جامعة إب
مجلة الباحث الجامعي
ISSN: 2079-5068 ISSN(online): 2663-

انموذج إحصائي مقترح لجدول حياة للسكان في اليمن باستخدام نماذج توزيعات Tweedie الاحتمالية (دراسة تطبيقية)

حسن حسن عبد الملك

قسم المحاسبة، كلية العلوم الإدارية، جامعة إب، اليمن

Email: hasmalik5@hotmail.com

الكلمات المفتاحية:

انموذج إحصائي، جدول حياة للسكان، توزيعات Tweedie، **اليمن**

الملخص:

يكتسب التوزيع العمري لسكان دولة ما أهمية بالغة في التخطيط وإعداد الاستراتيجيات المتعلقة بالتعليم والصحة والدفاع وخلافه. وحتى يكون التخطيط مبنيًا على أسس صحيحة فإن الإسقاطات السكانية يجب أن ترتكز على توزيعات عمرية ونوعية دقيقة بقدر الإمكان. تهدف هذه الدراسة إلى تصحيح التوزيع العمري للجمهورية اليمنية طبقًا لما هو مسجل لعام (2004) من الأخطاء التي تشوب التعدادات السكانية، وذلك باستخدام الأساليب الإحصائية السكانية كمرحلة أولى، كما تهدف أيضًا إلى إعداد جدول حياة مختصر للجمهورية اليمنية وذلك باستخدام توزيعات Tweedie الاحتمالية، حيث تعتبر جداول الحياة أداة من أدوات التقويم وهذه تعتبر المرحلة الثانية. وتوصلت الدراسة إلى أن توقع الحياة عند الميلاد في الجمهورية اليمنية حسب تعداد 2004 كان (46، 47) سنة لكل من الذكور والاناث على الترتيب.

انموذج إحصائي مقترح لجدول حياة للسكان في اليمن باستخدام نماذج توزيعات Tweedie الاحتمالية (دراسة تطبيقية)

A Statistical Model for a Life Scheduling Population in Yemen by Using Tweedie Distribution Model (Applied Study)

Hasan hasan abdualmalik

Department of Accounting, Faculty of Administrative Sciences, Ibb University, Yemen

Email: hasmalik5@hotmail.com

Keywords

Statistical Model, Life Scheduling Population, Tweedie Distribution, Yemen

Abstract:

This study aims to correct the age distribution of the Republic of Yemen according to what was recorded for the year (2004) of errors that affect population censuses, by using population statistical methods as a first stage. It also aims to prepare a brief life schedule for the Republic of Yemen by using the Tweedie distributions model, where the life schedule is considered as a correction tool, and this is the second stage. The study found that life expectancy at birth in the Republic of Yemen according to the 2004 census was (46, 47) years for both males and females, respectively.

أولًا: السياق العام للدراسة والمشكلة

1. المقدمة:

تُستخدم البيانات التوزيع العمري والنوعي أساسًا لعمل جداول الحياة أو الوفاة، وهي جداول يمكن بواسطتها قياس احتمالات الحياة أو الوفاة عند كل عمر من الأعمار. تُستخدم هذه الجداول كأداة علمية وأساسيه لتقدير عدد السكان في المستقبل ويُبنى على هذا التقدير غالبية الدراسات السكانية والديموغرافية. ويمكن تقسيم جداول الحياة أو "الوفاة" بحسب طول الفئة العمرية إلى نوعين من الجداول الأول جداول الحياة الكاملة: وهي جداول تتضمن بيانات عمرية لكل الأعمار الأحادية منذ الميلاد وحتى آخر عمر يمكن أن يصل إليه الفرد في هذه المجموعة، والثانية جداول الحياة المخُتصرة: وهي جداول تتضمن بيانات عُمرية لكل مجموعة من الفئات العمرية والتي قد تكون خمس سنوات أو عشر سنوات مثلا (فئات عمرية) ابتداءً من الميلاد وحتى آخر فئة عمرية يمكن أن يصل إليها الفرد في هذه المجموعة [10]. ويهدف كلا النوعين الي تحديد (النواحي الديموغرافية) و(الاختلافات البينية) في معدلات الوفاة داخل فئات الأعمار [11]. كما يتم تقسيم جداول الحياة بحسب مصدر البيانات اللازمة لعمل الجداول إلى نوعين: النوع الأول: مبني على أساس الإحصاءات العامة للسكان وعلى وجه التحديد التعدادات السُكانية [2]. والإحصاءات الحيوية وتُسمى جداول الحياة القومية [11]. أما النوع الثاني فهو يبنى على الإحصاءات المتجمعة لدى شركات التأمين عن جمهور المستأمنين لديها لفترات زمنية تسمح بتطبيق قانون الاعداد الكبيرة [12].

والفرق بين المصدرين –التعداد من ناحية وشركات التأمين من ناحية أخرى–في عمل جداول الحياة هو أن معدلات الوفاة في التعداد السكاني أكبر نسبيًا من معدلات الوفاة المناظِرة لها في شركات التأمين [8]، وذلك لأن مجتمع شركات التأمين مجتمع منتقاة تم اختياره بناء على كشف طبي فيكون معدلات وفياته أقل على خلاف تعداد السكان الذي يشمل جميع فئات السكان سواء منهم المريض أم غير ذلك فتكون معدلات وفياته أكبر. ويُلاحظ أن جميع بيانات جداول الحياة (لكلا النوعين) لا يوجد فيهما ما هو مستمد من بيانات إحصائية سوى احتمالات الوفاة. ويُلاحظ أن الجداول المُعدة على أساس إحصاءات السكان والتي تظهر بياناتها لكل عمر من الأعمار أو قد تقتصر على الفئات العُمرية [12]. يظهر بها كثيرٌ من الأخطاء التي تتعرض لها البيانات الإحصائية الخاصة بالسكان، وهي: أخطاء عدم الالتزام بالتعريف الدقيق للوفاة (في الإحصاءات الحيوية). وأخطاء عدم شمول الحصر، كما توجد أخطاء خاصة بعدم الدقة في توزيع الوفيات زمانيًا [2] ومكانيًا (في التعدادات السكانية)، وأخطاء أخرى تتعرض لها بيانات الأعمار [2]. ونظرًا لأهمية بيانات التعدادات السُكانية في عمل جداول الحياة سوف تركز هذه الدراسة على التعرف إلى هذه الأخطاء وتحديدها بدقة وإيجاد سبل علاجها.

لذا تهتم الكثير من الدراسات السكانية بدراسة الوفيات كخطوة أساسية لتقدير احتمالات الوفاة والتي تكون أساس في عمل جداول الحياة بمختلف أنواعها. حيث تستخدم في عملية التخطيط في كثير من القطاعات مثل التعليم والصحة وغير ذلك، حيث جداول الحياة تمر بثلاثة مراحل: التقييم وتقويم وإعداد جدول حياة.

وبهذا الصدد هناك بعض البحوث العلمية ذات الصلة بتقييم وتقويم البيانات السكانية وأعداد جدول حياة في عام(1999) [13] وتم فيها استخدام بيانات التعداد السكاني لجمهورية مصر العربية لعام (1996)، وذلك لمعرفة أهم فروض جداول الحياة، الأخطاء التي تتعرض لها البيانات الإحصائية الخاصة بالسكان وطرق معالجة تلك الأخطاء، وكانت أهم نتائج هذه الدراسة هي ملاحظة ارتفاع توقع الحياة للإناث عن الذكور بصفة عامة في جميع الأعمار، هذا بالإضافة لإعداد جدول حياة قومي مختصر، وفي عام(2000)[12] وتم فيها دراسة أهم طرق الإسقاط السكاني، كما تم دراسة معدلات الوفاة والخصوبة المتوقعة خلال فترة الإسقاط، أيضًا إسقاط السكان والهيكل الديموغرافي المتوقع لبيانات تعداد 1996، وكانت أهم نتائج هذه الدراسة هي أن دليل العمر من المتوقع أن يتضاعف ثلاث مرات خلال فترة الدراسة، وهذا يعني أن المجتمع يتحول تدريجيا من مرحلة الشباب الي مرحلة الكهولة". وفي عام (2001) [14] حيث ناقش الأخطاء التي تتعرض لها البيانات الإحصائية في التعداد من حيث دقتها ودرجة شمولها، كما أستعرض بعض طرق تمهيد بيانات التعداد السكاني ومفاضلة هذه الطرق، وكان ترتيب هذه الدراسة من حيث الأهمية هي طريقة (الأمم المتحدة) في المركز الأول ثم طريقة (كارير – فراج) في المركز الثاني، طريقة (نيوتن) في المركز الثالث ثم طريقة (اللوجيت) في المركز الرابع وأخيرا طريقة التمهيد باليد". وفي عام (2012) [5]. وتم فيها استخدام بيانات التعداد السكاني للجمهورية اليمنية لعام (2004). وأقترح الباحث أن التوزيعات الاحتمالية هي مقترح بديل للمسح البعدي حيث يمكن الاعتماد عليه في إجراء الإسقاطات في القطاعات الاقتصادية والتعليمية والقوة العاملة والصحة، ويمكن تقدير بعض المعالم الديموغرافية في ضوء التعداد المُصحح. وكانت النتيجة الرئيسية لهذه الدراسة هو أن توزيع جاما وويبل هما التوزيعان المناسبان لطبيعة البيانات في التعداد السكاني لجمهورية اليمن. وفي عام (2008) [7]. وتم فيها تصحيح التوزيع العمري في التعداد اليمني لعام 2004. حيث تم استخدام 13 نموذجا، أدت المفاضلة بينهما إلى أن نموذج معادلة الدرجة الثانية ذات المجهول الواحد، النموذج (اللوغاريتمي) هما أفضل نموذجين لتقويم بيانات تعداد السكان في (الجمهورية اليمنية)، كذلك أكثر من 55% من السكان يقعون تحت عمر 20 سنه، الأمر الذي يجعل مُعدل النمو السكاني مرتفعًا لعقود قادمة. مما سبق لاحظ الباحث من الدراسات السابقة أن:

- هناك فرق بين تقييم البيانات وتقويم البيانات، حيث يقصد بتقييم البيانات دراسة مدى دقة البيانات السكانية الناتجة من التعداد السكاني، أما تقويم البيانات يقصد به تعديل وتصحيح البيانات السكانية الناتجة من الأخطاء التي تتعرض لها التعدادات السكانية.
- هناك فرق بين التعدادات السكانية والمسوحات السكانية، حيث تعد المسوحات السكانية وسيلة من وسائل التقييم لتعدادات السكانية، ونظراُ لتكلفة العالية لهذه المسوحات يلجأ بعضهم لاستخدام طرق أخرى علمية لتقييم التعدادات السكانية.
- جداول الحياة تعد وسيلة من وسائل تقويم البيانات لتعدادات السكانية.
- جميع طرق التقويم واعداد جدول الحياة كانت نتائجها بشكل منحنى وبالتحديد منحنى تنازلي من الفئة الأولى الى الفئة الأخيرة وهذا غير صحيح (غير مطابق) على ارض الواقع من الناحية النظرية، حيث الواقع بالمجتمعات النامية يكون شكل المنحنى على شكل تذبذبات غير منتظمة بين الفئات العمرية.

نتيجة للملاحظات السابقة التي ذكرت أنفًا نستطيع القول بأن عتد اقتراح نموج احصائي من الضروري أن يمر هذا التموج بثلاثة مراحل وهي توفيق والتقييم وتقويم البيانات السكانية (من أجل التحقق من سلامة واختبار النموذج Tweedie في عملية التقويم تم نشر بحثين في مجلة الناصر ومجلة الجزير في عام (2021) عن توفيق البيانات السكانية ودراسة مقارنة لبعض طرق تقويم البيانات السكانية باستخدام نموذج Tweedie.)، بالتالي تظهر لدينا بعض التساؤلات.

2. مشكلة الدراسة:

تتمثل مشكلة الدراسة في الاتي: (1) ما مدى دقة التعداد السكاني في الجمهورية اليمنية ومدى الاعتماد عليه كأساس لإعداد جداول الحياة وكيفية تقويم بيانات التعداد السكاني باستخدام الأساليب الإحصائية السكانية؟ (2) ما هي الخطوات اللازمة لإنشاء جدول حياة باستخدام نماذج توزيعات Tweedie الاحتمالية؟

3. أهداف الدراسة:

تهدف الدراسة إلى التعرف على واقع التعداد السكاني في الجمهورية اليمنية لعام (2004)، وتحديد أوجه القصور فيه ومعالجتها، حيث تتم المعالجة بمرحلتين: المرحلة الأولى مرحلة تقويم بيانات التعداد باستخدام الأساليب الإحصائية السكانية (طريقة كاير – فراج) والمرحلة الثانية هي إعداد جدول حياة باستخدام نماذج توزيعات Tweedie الاحتمالية، حيث إنّ جداول الحياة تعد أداة من أدوات التقويم، وعلى وجه التحديد:

1- دراسة تحليلية لمدى دقة التعداد السكاني لدى الجمهورية اليمنية.

2- دراسة أوجه القصور بالتعداد(ومعالجاتها) باستخدام الأساليب الإحصائية السكانية.

3- دراسة أنماط الوفاة العمرية للسكان وتحليل الخصائص السكانية الأخرى المتعلقة

بالوفاة باستخدام نماذج توزيعات Tweedie الاحتمالية.

4. أهمية الدراسة:

تتمثل أهمية الدراسة في:

1. تقديم دراسة تحليلية لتقييم بيانات التعداد السكاني وأيضًا معالجة القصور من أخطاء هذا التعداد لدى الجمهورية اليمنية، وبالتالي يساعد في التخطيط الجيد وإعداد الاستراتيجيات المتعلقة بالنواحي التعليمية والصحية والدفاعية بصورة فعالة وهو ما يساعد في النمو الاقتصادي والاجتماعي لدى السكان.
2. إعطاء القيادات ومتخذي القرار والمخططين في الجهاز المركزي للإحصاء في الجمهورية اليمنية أنموذجًا مقترحًا لجدول حياة للسكان في اليمن من خلال تزويدهم بالمعلومات العلمية التي تساعدهم في كيفية تقويم هذا التعداد.
3. تقديم دراسة تحليله أنماط الوفاة العمرية للسكان وتحليل الخصائص السكانية الأخرى باستخدام نماذج توزيعات Tweedie الاحتمالية.
4. إكساب الباحثين في التخصصات الأخرى طريقة في كيفية التعامل مع الطرق الإحصائية لتقويم البيانات بشكل عام، وبصفة خاصة النماذج الاحتمالية نماذج توزيعات Tweedie الاحتمالية وعلى ماذا تقوم وكيفية التقدير، والتي ستضيف معرفة علمية تكاد تكون غير متوفرة في المكتبات اليمنية.

5. حدود ومنهجية الدراسة:

تتناول الدراسة تحليل وتقيم التعداد السكاني لجمهورية اليمن، وذلك بتطبيق الأساليب الإحصائية السكانية وعلى وجه التحديد نسبة العمر، ونسبة النوع ومقياس سكرتارية الأمم المتحدة، وكذلك معالجة وتصحيح بيانات التعداد، حيث تتم المعالجة بمرحلتين، وذلك بتطبيق طريقة كارير – فراج كمرحلة أولى ومن ثم إعداد جدول حياة باستخدام نماذج توزيعات Tweedie الاحتمالية ذات التشتت الزائد، حيث تم الحصول على بيانات الدراسة عن الأعداد الخام للأحياء من سكان الجمهورية اليمنية بحسب فئات الاعمار الخمسية والنوع لعام 2004 (وذلك هو المتاح حاليًا) من كتاب الإحصاء السنوي للجهاز المركزي للإحصاء. اعتمدت الدراسة على منهجين هما:

1. المنهج الوصفي: المتمثل في تحليل المصادر والدراسات العلمية السابقة والمتخصصة بهدف تحديد الجانب النظري، ومدى دقة بيانات التعداد واخيرًا درجة التطابق بين نتائج طريقة التقويم وتوزيع الأعمار والنوع لسكان الجمهورية اليمنية باستخدام الأساليب الإحصائية السكانية.
2. المنهج الاستنتاجي (الاستدلالي): المتمثل في تقدير معالم نماذج توزيعات Tweedie الاحتمالية وكيفية إعداد جدول حياة للسكان الجمهورية اليمنية.

6. خطة الدراسة:

تتكون خطة الدراسة من مرحلتين:

أولًا: مرحلة الدراسة المكتبية: وتتناول الدراسة العلمية المكتبة من خلال المراجع المتخصصة والدراسات السابقة الخاصة بالأخطاء التي تتعرض لها البيانات الإحصائية وكيفية تقييمها وتقويمها باستخدام الأساليب الإحصائية السكانية وكذلك دراسة توزيعات Tweedie الاحتمالية وكيفية تقديرها وتطبيقها.

ثانيًا: مرحلة الدراسة التطبيقية: سيقوم الباحث بدراسة التعداد العام للسكان في الجمهورية اليمنية لعام 2004 والبحث عن الأخطاء التي تعرض لها هذا التعداد ومعالجتها علميًا باستخدام الأساليب الإحصائية السكانية (كاير – فراج) كمرحلة أولى، ومن ثم إعداد جدول حياة باستخدام توزيعات Tweedie الاحتمالية كمرحلة ثانية.

ثانيًا: الأخطاء التي تتعرض لها الاحصاءات الديموغرافية وأساليب أكتشافها وتقويمها

قد تحتوي الإحصاءات التي يتم تقديرها واستخلاصها من البيانات الاحصائية الخاصة بالسكان إلى اخطاء، والتي تؤدي بدورها إلى أخطاء في حسابات تقدير السكان، وقصور في عمل أو بناء جدول الحياة. لذا سنتناول بشرح بعض المفاهيم للأخطاء التي تتعرض لها الإحصاءات الديموغرافية وأساليب أكتشافها وتقويمها.

1. الأخطاء التي تتعرض لها الإحصاءات الديموغرافية:

تعد بيانات التوزيع العمري والنوعي للسكان في المجتمع ذات أهمية كُبرى في عمليات التخطيط للتنمية الاقتصادية والاجتماعية. وحتى يكون التخطيط سليمًا فإن هذه الإسقاطات يجب أن تكون دقيقة بقدر الإمكان، ويتطلب الأمر أن تكون نقطة ارتكازها مبنية على توزيع عمري او نوعي صحيح. حيث أن هذه التوزيعات، ومصدرها عادة هو التعدادات السكانية (الحصر الشامل)، فإن هذه العملية غالبًا ما تتعرض الي وجود بعض الأخطاء والعيوب في البيانات، ليس من المبالغ في شيء أن يُقال إنه لا يوجد من بين تعدادات العالم ما يمكن ان يُسمى تعدادًا صحيحًا أو كاملا[9]. ففي كل دولة لابد من وجود أخطاء في بيانات التعداد نتيجة لتعدد مراحل التعداد وتداخلاتها، من أهم هذه المراحل هي الإعدادية والميدانية، التجهيزية، التحليلية[11]. لذا تظهر لدينا أخطاء نتيجة هذه المراحل وتختلف هذه الأخطاء بتقدم أو تأخر الدول في إجراء العمليات الإحصائية، وتنشأ الأخطاء من مصادر متنوعة تختلف بحسب الدور أو النشاط او المرحلة. فقد تأتي في المراحل الإعدادية وبسبب قصور في تصميم استمارة التعداد، وقد تأتي الأخطاء من البيانات التي يُدلي بها الأفراد أو بسبب تعمد الاجابة الخاطئة أو غير الدقيقة. كما يكون العداد نفسه مصدرًا للأخطاء من ناحية التسجيل أو عدم إيصال السؤال بشكل دقيق للمبحوث، وقد تنشأ

الأخطاء عند تجهيز البيانات أو عند إعدادها للنشر.

مما سبق ليس من السهل سرد كل ما يمكن أن يصيب الإحصاءات الديموغرافية من أخطاء لذا تم تصنيف أهم أنواع الأخطاء التي يُمكن أن تتعرض لها البيانات الديموغرافية في ثلاث مجموعات:

أ- أخطاء التغطية أو الشمول" coverage Errors":

تتمثل أخطاء الشمول في الزيادة أو النقص في عد السكان و يرجع ذلك للأسباب التالية:(1) قصور عد undercount[9]، (2) تكرار عد over count [11]، (3)عيوب في المراحل الإدارية[4].

ب- أخطاء المحتوى او المضمون" content Errors"

قد يكون العدادون أو الموظفون مصدرًا لهذه الأخطاء وقد يكون أفراد المبحوثون أنفسهم مصدرًا ثانيًا لهذا النوع من الأخطاء. لذا سنوجز هذا النوع من الأخطاء بالنقاط:(1) أخطاء الإجابة Response Errors[4]،(2) أخطاء العدادين Enumerators Errors[4]، (3) أخطاء تجهيز أو إعداد البيانات Date Processing [3] Errors.

وتنشأ هذه الأخطاء أثناء التجهيز الفني للبيانات لاسيما أثناء عملية الترميز أو وضع الدليل (Coding) ، والتثقيب (Punching) وتبويب البيانات للنشر (Tabulation)[4].

ت- الاخطاء التي تتعرض لها بيانات الإعمار الخاصة بالسكان:

تتعرض البيانات الإحصائية السكانية وبصفة خاصة في الدول النامية لعدة أخطاء، تنشأ بسبب قلة الخبرة الإحصائية الفنية لدى تلك الدول، وشيوع الأمية بها، وكذلك شيوع بعض القيم الاجتماعية التي تؤثر على دقة وشمول تلك البيانات وتتلخص هذه الاخطاء في:

- **أخطاء عدم الشمول Incompleteness**

ويُقصد بذلك الحصر غير الدقيق للسكان، ويرجع ذلك لأسباب: (1) أن التعداد لا يشمل حصر جميع السكان على وفق إجراء التعداد وبخاصة أذا كان المجتمع يتعرض لحركة هجره داخلية أو كان التعداد يتم على أساس فعلي[14].(2) عدم الالتزام بالتعريف الدقيق للوفاة طبقًا للأمم المتحدة[13].(3) أخطاء عدم الشمول للوفيات[14].(4)أخطاء خاصة بعدم الدقة في التوزيع الوفيات زمانيا ومكانيا[13].

- **التبليغ الخاطئ عن العمر" age misstatement"**

يحدث هذا بالنسبة للأحياء وقت التعداد عند الإبلاغ عن العمر أو بالنسبة للوفيات عند الإبلاغ عن الوفاة وتنشأ هذه الأخطاء من أحد المصدرين: 1) المصدر الأول: التحيز أو التفضيل لأرقام عمرية معينة للأحياء من السكان او الوفيات عند الإبلاغ عن العمر وهي صفر وخمسة ومضاعفاتها ويؤدي هذا التحيز إلي تراكم أعداد

السكان أو المتوفين منهم عند تلك الأعمار على حساب الأعمار الأخرى[14].

2) المصدر الثاني"age shifing" وهو الانتقال بالأعمار من فئاتها الأصلية الي فئات أخرى أكثر شبابًا او كهولة، مما يؤدي هذا الانتقال لفئات عمرية خاطئة إلي ظهور وطأة الوفاة أعلى من حقيقتها عند بعض الفئات و أقل من حقيقتها عند بعض الفئات[12].

2. الأساليب الاحصائية المستخدمة لتقويم أخطاء البيانات الديموغرافية:

ذكرنا سابقًا بأن هناك فرقًا بين مفهوم التقويم والتقييم لبيانات التعدادات السكانية حيث يعد الأخير (التقييم) مرحلة أولية لاكتشاف الأخطاء وتقييمها ـ ونتيجة لذلك (وجود الأخطاء في البيانات) وجب علينا تقويم البيانات(تصحيحها).

أ- الأساليب الإحصائية المستخدمة لاكتشاف وتقييم أخطاء العمر والنوع:

هناك عدة أساليب لاختبار جودة بيانات التعداد، هذه الاساليب تعد مؤشرًا لقياس الأخطاء في التعدادات السكانية، يمكن استخدامها في اختبار جودة البيانات السكانية سواء كان الموضوع معالجة بيانات تعداد سكاني أو إحصاءات حيوية أو بيانات الهجرة. ويرجع هذا التشابه في المعالجة إلي ارتباط الظواهر الديموغرافية فيما بينها وأيضا الظواهر الاجتماعية والاقتصادية الأخرى. وسيتم التركيز في هذه الدراسة على اهم أساليب الاتساق في اختبار جودة بيانات التعداد وذلك للأسباب التالية:(1) أن أساليب المقارنة الخارجية تعتمد بصفة أساسية على المقارنة بتوزيع خاص بمجتمع مشابه لمجتمع الدراسة. ولا شك أن تحديد المجتمع المشابه لا يخلو تماما من التحكمية أو التحيز. (2) حتى وإن توفرت مسوح تقييم البيانات (من أجل التحقق المباشر) فإن الأخطاء المرافقة لبيانات الأعمار تظل قائمة، وإن كانت أقل حدةً، لاسيما في المجتمعات الزراعية ذات نسبة الأمية المرتفعة. وعليه فإن أهم أساليب الاتساق الداخلي أو شمول ودقة بيانات التعداد التي تستخدم في تقييم العمر والنوع من حيث المحتوى هي دراسة التوزيع العمري ومدى توافق نسبة النوع والعمر مع الأنماط السائدة لهذه النسبة [6].

- **نسبة العمر Age Ratio:**

وتعد هذه النسبة أحد المؤشرات الأساسية التي يمكن حسابها بالاعتماد على توزيع السكان حسب السن والنوع وهي الناتج من قسمة عدد السكان في عمر معين ($_5P_a$) على متوسط عدد السكان السابق ($_5P_{5-a}$) واللاحق ($_5P_{5+a}$) للعمر المعين[6].وتحسب هذه النسبة لكل من الذكور والإناث كل على حدة، وتُحسب هذه النسبة لكل الفئات العمرية عدا الفئتين الأولى والأخيرة مصنفة حسب الجنس وذلك طبقا للصيغة التالية [6]:

$$Age\ Ratio = \frac{5\,P_a}{1/2\,(\,{}_5P_{5-a} + {}_5P_{5+a})}$$

إذ ينبغي أن يكون انحراف نسبة العمر عن الرقم 100 قليلًا جدا وذلك باستثناء الأعمار المتقدمة، فإذا كان فرق النسب عن الرقم 100

صغيرًا أمكننا القول أن حجم الخطأ في هذه الفئة قليل والعكس صحيح[5].

جدول(1): نسبة العمر لكل من الذكور والإناث للجمهورية اليمنية لعام (2004)

الفئات العمرية	النسبة لعدد السكان %		الانحرافات عن 100	
	الذكور %	الاناث %	الذكور	الإناث
0-4	-	-	-	-
5-9	103.94	106.51	3.94	6.51
10-14	106.92	100.44	6.92	0.44
15-19	102.01	106.22	2.01	6.22
20-24	96.96	94.78	-3.04	-5.22
25-29	100.34	105.55	0.34	5.55
30-34	84.07	78.69	-15.93	-21.31
35-39	102.25	112.08	-2.25	12.08
40-44	99.30	95.76	-0.7	-4.24
45-49	93.11	97.57	-6.89	-2.43
50-54	117.51	111.42	7.51	11.42
55-59	70.42	68.84	-29.58	-31.16
60-64	138.46	137.5	38.46	37.5
65-69	68.07	67.18	-31.93	-32.82
70-74	154.05	161.54	54.05	61.54
75-79	49.04	45.26	-50.96	-54.74
80+	-	-	-	-
مجموع الانحرافات(بغض النظر عن الإشارة)			264.51	293.18
مقياس دقة العمر (المتوسط)			17.634	19.545

* عدد الاناث في تعداد (2004) هو (9648208)، وعدد الذكور في تعداد (2004) هو (10036953).

ويوضح جدول(1) نسب العمر لكل من الذكور والإناث حسب تعداد 2004، حيث يتضح بأن الانحرافات لنسب العمر عن الرقم 100 بلغ 264.51 للذكور و (293.18) للإناث مما يدل على وجود أخطاء في بيانات الأعمار، وأن هذه الأخطاء كانت أكبر مما هي عليه عند الذكور، حيث بلغت قيمة مقياس دقة العمر للذكور (17.63) وللإناث(19.55)مما يدل على أن الإناث اقل دقة في الإدلاء بأعمارهن من الذكور في تعداد 2004.

- **نسبة النوع Age Ratio:**

تعرف نسبة النوع بأنها عدد الذكور لكل مائة أنثى وتحسب وفق الصيغة التالية[6]: $Sex\ Ratio = \frac{5M_a}{5F_a}$ ويُستفاد من دراسة نسبة النوع في الكشف عن أخطاء توزيع السكان بحسب العمر والنوع عند مقارنتها بالنمط العام لها . فالنمط العام أن نسبة النوع عند الميلاد تتراوح بين 95 و105 تقريبًا. والنسبة التي تخرج عن هذا المدى غالبًا ما تعود إلى النقص أو الزيادة في أحد الجنسين، ما لم يكن هناك تفسير منطقي لأوجه عدم الاتساق وتأتي الهجرة وخسائر الحروب والأوبئة أسبابًا لذلك[6].

جدول(2): نسبة النوع (عدد الذكور مقابل لكل 100 أنثى) للجمهورية اليمنية لعام (2004)

الفئات العمرية	نسبة النوع	فروق نسب النوع
0-4	104.45	-
5-9	105.38	0.93
10-14	111.73	6.35
15-19	104.46	-7.27
20-24	104.43	-0.03
25-29	98.38	-6.05
30-34	101.67	3.29
35-39	90.17	-11.5
40-44	95.17	5.00
45-49	94.35	-0.82
50-54	104.51	10.16
55-59	109.49	4.98
60-64	111.04	1.55
65-69	111.49	0.45
70-74	108.57	-2.92
75-79	118.60	10.03
80+	110.59	-8.01
مجموع الانحرافات(بغض النظر عن الإشارة)	.	79.34
متوسط الفروق	.	4.96

ويوضح جدول (2) نسبة النوع بحسب تعداد 2004 حيث تشير النتائج الي وجود انحرافات في نسب النوع عن النمط العام لها حيث يظهر بأن نسبة النوع كانت مرتفعة عن الرقم (100) في الأعمار الصغيرة (قبل 25-29) ثم انخفضت حتى وصلت 98.38 عند الأعمار الشابة (25-

29) وكانت متذبذبة في الفئات العمرية من (30–34) إلى (45–49) ثم ارتفعت بشكل مفاجئ عن الرقم (100) في الفئات الأخيرة المتبقية. مما يشير إلى وجود أخطاء في الأدلاء بالأعمار من الفئة (50–54) الي 80+ أدى إلى ارتفاع مفاجئ في قيمة نسبة النوع عند هذه الفئات.

ب- الأساليب الإحصائية لقياس وتقييم دقة بيانات العمر

إن استخدام الأرقام القياسية في قياس دقة الأعمار يفيد في أغراض المقارنات بين فئات مختلفة من نفس المجتمع كالحضر والريف أو بين الذكور والإناث أو لنفس المجتمع ولكن على نقاط زمنية مختلفة لرصد التطور الذي حدث في قياس دقة الأعمار خلال فترات سابقه. وسنستعرض فيما يلي طريقة لحساب الرقم القياسي، أن هذه الطريقة تقيس بصفة خاصة صافي خطأ الإدلاء عن العمر.

مقياس سكرتارية الامم المتحدة:

تتطلب هذه الطريقة توزيع السكان بحسب الفئات الخمسية حيث إن توزيع السكان بحسب الفئات الخمسية يساعد في تقليل صافي خطأ الإبلاغ عن العمر ولكنها قد تبتعد عن الصورة الحقيقية، بمعنىً فيما لو أن الأخطاء تمثلت في أخطاء الشمول على وجه التحديد في الأعمار مختارة. بالتالي هذه الطريقة تعكس صورة أوضح لمدى التكوين العمري عامة فضلًا عن تأثر هذه الطريقة بأخطاء شمول التعداد [3]. ويتم حساب هذا المقياس طبقًا للقانون التالي [3]"مقياس سكرتارية الأمم المتحدة= 3(متوسط القيم المطلقة لفروق النسب الخاصة بالنوع) + متوسط القيم المطلقة لفروق العمر للذكور+ متوسط القيم المطلقة لفروق العمر للإناث". ومقياس سكرتارية الأمم المتحدة لا يُعطي تقديرًا لزيادة العد أو النقص فيه ولكنه يعطى مؤشرًا صافيًا لأخطاء الإدلاء بالأعمار. فاذا كان المقياس أقل من 20 دل ذلك على أن بيانات العمر والنوع دقيقة، أما إذا تراوحت قيمته بين (20–40) فيدل على أن بيانات العمر والنوع غير دقيقة أما إذا زادت عن 40 فيدل على أن البيانات غير دقيقة على الاطلاق[3].

جدول(3): حساب مقياس سكرتارية الأمم المتحدة لقياس التراكم العمري للجمهورية اليمنية لعام (2004)*

الفئات العمرية	نسبة اعداد الذكور	نسبة اعداد الإناث	نسبة النوع	الانحرافات	نسبة العمر		انحرافات نسبة العمر عن 100	
					للذكور	للإناث	ذكور	اناث
0-4	15.02	14.38	104.45	-	-	-	-	-
5-9	15.68	14.88	105.38	0.93	103.94	106.51	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	5.00	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
59-55	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	2.25	2.33	96.57	14.92	-	-	-	-
المجموع	100.38	96.48		73.3			178.74	217.23

* عدد الاناث في تعداد (2004) هو (9648208)، وعدد الذكور في تعداد (2004) هو (10036953). لاستخراج اعداد السكان على حسب الفئات نضرب النسب في اعداد الذكور والاناث.

من جدول(3)،

$$\text{المقياس} = 3 \times \left(\frac{73.3}{13}\right) + \left(\frac{178.74}{13}\right) + \left(\frac{217.23}{13}\right)$$

= 47.37

ويتضح من الجدول أن قيمة مقياس سكرتارية الأمم المتحدة المحسوبة باستخدام بيانات الفئات العمرية لتعداد 2004 قد تجاوزت الحد الذي تعد عنده بيانات العمر والنوع دقيقة وهو أقل من 20، بل وقعت في ضمن المدى الذي تعد عنده بيانات العمر والنوع غير دقيقة على الإطلاق حيث بلغت قيمة سكرتارية الأمم المتحدة 47.37.

وسيتم الاكتفاء بالطرق السابقة وهي نسبة العمر، نسبة الجنس، ودليل سكرتارية الأمم المتحدة في هذه الدراسة للسببين الآتيين: (1) أن مقياس العمر والنوع هو أحد مكونات هذا الدليل.(2) أن تقييم التوزيع العمري ليس هدفًا في حد ذاته، لاسيما مع التوقع القوي بأن هذا التوزيع يحتوي على أخطاء في اغلب الدول النامية عامة،

وفي المجتمعات الأمية منها بصفة خاصة. وعلى ذلك فأن دليل سكرتارية الأمم المتحدة الذي يقدر للسكان في العمر أقل من 75 سنة، يشير الي عدم دقة التوزيع العمري إذا كانت قيمته 20 فأكثر.

(ج) طُرق تقويم بيانات التوزيع العمري للسكان

سبق أن ذكرنا أن بيانات التعداد السكاني تحتوي على الكثير من الأخطاء وقد ذكرنا سابقًا بعض هذه الأخطاء ولكن هناك عدة طرق لتقويم هذه البيانات وجعلها أكثر ملائمةً وممهدةً او بشكلٍ أدق أكثر تهذيبًا للبيانات ومن هذه الطرق[3](كارب- كنك- نيوتن)، أريجا، الأمم المتحدة، سترونك وسنركز على طريقة كارير- فراج، وذلك بسبب أن الطريقة أهتمت بمعالجة بيانات الفئات العمرية كلها ما عدا الفئات العمرية أقل من 10 سنوات والفئات العمرية أكبر من 70 عامًا مما يجعل هذه الطريقة أكثر دقة من بين الطرق الأخرى.

طريقة (كارير فراج) Carrier – Farrag Formula تستند على العلاقة بين توزيع السكان بحسب الفئات العمرية الخمسية مع توزيع السكان بحسب فئات العمر العشرية حيث تكون كمعدل مشابه إلى علاقة ثلاثة من الفئات المتتالية[6].

وهذه الطريقة لا تصحح الفئة (0-4)، (5-9)، (70 فأكثر). ووفق هذه الطريقة يتم تعديل أعداد السكان في الفئات العمرية التي تبدأ بالعدد (0) حسب الصيغة الآتية [6]:

$$_5P_x = {}_{10}P_x - {}_5P_{x+5}$$

اما الفئات العمرية الخمسية التي تبدأ بالعدد (5) فيتم تعديلها كالاتي :

$$_5P_{x+5} = \frac{{}_{10}Px}{1+({}_{10}Px-10/{}_{10}Px+10)^{1/4}}$$

حيث ان $_5P_{x+5}$ هو العدد الذي تم تعديله في معادلة السابقة.

جدول(4): البيانات النسبية قبل وبعد التصحيح للجمهورية اليمنية لعام (2004) باستخدام طريقة (كارير فراج)

فئات الأعمار	نسبة عدد الذكور قبل التصحيح	نسبةعدد الذكوربعد التصحيح	نسبةعدد الإناث قبل التصحيح	نسبةعدد الاناث بعد التصحيح
0-4	15.02	15.02	14.38	14.38
5-9	15.68	15.68	14.88	14.88
10-14	15.15	14.93	13.56	13.74
15-19	12.66	12.88	12.12	11.94
20-24	9.67	9.65	9.26	9.36
25-29	7.30	7.32	7.42	7.32
30-34	4.88	5.15	4.80	5.33
35-39	4.31	4.04	4.78	4.25
40-44	3.55	3.52	3.73	3.76

فئات الأعمار	نسبة عدد الذكور قبل التصحيح	نسبةعدد الذكوربعد التصحيح	نسبةعدد الإناث قبل التصحيح	نسبةعدد الاناث بعد التصحيح
45-49	2.84	2.87	3.01	2.98
50-54	2.55	2.24	2.44	2.14
55-59	1.50	1.81	1.36	1.66
60-64	1.71	1.49	1.54	1.35
65-69	0.97	1.19	0.87	1.06
70-74	1.14	1.14	1.05	1.05
75-79	0.51	0.51	0.43	0.43
80+	0.94	0.94	0.85	0.85

ويوضح جدول(4) أن الطريقة أهتمت بمعالجة بيانات الفئات العمرية كلها ما عدا الفئات العمرية أقل من 10 سنوات والفئات العمرية أكبر من 70 عامًا، كما يلاحظ بأن هناك تقاربًا شديدًا بين القيمة الحقيقية والقيمة المصحح لكل من (الذكور والإناث) على حد سواء.

رابعًا: بناء جدول حياة للجمهورية اليمنية وفقًا لآخر تعداد متاح

تستخدم البيانات الإحصائية السكانية أساسًا لعمل جداول الحياة أو الوفاة، وهي جداول يمكن بواسطتها قياس احتمالات الحياة أو الوفاة عند كل عمر من الأعمار [10]. ويمكن تقسيم جداول "الحياة "أو "الوفاة" حسب طول الفئة العمرية إلى نوعين من الجداول هما جداول الحياة الكاملة: وهي جداول تتضمن بيانات عمرية لكل سنة للأعمار منذ الميلاد وحتى آخر عمر يمكن أن يصل إليه الفرد في هذه المجموعة، وجداول الحياة المختصرة: وهي جداول تتضمن بيانات عمرية لكل مجموعة من السنوات تكون خمس سنوات أو 10 سنوات مثلا(فئات عمرية) ابتداء من الميلاد وحتى آخر فئة عمرية يمكن أن يصل إليها الفرد في هذه المجموعة. وهُناك عدة افتراضات في جدول الحياة[10]: (1) الفوج مغلق بالنسبة للهجرة منه واليه والفقدان يكون بالموت فقط.(2) يموت السكان في كل عمر وفق جدول حياة محدد مُقدمًا ولا يتغير.(3) أصل الفوج عدد معياري من المواليد وهو عادة 1000، 10000، 100000 ويُسمى أساس جدول الحياة (The Radix). (4) عند كل عمر(فيما عدا السنوات القليلة الأُولى من الحياة) تكون الوفيات موزعة بانتظام.(5) يُحسب عادة جدول لكل نوع على حدة سواء للذكور او الاناث نظرًا لاختلاف الوفيات فيما بينهما.

- **تعريف أهم مصطلحات جدول الحياة:**

ولتحديد أهم مصطلحات جداول الحياة وتعريفها، اقتصرت جميع رموز جدول الحياة على (X) ويدل هذا على العمر بينما (n) يدل على طول الفئة سواء كانت أحادية أو خماسية، $_xl$ايدل على عدد الأشخاص الباقيين على قيد الحياة عند العمر (X) بالضبط، $_nd_x$ يدل على عدد

الأشخاص المتوفين بين العمر (x) والعمر (x+n) ، ${}_nP_x$ يدل على احتمال البقاء على قيد الحياة من العمر (x) إلي العمر (x+n) ، ${}_nq_x$ يدل على احتمال الوفاة بين العمر (x) و العمر (x+n)، ${}_nL_x$ يدل على عدد السنوات التي عاشها الفوج بين العمر (x) إلي العمر (x+n)، T_X يدل على جملة السنوات التي عاشها الفوج بعد المضبوط (X) أو هي عدد الاشخاص الذين عمرهم (X) سنة فأكثر إذا كان المجتمع متوقفا، $e_{(X)}$ يدل على توقع الحياة أي متوسط عدد السنوات المتوقع أن يعيشها الفرد اعتبارا من تمام العمر (X) [10].

ومن أجل بناء جداول الحياة فأن الخطوة الأولى في بناء جدول الحياة المختصر ستكون الحصول على مُعدلات الوفاة المركزية، ونرمز لمعدل الوفاة المركزي الخاص بالفئة العمرية (x,x+n) بالرمز (${}_nm_x$) ويمكن الحصول عليه من الصيغة التالية:

$$ {}_nm_x = \frac{{}_nd_x}{{}_nL_x} $$

ولتحويل معدلات الوفاة المركزية(${}_nm_x$) إلي معدلات فئوية تستخدم علاقة تربط بين معدلات الوفاة(${}_nq_x$) للفئة العمرية (x,x+n) ومُعدل الوفاة المركزي(${}_nm_x$) داخل هذه الفئة العمرية [8]. هناك العديد من الطرق التي تم تصميمها وتطويرها لإيجادالعلاقة بين الاحتمالات الوفاة (${}_nq_x$) والمعدل المركزي للوفيات(${}_nm_x$) ومن اشهر هذه الطرق[8]طريقة جريفل Grevilles (method) و طريقة الرجوع الي جدول معياري (Method of reference to a standard table) وطريقة ريد – ميريل (The Reed–Merrell method)و أيضًا توزيعات Tweedie الاحتمالية. وسنركز على طريقةTweedie الاحتمالية كمقترح لإعداد جدول حياة.

- **توزيعات Tweedie الاحتمالية ذات التشتت الزائد:**

ينتمي توزيعات Tweedie إلي عائلة التشتت الاسي، فهي تحتوي على معلمة العائلة الأسية الخطية بالإضافة الي معلمة التشتت[18]، 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[16]). وبشكل عام دالة الكثافة للتوزيعات التشتت الأسي الزائد تكتب بالشكل التالي[17]:

$$ f_Y(y \mid \theta, \varphi) = h(y, \varphi)\exp\left(\tfrac{1}{\varphi}(\theta y - \eta(\theta))\right) $$

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، $h(y, \varphi)$ تمثل دالة غير محددة لقياس استقلالية المعلمة θ [18]. 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

$$\eta''(\theta) = Var(\mu)$$ ،(15)

وبحسب تباين التشتت الاسي و التي تسمى ايضًا 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 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:

$$Var(Y) = E(Y)^{\rho} = \mu^{\rho}$$

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:

$(\rho - 1)(1 - \alpha)$.(18)

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 سابقًا – بالحالات الخاصة لتوزيعات Tweedie [18].لذا توزيعات Tweedie هي توزيعات بواسون المركبة أو بتوزيعات المختلطة بواسون مع توزيعات جاما[17]. ويمكن تمثيل توزيعات Tweedie التجمعية على الصورة التالية:

$$Y = X_1 + X_2 + \cdots + X_N$$

حيث إن 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 $Y > 0$ على 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 إن λ, α, β ترتبط بمعالم توزيعات Tweedie على وفق العلاقات التالية[18]:

$$\lambda = \frac{\mu^{2-\rho}}{\varphi(2-\rho)} \quad ، \quad \alpha = \frac{\rho-2}{(\rho-1)} \quad ، \quad \beta = \varphi(\rho-1)\mu^{\rho-1}.$$

وبالتالي معالم توزيع Tweedie على وفق معالم التوزيع المختلط بين توزيع بواسون وجاما تكون بالشكل التالي[18]:

$$\mu = \lambda\alpha\beta \quad ، \quad \rho = \frac{\alpha+2}{\alpha+1} \quad ، \quad \varphi = \frac{\lambda^{1-\rho}(\alpha\beta)^{2-\rho}}{2-\rho}.$$

وبالتالي فإن قيمة $h(y,\varphi)$ تكون وفقاً للعلاقات التالية:

$$h(y,\varphi)=\begin{cases}\frac{\mu^{2-\rho}}{1-\rho}, & if\ \ y=0\\ \frac{1}{y}\sum_{n=1}^{\infty}h_n(y,\varphi,\rho)\ , & \text{if } y>0\end{cases}$$

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 لـ(Wright`s generalized Bessel) وهي ترتبط بتوزيع Tweedie الاحتمالية. بينما قيمة θ و $\eta(\theta)$

$$\theta=\begin{cases}\log\mu\ , & if\ \ \rho=1\\ \frac{\mu^{1-\rho}}{1-\rho}\ , & \text{if } \rho\neq 1\end{cases}$$

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

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

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

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 بـ 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[15]، وكذلك باستخدام طريقة الإمكان الأعظم المعدلة حيث تعتمد هذه الطريقة المعاملات المتعامدة على معالم φ و ρ[18].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 $h(y,\varphi)$.

- **إنشاء جدول حياة مختصر باستخدام توزيعات Tweedie الاحتمالية:**

سبق وناقشنا أن البيانات التوزيع العمري والنوعي والأخطاء التي تتعرض لها وتقويمها باستخدام طريقة كارير فراج، لذا تم استخدام بيانات هذه الطريقة– كارير فراج– لاحتساب معدلات الوفاة مع العلم بأن معدل وفيات الأطفال الرضع (اقل من سنة) ومعدل وفيات الأطفال اقل من خمس سنوات تم استخراج نتائجها من كتاب الجهاز المركزي للإحصاء(2006) نظرًا لعدم توفر بيانات الإحصاءات الحيوية في الكتاب الخاص للجهاز المركزي للإحصاء للجمهورية اليمنية. بينما بقية الفئات تم استخراج عدد الوفيات وذلك بالطرح المتتالي للفئات، ولأجل استخراج معدل الوفاة فإننا نقوم بقسمة عدد الوفيات على متوسط عدد الأفراد بين العمر (x) و العمر (x+n) مضروب في خمسة، النتيجة ستكون عبارة عن معدلات الوفاة ($_nm_x$). وبالنسبة للفئة العمر الأخير(80) فأكثر فإننا سوف نفترض أن احتمال الوفاة يكون كبيرا جدًا، تقترب من الواحد الصحيح.

جدول (5): نسبة عدد الوفيات ومعدلات الوفاة للجمهورية اليمنية لعام (2004)

الفئات	الذكور			الإناث		
	نسبة عدد الذكور	نسبة عدد الوفيات	معدلات الوفاة	نسبة عدد الإناث	نسبة عدد الوفيات	معدلات الوفاة
0-1	2.68	-	0.07910	2.53	-	0.07504
1-4	12.34	-	0.09289	11.85	-	0.09166
5-9	15.68	0.92	0.012089	14.88	1.14	0.015933
10-14	14.93	1.71	0.024594	13.74	1.80	0.028037
15-19	12.88	3.40	0.059912	11.94	2.85	0.053521
20-24	9.65	2.33	0.054914	9.36	2.04	0.048921
25-29	7.32	2.06	0.065501	7.32	1.99	0.062915
30-34	5.15	1.33	0.057876	5.33	1.08	0.045094
35-39	4.04	0.41	0.022008	4.25	0.49	0.024463
40-44	3.52	0.65	0.040676	3.76	0.78	0.046291
45-49	2.87	0.61	0.047545	2.68	0.84	0.065625
50-54	2.24	0.47	0.046397	2.14	0.48	0.050526
55-59	1.81	0.30	0.036585	1.66	0.31	0.041169
60-64	1.49	0.30	0.044776	1.35	0.29	0.048093
65-69	1.19	0.05	0.008576	1.06	0.01	0.001894
70-74	1.14	0.63	0.152542	1.05	0.62	0.167568
75-79	0.51	0.51	0.140496	0.43	0.43	0.134375
80+	0.94	0.94	1	0.85	0.85	1

بعد الحصول على معدلات الوفاة $_nm_x$ ، وباستخدام بيانات معدلات الوفاة (جدول 5) سنستطيع تحويل معدلات الوفاة المركزية($_nm_x$) إلي معدلات فئوية تستخدم علاقة تربط بين احتمالات الوفاة($_nq_x$) للفئة العمرية (x,x+n) ومُعدل الوفاة المركزي($_nm_x$) داخل هذه الفئة العمرية، وذلك باستخدام توزيعات Tweedie الاحتمالية، وللحصول على احتمالات الوفاة لابد لنا من تحديد المتغيرات المستقلة والمتغيرات التابعة الخاص بتوزيعات Tweedie الاحتمالية.حيث X_1تمثل أعداد الوفيات الذكور، X_2 يمثل تمثل أعداد الوفيات الاناث وهما متغيران يتبعان توزيع جاما، N_i تمثل أجمالي الوفيات لأعداد السكان للفئات العمرية و i تمثل مجموعة الذكور والاناث حيث N_iتتبع توزيع بواسون، والمتغير التابع (Y) ويمثل مجموع اعداد الوفيات

الذكور والاناث حيث Y يتبع توزيع Tweedie. لتوفيق بيانات معدلات الوفاة بالتوزيعات الاحتمالية المذكورة أنفًا، تم استخدام برنامج (Easy Fit)، الجدول التالي يوضح جودة المطابقة لبيانات الأعمار في الجمهورية اليمنية.

جدول(6): جدول يبين جودة التوفيق لبيانات معدلات الوفيات في الجمهورية اليمنية لعام 2004

جودة المطابقة	التوزيعات الاحتمالية			
	اعداد الوفيات الذكور (X_1)	اعداد الوفيات الاناث (X_2)	اعداد الوفيات الذكور والاناث معا	اختبار Kolmogorov–smirnov
Sample Size	18	18	18	
Statistic	0.30278	0.30158	0.55546	
P–value	0.05835	0.05999	0	
Rank	10	10	5	
قيمة معالم التوزيع	α=2.5195 β=25.558	α=2.5599 β=25.187	λ= 62.151	

يلاحظ من الجدول (6) أن بيانات معدلات الوفيات لكل من الذكور والاناث توافقت مع توزيع جاما الاحتمالي من الناحية النظرية والاحصائية، بينما وفيات 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 ويتم الحصول على قيم 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(7): 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)	4002.1130	1.28413	8.50683
اعداد الاناث (X_2)	4006.9422	1.28091	8.72000

يلاحظ من الجدول (7) أن قيمة ρ تراوحت بين القيمتين (1,2) وهذا يعني بالفعل أن بيانات الوفيات لكل من الذكور والإناث هي بيانات مستمرة ذات قيم متقطعة عند الصفر، وكذلك قيم μ تراوحت (4006.9422 , 4002.1130) 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، بذلك يمكن احتساب احتمالات الوفاة بعد تصحيحها من الوفيات

الاعمار غير المبينه، وباستخدام تقدير متسلسلة لا برنامج R وكانت النتائج التقديرية كما يلي:
نهائيــة لتوزيعــات Tweedieوذلــك عــن طريــق

جدول(8): احتمالات الوفاة $_nq_x$ لبيانات معدلات الوفيات في الجمهورية اليمنية لعام 2004 باستخدام توزيعات Tweedie

الفئات العمرية	احتمالات الوفاة لذكور	احتمالات الوفاة نسبة الاناث
0-1	0.078137	0.074208
1-4	0.091978	0.090851
5-9	0.012086	0.015929
10-14	0.024569	0.028013
15-19	0.059401	0.053113
20-24	0.008316	0.007212
25-29	0.009919	0.009274
30-34	0.05688	0.044366
35-39	0.021668	0.024113
40-44	0.040642	0.046259
45-49	0.047034	0.064987
50-54	0.007026	0.007448
55-59	0.035939	0.040473
60-64	0.044198	0.047526
65-69	0.008538	0.001887
70-74	0.15107	0.166114
75+	0.138093	0.132223

وبعد الحصول على احتمالات الوفاة ($_nq_x$) توزيعـات Tweedie الاحتماليـة، والجـدول التـالي صار بالإمكان عمل جدول حياة مختصر لكل من يوضح جداول الحياة لكل من الذكور والاناث على وفق تعداد 2004 وباستخدام

جدول(9): جدول حياة مختصر للذكور والاناث للجمهورية اليمنية لعام 2004 باستخدام توزيعات Tweedie

الفئات العمرية	الذكور				الاناث			
	$_nm_x$	$_nq_x$	l_x	e(x)	$_nm_x$	$_nq_x$	l_x	e(x)
0-1	0.07910	0.078137	100000	46.17	0.07504	0.074208	100000	47.16
1-4	0.09289	0.091978	92186	49.05	0.09166	0.090851	92579	49.91

5-9	0.012089	0.012086	83707	49.81	0.015933	0.015929	84168	50.72
10-14	0.024594	0.024569	73590	51.31	0.028037	0.028013	82827	46.79
15-19	0.059912	0.059401	71782	47.54	0.053521	0.053113	80507	43.62
20-24	0.054914	0.008316	67518	45.39	0.048921	0.007212	76231	41.50
25-29	0.065501	0.009919	66848	40.82	0.062915	0.009274	75681	37.36
30-34	0.057876	0.05688	66185	36.20	0.045094	0.044366	74979	33.28
35-39	0.022008	0.021668	52420	40.05	0.024463	0.024113	71653	30.68
40-44	0.040676	0.040642	51284	35.89	0.046291	0.046259	69925	27.73
45-49	0.047545	0.047034	49200	32.30	0.065625	0.064987	66690	25.31
50-54	0.046397	0.007026	46886	28.77	0.050526	0.007448	62356	23.22
55-59	0.036585	0.035939	46557	23.95	0.041169	0.040473	61892	19.62
60-64	0.044776	0.044198	44884	19.75	0.048093	0.047526	59387	16.60
65-69	0.008576	0.008538	42900	15.55	0.001894	0.001887	56565	13.55
70-74	0.152542	0.15107	42534	10.66	0.167568	0.166114	56458	9.79
75+	0.140496	0.138093	36108	7.12	0.134375	0.132223	47080	7.56

يلاحظ من الجدول (9) أن توقع الحياة عند الميلاد في الجمهورية اليمنية حسب تعداد 2004 كان (46، 47) سنة لكل من الذكور والاناث على الترتيب، ويعني ذلك أن متوسط عدد سنوات الحياة المتوقعة لمولود جديد (46، 47) لكل من الذكور والاناث على الترتيب. كذلك تشير نتائج الجدول إلي أن معدلات الوفيات العمرية السنوية في الفئة (20-24) و (25-29) و (50-54) سنة ليست متقاربة مع احتمالات الوفاة المستخرجة باستخدام Tweedie الاحتمالية بينما بقية الفئات متقاربة جدًا. يرجع أسباب عدم التقارب الي عدم دقة نتائج التعداد السكاني لعام 2004 بشكل عام.

خامسًا: الخلاصة والتوصيات

يكتسب التوزيع العمري للسكان أهميةً بالغةً في مجالات التخطيط الاقتصادي والاجتماعي لرسم الخطط العامة للتنمية ووضع البرامج الطويلة الأمد أو القصيرة الأمد، وإجراء التنبؤات الديموغرافية حول مستقبل السكان، وتحديد الاتجاهات الرئيسية في تطور العوامل والظواهر السكانية، حيث إن قبول النتائج التي نتوصل إليها يعتمد بالدرجة الرئيسة على مدة الثقة في البيانات الأساسية وعلى مدى دقة هذه البيانات. وعليه فتقييم هذه البيانات والتحقق من نوعيتها يعد خطوة أساسية لابد منها لإنجاح أية دراسة نستعمل فيها تلك البيانات، وهنا تظهر أهمية تقويم البيانات الإحصائية السُكانية، لذا سيتم استعراض أهم النتائج التي أسفرت عنها الدراسة بما يتفق مع أهداف البحث، لاسيما في تقييم وتقويم بيانات التعداد السكاني لعام (2004) للجمهورية اليمنية، ومن ثم عمل جدول حياة في ضوء بيانات التعداد

المصُـحـحة التـي تسـاعد فـي عمليـة التخطـيط والتنميـة فـي جميـع القطاعـات الاقتصـادية و التعليمية و القوى العاملة والصحة وتنظيم الأسرة.

- **الخلاصة:**

لقد اتضح من خلال النتائج التي تم التوصل إليها من تطبيق توزيعات Tweedie الاحتمالية علـى بيانـات التوزيـع العمـري والنـوعي لسـكان الجمهورية اليمنية على وفق تعداد 2004 لأعداد جدول حياة مختصر، تتلخص في عدة نقاط:

1. أن بيانـات التركيـب العمـري والنـوعي للجمهورية اليمنية كانت مشوبة بالأخطاء، بحسب ما تم التوصل إليها من تطبيق ثلاث مقايس (نسبة العمـر – نسـبة النـوع– مقيـاس سـكرتارية الأمـم المتحـدة) لتقيـيم بيانـات العمـري والنـوعي للسـكان وفقًا لتعداد (2004).

2. أشار مقياس سكرتارية الأمم المتحدة الي أن هنـاك أخطـاء تتجـاوز الحـد الـذي تعـد عنـده بيانـات التركيـب العمـري و النـوعي دقيقـة، حيـث بلغـت قيمـة مقيـاس سـكرتارية الأمـم المتحـدة المحسوبة بالاعتماد على الفئات العمرية من (5–9) إلي (70+) سنة،(47.37).

3. أهتمت طريقة كارير فراج بمعالجة بيانات الفئات العمرية كلها ما عدا الفئات العمرية أقل من 10 سنوات والفئـات العمريـة أكبـر مـن 70 عامًـا، كمـا يلاحـظ بـأن هنـاك تقاربًـا شـديدًا بـين القيمـة الحقيقيـة والقيمـة المصـحح لكـل مـن (الـذكور والإناث) على حد سواء.

4. ان توقـع الحيـاة للـذكور عنـد المـيلاد (46.17) سـنة بينمـا توقـع الحيـاة للإنـاث عنـد الميلاد (47.16) سنة.

5. أن معدلات الوفيات العمرية السنوية في الفئـة (20–24) و (25–29) و (50–54) سـنة ليسـت متقاربـة مـع احتمـالات الوفـاة المسـتخرجة باستخدام Tweedie الاحتمالية بينما بقية الفئات متقاربـة جدًا. يرجـع أسباب عدم التقارب الي عدم دقة نتائج التعداد السكاني لعام 2004 بشكل عام وعلـى وجـه التحديـد ضـمن الفئـات المـذكورة آنفًـا، وهذا ما يؤكد الفقرة(1) و(2).

6. يمكـن اسـتخدام توزيعـات Tweedie الاحتماليـة لعمـل جـداول حيـاة لمـا يتمتـع مـن خصائص ومميزات بشرط إذا كانت بيانات التعداد دقيقة.

- **التوصيات:**

1. الأخـذ بعـين الاعتبـار الإحصـاءات الوصـفية كالمتوسـط والتبـاين ومعامـل الاخـتلاف والتفـرطح والالتـواء عنـد دراسـة بيانـات التوزيـع العمري والنوعي للسكان.

2. استخدام التوزيعات الاحتمالية وعلـى وجـه التحديـد توزيـع Tweedie الاحتمـالي فـي إعـداد جداول حياة مختصرة او كاملة لسكان الجمهورية اليمنية لما يتمتع به من خصائص ومميزات.

3. التهيئة الجيدة لإجراء التعداد السكاني في موعـد ملائـم مـن خـلال التـدريب الجيـد للعـدادين وتـوفير كـل مـا يتطلبـه التعـداد السـكاني مـن مسـتلزمات فنيـه وماديـه وكـوادر، عـلاوة علـى

توظيف 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. التنسيق بين المستشفيات والوحدات الصحية التابعة لوزارة الصحة والجهاز المركزي من أجل تزويد الجهاز بأعداد الولادات و الوفيات السنوية التفصيلية فضلا عن تلك المصنفة بحسب الفئات العمرية الخمسية والتي تفتقر حاليًا المجموعات الإحصائية السنوية مما يسهم في تسهيل مهمة الباحثين في الدراسة.

5. الاستمرار في استخدام النماذج الإحصائية والاعتماد عليها في توفيق بيانات التوزيع العمري والنوعي لسكان الجمهورية اليمنية.

المراجع العربية:

1. الجهاز المركزي للإحصاء" كتاب الإحصاء السنوي2005 "وزارة التخطيط والتعاون الدولي، صنعاء، الجمهورية اليمنية، نوفمبر 2006.
2. الجهاز المركزي للتعبة العامة والإحصاء، السكان وبحوث ودراسات "تقديرات جداول الحياة 1976-2001"جمهورية مصر العربية، القاهرة-مدينة نصر-مركز الأبحاث والدراسات الاحصائية والدراسات السكانية، العدد35.سنة1987.
3. الحمداوي، نادية علي عايد" الاسقاطات السكانية لمحافظة البصرة للفترة 1997-2022 باستخدام نتائج التعداد العام لسنة 1997 في العراق"رسالة ماجستير غير منشوره، جامعة بغداد، العراق، 2005.
4. الشلقاني، مصطفى " طرق التحليل الديموغرافي" كلية التجارة والاقتصاد والعلوم السياسية، الطبعة الثانية، جامعة الكويت، 1994.
5. العلفي، سمير أحمد محمد" تطبيق توزيع ويبل على بيانات التعداد السكاني للجمهورية اليمنية"المجلة المصرية للدراسات التجارية، المجلد السادس والثلاثون العدد الثاني2012.جامعة المنصورة ، ص 491-506.
6. 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.
7. المنصوب، عبدالحكيم عبدالرحمن المنصوب"ا لنماذج الإحصائية في تصحيح التوزيع العمري لسكان اليمن" مجلة الباحث الجامعي، اب، الجمهورية اليمنية، العدد 17، يونيو 2008، ص 220-232.
8. الأشقر، السيد الشربيني وهبه "بناء جدول حياة من واقع خيرة شركات التامين المصرية" رسالة غير منشورة، قسم الاحصاء التطبيقي، كلية التجارة، جامعة المنصورة، جمهورية مصر العربية، 2008.

9. أمين، ندى احمد " اسقاطات القوى العاملة، التعليم والصحة للتركيب العمري-النوعي لسكان العراق لمدة 1997-2017 "بحث دبلوم عالي في الاحصاء التطبيقي. كلية الادارة والاقتصاد، جامعة بغداد سنة2011.
10. خواجه خالد زهدي "جداول الحياة" المعهد العربي للتدريب والبحوث الاحصائية.
11. فراج،عبدالمجيد" الأسس الإحصائية للدراسات السكانية"، دار النهضة العربية، القاهرة، 1980.
12. فوده، محمد عبدالفتاح " استخدام الاسقاطات السكانية في تقدير الهيكل الديموغرافي ج.ع.م.-وتحديد الاتجاهات المتوقعة لعبء تأمين معاش الشيخوخة "، مجلة الدراسات المالية والتجارية-كلية التجارة-بنى سويف-جامعة القاهرة، يناير 2000.
13. فوده، محمد عبدالفتاح " تقييم التعداد السكاني لعام 1996لإعداد جدول حياة قوميه مختصر" مجلة الدراسات المالية والتجارية (للعلوم الإدارية)،كلية التجارة- بني سويف- جامعة القاهرة، المجلد الثاني- ملحق 1999.
14. فوده، محمد عبدالفتاح " تقييم بعض طرق تهذيب البيانات الإحصائية السكانية-مع تطبيق على بيانات تعداد مصر1996"المجلة المصرية للدراسات التجارية-كلية التجارة - جامعة المنصورة، مارس2001.

المراجع الاجنبية:

15. 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}.
16. Ridout, M. and Besbeas, p. (2004) " An empirical model for underdispersed count data" statist. model.4:77-89.
17. 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
18. 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}.