مجلة كلية التجارة للبحوث العلمية
كلية التجارة
جامعة الإسكندرية

العدد الثاني– المجلد الرابع والخمسون – يوليو ٢٠١٧

أ/ هبة أحمد حسن أحمد
معيدة بقسم الإحصاء والرياضة والتأمين
كلية التجارة – جامعة دمنهور

تحليل البيانات غير الكاملة في ظل وجود الأخطار المتنافسة:دراسة محاكاة[1]

Analysis of Incomplete Data in the Presence of Competing Risks :A Simulation Study

مخلص البحث

يهدف هذا البحث إلي تقييم عدد من المقدرات المعلمية واللامعلمية المستخدمة في تقدير نماذج الأخطار المتنافسة ذات البيانات غير الكاملة باستخدام "**دراسة المحاكاة**".

وتشمل هذه المقدرات اللامعلمية: **مقدر(KM) Kaplan–Meier التقليدي، ومقدر "KM" المرجح المعدل(MWKM)** وقد استخدمت لتقدير دالة المأمونية في ظل وجود الأخطار المتنافسة والبيانات المراقبة من جهة اليمين في دراسة المحاكاة. بينما تشمل المقدرات المعلمية: **المقدر الأسي ومقدر وايبل.**

وقد تم مقارنة نتائج دراسة المحاكاة بمعيار "**الكفاءة النسبية**".وقد أوضحت نتائج دراسة المحاكاة أن التوفيق المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) "**أفضل**" مقارنةً بالتوفيقات اللامعلمية، كما أن التوفيقات اللامعلمية باستخدام **"KM"** "**أفضل**" مقارنةً بالتوفيقات اللامعلمية باستخدام **"MWKM"**، كما أنه مع زيادة حجم العينة ونسبة الوقف "**لا تتغير**" كفاءة المقدر المعلمي من توزيع وايبل (التوزيع الخطأ)، بينما "**تتحسن**" كفاءة المقدرات اللامعلمية باستخدام **"KM"**، في حين "**تسوء**" كفاءة المقدرات اللامعلمية باستخدام **"MWKM"**.

الكلمات الاستفتاحية: البيانات غير الكاملة، الأخطار المتنافسة، دالة المأمونية، دراسة محاكاة، اللغة **"R"** البرمجية.

Abstract

This research aims at assessing a number of parametric and nonparametric estimators that are used in the estimation of competing risks models in the presence of incomplete data using a “simulation study”.

The nonparametric estimators used to estimate the reliability function in the presence of competing risks and right censored data in the simulation study are kaplan-meier estimator (KM) and modified weighted kaplan-meier estimator (MWKM). In addition, the parametric estimators are exponential estimator and weibull estimator.

The “relatively efficient”was used to compare the results of the simulation study.

The findings showed that the parametric estimator using weibull distribution (incorrect distribution) is better comparaed to the nonparametric estimators.Moreover, the nonparametric estimator using “KM” is better comparaed to the nonparametric estimator using “MWKM”. Additionally,the estimated parametric efficiency of the weibull distribution (incorrect distribution) isnot enhanced as the sample size and the censored ratio are increasing. Furthermore, the efficiency of the nonparametric esti-mators using “KM” is improved; however it dete-riorates when using “MWKM”.

Key words: incomplete data, competing risks, reliability function, simulation study, R program.

[1]- هذا البحث مشتق من رسالة ماجستير تعدها الباحثة بعنوان **"تحليل البيانات غير الكاملة في ظل وجود الأخطار المتنافسة: دراسة محاكاة"**، تحت إشراف كلاً من الأستاذ الدكتور/ محمد علي محمد أحمد، والدكتورة / لبيبة حسب النبي العطار، قسم الإحصاء والرياضة والتأمين - كلية التجارة - جامعة الإسكندرية.

١. مقدمة

ترجع الجذور التاريخية "**لتحليل الأخطار المتنافسة**" "Competing Risks Analysis" إلي القرن الثامن عشر، حيث أجري **Bernoulli (1766)** العديد من التجارب لمحاولة فصل حالات الوفاة (الحدث) من"**الجدري**" "Smallpox" (سبب الوفاة الرئيسي) عن حالات الوفاة من "**أسباب أخري متنافسة**" (**Ma and Krings, 2007**).

ويمكن تعريف "**الأخطار المتنافسة**" بأنها مجموعة العوامل المستقلة "Independent" التي تؤدي إلي وقوع الحدث "event" (الفشل أو الوفاة)، ويطلق علي هذه العوامل "**أخطاراً متنافسة**" "Competing Risks" قبل وقوع الحدث أما بعد وقوعه فإن العامل المتسبب في ذلك يطلق عليه "**سبباً**" "Cause" (**Deshpande and Purohit, 2001**). فعلي سبيل المثال، قد يواجه "طفل بلا مأوي" بعددٍ من الأخطار المتنافسة التي تؤدي إلي وفاته مثل الإدمان أو القتل أو حادثة سيارة أو أصابته بوباء نتيجة معيشته غير الصحية، ولكن في النهاية يتوفي هذا الطفل نتيجة سبب واحد فقط من هذه الأسباب.

ويُعد **تحليل الأخطار المتنافسة امتداداً لتحليل البقاء**[1] "Survival Analysis" وهذا راجع إلي تَمكن الملاحظ من مشاهدة وتسجيل الوقت التي تستغرقه المفردة إلي وقوع الحدث النهائي من أسرع (أول) سبب فقط، حيث تتعرض المفردة للعديد من الأخطار المتباينة المستقلة التي بدورها تؤدي لوقوع الحدث (**Beyersmann *et al.*, 2009**).

ويظهر "**الخطر المتنافس**" عندما يقع الفشل نتيجة "**سبب**" واحد فقط من بين العديد من الأسباب وهذا "**السبب**" يمنع وقوع الأسباب الأخري (**Klein, 2010**). بمعني أن "**الخطر المتنافس**" يظهر عندما لا يمكن مشاهدة الحدث محور اهتمام الدراسة نتيجة وقوع "**الحدث المانع**" "Preceding Event" والذي يطلق عليه "**الحدث المتنافس**" "Competing Event" (**Gondara, 2015**).

وتستخدم "**نماذج الأخطار المتنافسة**" "Competing Risks Models" علي نطاق واسع في "**الإحصاء الحيوي**" "Biostatistics"، و"**اقتصاديات الصحة التجريبية**" "Empirical Health Economics"، و"**اقتصاديات العمل**" "Labor Economics"، و"**العلوم الأجتماعية**" "Social Sciences"، وفي "**التطبيقات الهندسية**" (**Dewan and Naik-Nimbalkar, 2013; Effraimidis and Dahl, 2014**).

في كثير من "**تطبيقات تحليل البقاء ذات الأخطار المتنافسة**"، يصطدم الباحث بالعديد من الصعوبات التي تعوقه عن المضي قدماً في بحثه وهي: **إنتهاء وقت الدراسة، وانسحاب بعض المفردات من الدراسة، وفقد متابعة بعض المفردات، وعدم كفاءة الأساليب، وفقد المستندات، وضآلة الإمكانيات المادية المتاحة للباحث** (**Abd El Massieh *et al.*, 2013; Park and Padgett, 2004; Sarhan *et al.*, 2013**).

---

[1] **تحليل البقاء"** هو "مجموعة من الإجراءات الإحصائية لتحليل بيانات المتغير العشوائي محور اهتمام الدراسة وهو **"الوقت"** والذي يُقاس من نقطة محددة إلي وقت وقوع الحدث أو انتهاء فترة الدراسة". ووفقاً لهذا التحليل **تتعرض** المفردة للحدث من سبب واحد فقط . **(Kleinbaum and Klein, 2012)**

ونتيجة للصعوبات السابقة يكون لدينا، في نهاية أي دراسة:

أولاً: مشاهدات كاملة "Complete Obser-vations":- تحتوي بياناتها علي وقت وسبب وقوع الحدث. وبطبيعة الحال، لا يواجه الباحث أي صعوبات عند التحليل الإحصائي لهذه المشاهدات.

ثانياً: مشاهدات غير كاملة "Incomplete Observations":- لا تحتوي بياناتها علي أوقات و/أو أسباب وقوع الحدث. وبطبيعة الحال يواجه الباحث صعوبات عند التحليل الإحصائي لهذه المشاهدات. **وتنقسم المشاهدات غير الكاملة إلي:**

أ. **المشاهدات المراقبة "Censored Obs-ervations"**: تحتوي فقط علي "**معلومات جزئية**" "Partial Information" عن المتغير العشوائي الذي نهتم بدراسته (**Miller,1981**).

ب. **المشاهدات المفقودة "Missing Obse-rvations"**: هي مشاهدات ذات أوقات فشل (أو وفاة) معلومة وأسباب فشلها (أو وفاتها) غير معلومة أو العكس (**Andersen *et al.*, 1996**).

ج. **المشاهدات المحجوبة "Masking Obs-ervations"**: هي مشاهدات ذات أوقات فشل (أو وفاة) معلومة ولكن أسباب فشلها (أو وفاتها) غير معلومة، مع العلم بانتماء أسباب فشلها (أو وفاتها) إلي "**الفئة المحجوبة**" "**Masking Group**" (**Flehinger *et al.*, 1998**).

مشكلة البحث

أن استخدام "**المدخل اللامعلمي**" لتحليل البيانات غير الكاملة لا يكون دائماً "**أفضل**" من "**المدخل المعلمي**"، حيث يتوقف ذلك علي طبيعة البيانات التي نتعامل معها. علي الجانب الآخر، التوصيف الخاطيء للبيانات باختيار "**توزيع معلمي**" لا يلائم البيانات قد يؤدي إلي الحصول علي مقدرات لدالة المأمونية لا تتمتع بخصائص جيدة. بالإضافة إلي أن بعض الظواهر التي يختص بها تحليل المأمونية قد لا يكون من السهل التعرف علي توزيعاتها، وبالتالي لا يكون من السهل تقدير دالة المأمونية معلمياً.

ولذلك يمكننا تلخيص مشكلة البحث في التساؤل التالي:

ما هو المدخل الأفضل لتحليل البيانات غير الكاملة المراقبة من جهة اليمين في ظل وجود الأخطار المتنافسة ؟

هدف البحث

يهدف هذا البحث إلي الإجابة عن التساؤل السابق من خلال المقارنة بين **المدخل المعلمي واللامعلمي** لتحليل البيانات غير الكاملة المراقبة من جهة اليمين في حالة وجود الأخطار المتنافسة لاختيار أفضلهما اعتماداً علي "**دراسة المحاكاة**" "Simulation Study"، وفقاً لمعيار "**الكفاءة النسبية**" (RE) "Relatively Efficient"، مع الأخذ في الاعتبار بعض العوامل وهي: **حجم العينة، ونسبة الوقف (cr) "Censored Ratio" في العينة، وسبب وقوع الحدث.**

أهمية البحث

تتمثل أهمية البحث من "**الناحية النظرية**" في استعراض أهم المفاهيم والمصطلحات المتعلقة بالبيانات غير الكاملة. **بينما تتمثل أهمية البحث من "الناحية العملية" في الآتي:**

أ. دراسة الأخطار المتنافسة في ظل وجود البيانات غير الكاملة خاصة في "**المجالات الصناعية**" تساعد الباحث علي تحديد السبب الحقيقي (أحد مكونات النظام) لفشل النظام.

ب. دراسة الأخطار المتنافسة في ظل وجود البيانات غير الكاملة خاصة في "**البحوث الطبية**" تساعد علي تحديد السبب الحقيقي وراء وفاة المفردة.

ج. إعداد برنامج إحصائي مصمم باللغة البرمجية **"R"** يقوم بتوليد بيانات معرضة للفشل من سببين ومعرضة للوقف العشوائي من جهة اليمين، كما يقوم هذا البرنامج بتقدير دالة المأمونية باستخدام المقدرات المعلمية واللامعلمية وعقد مقارنة بينها.

٢. تحليل الأخطار المتنافسة

يتحدد "**نموذج الخطر المتنافس**" من خلال "**التوزيع المشترك**" "Joint Distribution" للمتغيرين **T**، **C** حيث يشير **T** إلي "**وقت الحدث**" بينما يشير **C** إلي "**سبب وقوع الحدث**". وبالتالي تتكون البيانات من وقت الحدث **T** ومؤشر يعرف سبب وقوع الحدث **δ**. ويتحدد "**التوزيع المشترك للمتغيرين C, T**" من خلال إحدي الدالتين الآتيتين: (أ) "**دالة الخطر محدد السبب**" "Cause-Specific Hazard Function" والتي تُعرف علي أنها معدل التغير اللحظي لوقوع الحدث من السبب **"j"** عند تغير الزمن بمقدار طفيف بشرط بقاء المفردة علي قيد الحياة حتي الزمن **t**، **(ب)** "**دالة التوزيع التراكمية محددة السبب**" "Cause-Specific Cumulative Distribution Function" والتي تُعرف علي أنها احتمال أن لا يتجاوز زمن حدوث الحدث من السبب **"j"** وقت معين (**Porta *et al.*, 2007**).

ويعتبر "**نموذج الخطر المتنافس**" "Competing Risk Model" حالة خاصة من "**نموذج متعدد الحالات**" "Multi – State Model" يحتوي علي حالة "**عابرة**" (انتقالية) "Transient" واحدة، وعدد **"J"** من الحالات "**الممتصة**" "Absorbing" المناظرة للوفاة من السبب **"j"**، حيث $1 \le j \le J$ (**Lin, 2011**).

٣. البيانات المراقبة "Censored Data"

يطلق علي البيانات أنها "**مراقبة**" إذا احتوت فقط علي معلومات جزئية عن المتغير العشوائي (وقت الفشل أو الوفاة) الذي نهتم بدراسته (**Gijbels, 2010; Miller, 1981**). وهناك العديد من **الأسباب** لظهور ما يُعرف "**بالبيانات المراقبة في ظل تحليل الأخطار المتنافسة**" وهي:

أ. **إنتهاء وقت الدراسة "Termination of Study"**: ويقصد به عدم كفاية المدة المحددة للدراسة لرصد وتسجيل وقت وقوع الحدث لجميع مفردات الدراسة.

ب. انسحاب بعض المفردات من الدراسة: ويقصد به ترك بعض المفردات الدراسة لعدم الرغبة في الاستمرار، أو للآثار السلبية لطرق العلاج المستخدمة في الدراسة.

ج. فقد متابعة بعض المفردات: ويقصد به عدم قدرة الباحث علي متابعة بعض المفردات خلال فترة الدراسة. فمثلاً، قد تُغير بعض المفردات

(الأشخاص) محل إقامتها، أو قد يسافر البعض الآخر، أو قد يتوقف البعض عن استكمال الدراسة.

وتُصنف "البيانات المراقبة" من حيث "آلية الوقوع" إلي ثلاثة أنواع هي:

أ. **البيانات المراقبة من جهة اليسار** "Left Censored Data": وهي بيانات المشاهدات التي "**حدث**" لها فشل أو وفاة "**قبل بداية الدراسة**".

ب. **البيانات المراقبة خلال فترة** "Interval Censored Data": وهي بيانات المشاهدات التي "**حدث**" لها فشل أو وفاة "**خلال فترة غير محددة من الدراسة**".

ج. **البيانات المراقبة من جهة اليمين** "Right Censored Data": وهي بيانات المشاهدات التي "**لم يحدث**" لها فشل أو وفاة "**أثناء فترة الدراسة**" (Gijbels, 2010).

بينما تُصنف "البيانات المراقبة" من حيث "النوع" إلي ثلاثة أنواع هي:

أ. **البيانات المراقبة من النوع الأول** "**Type-I Censoring**"

تُعد الأكثر شيوعاً، نظراً لأن إيقاف التجربة يتم بعد مرور زمن معين أو عند نقطة زمنية معينة. ولذلك تُوصف البيانات بأنها "**مراقبة من النوع الأول**" إذا كان الوقت الموقوف "**ثابتاً**"، وكان عدد المفردات المعرضة للحدث "**عشوائياً**" والذي يشار إليه بالرمز **"n"** (**Miller, 1981**).

ب. **البيانات المراقبة من النوع الثاني** "**Type-II Censoring**"

تُعد البيانات "**مراقبة من النوع الثاني**" إذا كان الوقت الموقوف "**عشوائياً**"، وكان عدد المفردات المعرضة للحدث (**n**) "**ثابتاً**". وبصفة عامة، فإن إيقاف التجربة يتم بعد فشل عدد معين من الوحدات والذي يشار إليه بالرمز **"r"** (**Miller, 1981**).

وينتشر استخدام البيانات المراقبة من النوعين الأول والثاني في "**التطبيقات الهندسية**"، حيث يمكن تحديد المفردات المراقبة من النوعين الأول والثاني، وحيث تدخل المفردات جميعها إلي الدراسة عند الوقت (**t** = 0).

ج. **البيانات المراقبة عشوائياً** "**Random Censoring**"

ينتشر استخدام البيانات المراقبة عشوائياً في "**المجالات الطبية**" حيث يكون من الصعب تحديد المفردات المراقبة مما يتطلب تعريف $\boldsymbol{\delta_i}$ (مؤشر الحدث) لكل مفردة كما أن المفردات تدخل إلي الدراسة عند أوقات مختلفة.

وتندرج البيانات المراقبة من النوعين الأول، والثاني، والمراقبة عشوائياً تحت "البيانات المراقبة من جهة اليمين"، نظراً لأن المتغير الذي نهتم به يستغرق وقتاً كبيراً لحدوثه، وبالتالي لا نستطيع مشاهدته أثناء فترة الدراسة. **وتُعد "البيانات المراقبة من اليمين"** من أكثر الأنواع شيوعاً، ولذلك سنقتصر في هذا البحث علي هذا النوع فقط (**Miller, 1981**).

٤. تحليل الأخطار المتنافسة في ظل وجود بيانات مراقبة من جهة اليمين

أولاً: دالة المأمونية "Reliability Function"

تُعرَّف "**دالة المأمونية**" بإنها احتمال أن يتعدي زمن حدوث الفشل قيمة معينةً (**Miller, 1981; Shafiq and Viertl, 2014**).

$$\boldsymbol{R(t) = S(t) = \Pr(\,T\, > t\,), \qquad t \geq 0.} \tag{1}$$

حيث يشير الرمز **T** إلي وقت وقوع الحدث، بينما يشير $R(t)$ أو $S(t)$ إلي **دالة المأمونية**.

وجدير بالذكر أن الاسم الذي يطلق علي "**دالة المأمونية**" قد **يختلف** باختلاف المجال التطبيقي محل الدراسة. فعلي سبيل المثال، يطلق علي هذه الدالة في "**مجال التطبيقات الهندسية**" اسم "**دالة المأمونية**" ويرمز لها بالرمز $R(t)$ في حين تسمي "**بدالة البقاء**" في "**المجال الطبي**" ويرمز لها بالرمز $S(t)$ (Miller, 1981; Zhao, 2008).

ولقد اهتمت العديد من الدراسات الإحصائية بتقدير "دالة المأمونية" نظراً لمايلي:

١. أنها **الأكثر** استخداماً لتحليل المتغير العشوائي **T**، كما أنها توفر معلومات عن احتمال وقوع الحدث (محور اهتمام الدراسة).

٢. أنها "**مكمل**" دالة التوزيع التراكمية[٢] "Cumulative Distribution Function" $\mathbf{F(t)}$، حيث:

$$\mathbf{S}(t) = \mathbf{1} - \mathbf{F(t)}. \tag{2}$$

٣. وجود **علاقة** بين دالة **المأمونية** $\mathbf{S}(t)$ ودالة **الخطر**[٣] "Hazard Function" $\mathbf{h}(t)$ ودالة **كثافة الاحتمال**[٤] "Probability Density Function" $\mathbf{f}(t)$، حيث:

$$\mathbf{h}(t) = \frac{\mathbf{f}(t)}{\mathbf{S}(t)}. \tag{3}$$

٤. وجود **علاقة** بين دالة **المأمونية** ودالة **كثافة الاحتمال**، حيث:

$$\frac{\mathbf{d}}{\mathbf{d}t}\mathbf{S}(t) = -\mathbf{f}(\mathbf{t}). \tag{4}$$

٥. وجود **علاقة** بين دالة **المأمونية** ودالة **الخطر** ودالة **الخطر التراكمية** "Cumulative Hazard Function" $\Lambda(t)$، حيث:

$$\mathbf{S}(t) = \mathbf{exp}\left\{-\int_{0}^{t} \mathbf{h}(u)\,du\right\} = \mathbf{exp}\{-\Lambda(t)\}\,. \tag{5}$$

ثانياً: تقدير دالة المأمونية "Reliability Function Estimation"

أ.المدخل المعلمي

يفترض "**المدخل المعلمي**" وجود شكل دالي معين لتوزيع المتغير العشوائي T . وهذا الشكل الدالي يستند إلي معلمة أو أكثر يجب تقديرها بإحدي طرق التقدير المعلمية. **ولاشك** أن أكثر التوزيعات المعلمية شيوعاً في الأدبيات الإحصائية والمتعلقة بتحليل البيانات المراقبة من جهة اليمين

[٢] هي احتمال أن لا يتجاوز زمن حدوث الفشل وقت معين.
[٣] هي معدل التغير اللحظي لوقوع الحدث عند تغير الزمن بمقدار طفيف بشرط بقاء المفردة علي قيد الحياة حتي الزمن **t**.
[٤] هي احتمال أن يكون الزمن اللازم لحدوث الفشل مساوياً t.

هي التوزيع الأسي "Exponential Distribution"، وتوزيع وايبل "Weibull Distribution"، نظراً لأهميتهم التاريخية وبساطتهم الحسابية وخصائصهم الهامة **(Kundu and Basu, 2000)**.

- **التوزيع الأسي**

يعتبر التوزيع الأسي من أكثر التوزيعات المعلمية شيوعاً في الأدبيات الإحصائية المتعلقة بتحليلي البقاء والمأمونية. **وتُعرف "دالة كثافة الاحتمال للتوزيع الأسي" علي النحو التالي:**

$$f(t) = \lambda[\exp(-\lambda t)], \quad \lambda > 0, t > 0. \quad (6)$$

حيث تشير λ إلي معلمة القياس "Scale Parameter".

وتكون "دالة التوزيع التراكمية للتوزيع الأسي" علي النحو التالي:

$$F(t) = 1 - exp[-\lambda t]. \quad (7)$$

بينما تُعرف "دالة المأمونية للتوزيع الأسي" علي النحو التالي:

$$S(t) = \exp[-\lambda t]. \quad (8)$$

ويتميز التوزيع الأسي بأنه التوزيع الوحيد الذي له "**دالة خطر ثابتة**" "Constant Risk Function"، حيث تُعرف "**دالة الخطر للتوزيع الأسي**" علي النحو التالي:

$$h(t) \equiv \lambda. \quad (9)$$

- **توزيع وايبل**

يعتبر توزيع وايبل "**تعميم**" للتوزيع الأسي، **وتكون "دالة كثافة الاحتمال لتوزيع وايبل" علي النحو التالي:**

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}[\exp(-\lambda t)^{\alpha}], \quad for\ t, \lambda, \alpha \geq 0. \quad (10)$$

حيث تشير λ إلي معلمة القياس، بينما تشير α إلي معلمة الشكل "Shape Parameter".

وتُعرف "دالة التوزيع التراكمية لتوزيع وايبل" علي النحو التالي:

$$F(t) = 1 - exp[-(\lambda t)^{\alpha}]. \quad (11)$$

بينما تُعرف "دالة المأمونية لتوزيع وايبل" علي النحو التالي:

$$S(t) = \exp[-(\lambda t)^{\alpha}]. \quad (12)$$

كما تُعرف "دالة الخطر لتوزيع وايبل" علي النحو التالي:

$$h(t) = \lambda\alpha(\lambda t)^{\alpha-1}. \quad (13)$$

وتتميز دالة الخطر لتوزيع وايبل بأنها "**متناقصة**" عندما تكون $(0 < \alpha < 1)$، و"**ثابتة**" عندما تكون $(\alpha = 1)$، "**ومتزايدة**" عندما تكون $(\alpha > 1)$.

وبناءً علي ما سبق، يُعتبر التوزيع الأسي "**حالة خاصة**" من توزيع وايبل وذلك عندما $(\alpha = 1)$.

ب. المدخل اللامعلمي

لا يفترض "**المدخل اللامعلمي**" وجود شكل دالي معين لتوزيع المتغير العشوائي T (وقت وقوع الحدث) **(Kundu and Basu,2000)**. وبالتالي يتم الحصول علي تقديرات **لامعلمية** لدالة المأمونية

مباشرة باستخدام أحدي الطرق اللامعلمية التالية: **مقدر "KM" (1958) التقليدي في ظل وجود الأخطار المتنافسة، ومقدر "KM" (2007) المرجح المعدل في ظل وجود الأخطار المتنافسة.**

طريقة "KM" في ظل وجود الأخطار المتنافسة:

يُعد كلاً من Kaplan and Meier (1958) أول من اقترحا طريقة **لامعلمية** لتوفيق "**دالة بقاء المجتمع**" في حالة وجود "**بيانات مراقبة وسبب وحيد لوقوع الحدث**". ويطلق علي هذه الطريقة **مقدر "KM"**، أو يطلق عليها مقدر "Product – limit" (P–L).

وقام Gooley *et al.* (1999) **بتعميم مقدر دالة المأمونية "KM" في ظل وجود الأخطار المتنافسة، علي النحو التالي:**

$$\hat{S}_{KM,j}(t) = \prod_{i:\, y_{(ij)} \le t} \left(\frac{n - R_i}{n - R_i + 1}\right)^{\delta_{ij}} = \prod_{i:\, y_{(ij)} \le t} \left(1 - \frac{d_{ij}}{n_i}\right)^{\delta_{ij}} \qquad (14)$$

حيث:

- يشير $\hat{S}_{KM,j}(t)$ إلي **مقدر "KM" لدالة المأمونية محددة السبب "j".**
- يشير الرمز $y_{(ij)}$ إلي وقت الحدث المشاهد المرتب المناظر للمفردة **i** من السبب **"j"**.
- يشير الرمز n إلي عدد المفردات المعرضة للحدث أو للانسحاب من الدراسة.
- يشير الرمز R_i إلي رتبة $y_{(ij)}$.
- يشير الرمز δ_{ij} إلي مؤشر الحدث المناظر للمفردة **i** من السبب **"j"**.
- يشير الرمز n_i إلي عدد المفردات المعرضة للحدث أو للانسحاب من الدراسة عند الوقت $y_{(i)}$.
- يشير الرمز d_{ij} إلي عدد حالات الحدث (الفشل أو الوفاة) من السبب **"j"** عند الوقت $y_{(ij)}$.

كذلك عرَّف Gooley *et al.* (1999) "مقدر دالة المأمونية الكلية[°]" "Overall Reliability Function Estimation" $\hat{S}_{KM}(t)$ **علي النحو التالي:**

$$\hat{S}_{KM}(t) = \prod_{j=1}^{J} \hat{S}_{KM,j}(t). \qquad (15)$$

ويكون مقدر تباين $\hat{S}_{KM}(t)$ كالتالي:

$$\widehat{Var}\left(\hat{S}_{KM}(t)\right) = [\hat{S}_{KM}(t)]^2 \sum_{i=1}^{n} \left(\frac{d_i}{n_i(n_i - d_i)}\right). \qquad (16)$$

حيث:

- يشير $\widehat{Var}\left(\hat{S}_{KM}(t)\right)$ إلي **مقدر تباين** $\hat{S}_{KM}(t)$.

يشير الرمز d_i إلي عدد حالات الحدث (الفشل أو الوفاة) من جميع الأسباب عند الوقت $y_{(i)}$.

طريقة Modified Weighted Kaplan–Meier (MWKM) في ظل وجود الأخطار المتنافسة:

° يقصد بالكلية تقدير دالة المأمونية من جميع الأسباب المسببة لوقوع الحدث.

اقترح Shafiq *et al.* (2007) "المقدر KM المرجح المعدل" (MWKM) والذي يعطي وزناً جديداً "غير صفري" لآخر مشاهدة إذا كانت "مراقبة" بحيث يصبح احتمال البقاء لهذه المشاهدة صغيراً جداً.

وقام Zare and Mahmoodi (2013) **بتعميم مقدر دالة المأمونية "KM" المرجح المعدل في ظل وجود الأخطار المتنافسة كالتالي:**

$$\hat{S}_{MWKM,j}(t) = \prod_{i:y_{(ij)} \le t} W_{ij}^{*} \left\{\frac{n_{ij} - d_{ij}}{n_{ij}}\right\}, \quad y_{(ij)} \le t < y_{(ij+1)}. \qquad (17)$$

حيث:

- يشير $\hat{S}_{MWKM,j}(t)$ إلي **مقدر "KM" المرجح المعدل محدد السبب "j".**
- يشير الرمز W_{ij}^{*} إلي الوزن المرجح المعدل للمفردة i من السبب "j"، ويأخذ ثلاث قيم (يأخذ قيمة لا تساوي "0" إذا كانت آخر مشاهدة مراقبة، بينما يأخذ القيمة "1" إذا كانت المشاهدة كاملة، وأخيراً يأخذ قيمة "**أقل من 1**" إذا كانت المشاهدة مراقبة عند الوقت ($y_{(ij)}$)، حيث:

$$W_{ij}^{*} = 1 - \mathrm{Sin}\left\{\frac{n_{V_{ij}} \times P_{ij}^{*}}{n_{ij}}\right\}. \qquad (18)$$

حيث يشير الرمز P_{ij}^{*} إلي احتمال البقاء للمفردة i من السبب "j".

$$P_{ij}^{*} = \left\{\frac{n_{ij} - d_{ij}}{n_{ij}}\right\}. \qquad (19)$$

ويكون "**مقدر دالة المأمونية الكلية**" $\widehat{S}_{MWKM}(t)$ **كالتالي:**

$$\widehat{S}_{MWKM}(t) = \prod_{j=1}^{J} \hat{S}_{MWKM,j}(t). \qquad (20)$$

ويكون مقدر تباين $\hat{S}_{MWKM}(t)$ **كالتالي:**

$$\widehat{Var}\left(\hat{S}_{MWKM}(t)\right) = \left[\hat{S}_{MWKM}(t)\right]^{2} \sum_{i=1}^{n} \left(\frac{d_i (W_i^{*})^2}{n_i (n_i - d_i)}\right). \qquad (21)$$

حيث:

- يشير $\widehat{Var}\left(\hat{S}_{MWKM}(t)\right)$ إلي **مقدر تباين** $\hat{S}_{MWKM}(t)$.
- يشير الرمز W_i^{*} إلي الوزن المرجح للمفردة i، ويأخذ ثلاث قيم (يأخذ قيمة لا تساوي "0" إذا كانت آخر مشاهدة مراقبة، بينما يأخذ القيمة "1" إذا كانت المشاهدة كاملة، وأخيراً يأخذ قيمة "**أقل من 1**" إذا كانت المشاهدة مراقبة عند الوقت ($y_{(i)}$)، حيث:

$$W_i^{*} = 1 - \mathrm{Sin}\left\{\frac{n_{V_i} \times P_i^{*}}{n_i}\right\}. \qquad (22)$$

حيث يشير الرمز P_i^{*} إلي احتمال البقاء للمفردة i من جميع الأسباب.

$$P_i^{*} = \left\{\frac{n_i - d_i}{n_i}\right\}. \qquad (23)$$

٥. دراسة محاكاة

يتم في "**دراسة المحاكاة**" توليد البيانات من التوزيع الأسي ذو معلمة القياس $(\lambda = 0.9)$، **وذلك في ظل ثلاثة افتراضات هي:**

١. وجود سببين لوقوع الفشل.

٢. استقلال وقت الفشل ووقت الوقف.

٣. ثبات المخاطر محددة السبب $\propto_{01}(t)$ (احتمال الانتقال من الحالة "0" إلي الحالة "1")، $\propto_{02}(t)$ (احتمال الانتقال من الحالة "0" إلي الحالة "2").

وقد تم تنفيذ "500" معاودة عند بعض قيم أوقات الفشل المشاهدة المناظرة لمستويات المأمونية "**المنخفضة**" (35%,20%)، "**المتوسطة**" (50%)، "**العالية**" (90%, 75%) وذلك عند عدة مستويات مختلفة لحجم العينة لمعرفة أي من الأساليب التي تم المقارنة بينها "**أفضل**" في حالة أحجام العينات "**الصغيرة**" (n = 25)، "**المتوسطة**" (n = 50)، "**الكبيرة**" (n = 100, 200, 250, 300). كذلك تم استخدام عدة نسب مختلفة للوقف "**الصغيرة**" (cr = 5%)، "**المتوسطة**" (cr = 10%)، "**الكبيرة**" (cr = 25%)، "**الكثيفة**" (cr = 50%).

ولتوليد البيانات تم أتباع الخطوات التالية:

أ. تحديد المخاطر محددة السبب= $\propto_{01}(t)$ ، 0.3 $\propto_{02}(t) = 0.6$.

ب. توليد المتغير العشوائي T (وقت الفشل) من التوزيع الأسي من كل المخاطر بمعلمة λ حيث: $(\lambda = \propto_{0.}(t) = 0.9)$.

ج. إجراء تجربة ذي الحدين "A Binomial Experiment" لوقت الفشل المولد، بحيث يكون احتمال وقوع الفشل هو $\left(\propto = \frac{\propto_{0j}(t)}{\propto_{01}(t)+\propto_{02}(t)}\ ,\ j = 1,2\right)$.

د. توليد المتغير العشوائي C (سبب الفشل) من توزيع ذي الحدين"Binomial Distribution"، بحيث يكون احتمال النجاح هو احتمال وقوع الفشل من السبب الأول. وبالتالي تكون قيمة هذا الاحتمال$\left(\propto = \frac{0.3}{0.3+0.6} = \frac{1}{3}\right)$.ولذا تكون قيمة احتمال عدم النجاح (احتمال وقوع الفشل من السبب الثاني) تساوي $\left(1-\propto = \frac{2}{3}\right)$.

هـ. توليد المتغير العشوائي V والذي يرمز لوقت الوقف من التوزيع المنتظم "Uniform Distribution" ذي المعلمتين $(a\,,b)$. وتجدر الإشارة هنا إلي أن قيمة المعلمة $(a = 0)$. أما بالنسبة للمعلمة b فقيمتها سوف تختلف تبعاً لنسبة الوقف، بحيث يتم تحديد قيمتها تبعاً للمعادلة $\left(b = \frac{\lambda \text{ متوسط التوزيع الأسي}}{\text{نسبة الوقف}}\right)$.

و. توليد المتغير العشوائي Y والذي يرمز لوقت الفشل المشاهد عن طريق مقارنة وقت الفشل T بوقت الوقف V.

٦. نتائج دراسة المحاكاة

أولاً: تأثير تغير حجم العينة علي تقديرات دالة المأمونية الكلية

النتائج الخاصة:

يتبين عند جميع نسب الوقف وجميع أحجام العينات مايلي:

أ. أن الكفاءة النسبية للمقدرات "**أفضل**" عند مستويات المأمونية "**المتوسطة والمنخفضة**" مقارنةً بمستويات المأمونية "**العالية**".

ب. أن التوفيق المعلمي باستخدام التوزيع الأسي (التوزيع الأصلي للبيانات) هو "**الأكثر كفاءة**" مقارنةً بالتوفيق اللامعلمي أو المعلمي باستخدام

توزيع وايبل (التوزيع الخطأ) وذلك عند "**جميع**" مستويات المأمونية.

ج. أن التوفيق المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) هو "**الأكثر كفاءة**" مقارنةً بالتوفيقات اللامعلمية وذلك عند "**جميع**" مستويات المأمونية.

د. التوفيقات اللامعلمية غير المرجحة **"KM"** "**أفضل**" من التوفيقات اللامعلمية المرجحة المعدلة **"MWKM"** وذلك عند "**معظم**" مستويات المأمونية.

هـ. لا يترتب علي زيادة حجم العينة أي "**تغير ملموس**" في قيمة الكفاءة النسبية للمقدرات اللامعلمية غير المرجحة **"KM"** وذلك عند "**جميع**" مستويات المأمونية.

النتائج العامة:

١. عند جميع أحجام العينات يكون:

أ. التوفيق المعلمي باستخدام التوزيع الأسي "**أكثر كفاءة**" من التوفيقات اللامعلمية والمعلمية من توزيع وايبل.

ب. التوفيق المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) "**أكثر كفاءة**" من التوفيقات اللامعلمية.

ج. التوفيقات اللامعلمية غير المرجحة **"KM"** "**أكثر كفاءة**" من التوفيقات اللامعلمية المرجحة المعدلة **"MWKM"**.

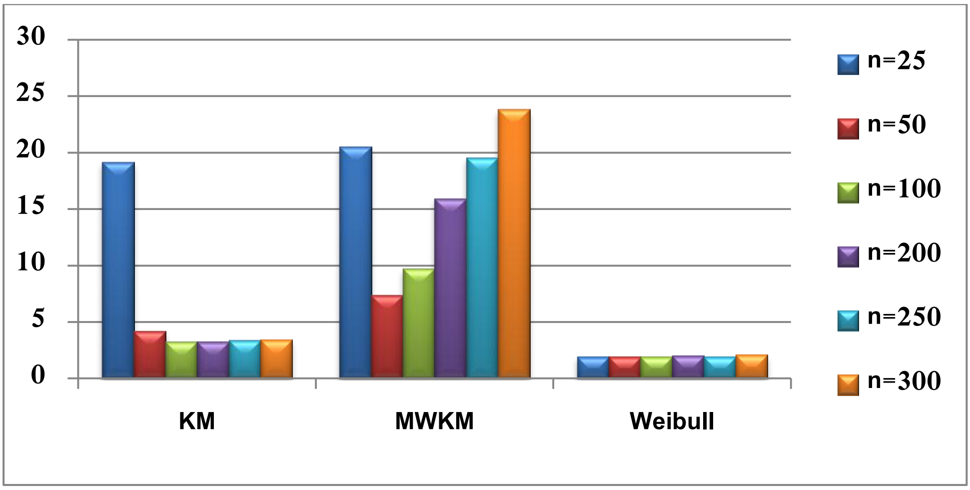
٢. يترتب علي زيادة حجم العينة مايلي:

أ. "**عدم تغير**" كفاءة المقدر المعلمي من توزيع وايبل (التوزيع الخطأ) "**تقريباً**".

ب. "**تحسن**" كفاءة المقدرات اللامعلمية غير المرجحة **"KM"**، بينما "**تسوء**" كفاءة المقدرات اللامعلمية المرجحة المعدلة **"MWKM"** (**جدول ١**).

جدول(١):متوسطات الكفاءة النسبية لتقديرات دالة المأمونية الكلية من بيانات التوزيع الأسي ($\lambda = 0.9$) عند أحجام مختلفة من العينات.

جحم العينة	KM	MWKM	Weibull
25	19.16986	20.54783	1.937199
50	4.201263	7.386242	1.929881
100	3.269486	9.730158	1.955943
200	3.275081	15.93567	2.022787
250	3.386651	19.57038	1.918029
300	3.405085	23.84294	2.106822



شكل (١): متوسطات الكفاءة النسبية لجميع مقدرات دالة المأمونية الكلية عند أحجام العينات المختلفة.

ثانياً: تأثير تغير نسبة الوقف علي تقديرات دالة المأمونية الكلية

النتائج الخاصة:

١. يتبين عند جميع أحجام العينات وجميع نسب الوقف مايلي:

أ. أن الكفاءة النسبية للمقدرات "**أفضل**" عند مستويات المأمونية "**المتوسطة والمنخفضة**" مقارنةً بمستويات المأمونية "**العالية**".

ب. التوفيق المعلمي باستخدام التوزيع الأسي (التوزيع الأصلي للبيانات) "**أكثر كفاءة**" مقارنةً بالتوفيق اللامعلمي أو المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) وذلك عند "**جميع**" مستويات المأمونية.

ج. التوفيق المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) "**أكثر كفاءة**" مقارنةً بالتوفيقات اللامعلمية وذلك عند "**جميع**" مستويات المأمونية.

٢. عند أحجام العينات "**الكبيرة**"، نلاحظ أنه مع زيادة نسبة الوقف عند جميع مستويات المأمونية "تسوء" توفيقات المقدرات اللامعلمية المرجحة المعدلة "MWKM".

النتائج العامة:

١. عند جميع نسب الوقف:

أ. أن التوفيق المعلمي باستخدام التوزيع الأسي هو "**الأكثر كفاءة**" من جميع التوفيقات اللامعلمية أو المعلمية من توزيع وايبل (التوزيع الخطأ).

ب. أن التوفيق المعلمي باستخدام توزيع وايبل (التوزيع الخطأ) هو "**الأفضل**" مقارنةً بالتوفيقات اللامعلمية.

ج. التوفيقات اللامعلمية غير المرجحة **"KM"** "**أفضل**" مقارنةً بالتوفيقات اللامعلمية المرجحة المعدلة **"MWKM"**.

٢. مع زيادة نسبة الوقف:

أ. لا تتغير "**تقريباً**" قيمة الكفاءة النسبية للمقدر المعلمي من توزيع وايبل (التوزيع الخطأ)، مما

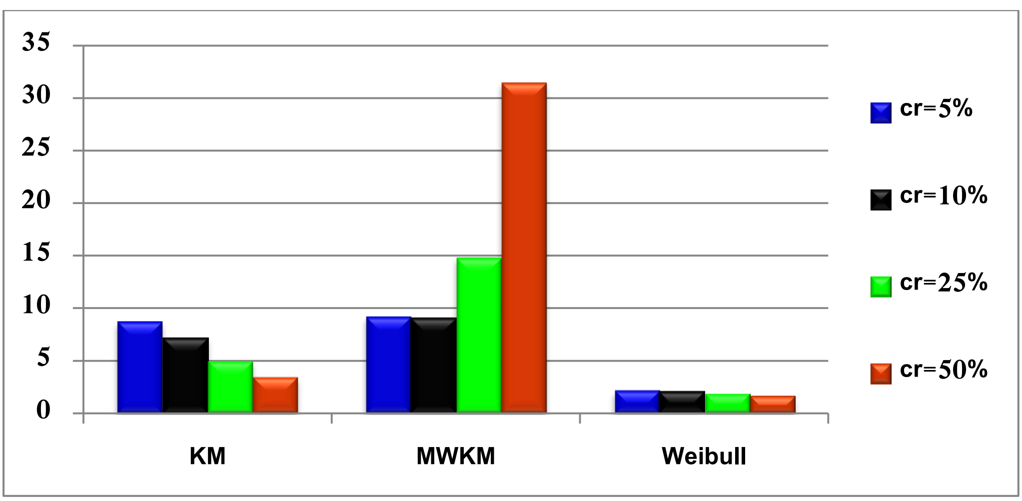
يؤدي إلي "**عدم وجود تأثير ملموس**" علي نتائج هذا المقدر المعلمي من توزيع وايبل (التوزيع الخطأ).

ب. "**اقتراب**" قيمة الكفاءة النسبية للمقدرات اللامعلمية غير المرجحة **"KM"** من الواحد الصحيح، مما أدي إلي "**تحسن**" نتائج هذه المقدرات.

ج. "**ابتعاد**" قيمة الكفاءة النسبية للمقدرات اللامعلمية المرجحة المعدلة **"MWKM"** عن الواحد الصحيح، مما أدي إلي"**سوء**" نتائج هذه المقدرات.

جدول (٢):متوسطات الكفاءة النسبية لتقديرات دالة المأمونية الكلية من بيانات التوزيع الأسي عند نسب مختلفة من الوقف. $(\lambda = 0.9)$

نسبة الوقف	KM	MWKM	Weibull
5%	8.755201	9.220051	2.210502
10%	7.206386	9.10264	2.110243
25%	5.028455	14.84541	1.900423
50%	3.481575	31.50738	1.692606



شكل (٢): متوسطات الكفاءة النسبية لجميع مقدرات دالة المأمونية الكلية عند نسب الوقف المختلفة.

قائمة المراجع

1- Abd El Massieh, P. M., Aboul Ezz, E. H. A. and Mousa, A. M. (2013), "Competing Risks Survival Analysis Applied to Ora-Dental Cancer Patients," ***48th The Annual Conference on Statistics, Computer Science and Operation Research***, PP. 1 – 16.

2- Anderson, J., Goetghebeur, E., and Ryan, L. (1996), "Missing Cause of Death Information in the Analysis of Survival Data," ***Statistics in Medicine***, Vol. **15**, PP. 2191-2201.

3- Bernoulli, D. (1766), "An Attempt at a New Analysis of the Mortality Caused by Smallpox and of the Advantages of Inoculation to Prevent it," ***Mem. Math. Phys. Acad. Roy. Sci. Paris, 1-45***. (Reprinted in 1971 in Reviews in Medical Virology, Vol. **14**, No. 5, 275-288).

4- Beyersmann, J., Latouche, A. Buchholz, A., and Schumacher, M. (2009), "Simulating Competing Risks Data in Survival Analysis," ***Statistics in Medicine***, Vol. **28**, PP. 956 - 971.

5- Deshpande, J. V. and Purohit, S. G. (2001), "Survival, hazard and competing risks," ***Current Science***, Vol. **80**, No. 9, PP.1191-1202.

6- Dewan, I. and Naik-Nimbalkar, U. V. (2013), "Statistical Analysis of Competing Risks with Missing Causes of Failure," Available@ ***http://isiproceedings.org/Files/STS009-P2-S.pdf.***

7- Effraimidis, G. and Dahl, C. M. (2014), "Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure," ***Statistics and Probability Letters***, Vol. **89**, PP.1-7.

8- Flehinger, B. J., Reiser, B. and Yaschin, E. (1998), "Survival with Competing Risks and Masked Ca-uses of Failures," ***Biometrika Trust***, Vol. **85**, No. 1, PP. 151- 164.

9- Gijbels, I. (2010), "Censored Data," ***John Wiley and Sons***, Vol. **2**, PP. 178-188.

10- Gondara, L. (2015), "Competing Risk Survival Analysis Using SAS When, Why and How," Available@ ***https://support.sas.com/resources/papers. pdf.***

11- Gooley,T. A.,Leisenring,W., Crowley, J. and Storer, B. E. (1999), "Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators," ***Statistics in Medicine***, Vol. **18**, PP. 695- 706.

12- Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," ***Journal of the American Statistical Association***, Vol. **53**, No. 282, PP. 457-481.

13- Klein, J. (2010), "Competing Risks," ***Wiley Interdisciplinary Reviews: Computational Statistics***, Vol. **2**, No. 3, PP. 333-339.

14- Kleinbaum, D. G. and Klein, M. (2012), Survival Analysis: A Self Learning Text (3rd edition). ***New York: Springer***.

15- Kundu, D. and Basu, S. (2000), "Analysis of Incomplete Data in Presence of Competing Risks," ***Journal of Statistical Planning and Inference***, Vol. **87**, PP. 221- 239.

16- Lin, Y. (2011), "Parametric Estimation in Competing Risks and Multistate Models," Available@ ***http:l//***

uknowledge.uky.edu/statistics_etds/1.

17- Ma, Z. S. and Krings, A. W. (2007), "Competing Risks Analysis of Reliability, Survivability, and Prognostics and Health Management (PHM)," ***In Aerospace Conference, LEEE***, PP. 1-21.

18- Miller, R. G. (1981), Survival Analysis. ***Biostatistics Casebook, John Wiley & Sons***.

19- Park, C. and Padgett, W. J. (2004), "Analysis of Strength Distributions of Multi-Modal Failures Using the EM Algorithm," ***Journal of Statistical Computation and Simulation***, Vol. **76**, No. 07, PP. 619-636.

20- Porta, N., Gómez, G., Calle, M. L., and Malats, N. (2007), "Competing Risks Methods". Available @ ***http: // upcommons.upc.edu/e-print/bitstream//2117/2201/1/TR_CR.pdf.***

21- Sarhan, A. M., Alameri, M., and Al-Wasel, I. (2013), "Analysis of A Competing Risks Model with Generalized Weibull Distributions," ***Pakistan Journal of Statistics***, Vol. **29**, No. 3, PP. 271-281.

22- Shafiq, M., Shah, S., and Alamgir. (2007), "Modified Weighted Kaplan-Meier Estimator," ***Pakistan Journal of Statistics Operation Research***, Vol. **3**, No. 1, PP. 39-44.

23- Shafiq, M., and Viertl, R. (2014), "Empirical Reliability Functions Based on Fuzzy Life Time Data," ***Journal of Intelligent and Fuzzy Systems***, Vol. **28**, No. 2, PP.707 – 711.

24- Zare, A. and Mahmoodi, M. (2013), "Modified Kaplan-Meier Estimator Based on Competing Risks for Heavy Censoring Data," ***International Journal of Statistics in Medical Research***, Vol. **2**, No. 4, PP. 297-304.

25- Zhao, G. M. A. (2008), "Nonparametric and Parametric Survival Analysis of Censored Data with Possible Violation of Method Assumptions," ***Master of Arts, Greensboro, North Carolina***.