 4. Three normal distributions with variances (σ^2) of 0.5, 1, and 2.

شكل(2−6) أيضا يوضح درجات التفرطح المختلفة لعدد من التوزيعات الاحصائية بما فيها التوزيع الطبيعي. وكما هو موضح فإن كل توزيع له درجة تفرطح قد تزيد أو تقل تبعا لطبيعة البيانات المستخدمة.

شكل (2−6) التفرطح للتوزيعات المختلفة

!Error



D :Laplace distribution

S :hyperbolic secant distribution

L :logistic distribution

N normal distribution

C :raised cosine distribution.

W :Wigner semicircle distribution

U :uniform distribution.

الفصل الثالث

اختبارات جودة المطابقة والاختبارات اللامعلمية

3-1 اختبارات جودة المطابقة:

اختبارات جودة المطابقة للنموذج الاحصائي تصف مدى جودة أو قوة مطابقته لمجموعة من المشاهدات، كما تلخص الوصف بين القيم المشاعدة والقيم المتوقعة وذلك تحت نموذج احصائي معين. احصاءات اختبارات جودة المطابقة يتم التحصل عليها عادة بالطرق المعروفة في اختبارات الفروض.

وكما هو معروف ان اقتراب العينات الكبيرة من التوزيع الطبيعي لا يظهر جليا في العينات الصغيرة، لذلك قام عدد لا يستهان به من العلماء بتكريس الوقت وباستخدام اسلوب المحاكاة لدراسة مدى دقة قيم الاحصاءات، بمعنى اخر كم يجب ان يكون حجم العينة لمختلف العينات ولمختلف الحالات.

ان تقييم النموذج الاحصائي، أي قياس التناقض بين النموذج والبيانات أمر غاية في الاهمية في التطبيقات، اذ ان الاستدلالات المتوصل اليها قد تكون مضلله وغير حاسمة على الاطلاق. ويجب على الباحثين في المجال التطبيقي أن لا يدرسو ا التطابق الكلي للنموذج فحسب بل يجب أن يشمل ايضا دراسة التطابق الجزئي كذلك، اذ انه في بعض الحالات يظهر أن النموذج مطابق ككل، ولكن عند الدراسة الجزئية له يتضح جليا عدم التطابق، كما تساعد الدراسة الجزئية على تحديد مصدرعدم التطابق للمعالجات اللازمة ان أمكنٍ .

هكذا، يمكن تصنيف تقييم جودة المطايقة باستخدام تقسيمان مفيدان: مؤشرات جودة المطابقة مقابل إحصاءات جودة المطابقة، و المطابقة المطلقة مقابل المطابقة النسبية. وبذلك، يمكن تصنيف إحصاءات ومؤشرات جودة المطابقة الى كلي أو

جزئي. وهناك أيضا تقسيم ثالث مفيد لتصنيف تقييم جودة المطابقة يستند إلى طبيعة البيانات المتحصل عليها، منفصلة مقابل مستمرة.

تاريخيا، تقييم جودة المطابقة للبيانات المنفصلة متعددة المتغيرات وتلك للمتغيرات المتعددة للبيانات المستمرة تم تقديمهم بطر ق مختلفة تماما ويرجع ذلك الى التطورات الجديدة في مجال المعلومات.

3-2 اختبار كولموقروف – سميرنوف للعينة الواحدة

المشكلة في اختبار χ^2 لجودة التوفيق أن تقريب توزيع إحصائية الاختبار بتوزيع χ^2 لن يكون جيداً إذا كان حجم العينة صغيراً أو كانت التكرارات المتوقعة في بعض الخلايا أقل من قيم معينة كما رأينا.فإذا لم نتمكن من استخدام اختبار χ^2 لجودة التوفيق لأحد هذه الأسباب فإنه يمكننا اللجوء لاختبار كولموقروف– سميرنوف، أو اختبار كولموقروف كما يطلق عليه أحيانا أخرى لأن كولموقروف أول من أشار إليه.

ويتميز اختبار كولموقروف–سميرنوف بأن توزيع إحصائية الاختبار فيه معروف تماماً حتى عندما يكون حجم العينة صغيراً ما دامت ليست هناك معالم تحتاج لتقدير. كما يتميز بأنه يمكننا من انشاء فترة ثقة لـ $F(x)$. ولكن لم يتضح بعد أي الاختبارين (كلوموقروف – سميرنوف أو χ^2) أكثر قوة.

لنفرض أن لدينا العينة العشوائية $X_1 ,X_2 , \ldots, X_n$ من مجتمع دالة توزيعه $F(x)$ مجهولة* . نفرض أيضاً أن دالة التوزيع التراكمية المشاهدة emperical distribution function من العينة هي $S(x)$. هذه الدالة تعطى لأي قيمة x ، نسبة القيم في العينة التي تساوي أو تقل عن x، وتستخدم عادة لتقدير دالة التوزيع في المجتمع $F(x)$.

ليكن فرض العدم الذي نرغب في اختباره هو أن دالة التوزيع الحقيقية للمجتمع الذي جاءت منه العينة هي $F^*(x)$ أي أنه لكل x

$$H_o : F(x) = F^*(x)$$

وتحدد إحصائية الاختبار لاختبار كولموقروف – سمير نوف حسب الفرض البديل على النحو التالي:

1) إذا كان الفرض البديل هو:

$$H_1\ F^*(x) \neq F(x)$$

لـ x واحده على الأقل ، فإن إحصائية الاختبار تكون أكبر مسافة رأسيه بين $F^*(x)$ و $S(x)$ سواء كانت $F^*(x)$ أكبر من $S(x)$ أو أصغر. بمعنى آخر إحصائية الاختبار هي أكبر قيمة لكل x للفرق المطلق $| S(x)-F^*(x) |$ ونكتب هذا بالشكل:

$$T = \sup_x \left| F^*(x) - S(x) \right|$$

حيث $\sup_x$ تعني الأكبر لكل x .

2) إذا كان الفرض البديل هو:

$$H_1 : F(x) < F^*(x)$$

لـ x واحدة على الأقل ، فإن احصائية الاختبار هنا هي أكبر مسافة رأسية لـ $F^*(x)$ فوق $S(x)$، أي أكبر قيمة للمقدار $[S(x)-F^*(x)]$ لكل x ونكتب:

$$T^- = \sup_x \left[F^*(x) - S(x) \right]$$

لاحظ أن فرض العدم في هذه الحالة يأخذ الشكل:

$$H_o : F(x) \geq F^*(x)$$

3) إذا كان الفرض البديل هو:

$$H_1 : F(x) > F^*(x)$$

فإن إحصائية الاختبار تكون أكبر مسافة رأسية لـ $F^*(x)$ تحت $S(x)$ أي أكبر قيمة للفرق $[\, F^*(x)-S(x) \,]$ لكل x ونكتب

$$T^+ = \underset{x}{Sup}\left[S_{(x)} - F^*_{(x)} \right]$$

ويأخذ فرض العدم الشكل

$$H_o : F(x) \leq F^*(x)$$

حالة الاختبار ذو طرفين :

إذا كانت دالة التوزيع المفترضه متقطعة، فبما أن أكبر مسافة لابد أن تحدث عند إحدى قيم x، فكل الذي نحتاجه لإيجاد القيمة المشاهدة لـ T هو حساب $S(x)$ و $F^*(x)$ ثم إيجاد الفرق $| F^*(x) - S(x) |$ عند كل قيمة x وأكبر هذه الفروق هو القيمة المشاهدة للإحصائية T. أيضاً يمكن ايجاد قيمة T بيانياً برسم منحنى $S(x)$ ومنحنى $F^*(x)$ وتحديد أكبر مسافة بينهما وهي القيمة المشاهدة لـ T.

أما إذا كانت دالة التوزيع متصلة فينبغي توخي الحذر عند إيجاد قيمة T جبرياً .

الاختبار ذو طرف واحد :

في هذه الحالة تكون T^+ هي أكبر فرق بين $F^*(x)$ و $S(x)$ لقيم x التي تكون عندها $S(x)$ أكبر من $F^*(x)$ ، بينما T^- أكبر فرق لقيم x التي تكون عندها $S(x)$ أقل من $F^*(x)$.

من أهم مزايا اختبار كولموقروف – سميرنوف أنه يسمح لنا بإنشاء نطاق ثقه لدالة التوزيع $F(x)$. فبما أن $S(x)$ مقدر نقطه لـ $F(x)$ فيمكن استخدامها لانشاء نطاق يحوي $F(x)$ بكاملها، أي لجميع قيم x، بدرجة ثقة معينة.

3-3 الاختبارات اللامعلمية:

من المتعارف عليه ان الاختبارات الإحصائية المعلمية Parametric تتبع نوع ما من التوزيع سواء كان طبيعي، ذي الحدين، بواسون .. الخ. وان هذا التوزيع يكون اساسا لتقدير العينة من المجتمع الاحصائي، ولذلك فانه يجب ان نفترض دائما ان كل اختبار معلمي يتعلق فقط بتوزيع بيانات المجتمع.
لكن لايمكن التسليم بهذا الافتر اض في كل الاحو ال ، حيث هناك بعض الامثلة العملية او الحياتية التي من غير الممكن استخدام الطرق المعلمية لتحليلها، لذلك فمن الملائم تحليل البيانات بطرق يطلق عليها لا معلمية او خالية من التوزيع Distribution free Methods.

مثل هذه الاختبارات لاتتطلب معلمات عن توزيع المجتمع، علاوة على ذلك فانها لاتحتاج الى حسابات معقدة. لكن يجب عدم استعمال الاختبارات اللامعليمة لمجرد انها اسرع واسهل من الاختبارات المعلمية القياسية. ويكون استعمالها في حالة لايمكن عمل افتراضات بالنسبة لتوزيعات المجتمع ، او عند الحاجة لاجراء اختبار

لمجموعات صغيرة من البيانات مستمدة من دراسات صغيرة. وفي هذه الحالة الاخيرة يجب اعتبارها مقدمة لاختبار اكثر دقة بالطرق المعيارية القياسية.

كما ان الباحث الذي يحل دون تمييز اختبار لا معلمي محل اختبار معلمي قد يضحي بالفعالية،و الفعالية نعني بها هنا مقدرة اختبار ما على كشف الفرضيات غير الصحيحة، وبمعنى اخر احتمال حدوث خطا من النوع الثاني يكون اقل عند استخدام اختبار معلمي قياسي.و فيما يلي بعض النقاط التي يمكن الاسترشاد بها للاختيار بين طريقة معلمية وبين اختبار لا معلمي:

1- اذا كان من غير الممكن عمل فرضية تتعلق بنوع التوزيع الذي يتبعه متغير فانه يشار الى استخدام اختبار لامعلمي.

2- الاختبارات اللامعلمية تكون مفيدة للحصول على تقديرات سريعة من كميات صغيرة من البيانات التجريبية.

3- عند عمل الاستدلالات الاحصائية والتعميميات الواسعة من بيانات عينة فان الاختبارات المعلمية تكون ضرورية.

3-3-1 مزايا وعيوب الطرق اللامعلمية:

تتميز الطرق اللامعلمية بعدة مزايا مقارنة بالطرق المعلمية, ومن أهم هذه المزايا ما يلي:

1. قلة الافتراضات المطلوبة

لعل أهم مزايا الطرق اللامعلمية أنها لا تتطلب افتراضات كثيرة حول توزيعات المجتمع كما هو الحال في الطرق المعلمية. وهذه ميزة كبيرة لأن مستخدم الطرق

الإحصائية قد لا يعرف ما إذا كانت الافتراضات التي تقوم عليها متحققة في بياناته، أو قد يعرف أن بعضاً منها أو كلها غير متحققة.

استخدام طريقة معلمية دون التأكد من تحقق الافتراضات التي بنيت عليها يؤدي إلى تقليل دقتها. أما الطرق اللامعلمية فإن الافتراضات القليلة التي قد تتطلبها تكون عادة عامة ومن النوع الذي يتحقق في معظم التطبيقات مثل افتراض أن التوزيع متصل. كما أن بعض الطرق اللامعلمية لا تتطلب أي افتراضات حول التوزيع . على سبيل المثال اختبار t للفرق بين متوسطين من عينتين مستقلتين يفترض لاستخدامه أن يكون توزيع المجتمعين طبيعياً وأحياناً بتباينين متساويين .أما الشبيه اللامعلمي لهذا الاختبار فلا يتطلب من الافتراضات أكثر من أن يكون توزيعا المجتمعين متصلين.

2. إمكانية التطبيق على البيانات الوصفية والترتيبية

في كثير من الدراسات وخاصة في مجال العلوم الاجتماعية تكون لدينا بيانات طبيعة وصفية أو ترتيبية. فمن البيانات الوصفية تلك التي تمثل تصنيف ذات الوحدات حسب معيار معين, مثلاً تصنيف أشخاص حسب نوعهم أو جنسيتهم. ومن البيانات الترتيبية على سبيل المثال ترتيب مجموعة من العمال حسب درجة حماسهم أو رضائهم عن عملهم.

ولقد استحدثت معظم الطرق اللامعلمية لمعالجة مثل هذا النوع من البيانات. ذلك أن تطبيق الطرق المعلمية على بيانات وصفية أو ترتيبية يؤدي عادة لنتائج يصعب أو يستحيل تفسيرها أو إعطاؤها معنى.

3. السرعة في جمع البيانات وتحليلها:

بما أن البيانات المستخدمة في الطرق اللامعلمية تكون عادة بمقاييس دنيا مثل المقياس الأسمى أو الترتيبي, وبما أن معظم الطرق اللامعلمية لا تتطلب حجم عينة كبير أو حسابات معقدة فإنه يمكن جمع البيانات وتحليلها بسرعة أكبر. فمثلاً إذا كانت البيانات ستقاس بالمقياس الأسمى وحجم العينة صغير جداً فإن جمع البيانات لن يستغرق سوى القليل من الوقت.

4. سهولة الفهم:

تقوم غالبية الطرق اللامعلمية على مفاهيم بسيطة تستند عادة إلى فكرة التباديل أو العشوائية (Randomization) وهذا يسهل استيعابها وتفهم المنطق الذي يسندها.

5. إمكانية إعطاء عبارات احتمالية مضبوطة:

نتيجة للميزة السابقة فإن التوزيع الاحتمالي للكثير من الإحصائيات في الطريقة اللامعلمية يكون مضبوطاُ , وبالتالي فإن بعض العبارات الاحتمالية مثل قيمة p في اختبار الفروض تكون مضبوطة أي تعطى تماماً exact وليست تقريبية كما هو الحال عادة في الإحصاء المعلمي حيث يعتمد مدى تقريبها على تحقق الافتراضات التي تقوم عليها الطريقة.

6. سعة مجال التطبيق:

إن إمكانية تطبيق الطرق اللامعلمية على البيانات ذات مستويات القياس المنخفضة مثل البيانات الاسمية والترتيبية،وا إمكانية استخدامها في مسائل الاستدلال غير المرتبطة بمعالم كاختبار عشوائية تتالي مجموعة من القيم. إضافة لعدم تقيدها بافتراضات كثيرة جعل مجال التطبيق هذه الطرق أوسع منه في الطرق المعلمية. ومن المزايا الأخرى للطرق اللامعلمية أنها في معظم الحالات لا تتأثر بعدم تحقق

الافتراضات التي تقوم عليها والتي هي في الأصل افتراضات ضعيفة وقليلة كما ذكرنا. هذه الخاصية تمثل ما يطلق عليها في الإحصاء الاستدلالي بالإنكليزية Robustnessوتشير لقدرة الأداة الإحصائية للصمود في وجه عدم تحقق افتراض تقوم عليه. كذلك فإن بعض الاختبارات اللامعلمية أقوى من شبيهتها المعلمية خاصة في حالة العينات الصغيرة.

ومن ناحية أخرى تعاني الطرق اللامعلمية من عيوب ينبغي التنبه لها. ومن هذه العيوب مايلي:

1. الحسابات في الطرق اللامعلمية قد تغدو بالغة التعقيد إذا كان حجم العينة كبيراً .
2. الاختبارات بصفة عامة أقل قوة في الطرق اللامعلمية منها في شبيهاتها المعلمية خاصة في حالة العينات الكبيرة وعند تحقق الافتراضات التي تقوم عليها الطريقة المعلمية. ذلك أن هذه الافتراضات تحدد عادة في الطريقة المعلمية بحيث تحقق لها القوة.
3. قلة الافتراضات , عدم اشتراط عينات كبيرة، وسهولة الحسابات المطلوبة في معظم الطرق اللامعلمية يشجع الكثير من الباحثين أحياناً على تفضيل الطريقة اللامعلمية رغم وجود طريقة معلمية ذات قوة أكبر.

3-3-2 حالات الاستخدام:

يمكننا أن نتبين بعد تعرفنا على مزايا وعيوب الطرق اللامعلمية الحالات التي يمكن أن يلجأ فيها الباحث إليها. ومن أهم هذه الحالات:

1. عندما تكون البيانات مقاسة على المقياس الاسمي (Nominal Scale)أو المقياس الرتبي(Ordinal Scale) أو تكون بيانات عد (count data) ويقصد ببيانات العد تلك التي تكون على شكل تكرارات تعطي عدد الوحدات التي تقع في كل مستوى من مستويات متغير معين. فإذا حددنا فئات دخل معينة مثلاً (منخفض – متوسط – مرتفع) وصنفنا مجموعة من العاملين حسب هذه الفئات، فإن أعداد أو تكرارات العمال في الفئات الثلاث تمثل بيانات عد.

2. عندما يكون الباحث غير متأكد من تحقق افتراضات أساسية تعتمد عليها الطريقة المعلمية المناسبة (مثل فرض التوزيع الطبيعي أو كبر حجم العينة) أو يعلم أنها غير متحققة.

3. حين لا تستهدف الدراسة الاستدلال على معلم , كما هو الحال عند اختبار استقلال عاملين أو اختبار عشوائية تسلسل أحداث معينة.

4. عندما يكون المطلوب استخدام طريقة سريعة لا تتطلب حسابات معقدة. وهذا ينشأ عادة في حالة العينات الصغيرة.

وهناك العديد من الاختبارات اللامعلمية التي يمكن استخدامها لاختبار معنوية العلاقة بين المتغيرات المختلفة، على سبيل المثال اختبار كاندال تاو اللامعلمي والذي يعتبر من الاختبارات التي يكثر استخدامها كبديل جيد للاختبارات المعلمية، وفيما يلي سنتطرق لهذا الاختبار وسيتم استخدامه في هذه الدراسة لبيان جودته.

3-3-3 اختبار كندال تاو Kendall Tau:

يشبه اختبار كندال تاو معامل إسبيرمان في أنه يمكن تطبيقه أيضاً على البيانات التي تكون في شكل رتب أو تحول الى رتب. ورغم أن حسابه أكثر تعقيداً من معامل إسبيرمان، إلا أنه يتميز بأن توزيعه يتجه للتوزيع الطبيعي بشكل أسرع حينما يزيد حجم العينة؛ مما يجعله الخيار الأفضل في حالة العينات الكبيرة.

وكما هو الحال في معامل إسبيرمان تستخدم كندال تاو كمقياس لقوة العلاقة أو الارتباط بين متغيرين، لاختبار استقلال متغيرين ولاختبار وجود اتجاه عام عندما يمثل أحد المتغيرين الزمن. ومن حيث القوة يمكننا أن نقول بصفة عامة أنه لم تتوفر حتى الآن مبررات قوية تجعلنا نفضل أحد المعاملين على الآخر في اختبارات المعنوية. وكمقاييس لقوة العلاقة نجد معامل إسبيرمان يميل لأن يكون أكبر من معامل كندال تاو من حيث القيمة المطلقة. وسيتم شرح طريقة استخدام كاندال تاو في الجانب التطبيقي من الفصل الرابع.

الفصل الرابع

التحليل

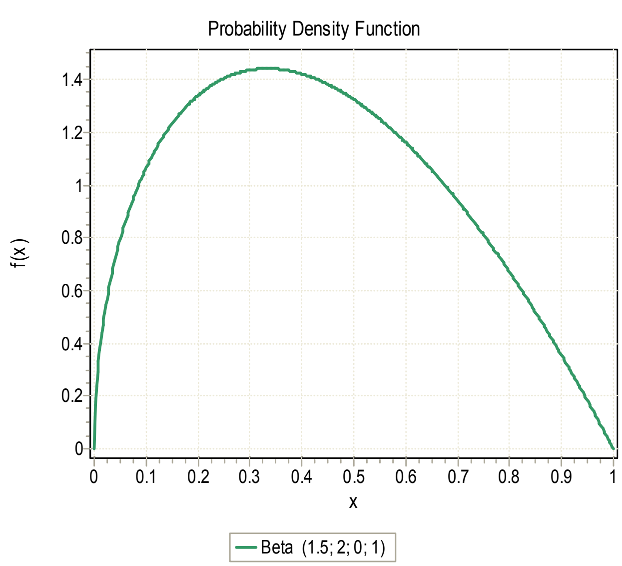
4-1 مقدمة :

في هذا الفصل تم اعتماد المحاكاة كاسلوب للتحليل، حيث تم توليد أربع عينات عشوائية باحجام مختلفة وهي 100 ، 200، 500، 800.

التوزيع الاحتمالي المستخدم في توليد البيانات هو توزيع بيتا والذي هو أحد التوزيعات الاحتمالية الملتوية الى اليمين ومن أسباب اختياره أن التفرطح يظهر جليا في رسمه البياني كما هو موضح في الشكل (4-1) ادناه.

تم استخدام المعادلتين(2-1) و (2-2) الموضحتين في الفصل الثاني كما تم استخدام قيم مختلفة للمعلمتين(β,α) وذلك للحصول على نتيجة يمكن من خلالها التحكم في شكل التوزيع عند استخدام أحجام العينات المختلفة.

شكل (4-1) تفرطح توزيع بيتا



4-2 العلاقة بين اختبار كلوموقروف سميرنوف وخاصية التفرطح :

تم حساب معامل التفرطح، ومن ثم حساب قيمة احصاءة اختبار كلوموقروف للبيانات لكل عينه على حده كما هو موضح في الجداول من (4-1) الي (4-4) وستتم المقارنة بين خاصية التفرطح وقيمة احصاءة كولموقروف باستخدام أحد الاختبارات اللامعلمية وهو اختبار كاندال تاو وذلك لمعرفة معنوية العلاقة بينهما.

جدول (4-1) العلاقة بين اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 100

Shape parameter	kurtosiss	k.s test
1.5 , 2	−0.6759	0.0842
2 , 3	−0.4540	0.0453
4 , 6	−0.2797	0.0488
5 , 8	0.0135	0.0506

جدول (4-2) العلاقة بين اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 200

Shape parameter	kurtosiss	k.s test
1.5 , 2	−0.8119	0.0476
2 , 3	−0.6297	0.0427
4 , 6	−0.4406	0.0404
5 , 8	−0.3471	0.0251

جدول (4-3) العلاقة بين قيم اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 500

Shape parameter	kurtosiss	k.s test
1.5 , 2	−0.8448	0.0277
2 , 3	−0.5540	0.0270
4 , 6	−0.3011	0.0256
5 , 8	−0.0698	0.0190

جدول (4−4) العلاقة بين اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة800

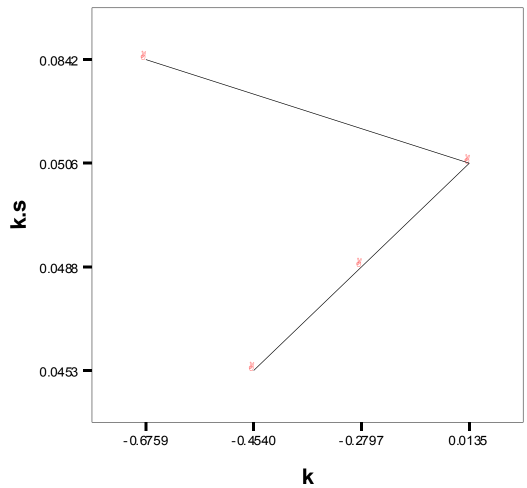
Shape parameter	kurtosiss	k.s test
1.5 , 2	−0.8752	0.0240
2 , 3	−0.8182	0.0228
4 , 6	−0.4451	0.0199
5 , 8	−0.2433	0.0195

القيم السالبة في قيم التفر طح للعينات المختلفة تدل على أن تفرطح توزيع بيتا أقل من التوزيع الطبيعي كما تم الاشارة اليه في الفصل الثاني.

بملاحظة قيم التفرطح وقيم اختبار كولموقروف نجد ان العلاقة بين القيمتين تبدا عشوائية عند استخدام حجم عينة صغير نسبيا 100 كما هو موضح في جدول (4−1) وتقل هذه العشوائية كلما زاد حجم العينة كما هو موضح في

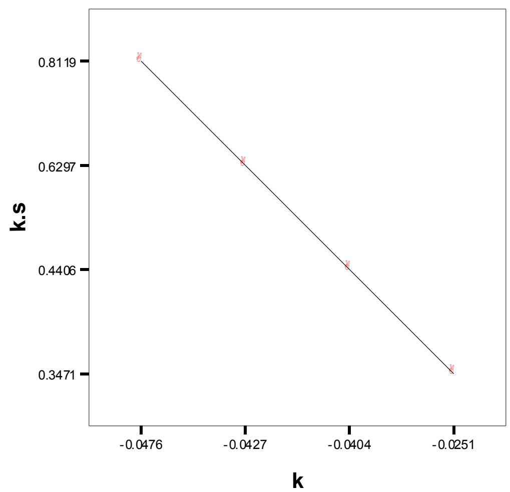
جداول (4-2)، (4-3) و (4-4) وتظهر علاقة عكسية واضحة بين القيمتين.يمكن ايضاح هذه العلاقة من خلال الرسم البياني في شكل (4-2) الى (4-5).

شكل (4-2) العلاقة بين قيم اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 100



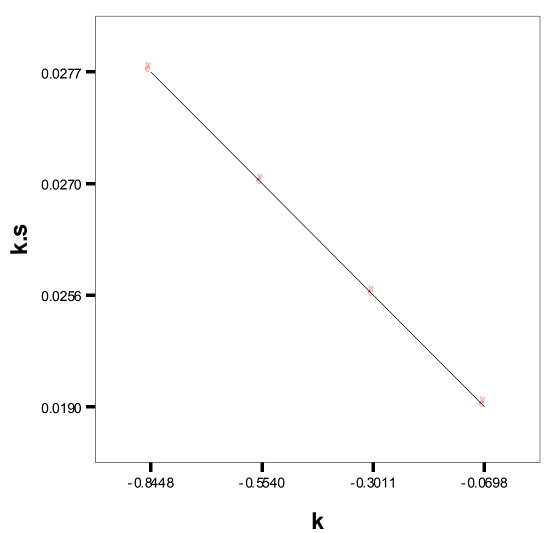
المصدر: من اعداد الباحث

شكل (4-3) العلاقة بين قيم اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة200



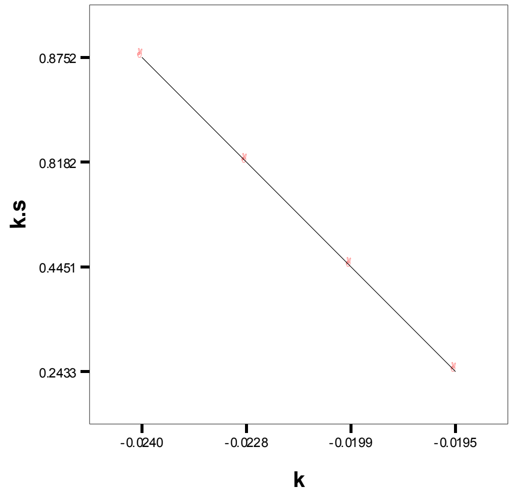
المصدر: من اعداد الباحث

شكل (4-4) العلاقة بين قيم اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 500



المصدر: من اعداد الباحث

شكل (4-5) العلاقة بين قيم اختبار كلوموقروف سميرنوف والتفرطح بإستخدام حجم عينة 800



المصدر: من اعداد الباحث

3-4 اختبار كانداول تاو

تم استخدام الاختبارات اللامعلمية وتطبيق اختبار كاتدال تاو لمعرفة معنوية العلاقة بين معامل التفرطح وقيمة اختبار كلوموقروف سميرنوف.

اجراءات اختبار كاندال تاو كالتالي:

(1) بإفتراض أن X تمثل قيم التفرطح Y تمثل اختبار كلوموقروف تم ترتيب قيم X وتحديد قيم Y المقابلة لها ووضعهما في جدول .

(2) تم مقارنة قيم y مع القيم التي تأتي بعدها لتحديد ماذا كانت اكبر أو اصغر منها، ونرمز ب c اذا كانت اكبر و d اذا كانت أصغر .

(3) تم حساب قيمة تاو بإستخدام القانون : C-D\C+D

حيث Cعدد قيم ال c الكلية و D عدد القيم الكلية ل d .

(4) تم حساب قيمة z ومن ثم استخراج قيمة P.value ومقارنتها مع قيمة اختبار كاندال تاو لمعرفة معنوية العلاقة. جدول(4–5) يوضح اجراءات اختبار كاندال تاو وقيم كل من c و d.

جدول(4−5) اجراءات اختبار كندال تاو :

x	Y	Rank y	Num 0f c's	Num of d's
−0.9947	0.0191	23	2	22
−0.8752	0.0240	14	10	13
−0.8448	0.0277	10	13	9
−0.8182	0.0228	15	9	12
−0.8119	0.0476	4	17	3
−0.7123	0.0207	17	7	12
−0.6759	0.0842	1	18	0
−0.6297	0.0427	6	14	3
−0.5540	0.0270	11	10	6
−0.4540	0.0453	5	13	2
−0.4451	0.0199	20	4	10
−0.4406	0.0404	8	10	3
−0.4077	0.0203	18	5	7
−0.3471	0.0251	13	6	5
−0.3011	0.0256	12	6	4
−0.2797	0.0488	3	8	1
−0.2433	0.0195	22	3	5
−0.1812	0.0200	19	3	4
−0.1126	0.0208	16	3	3
−0.1110	0.0163	25	0	5
−0.0698	0.0190	24	0	4
−0.0192	0.0197	21	0	3
0.0135	0.0506	2	2	0
0.1548	0.0351	9	0	1
0.1735	0.0409	7	0	0

من جدول (4−6) نجد ان قيمة D = 163 وقيمة C= 137.

تم تعويض هذه القيم في معادلة اختبار كاندال تاو كما يلي :

$$Tau=c-d\backslash c+d$$

$$=(137-163)\backslash(137+163)$$

$$=0.0866$$

تم ايجاد قيمة Z بالتعويض في المعادلة :

$$Z=3*tau((0.5n(n-1)\backslash 2n+5)^{1/2}$$

$$=-0.6067$$

ومن ثم استخدام z في استخراج قيمة p-value والتي وجد انها تساوي 0.272 بإستخدام مستوي معنوية 0.05 يلاحظ ان قيمة P.value أكبر من قيمة $\alpha/2$ وبالتالي نستنتج ان هنالك علاقة معنوية بين معامل التفرطح واختبارات جودة المطابقة المتمثلة في اختبار كولموقروف.

الفصل الخامس

النتائج والتوصيات

5-1 النتائج:

بعد استخدام أسلوب المحاكاة و توليد أربع عينات بأحجام مختلفة تتبع للتوزيع الاحتمالي بيتا،و استخدام قيم معالم مختلفة في كل عينة، كما تم استخدام اختبار كاندال تاو على النتائج المجدولة تم التوصل الى أن العلاقة بين قيمة معامل التفرطح والذي يمثل احد خصائص التوزيعات الاحتمالية هي علاقة معنوية مع اختبارات جودة المطابقة الممثلة باختبار كلوموقروف سميرنوف، هذه النتائج اثبتت ايضا ان هذه العلاقة علاقة عكسية، اي انه كلما زادت قيمة معامل التفرطح قلت قيمة احصاءة كلوموقروف وتتضح هذه العلاقة العكسية كلما زاد حجم العينة، هذا يدل على أن خواص التوزيعات الاحتمالية ذات أثر في اختبارات جودة المطابقة مما يستوجب مراعاه هذا الامر عند استخدام منطق الاستدلال الاحصائي من قبل الباحثين للحصول على نتائج يمكن الاعتماد عليها.

يمكننا أن نلخص النتائج والاستنتاجات التي تم التوصل اليها في هذا البحث في عدة نقاط كالتالي:

(1) ان حجم العينة يؤثر في شكل العلاقة بين قيمة معامل التفرطح واختبار جودة المطابقة في التوزيعات الملتوية الى اليمين.

(2) خواص التوزيعات الاحتمالية مثل الالتواء والتفرطح وغيرها تؤثر في قيمة اختبارات جودة المطابقة في التوزيعات موجبة الالتواء.

(3) هناك علاقة معنوية بين معامل التفرطح واختبار كلولموقروف سميرنوف، هذه العلاقة علاقة غكسية، حيث أنه كلما زادت قيمة معامل التفرطح قلت قيمة اختبار كولموقروف.

2-5 التوصيات:

من خلال النتائج التي توصلنا اليها انه لا يمكن للباحث الذي يستخدم منطق الاستدلال أن يهمل خواص التوزيعات الاحتمالية لان هذا الامر يودي الى الوصول الى نتائج يمكن القول انها مهزوزة او لا يمكن الاعتماد عليها، بل يجب عليه أن ينظر نظرة شمولية ويحيط بجميع خواصها عند استخدام أحدها.و كما توصلنا في هذا البحث الى أن خاصية التفرطح ذات اثر في اختبارات جودة المطابقة، فإننا نوصي بتوسيع الدراسة لتشمل الاتي:

*اختبار علاقة خاصية التفرطح باختبارات جودة المطابقة في التوززيعات سالبة الالتواء.

*اختبار خواص التوزيعات الاحتمالية الأخرى مثل خاصية الالتواء ومعرف علاقتها باختبارات جودة المطابقة.

5-3 المراجع والاوراق العلمية

1. كتاب الاستدلال الاحصائي ، تأليف أ.د.أحمد عودة عبدالمجيد عودة، زين العابدين عبدالرحيم البشير.

2. Browne, M. W. (1982). Covariance structures. In Hawkins, D. M. (ed.)

3. Browne, M. W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A. and Long, J. S. (eds.) Testing Structural Equation Models, pp 136–162. Newbury Park, CA: Sage.

4. Hosmer, D. W. and Lemeshow, S. (2000). Applied Logistic Regression. New York: Wiley.

5. MacCallum, R. C., Roznowski, M., and Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin 111, 490–504.

6. MacCallum, R. C., Wegener, D. T., Uchino, B. N., and Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. Psychological Bulletin 114, 185–199.

7. W. J. Conover(1999),"*Practical Nonparametric Statistical* " ,3rd edition, pp.428-433 (6.1), John Wiley & Sons, Inc. New York.