توفيق نماذج توزيعات Tweedie الاحتمالية مع بيانات التركيب العمري والنوعي للسكان في اليمن - دراسة تطبيقية

د. حسن حسن على عبد الملك

استاذ مساعد– قسم العلوم المالية والمصرفية– كلية التجارة – جامعة اب

hasmalik5@hotmail.com

الملخص

4

تهدف الدراسة الي توضيح عملية توفيق نماذج توزيعات Tweedie الاحتمالية مع بيانات التركيب العمري والنوعي للسكان في الجمهورية اليمنية بحسب ما أشار اليها نتائج التعداد السكاني لعام 2004، وذلك من خلال التعرف على: (1) دراسة مدى دقة بيانات التوزيع العمري والنوعي باستخدام نماذج توزيعات Tweedie الاحتمالية. (2) دراسة الخطوات المتبعة لتطبيق وتقدير معالم نماذج توزيعات Tweedie الاحتمالية. (3) اجراء مفاضلة لدرجة الأهمية النسبية لتوفيق بيانات نماذج توزيعات Tweedie الاحتمالية مع التوزيع العمري والنوعي للسكان الجمهورية اليمنية. وتوصلت الدراسة إلي أن جودة التوفيق بين بيانات نماذج توزيع Tweedie و بيانات التوزيع العمري و النوعي لسكان الجمهورية اليمنية كانت جيدة حيث بلغ المتوسط العام لجودة التوفيق لكل من الذكور والاناث وكلاهما معاً بطريقة MAPE نسبة 81%.

Fitting Tweedie Distributions Models with Data of the Age and Gender Composition for the Population in Yemen
(A Case Study)

Hassan Hassan Ali Abdulmalik
Assistant Professor - Department of Banking and Finance
College of Commerce - University of Ibb - Republic of Yemen

Abstract:

The study aims to clarify the process of fitting the models of Tweedie distributions with data of the age and gender composition for the population in the Republic of Yemen, according the results of the census of 2004, by identifying: (1) Studying the accuracy of the data of age and gender distribution using Tweedie Distribution Model. (2) Study the steps involved in applying and estimating the parameters of Tweedie distributions models. (3) Perform a comparison of the degree of relative importance for fitting the data of Tweedie distributions with age and gender distribution for the Republic of Yemen. The study found that the quality of Tweedie distribution models data fitting with the age and gender distribution data was good, as the overall average quality of fitting for males, females and both of them together, was 81% by MAPE.

أولاً: الإطار المنهجي العام للدراسة

1. المقدمة:

تعد بيانات التوزيع العمري والنوعي لسكان في المجتمع ذات أهمية كبرى في عملية التخطيط للتنمية الاجتماعية والاقتصادية[6]، كونها تساعد على إيجاد توازن في العلاقة بين السكان والموارد البيئية، التي ينتج عنها انشأ مجتمع يتمتع بالاستقرار المعيشي بكل أشكالها الاقتصادية والاجتماعية، ونظراً لذلك ولضمان التخطيط الجيد لذلك ينبغي أن تكون البيانات الإحصائية المعبرة عن تلك العلاقة أن تكون دقيقة بقدر الإمكان، وبالرغم من ذلك نجد انه يعتمد على هذه البيانات من المصدر الرسمي المتمثل التعدادات السكانية، وبالتالي فإن هذه العملية غالباً ما تتعرض إلي وجود أخطأ في بيانات التعداد نتيجة لتعدد مراحل التعداد وتداخلاتها[1]. وتتلخص هذه الأخطاء بمحورين هما أخطأ الشمول وأخطأ التغطية[5]. ولمعرفة وسائل اكتشاف هذه الأخطاء في بيانات التوزيع العمري والنوعي يتم استخدام العديد من الأساليب العلمية الإحصائية، من أهمها: الأساليب المباشرة والتي تهدف الي الاتصال المباشر بالمبحوثين الاصلين ومراجعة كل حالة على حدة. والأساليب غير المباشرة والتي تهدف الي مقارنة بيانات التعداد الحالي بنتائج تعداد سابق أو أي مصدر أخر للبيانات كالإحصاءات الحيوية. وأخيراً أساليب الاتساق والتي تهدف مقارنة بيانات التعداد الحالي ببعضها البعض[3]، إلا أن الطرق الثلاث السابقة لها عيوب من أهمها: التكلفة المادية لأجراء تعداد أخر وأيضاً عدم توفر الإحصاءات الحيوية وهذا غالباً لا تتوفر لدى البلدان النامية وأخيراً المقارنة بتعداد اخر أو سابق يكمن العيب فيما إذا كان التعداد السابق ايضاُ غير دقيق. لذا ظهرت طرق إحصائية مختلفة أخرى لقياس دقة الاعمار وتقيمها ومن أهمها: مقياس سكرتارية الأمم المتحدة (UN Age-Sex Accuracy)، ويبل (Weibull)، مايرز (Myers) وغيرها من الطرق الأخرى[7]. ولوجود الأخطاء في بيانات التوزيع العمري ظهرت أساليب إحصائية لتوفيق بيانات التوزيع العمري للسكان ومن أهم هذه الطرق طريقة كارير فراج (Carrier & Farrag)، طريقة كارب – كنك – نيوتن (Karup-King-Newton)، طريقة أريجا (Arriga)، وطريقة الأمم المتحدة (UN.Method 1/16)، وطريقة سترونج (Strong) [2]. لاحظ الباحث أن طرق التوفيق السابقة استندت على المتوسطات وهذا يعني النظر الي تمركز البيانات فقط، دون النظر الي مقدار التشتت الحاصل بين الفئات العمرية، وايضاً تم توفيق فئات معينة دون الاخرى وأخيراً كل طرق التوفيق ليس لها القدرة على اكتشاف الأخطاء. مهما يكن من أمر ذلك نجد أن لكل من هذه الطرق مزايا وعيوب. بشكل عام هناك طريقة يقترحها الباحث في توفيق بيانات التوزيع العمري والنوعي وذلك

باستخدام التوزيعات الاحتمالية وعلى وجه الخصوص نماذج توزيعات Tweedie الاحتمالية ذات التشتت الزائد، وذلك لأسباب التالية: أن النموذج يدرس تمركز البيانات وتشتتها، وايضاُ يقوم بدراسة التوزيعات الاحتمالية كنموذج مقترح لتوفيق البيانات، وأخيراً فأن النموذج له القدرة لاكتشاف الأخطاء التي تطرأ في بيانات التوزيع العمري والنوعي الناتج عن التعدادات.

2. مشكلة الدراسة:

ترتبط عملية توفيق البيانات بمجموعة من المبادئ والطرق الاحتمالية المناسبة للحصول على منحنى ممهد للعمليات المستنتجة، بهذا تتميز عمليات التوفيق بمدى اقتراب القراءات الممهدة من البيانات المشاهدة. لذا لا يمكن اعتبار عمليتي التمهيد والتوفيق غير مستقلتين عن بعضهما البعض. حيث أن توفيق سلسلة البيانات المشاهدة تعني أن سيتم تغييرها واستبدالها بقيم تختلف عن القيم المشاهدة، بمعنى آخر عندما تكون سلسلة البيانات المسواة قريبة من البيانات المشاهدة هذا يعني تحسن في التوفيق وبالتالي ينخفض دقة التمهيد بالضرورة والعكس صحيح. لذلك تعد عملية التمهيد والتوفيق متعارضتان بمعنى أنه يجب تسوية البيانات الخام في حدود المعقول والسليم بحيث لا نضحي بالتوفيق، ولكي تكون طريقة التوفيق طريقة عامة يمكن استعمالها، يجب أن يستطيع القائم بالتوفيق اختيار الأهمية النسبية لكل من التمهيد والتوفيق في السلسلة المسواة في كل حالة على حدى، ونظراً على لذلك قد تحتوي الإحصاءات التي يتم تقديرها واستخلاصها من التعدادات السكانية إلى وجود أخطأ، والتي تؤدي بدورها إلى أخطاء في حساب تقدير السكان، وقصور في دقة بيانات التوزيع العمري والنوعي لسكان. ونظراً الجهاز المركزي للإحصاء هي الجهة الفاعلة والرئيسية لتوفير التعداد السكاني، لذا أصبح من الأهمية تقييم هذا التعداد وعلى وجه التحديد تعداد 2004 باعتباره أخر تعداد اجري في اليمن والتأكد من خلو هذا التعداد من الأخطاء باستخدام نماذج توزيعات Tweedie الاحتمالية. وعلى وجه التحديد سيتناول الدراسة الآتي: (1) ما مدى دقة بيانات التوزيع العمري والنوعي باستخدام نماذج توزيعات Tweedie الاحتمالية. (2) ما هي الخطوات المتبعة لتطبيق وتقدير معالم نماذج توزيعات Tweedie الاحتمالية. (3) ماهي درجة الأهمية النسبية لتوفيق بيانات نماذج توزيعات Tweedie الاحتمالية مع التوزيع العمري والنوعي للسكان الجمهورية اليمنية.

3. أهداف الدراسة:

تهدف الدراسة إلى توفيق بيانات نماذج توزيعات Tweedie الاحتمالية مع التوزيع العمري والنوعي لسكان الجمهورية اليمنية، من خلال الأهداف الفرعية الآتية:

(1) دراسة مدى دقة بيانات التعداد السكاني للجمهورية اليمنية باستخدام نماذج توزيعات Tweedie الاحتمالية.

(2) التعرف على نماذج توزيعات Tweedie الاحتمالية وكيفية تطبيقها وتقدير معالمها وكيفية توفيق بياناتها مع بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية.

(3) إجراء المقارنة أو المفاضلة لاحتساب الأهمية النسبية لتوفيق البيانات باستخدام نماذج توزيعات Tweedie الاحتمالية.

4. أهمية الدراسة:

تكمن أهمية الدراسة في العديد من الجوانب منها:

(1) تبرز الدراسة توضيح المفاهيم العامة لبعض الأساليب الإحصائية المستخدمة في تقديم طريقة تقدير معالم توزيعات Tweedie الاحتمالية التي تلائم طبيعة هذه التوزيعات والتي ستضيف معرفة علمية تكاد تكون غير متوفرة في المكتبات اليمنية.

(2) إعطاء القيادات ومتخذي القرار والمخططين في الجهاز المركزي للإحصاء في الجمهورية اليمنية نموذج متكامل وبأسلوب واحد –نماذج توزيعات Tweedie الاحتمالية–وذلك من خلال تزويدهم بالمعلومات العلمية التي تساعدهم في كيفية اكتشاف الأخطاء في بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية وكيفية تقيمها وتوفيقها دون التنقل الي أساليب إحصائية أخرى كما هو المتعارف في الدراسات السكانية السابقة.

(3) تبرز الدراسة طريقة مستحدثة باستخدام التوزيعات الاحتمالية وعلى وجه التحديد نماذج توزيعات Tweedie الاحتمالية لاستخدامها لتقيم التعدادات السكانية وتمهيد بياناتها لتلافي مشاكل التكلفة المادية او عدم توفر البيانات الإحصاءات الحيوية.

(4) إعطاء الباحثين في التخصصات الأخرى كيفية التعامل مع النماذج الاحتمالية وعلى وجه التحديد نماذج توزيعات Tweedie الاحتمالية وعلى ماذا تقوم وكيفية التقدير.

5. حدود ومنهجية الدراسة:

يتناول الدراسة توفيق بيانات نماذج توزيعات Tweedie الاحتمالية مع بيانات التوزيع العمري والنوعي، وذلك بتطبيق مقياس سكرتارية الأمم المتحدة وهذا المقياس ليس بالهدف الرئيسي وانما لتأكد فقط بوجود الأخطاء في بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية، نماذج توزيعات Tweedie الاحتمالية، ومقايس المفاضلة بين البيانات المقدرة والحقيقية، حيث تم الحصول

على بيانات الدراسة عن الاعداد الخام للأحياء من سكان الجمهورية اليمنية حسب فئات الاعمار الخمسية والنوع لعام 2004 (وذلك هو المتاح حالياً) من الكتاب السنوي للجهاز المركزي للإحصاء.

اعتمدت الدراسة على منهجين هما:

(1) المنهج الوصفي: المتمثل في تحليل المصادر والدراسات العلمية السابقة والمتخصصة بهدف تحديد الجانب النظري، ودقة توزيع الاعمار والنوع لسكان الجمهورية اليمنية باستخدام مقياس الأمم المتحدة ومعامل التشتت والاختلاف.

(2) المنهج الاستنتاجي (الاستدلالي): المتمثل في تقدير معالم نماذج توزيعات Tweedie الاحتمالية وكيفية توفيقها مع بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية.

6. <u>خطة الدراسة</u>: خطة الدراسة ستتكون من مرحلتين:

أولاً: مرحلة الدراسة المكتبية (الجانب النظري): وتتناول الدراسة العلمية المكتبية من خلال المراجع المتخصصة والدراسات السابقة الخاصة بمفهوم الأخطاء التي تتعرض لها البيانات السكانية وكيفية تقييمها، وتوضيح مفهوم كل من التشتت الزائد كيفية تقدير معالم توزيعات Tweedie الاحتمالية.

ثانياً: مرحلة الدراسة التطبيقية (الجانب التطبيقي): تتناول دراسة التعداد العام لسكان الجمهورية اليمنية لعام 2004 (وذلك هو المتاح حالياً) والبحث عن الأخطاء التي تتعرض لها هذا التعداد وتوفيق بشكل عملي توزيعات Tweedie الاحتمالية ذات التشتت الزائد مع بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية والمفاضلة فيما بين البيانات المقدرة لتوزيع العمري والنوعي لكل من الذكور والاناث وكلاهما معاً باستخدام توزيعات Tweedie الاحتمالية.

<u>ثانياً: الجانب النظري:</u>

في هذا الجانب سنقوم بتوضيح مفهوم الأخطاء التي تتعرض لها البيانات السكانية وكيفية تقييمها باستخدام الطرق الإحصائية المتعارف عليها بالدراسات السكانية، وتوضيح مفهوم كل من التشتت الزائد كيفية تقدير معالم توزيعات Tweedie الاحتمالية.

2-1 الأخطاء التي تتعرض لها الإحصاءات السكانية:

يعد معرفة مصدر الخطأ ونوعه واتجاهه يساعد الباحثين على حصول مقياس لقياس هذه الأخطاء وعلى وجه التحديد الحصول على مقياس لتصحيح البيانات السكانية. وخصوصاً إذا كان من بين الأخطاء ما قد يكون ناجماً عن أسباب فنية أو أسباب يمكن تلافيها كنقص التدريب او عدم توفر المصادر المالية أو غيرها من الاسباب، لذلك ليس من السهل سرد كل ما يمكن أن يصيب الإحصاءات السكانية من أخطأ

لذا حاول الباحث تصنيف أهم أنواع الأخطاء التي يمكن أن تتعرض لها بيانات التوزيع العمري والنوعي لسكان الي مجموعتين هما:

1) أخطاء التغطية أو الشمول[8]: والتي تتمثل في زيادة أو نقص في عدد السكان ويرجع ذلك الي أحد الأسباب اما اسقاط بعض الافراد او تكرار في عدد السكان أو أثناء تجهيز البيانات السجلات.

2) أخطاء المحتوى أو المضمون[4]: التي تتمثل بالأخطاء التي يقع فيها الموظفون أو العدادون حيث تتلخص هذه الأخطاء بعدم تمكن توجيه الأسئلة بدقة، اثناء التجهيز الفني للبيانات اثناء الترميز أو ضع الدليل أو تبويب البيانات للنشر.

ونتيجة لمصادر ونوع الأخطاء في بيانات التوزيع العمري والنوعي يؤثر بالتالي على جودة بيانات التعداد، لذا وجب معرفة الأساليب الإحصائية لاكتشاف وتقيم أخطاء بيانات التوزيع العمري والنوعي.

2-2 الأساليب الإحصائية المستخدمة لتقييم أخطاء بيانات التوزيع العمري والنوعي:

هناك عدة أساليب لاختبار جودة بيانات التعداد، هذه الأساليب تعتبر مؤشراً لقياس الأخطاء في التعدادات السكانية. لذا تم تصنيف لاكتشاف وتقييم بيانات إلي ثلاثة مجاميع: الأساليب المباشرة و الأساليب غير مباشرة (المقارنات الخارجية)، أخيراً أساليب الاتساق (المقارنات الداخلية)، حيث أن الأساليب الثلاث المذكورة أنفاً لها خصائص معينة فعلى سبيل المثال الأسلوب الأول تعمل على اكتشاف الأخطاء والتي تنشا عن سقوط العد أو التكرار، اما الأسلوب الثاني تعمل على مراجعة درجة الشمول واكتشاف الاختلافات بين التعداد سابق والإحصاءات الحيوية، أما الأسلوب الثالث تعمل على المقارنة الاتساق الداخلي لدقة بيانات التوزيع العمري لسكان. وتحديد مسايرة نسبة النوع والعمر مع الأنماط السائدة لهذه النسبة[3]، هذا ما سيتم التركيز عليه في هذه الدراسة وذلك للأسباب الآتية: أن مقياس العمر هو أحد مكونات هذا الدليل، وايضاً أن تقييم التوزيع العمري ليس هدفاً في حد ذاته، وخاصة مع التوقع القوي بأن هذا التوزيع يحتوي على أخطاء في أغلب الدول النامية عامة، وفي المجتمعات الامية منها بصفة خاصة، ومن المقاييس الاحصائية التي تهدف إلى اكتشاف الأخطاء في بيانات التوزيع العمري والنوعي لسكان: 1) نسبة العمر إذ ينبغي أن يكون انحراف نسبة العمر عن الرقم 100 قليل جداً بمقدار(10%) بالموجب أو السالب، باستثناء الاعمار المتقدمة فإذا كان الفرق النسب بين (10 %) أمكننا القول أن حجم الخطأ في هذه الفئة والعكس صحيح[7]. 2) نسبة النوع وتتبع نسبة النوع نمطاً ثابتاً إذ

تكون مرتفعة عند الميلاد وفي الاعمار المبكرة، ثم تأخذ بعد ذلك في التناقص التدريجي مع تقدم العمر، أيضاً تقوم على أساس أن أعداد المواليد أكبر المواليد الاناث وأن وفيات الذكور أكثر من وفيات الاناث، وعليه فإن انحراف نسبة النوع عند النمط المحدد له يشير إلي وجود خطاء في البيانات[5]. 3) مقياس سكرتارية الأمم المتحدة هذا المقياس لا يعطي تقديراً لزيادة العد او النقص فيه ولكنه يعطي مؤشراً صافياً لأخطاء الإدلاء بالأعمار. فإذا كان المقياس أقل من 20 دل ذلك على أن بيانات العمر والنوع دقيقة، أما أذا تراوحت قيمته ما بين (20–40) فيدل على أن البيانات غير دقيقة، أما أذا زادت عن 40 فيدل على أن البيانات غير دقيقة على الإطلاق[4]. 4) مقياس ويبل هذه الطريقة محدودة وذلك بسبب انها تقيس درجة تفضيل الاعمار التي تنتهي بالصفر أو بالرقم خمسة، إذ تتراوح قيمة هذا المقياس بين (100,500) فإذا كانت القيمة 100 دل ذلك على عدم وجود انحراف أو تفضيل للرقمين صفر وخمسة. لكن أذا كانت القيمة 500 دل ذلك على أنه يوجد تحيز لهذه الرقمين[8]. 5) مقياس مايرز يقيس دقة بيانات الاعمار عند الإحدى عشر من (1–9) علماً بأنه يتشابه مع مقياس ويبل في أن نصيب كل عدد من أعداد الآحاد في الأعمار يساوي 10%[2].

مما سبق لاحظ الباحث أن الطرق السابقة في تقييم واكشاف الأخطاء في بيانات التوزيع العمري والنوعي لسكان لها مزايا وعيوب. كما لاحظ أيضاً أن الطرق لم تستند على مقايس إحصائية كمقاييس النزعة المركزية ومقايس التشتت.

من هذا المنطلق وعلى وجه التحديد تم استخدام الإحصاءات الوصفية في تحديد النزعة المركزية ومقايس التشتت لبيانات التوزيع العمري والنوعي لسكان، وأيضاً استخدام التوزيعات الاحتمالية. حيث يقترح الباحث استخدام توزيعات Tweedie الاحتمالية للأسباب التالية: 1) أن بيانات التوزيع العمري والنوعي لسكان تتبع التوزيعات الاحتمالية ذات الذيول العالي او الثقيلة بمعنى آخر ذات الالتواء الموجب. 2) بيانات التوزيع العمري والنوعي لسكان لكل من الذكور والاناث مجتمعين مختلفين كل منهما خصائص معينة من ناحية الإحصاءات الوصفية. لذا سيقوم الباحث بشرح توزيعات Tweedie الاحتمالية وكيفية تقديرها من الناحية النظرية.

2–3توزيعات Tweedie الاحتمالية ذات التشتت الزائد:

ينتمي توزيعات Tweedie إلي عائلة التشتت الاسي، فهي تحتوي على معلمة العائلة الأسية الخطية بالإضافة الي معلمة التشتت[15]، ويمكن تصنيف التشتت من حيث المنحنى إلي تشتت أسي أو غير أسي، ومن ناحية الدرجة فقد صنفت أنواع التشتت إلي التشتت الأقل أو التشتت الزائد–سيركز الباحث على

التشتت الزائد في العائلة الأسية–حيث يحدث التشتت الزائد في البيانات عندما يكون التباين أكبر من المتوسط أي عندما يكون مؤشر التشتت (نسبة التباين إلي المتوسط أكبر من الواحد الصحيح[13]. وبشكل عام دالة الكثافة للتوزيعات التشتت الأسي الزائد تكتب بالشكل التالي[14]: $f_Y(\mathrm{y}\,|\,\theta,\varphi) = h(\mathrm{y},\varphi)\exp(\frac{1}{\varphi}(\theta y - \eta(\theta))$ حيث θ تمثل معلمة العائلة الأسية الخطية وتسمى أيضا المعلمة المخروطية، φ تمثل معلمة التشتت الزائد، $\eta(\theta)$ تمثل الدالة التجميعي لمعلمة العائلة الأسية لضمان التوزيع الطبيعي، $h(\mathrm{y},\varphi)$ تمثل دالة غير محددة لقياس استقلالية المعلمة θ [15]. يمكن إيجاد المتوسط والتباين لتوزيعات التشتت الاسي باستخدام المشتقة الأولى والمشتقة الثانية للدالة التجميعية $\eta(\theta)$ بمعنى يكون الوسط الحسابي لتوزيعات التشتت الاسي بالشكل التالي: $E(Y) = \eta'(\theta) = \theta = \mu$ والتباين $\eta''(\theta) = Var(\mu)$ [9]، وبحسب تباين التشتت الاسي و التي تسمى ايضاً دالة تباين الوحدة (وتعني أن التباين يساوي 1) ، فكلما اختلف شكل هذه الدالة ظهرت لدينا مجموعة جديدة من التوزيعات التي تنتمي لتوزيعات التشتت الاسي وتأخذ صفاتها وخصائصها المميزة. لذا فإن الهدف الرئيسي من دالة التباين الوحدة في نماذج التشتت الأسي في تحديد كل مجموعة من مجموعات التوزيعات عن الأخرى.

لذا تظهر توزيعات Tweedie في توزيعات التشتت الأسي في حالة خاصة أذا كانت دالة تباين الوحدة تأخذ الشكل التالي: $Var(Y) = E(Y)^P = \mu^\rho$ حيث ρ هي المعلمة التي تحدد نوع التوزيع الذي ينتمي لمجموعة توزيعات Tweedie وانها تأخذ كل القيم تقع خارج الفترة (0,1) . ويرتبط معلمة المؤشر α الخاصة بالتوزيعات المستقرة بالمعلمة ρ المحددة لنوع التوزيع في توزيعات Tweedie بالعلاقة التالية: $(\rho - 1)(1 - \alpha)$ [15]. وتتضمن توزيعات Tweedie مجموعة من التوزيعات كل حسب قيمة ρ ، فإذا كانت قيمة $\rho = 0$ فإن دالة كثافة توزيعات Tweedie تكون دالة التوزيع الطبيعي، واذا كانت قيمة $\rho = 1$ فإن دالة كثافة توزيعات Tweedie تكون دالة بواسون، واذا كانت قيمة $\rho = 2$ فإن دالة كثافة توزيعات Tweedie تكون دالة جاما، واذا كانت قيمة $\rho = 3$ فإن دالة كثافة توزيعات Tweedie تكون دالة معكوس التوزيع الطبيعي، وتسمى هذه الحالات لقيمة ρ – المذكورة سابقاً – بالحالات الخاصة لتوزيعات Tweedie .لذا توزيعات Tweedie هي توزيعات بواسون المركبة أو بتوزيعات المختلطة بواسون مع توزيعات جاما[14]. ويمكن تمثيل توزيعات Tweedie التجمعية على الصورة التالية: $Y = X_1 + X_2 + \cdots + X_N$ حيث أن Y هو متغير الدارسة وهو عبارة عن مجموع يتبع توزيع بواسون لمجموعة من المتغيرات المستقلة X_i والتي تتبع كل منها توزيع جاما. و N تتبع توزيع بواسون بمتوسط φ ، بذلك يمكن كتابة المعادلة عندما تكون

المتغير $Y > 0$ وفق الصورة التالية[9]: $f(Y|\lambda,\alpha,\beta) = \frac{\lambda^n \exp(-\lambda)}{n!} \frac{Y^{n\alpha-1}\exp\left(-Y/\beta\right)}{\beta^{n\alpha}\Gamma(n\alpha)}$. حيث أن α,β ترتبط بمعالم توزيعات Tweedie وفق العلاقات التالية[15]:

$\lambda = \frac{\mu^{2-\rho}}{\varphi(2-\rho)}$ ، $\alpha = \frac{\rho-2}{(\rho-1)}$ ، $\beta = \varphi(\rho-1)\mu^{\rho-1}$. وبالتالي معالم توزيع Tweedie وفق معالم التوزيع المختلط بين توزيع بواسون وجاما تكون بالشكل التالي[15]: $\rho = \frac{\alpha+2}{\alpha+1}$ ، $\mu = \lambda\alpha\beta$ ، $\varphi = \frac{\lambda^{1-\rho}(\alpha\beta)^{2-\rho}}{2-\rho}$. وبالتالي فإن قيمة $h(y,\varphi)$ تكون وفق العلاقات التالية: $h(y,\varphi) =$

$$\begin{cases} \frac{\mu^{1-\rho}}{1-\rho}, & \text{if } y = 0 \\ \frac{1}{y}\sum_{n=1}^{\infty} h_n(y,\varphi,\rho), & \text{if } y > 0 \end{cases}$$

حيث أن $h_n(y,\varphi,\rho) = \frac{y^{n\alpha}}{y(\rho-1)^{n\alpha}\varphi^{n(1+\alpha)}(2-\rho)^n n!\,\Gamma(n\alpha)}$ تمثل الدالة الاحتمالية لـــ(Wright`s generalized Bessel) وهي ترتبط بتوزيع Tweedie الاحتمالية. بينما قيمة θ و $\eta(\theta)$ و

$$\theta = \begin{cases} \log\mu, & \text{if } \rho = 1 \\ \frac{\mu^{1-\rho}}{1-\rho}, & \text{if } \rho \neq 1 \end{cases} \text{ و } \eta(\theta) = \begin{cases} \log\mu, & \text{if } \rho = 2 \\ \frac{\mu^{2-\rho}}{2-\rho}, & \text{if } \rho \neq 2 \end{cases}.$$

مما سبق وباستخدام المعلومات السابقة فإن دالة التوزيع المشترك لـــ $p(Y,N\,|\,\mu,\varphi,\rho)$ لتوزيع Tweedie تكون العلاقة الاتية:

$$P(Y,N|\mu,\varphi,\rho) = \begin{cases} \exp\left(-\frac{\mu^{2-p}}{\varphi(2-\rho)}\right), & \text{if } n = 0 \\ \exp\left(n\left(-\frac{\log(\varphi)}{\rho-1} + \frac{2+\rho}{\rho-1}\log\left(\frac{y}{\rho-1}\right) - \log(2-\rho)\right) - log\Gamma(n+1) - \frac{1}{\varphi}\left(\frac{y\mu^{1-p}}{\rho-1} + \frac{\mu^{2-\rho}}{2-\rho}\right)\right. & \\ \left. -log\Gamma\left(n\frac{2-\rho}{\rho-1}\right) - \log(y)\right), & \text{if } n > 0 \end{cases}$$

نستنتج مما سبق أن تقدير ρ هو المشكلة الأصعب، حيث تحدد بشكل تحكمي مما يعطي القدرة على حساب مقدر الأمكان الأعظم φ بشرط معلومية ρ. فإذا كانت عدد المعالم صغيرة بالمقارنة بـــ n فإن ذلك سيؤثر على تقدير φ و ρ ، لذا يتم تقدير دوال كثافة هذه التوزيعات ببعض الطرق العددية وعلى وجه التحديد باستخدام تقدير متسلسلة لا نهائية[9]، وايضاً باستخدام طريقة الإمكان الأعظم المعدلة حيث تعتمد هذه الطريقة المعاملات المتعامدة على معالم φ و ρ [15]، وهناك طرق أخرى لتقدير سيتم مناقشتها في بحث آخر. وبعد التعرف على ماهية التشتت الزائد والتشتت الاسي مع التركيز على

توزيعات Tweedie من حيث دالة كثافتها الاحتمالية وتقدير قيمة المعلمة θ و $\eta(\theta)$ و $h(y,\varphi)$. لا بد لنا من المفاضلة بين البيانات الحقيقية والمقدرة.

4-2 المفاضلة بين القيمة الحقيقية والمقدرة:

يتم المفاضلة بين الحقيقية والمقدرة بعدة طرق، حيث أن هذه الطرق لا تأخذ بعين الاعتبار طبيعة النموذج أو تركيبه، أو كل ما يؤثر فيه، فقط تهتم بعنصر البواقي أو خطأ التقدير. حيث تنقسم طرق المفاضلة إلي نوعين رئيسين هما:

- المؤشرات التامة (المطلقة): وتنقسم هذه المؤشرات إلي نوعين: متوسط القيم المطلقة للأخطاء[12] تكتب بالصيغة التالية $MAE = \frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|}{n}$ حيث أن (y_i) تمثل القيمة الفعلية للظاهرة، ($\hat{y}_i$) تمثل القيمة المقدرة للظاهرة، n تمثل حجم العينة. وتكون المفاضلة لكل من القيم الحقيقية والتقديرية بين المجموعات هي المجموعة التي تتمتع بأقل قيمة لـ MAE . بينما متوسط مربع الأخطاء[10] وتعبر بالشكل التالي $MSE = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$ حيث تحسب MSE لكل مجموعة، وتعتبر المجموعة التي تتمتع بأقل قيمة لـ MSE هي الأفضل.
- المؤشرات النسبية: وتنقسم هذه المؤشرات إلي نوعين: النسبة المطلقة لمتوسط الأخطاء[11] وتكتب بالصيغة التالية: $MAPE = \frac{\sum_{i=1}^{n}\left(|y_i-\hat{y}_i|/y_i\right)}{n}$
- حيث تحسب مجموع متوسط نسبة كل قيمة مطلقة مقدره إلي القيمة الفعلية ونقسم المجموع على عدد بيانات الظاهرة، والمجموعة ذو الأقل نسبة مطلقة للأخطاء تعتبر المجموعة هي الأدق. بينما نسبة متوسط الأخطاء[11] و تكتب بالصيغة التالية: $MPE = \frac{\sum_{i=1}^{n}\left((y_i-\hat{y}_i)^2/y_i\right)}{n}$
- وتستخدم هذه الصيغة لمعرفة التحيز في الأخطاء نحو الموجب أو السالب، فكلما كانت القيمة قريبة من الصفر فإن هذا يشير إلي دقة المجموعة.

لذا فإن في التطبيق العملي سيتطرق الباحث الي كيفية اكتشاف الأخطاء في بيانات التوزيع العمري والنوعي وتوفيق البيانات باستخدام توزيعات Tweedie الاحتمالية مع التوزيع العمري

والنوعي لسكان الجمهورية اليمنية والمفاضلة فيما بين البيانات المقدرة لتوزيع العمري والنوعي لكل من الذكور والاناث وكلاهما معاً باستخدام توزيعات Tweedie الاحتمالية.

ثالثاً: الجانب التطبيقي:

في هذا الجانب سنقوم بتقييم التعداد السكاني للجمهورية اليمنية للتركيب العمري والنوعي لعام (2004) واكتشاف الأخطاء الحاصلة في البيانات، وايضاً سنعرض توفيق ومعالجة بيانات التوزيع التركيب العمري والنوعي باستخدام توزيعات Tweedie الاحتمالية ذات التشتت الزائد وكيفية تقديرها.

3-1 تقييم بيانات التعداد السكاني للجمهورية اليمنية لعام (2004):

تهدف هذه الأساليب في الوقوف على طبيعة الخطأ في ذكر بيانات التوزيع العمري والنوعي. وقد ناقشنا فيما سبق هناك طرق لاكتشاف هذه الأخطاء والوقوف عليها، حيث أن الباحث سيقتصر على البعض منها وليس على كلها لان ما يهم الباحث هو التحقق من وجود الأخطاء في بيانات التوزيع ى العمري والنوعي للسكان فقط، لذا ستقتصر الدراسة على طريقة سكرتارية لأنها معتمدة نسبة العمر ونسبة النوع. وتتطلب هذه الطريقة توزيع السكان حسب فئات خمسية حيث أن توزيع السكان حسب فئات خمسية يساعد في تقليل صافي خطاء الإبلاغ عن العمر وايضاً تعكس صورة أوضح لمدى التكوين العمري والنوعي عامة، يتم حساب هذا المقياس عن طريق مجموع كل من متوسط القيم المطلقة لفروق العمر للذكور ومتوسط القيم المطلقة لفروق العمر للإناث وثلاثة اضعاف متوسط القيم المطلقة لفروق النسب الخاصة بالنوع. ويوضح الجدول التالي حساب مقياس سكرتارية الأمم المتحدة لبيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية لعام (2004).

جدول(1)

حساب مقياس سكرتارية الأمم المتحدة لقياس التراكم العمري والنوعي النسبي للجمهورية اليمنية لعام 2004

الفئات العمرية	نسبة الذكور	نسبة الاناث	نسبة النوع	الانحرافات للنوع عن 100	نسبة العمر		انحرافات نسبة العمر عن 100	
					اناث	ذكور	اذكور	اناث
0–4	15.02	14.38	104.45	–	–	–	–	–
5–9	15.68	14.88	105.38	0.93	103.94	106.61	3.94	6.51
10–14	15.15	13.56	111.73	6.35	106.92	100.44	6.92	0.44
15–19	12.66	12.12	104.46	7.27	102.01	106.22	2.01	6.22
20–24	9.67	9.26	104.43	0.03	96.89	94.78	3.11	5.22
25–29	7.30	7.42	98.38	6.05	100.43	105.55	0.43	5.55
30–34	4.88	4.80	101.67	3.29	84.07	78.69	15.93	21.31
35–39	4.31	4.78	90.17	11.5	102.25	112.08	2.25	12.08
40–44	3.55	3.73	95.17	5	99.30	95.76	0.70	4.24
45–49	2.84	3.01	94.35	0.82	93.11	97.57	6.89	2.43
50–54	2.55	2.44	104.51	10.16	117.51	111.42	17.51	11.42
55–59	1.50	1.37	109.49	4.98	20.42	68.84	29.58	31.16
60–64	1.71	1.54	111.04	1.55	138.46	137.50	38.46	37.50
65–69	0.79	0.78	111.49	0.45	48.99	26.85	51.01	73.15
70–74	1.14	1.05	96.57	14.92	–	–	–	–
75–79	0.51	0.43						
80+	0.94	0.85						
المجموع	100	100		73.3			178.74	217.23

يلاحظ من الجدول (1) الاتي:

- نسبة النوع حسب تعداد 2004 حيث تشير النتائج إلي وجود انحرافات في نسبة النوع عن النمط العام لها حيث يظهر بأن نسبة النوع كانت مرتفعة عن الـــ (100) في الأعمار الصغيرة قبل الفئة (25–29) ثم انخفضت حتى وصلت 98.38 عند الاعمار الشابة (25–29) وكانت متذبذبة في فئات العمرية (30–34) الي (45–49) ثم ارتفعت بشكل مفاجئ عن الـــ (100) في الفئات الأخيرة المتبقية. مما يشير إلي وجود أخطاء في الإدلاء بالأعمار من الفئة (50–54) إلي 80+ أدى إلي الارتفاع مفاجئ في قيمة نسبة النوع عند هذه الفئات.
- نسبة العمر لكل من الذكور والاناث حسب تعداد 2004 حيث يتضح بأن مجموع الانحرافات لنسب العمر عن الرقم 100 يبلغ 178.74 و217.23 للإناث مما يدل على وجود أخطاء في بيانات الأعمار، وأن هذه الأخطاء كانت أكبر مما عليه عند الاناث.
- مقياس سكرتارية الأمم المتحدة حسب تعداد 2004 يتضح أن قيمة هذا المقياس 47.37 وهذا الرقم قد تجاوز الحد الذي تعد عنده بيانات العمر والنوع دقيقة.

بصفة عامة فإن استخدام مقياس سكرتارية الأمم المتحدة لقياس دقة بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية ليس بالهدف الرئيسي في البحث وانما للتأكد فقط من وجود الأخطاء في هذه البيانات من أجل مطابقتها مع نتائج نماذج توزيعات Tweedie الاحتمالية.

2-3 توفيق بيانات التوزيع العمري والنوعي باستخدام توزيعات Tweedie:

مما سبق وبشكل عام فإن بيانات التوزيع العمري والنوع تحتوي على كثير من الأخطاء لذا وجب علينا توفيق هذه البيانات وجعلها أكثر ملائمة وذلك باستخدام توزيعات Tweedie الاحتمالية. وقبل التطرق الي كيفية تقدير وتطبيق توزيعات Tweedie على التوزيع العمرية والنوع لسكان الجمهورية اليمنية لا بد لنا من التأكد من بيانات التوزيع العمري والنوعي تحتوي على تشتت زائد. الجدول التالي يوضح الإحصاءات الوصفية للتركيب العمري والنوعي لسكان الجمهورية اليمنية لعام 2004.

جدول(2)

الإحصاءات الوصفية لبيانات التوزيع العمري والنوعي للجمهورية اليمنية لعام 2004

	Sum	St,Dev	Mean	N
اعداد الذكور (X_1)	10036953.00	520327.40309	501847.6500	17
اعداد الاناث (X_2)	9648208.00	489226.05422	482410.4000	17
اعداد الذكور والاناث معا $(X_1 + X_2)$	19685161.00	1009122.24433	984258.0500	17

يلاحظ من الجدول (2):

- عند استخدام معامل الاختلاف لقياس التشتت في الاعمار يتضح أن الذكور أكثر تشتت من الاناث حيث بلغ معامل الاختلاف لكل من الذكور والاناث على الترتيب (103% و 101%).
- أن كل من بيانات التوزيع العمري لكل من الذكور والاناث تحتوي على تشتت زائد، حيث بلغ قيمة مؤشر التشتت لكل من الذكور والاناث وكلاهما معاً هي (539487.64) و(496138) و (1034614.55) على الترتيب. وهي أكبر من الواحد الصحيح وهذا يعني أن التوزيع العمري والنوعي لكل من الذكور والاناث وكلاهما معاً تحتوي على تشتت زائد.

بعد فحص البيانات واكتشاف التشتت الزائد في التوزيع العمري والنوعي لسكان الجمهورية اليمنية، وجب علينا تحديد المتغيرات المستقلة والتي تمثل أعداد الذكور X_1 واعداد الاناث X_2 حيث إن المتغيرين يتبعان توزيع جاما، N_i تمثل عدد السكان الاجمالي للفئات العمرية و i تمثل مجموعة الذكور والاناث حيث N_i تتبع توزيع بواسون، والمتغير التابع ويمثل مجموع اعداد الذكور والاناث حيث Y يتبع توزيع Tweedie. لتوفيق بيانات التوزيع العمري والنوعي بالتوزيعات الاحتمالية المذكورة أنفاً، تم استخدام برنامج (Easy Fit)، الجدول التالي يوضح جودة المطابقة لبيانات الأعمار في الجمهورية اليمنية.

جدول(3)

جدول يبين جودة التوفيق لبيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية لعام 2004

جودة المطابقة	التوزيعات الاحتمالية				
	اعداد الذكور والاناث معا	اعداد الذكور (X_1)	اعداد الاناث (X_2)	اعداد الذكور والاناث معا	اختبار Kolmogorov-smirnov
Sample Size	17	17	17	17	
Statistic	0.10303	0.10486	0.10885	0.42401	
P-value	0.985	0.98208	0.9744	3.3773E-17	
Rank	12	8	12	5	
قيمة معالم التوزيع	α=1.421 β=15.58	α=1.3858 β=15.27	α=1.4394 β=14.885	λ= 22.143	

يلاحظ من الجدول (3) أن بيانات التوزيع العمري لكل من الذكور والاناث وكلاهما توافقت مع توزيع جاما الاحتمالي من الناحية النظرية والاحصائية، بينما الأعمار لكلا الجنسين توافق مع توزيع بواسون الاحتمالي من الناحية النظرية فقط. بعد التحقق من شروط توزيع Tweedie لابد من إيجاد قيمة كل من المعلمة φ و ρ و μ قبل تقدير دالة الكثافة لتوزيعات Tweedie، ويتم الحصول على قيمتي φ و ρ و μ باستخدام العلاقات المذكورة في الاطار النظري، باستخدام برنامج (R) كانت نتائج تقدير φ و ρ و μ كما يلي:

جدول(4)

تقدير قيمة φ و ρ و μ لبيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية لعام 2004

بيانات التوزيع العمري	μ	ρ	φ
اعداد الذكور (X_1)	468.572	1.41915	2.76707
اعداد الاناث (X_2)	474.424	1.40993	2.90393
اعداد الذكور والاناث معا $(X_1 + X_2)$	490.228	1.41305	2.91939

يلاحظ من الجدول (3) أن قيمة ρ تراوحت بين القيم (1,2) وهذا يعني بالفعل أن بيانات التوزيع العمري لكل من الذكور والإناث وكلاهما معاً هي بيانات مستمرة ذات قيم متقطعة عند الصفر، ايضاً قيم μ تراوحت (469 ، 490) لكن توزيعات Tweedie تشترط أن قيمة $\mu = 1$ و $Var(Y) = 1$ لذا سنقوم بالتعويض عن قيمة μ المستخرجة بالواحد، بعد الحصول على تقدير القيم كل من φ و ρ و μ يمكن تقدير دالة الكثافة Y_i باستخدام تقدير متسلسلة لا نهائية وذلك عن طريق برنامج R وكانت النتائج التقدير كما يلي:

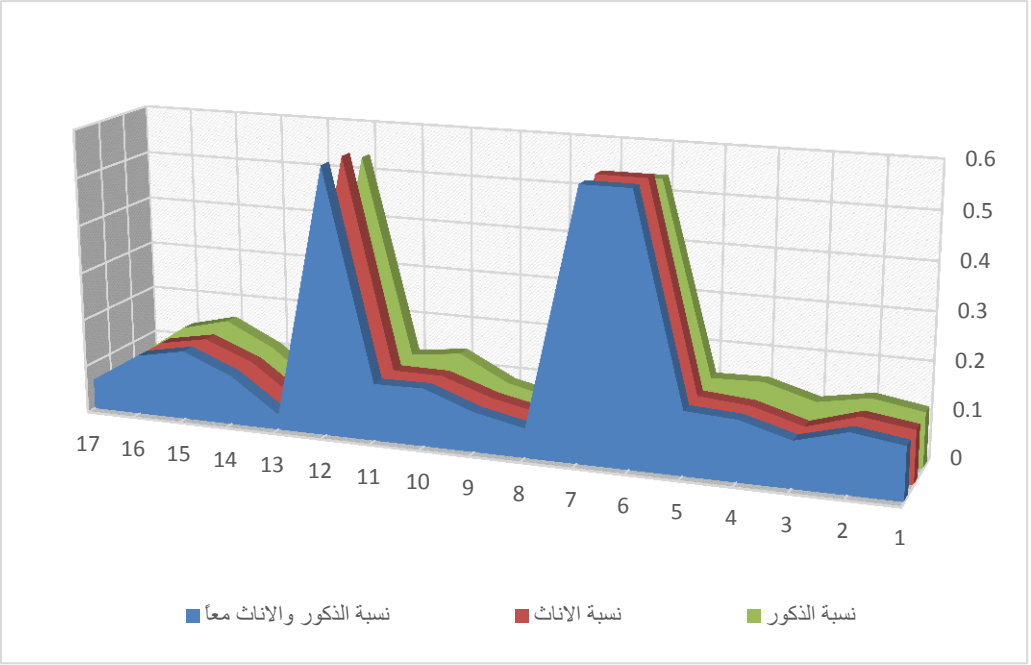
جدول(5)

تقدير دالة الكثافة لتوزيع Tweedie لبيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية لعام 2004

الفئات العمرية	نسبة الذكور	نسبة الاناث	نسبة الذكور والاناث معاً
0−4	0.1156	0.1105	0.1099
5−9	0.1330	0.1257	0.1251
10−14	0.1138	0.0936	0.0964
15−19	0.1438	0.124	0.1259

20–24	0.1413	0.1326	0.132
25–29	0.5368	0.5579	0.5579
30–34	0.5368	0.5579	0.5579
35–39	0.0629	0.0621	0.0618
40–44	0.0859	0.0836	0.0831
45–49	0.1392	0.1191	0.1212
50–54	0.1264	0.12	0.1194
55–59	0.5368	0.5579	0.5579
60–64	0.0314	0.0319	0.0318
65–69	0.1097	0.1053	0.1047
70–74	0.1595	0.1455	0.1456
75–79	0.1341	0.1266	0.126
80+	0.0648	0.0639	0.0636

يلاحظ من الجدول (5) أن احتمالات التشتت في الأعمار لبيانات التوزيع العمري والنوعي لكل من الذكور والاناث وكلاهما معاً لكل فئة والمقدرة باستخدام توزيع Tweedie تمركزت في أربع مجموعات، المجموعة الأولى عند الفئات (4–0) و (9–5) و(14–10) و (19–15) حيث كانت نوعا ما شبة متقاربة، ثم حدث ارتفاع مفاجئ عند المجموعة الثانية عند الفئات (24–20) و(29–25) و(34–30) و(39–35) وثم انخفاض حتى المجموعة الثالثة التي ارتفعت مرة أخرى عند الفئات (54–50) و(59–55) و (64–60) و اخيراً المجموعة الرابعة التي كانت من الفئة (69–65) الي الفئة الأخيرة، الشكل التالي يوضح احتمالات التشتت في كل فئات بيانات التوزيع العمري والنوعي المقدرة باستخدام توزيع Tweedie.



الشكل(1)

الشكل يبين احتمالات التشتت في كل فئات بيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية لعام 2004 المقدرة باستخدام توزيع Tweedie

وبشكل عام من الجدول (5) والشكل (1) نستنتج أن الاحتمالات المقدرة من توزيع Tweedie لبيانات التوزيع العمري والنوعي في الجمهورية اليمنية ما هي إلا احتمالات الاعمار التي لا تقع ضمن الفئات العمرية لكل من الذكور والاناث وكلاهما معاً، لذا لابد من التخلص من هذه الاحتمالات لتوفيق بيانات التوزيع العمري والنوعي لجمهورية اليمن، وباستخدام برنامج (SPSS) كانت النتائج التوفيق كما يلي:

جدول(6)

توفيق بيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية لعام 2004 باستخدام توزيع Tweedie

الفئات العمرية	نسبة الذكور	نسبة الاناث	نسبة الذكور والاناث معاً
0-4	13.28	12.79	13.26
5-9	13.59	13.01	13.56
10-14	13.43	12.29	13.19
15-19	10.84	10.62	11.01
20-24	8.30	8.03	8.33
25-29	3.38	3.28	3.32
30-34	2.26	2.12	2.17
35-39	4.04	4.48	4.32
40-44	3.24	3.42	3.39
45-49	2.44	2.65	2.64
50-54	2.23	2.15	2.20
55-59	0.69	0.61	0.66
60-64	1.66	1.49	1.55
65-69	0.70	0.70	0.81
70-74	0.96	0.90	0.94
75-79	0.44	0.38	0.44
80+	0.88	0.80	0.84

بعد توفيق بيانات التوزيع العمري والنوعي باستخدام Tweedie يتم المفاضلة بين هذه الطريقة وبيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية الاصلية وذلك بأحدى الطرق الاتية: MAE ، MSE، MAPE ، MPE، وباستخدام بيانات التوزيع العمري والنوعي في الجمهورية اليمنية لعام

2004 وعلى وجه التحديد بيانات جدول(6) و(1)، وباستخدام برنامج (SPSS) كانت النتائج المفاضلة كما يلي:

جدول(7)

يبين المفاضلة بين البيانات المقدرة باستخدام توزيع Tweedie وبيانات التوزيع العمري والنوعي النسبي في الجمهورية اليمنية الاصلية لعام 2004

	نسبة الذكور	نسبة الاناث	نسبة الذكور والاناث معاً	المتوسط العام
MAE	1.04	0.96	1.01	1
MSE	2.29	2.17	2.28	2.25
MAPE	0.20	0.19	0.19	0.19
MPE	0.30	0.31	0.31	0.31

يلاحظ من الجدول (7) أن جودة التوفيق لبيانات التوزيع العمري والنوعي باستخدام توزيع Tweedie كانت النتائج متقاربة جداً والاختلافات بينها بسيطة، حيث بلغ المتوسط النسبي للأخطاء لكل من طريقة MAPE و MPE (0.19 ، 0.31) على الترتيب وهي قيمتان قريبة من الصفر، وهذا يعني أن جودة التوفيق بين توزيع Tweedie و بيانات التوزيع العمري و النوعي لسكان الجمهورية اليمنية كانت جيدة حيث بلغ المتوسط العام لجودة التوفيق لكل من الذكور والاناث وكلاهما معاً بطريقة MAPE نسبة 81%.

رابعاً: الخلاصة والتوصيات:

في هذه النقطة سيقوم الباحث باستعراض أهم النتائج التي أسفرت عنها الدراسة بما يتفق مع أهداف الدراسة، وخاصة في تحديد النموذج الإحصائي المناسب المقترح لتوفيق بيانات توزيعات Tweedie الاحتمالية مع بيانات التوزيع العمري والنوعي للسكان الجمهورية اليمنية وفق تعداد 2004.

- **الخلاصة:**

لقد اتضح من خلال النتائج التي تم التوصل إليها من تطبيق النموذج الاحصائي المقترح لتوفيق توزيعات Tweedie الاحتمالية مع بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية وفق تعداد 2004 لتأكد من صلاحية النموذج لتقدير والمطابقة عدة نقاط:

(1) أن نسبة النوع حسب تعداد 2004 حيث تشير النتائج إلي وجود انحرافات في نسبة النوع عن النمط العام لها حيث يظهر بأن نسبة النوع كانت مرتفعة عن الـ (100) في الأعمار الصغيرة قبل الفئة

(25–29) ثم انخفضت حتى وصلت 98.38 عند الاعمار الشابة (25–29) وكانت متذبذبة في فئات العمرية (30–34) الي (45–49) ثم ارتفعت بشكل مفاجئ عن الـــ (100) في الفئات الأخيرة المتبقية. مما يشير إلي وجود أخطاء في الإدلاء بالأعمار من الفئة (50–54) إلي 80+ أدى إلي الارتفاع مفاجئ في قيمة نسبة النوع عند هذه الفئات. اما من ناحية نسبة العمر لكل من الذكور والاناث حسب تعداد 2004 حيث يتضح بأن مجموع الانحرافات لنسب العمر عن الرقم 100 يبلغ 178.74 و217.23 للإناث مما يدل على وجود أخطاء في بيانات الأعمار، وأن هذه الأخطاء كانت أكبر مما عليه عند الاناث.

(2) أن بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية لكل من الذكور والاناث تحتوي على تشتت زائد، حيث بلغ معامل الاختلاف لكل من الذكور والاناث على الترتيب (103% و 101%) و هذا يعني وجود الأخطاء عن تبليغ الاعمار لكل من الذكور والاناث، على الرغم من النتيجة التي توصل اليها الباحث عن طريق معامل الاختلاف أن الذكور اكثر أخطاء في الإدلاء بالأعمار وهذا مخالف لما في النقطة رقم(1) إلا أن الهدف من النموذج هو معرفة وجود الأخطاء فقط وليس تحديد ايهما اكثر عند الذكور أم الاناث. لذا يمكن اعتبار هذه الطريقة أسلوب بديل لاكتشاف الأخطاء في بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية.

(3) أن تقدير توزيعات Tweedie الاحتمالية بالأساس قائمة على توزيعين احتماليين توزيع جاما وتوزيع بواسون، ايضاً قيمة ρ تراوحت بين القيم (1,2) وهذا يعني بالفعل أن بيانات التوزيع العمري لكل من الذكور والإناث وكلاهما معاً هي بيانات مستمرة ذات قيم متقطعة عند الصفر.

(4) أن احتمالات التشتت في الأعمار لبيانات التوزيع العمري والنوعي لكل من الذكور والاناث وكلاهما معاً لكل فئة والمقدرة باستخدام توزيع Tweedie تمركزت بشكل كبير وملحوظ عند مجموعتين وهي: المجموعة الاولى الفئات (20–24) و(25–29) و(30–34) و(35–39) والمجموعة الثانية (50–54) و(55–59) و (60–64)، يعني هذا أن النموذج له القدرة على اكتشاف مواطن الأخطاء في الاعمار بشكل دقيق، وهذا ما يؤكد الفقرة رقم (1) و(2).

(5) أن جودة التوفيق بين بيانات توزيع Tweedie و بيانات التوزيع العمري و النوعي لسكان الجمهورية اليمنية كانت جيدة حيث بلغ المتوسط العام لجودة التوفيق لكل من الذكور والاناث وكلاهما معاً بطريقة MAPE نسبة 81%.

- **التوصيات:**

(1) الأخذ بعين الاعتبار الإحصاءات الوصفية كالمتوسط والتباين و معامل الاختلاف والتفرطح و الالتواء عند دراسة بيانات التوزيع العمري والنوعي للسكان.

(2) استخدام التوزيعات الاحتمالية وعلى وجه التحديد توزيع Tweedie الاحتمالي في توفيق بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية لما يتمتع به من خصائص ومميزات.

(3) التهيئة الجيدة لإجراء التعداد السكاني في موعد ملائم من خلال التدريب الجيد للعدادين وتوفير كل ما يتطلبه التعداد السكاني من مستلزمات فنيه وماديه وكوادر، إضافة الي استغلال كل وسائل الإعلام من أجل توضيح أهمية التعداد السكاني والإدلاء بالمعلومات الدقيقة من قبل المواطنين بهدف جعل بيانات التعداد القادم أكثر دقةً من التعدادات السابقة.

(4) يوصي الباحث بالاستمرار في استخدام النماذج الإحصائية والاعتماد عليها في توفيق بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية.

(5) **المراجع العربية:**

1. الجهاز المركزي للتعبئة العامة والاحصاء – السكان والبحوث والدارسات " تقدير جداول الحياة من 1976–2001" جمهورية مصر، القاهرة – مدينة نصر– مركز الأبحاث والدراسات السكانية، العدد 35، 1987.
2. الحمداوي، نادية علي عايد " الإسقاطات السكانية لمحافظة البصرة للفترة 1997–2002 باستخدام نتائج التعداد 1997 في العراق" رسالة ماجستير غير منشورة، جامعة بغداد، العراق، 2005.
3. الشلقاني، مصطفى " طرق التحليل الديموجرافي" كلية التجارة والاقتصاد والعلوم السياسية، جامعة الكويت.
4. العلفي، سمير أحمد محمد" نموذج إحصائي متعدد المتغيرات للتنبؤ بالمسح البعدي بالجمهورية اليمنية" رسالة دكتوراه غير منشوره، قسم الإحصاء التطبيقي والتأمين، كلية التجارة جامعة المنصورة، 2012.
5. المنجي، هشام محمد " تطوير بعض أساليب التحليل الاحصائي لرفع كفاءة المسح البعدي لسكان" رسالة دكتوراه غير منشورة، قسم الإحصاء التطبيقي والتامين، كلية التجارة، جامعة المنصورة، 2012.

6. أمين، ندى أحمد " إسقاطات القوى العاملة، والتعليم والصحة للتركيب العمري والنوعي لسكان العراق لمدة 1997–2017" بحث دبلوم عالي في الإحصاء التطبيقي، كلية الإدارة والاقتصاد، جامعة بغداد، 2011.

7. فوده، محمد عبد الفتاح" تقييم التعداد السكاني لعام 1996 لإعداد جدول حياة قومية مختصر" مجلة الدراسات المالية والتجارية للعلوم الإدارية، كلية التجارة – بني سويف –جامعة القاهرة، المجلد الثاني–ملحق 1999.

8. فوده، محمد عبد الفتاح" تقييم بعض طرق تهذيب البيانات الإحصائية السكانية مع تطبيق على بيانات تعداد مصر 1996" المجلة المصرية للدراسات التجارية، كلية التجارة، جامعة المنصورة، 2001.

المراجع الاجنبية:

9. Chritoph Kurz (2016) " Tweedie distributions for fitting semicontinuous health care utilization cost data" American International Group (AIC).Science Team, The AIC Building, 58 Fenchurch Street, London EC3M4AB, U.K. arXiv: 1608.04910v1 {stat.AP}.
10. David Anderson, Dennis Sweeney and Thomas William (2001)" Quantitative Methods for Business" South Western college publishing, Ohio.
11. Donald Harnett and James Horriel (1998) " Data statistics and decision models with Excel" John and Sons, New York
12. Geral Keller & Brian worrck (1997) "Statistics for Management and Economic" Cole publishing company, New York.
13. Ridout,M. and Besbeas,p. (2004) " An empirical model for underdispersed count data" statist.model.4:77-89.
14. S. Santhana Lakshmi,R. Geetha (2016) " Tweedie Modelling For the Determinants of Child Mortality" IOSR Journal Of Humanities And Social Science (IOSR-JHSS) Volume 21, Issue 4, Ver. PP 35-40

Yaodong Yang and Rui Luo, Reza Khorshidi, Yuanyuan (2018)" Inferring Tweedie compound Poisson Mixed Models With Adversarial Variational Bayes" American International Group (AIC).Science Team, The AIC Building, 58 Fenchurch Street, London EC3M4AB, U.K. arXiv: 1706.05446v4 {stat.ML