-

أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap واختبار "ت" ذو التباين الممزوج : دراسة محاكاة

د/ محسوب عبد القادر الضوى

أستاذ مساعد علم النفس التربوى

كلية التربية بقنا – جامعة جنوب الوادى

ملخص الدراسة

هدفت الدراسة إلى فحص أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة الإحصائية لإجراء Bootstrap واختبار "ت" ذو التباين الممزوج تحت الشروط التالية :

١. أحجام العينات : استخدمت أربع أزواج من العينات [(٢٣ ، ١٦) ؛ (٦٨ ، ٥١) ؛ (١٣٤ ، ١١٩) ؛ (٢٨٩ ، ٢٦٠)] لتمثل العينات الصغيرة والمتوسطة والكبيرة والكبيرة جداً على الترتيب ، وبلغ عدد العينات المولدة ست وخمسون عينة .

٢. التوزيع Distribution : مجتمعى الأصل موزعين توزيعاً اعتدالياً .

٣. عملية Bootstrapping : تم توليد (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap .

٤. التباينات Variances : استخدمت أزواج التباينات [(١ ، ١)] ؛ [(٤ ، ١) ؛ (١ ، ٤)] ؛ [(٩ ، ١) ؛ (١ ، ٩)] ؛ [(٢٥ ، ١) ؛ (١ ، ٢٥)].

وتوصلت الدراسة الحالية إلى النتائج الآتية :

- يتأثر تقدير الخطأ من النوع الأول α_e لاختبار "ت" ذو التباين الممزوج بعدم تجانس التباين مقارنة بمستوى الدلالة الإسمى α_n عندما تشتق العينات الأصغر من المجتمعات الأكثر تبايناً .
- يتميز إجراء Bootstrap بالقوة الإحصائية مثل اختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة والكبيرة جداً ، لكنه لا يؤدى بنفس القوة فى حالة العينات الصغيرة والمتوسطة .
- بزيادة أزواج أحجام العينات تزداد القوة الإحصائية لإجراء Bootstrap واختبار "ت" ذو التباين الممزوج.
- يسيطر اختبار "ت" ذو التباين الممزوج وإجراء Bootstrap على تقديرات الخطأ من النوع الأول بشكل جيد فى حالة العينات الكبيرة والكبيرة جداً ، لكن تقديرات الخطأ من النوع الأول تتضخم بشكل طفيف فى حالة العينات الصغيرة والمتوسطة .
- لا توجد حاجة إلى استخدام عينات Bootstrap أكبر من ١٠٠٠ .

وأخيراً أوصت الدراسة الحالية بأنه على الباحثين فحص افتراض تجانس التباين عند محاولتهم مقارنة الفروق بين متوسطى مجموعتين ، كما أوصت بأن إجراء Bootstrap بديل مناسب لاختبار "ت" ذو التباين الممزوج فى غياب افتراض تجانس التباين .

أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap واختبار "ت" ذو التباين الممزوج : دراسة محاكاة

د/ محسوب عبد القادر الضوى
أستاذ مساعد علم النفس التربوى
كلية التربية بقنا – جامعة جنوب الوادى

مقدمة الدراسة

من مؤشرات جودة البحث فى المجال السيكولوجى – أو أى مجال بحثى آخر – قدرة الباحث على عمل معاينات صحيحة خلال تجربته واختيار الطريقة الإحصائية المناسبة لمعالجة البيانات التى يتم جمعها بقصد اختبار صحة الفروض الموضوعة ، فالمعاينة Sampling خطوة حرجة فى البحث ، إذ تعتمد الاستنتاجات الإحصائية بشكل أساسى على الطريقة التى يتم بها اختيار عينة البحث .

والتجربة الجيدة هى التى تماشى الطريقة الإحصائية وتوجهها وتقدم لها البيانات العددية التى تشكل المادة التى تعمل بها الطريقة الإحصائية ، ولا تقل أهمية الطريقة الإحصائية هى الأخرى فى توجيه التجربة ، وما تقدمه من بيانات عددية يساعد الإحصاء على تلخيص البيانات العددية ، وتصنيفها ، وتحديد نتائجها ودلالاتها وذلك انطلاقاً من مجموعة متكاملة من افتراضات الإحصاء الاستدلالى (ميخائيل أسعد ، ١٩٩٠ : ١٤٣) .

ويعتمد الإحصاء الاستدلالى فى جوهره على عملية المعاينة ، واستخدام إحصاء العينات يعتمد على توافر افتراضات معينة ، وإذا لم تتوافر فإن الخطأ المعيارى قد يؤدى إلى نتائج مضللة ، أو فى أحسن الأحوال يعطى تقديرات يمكن منها اتخاذ قرارات واستنتاج نتائج دون يقين كامل ، وإنما بدرجات متفاوتة من هذا اليقين يمتد من الشك الكبير إلى اليقين الكبير (فؤاد أبو حطب وآمال صادق ، ١٩٩٦ : ٣٠٨) .

والمشكلة المتكررة فى الإحصاء التطبيقى هى المقارنة بين متوسطى مجتمعين ، وعند محاولة تحرى ووصف الفروق بين مجموعتين مستقلتين ، يُعد استخلاص الاستنتاجات والاستدلالات حول الفروق بين المجموعات حدثاً شبه يومى فى الحياة البحثية للباحث التربوى أو السيكولوجى ، فمقارنة المجموعات هو صميم الأسئلة البحثية التى يتناولها هؤلاء الباحثون ، والاستراتيجية تعتمد إلى حد بعيد على استخدام المتوسط الحسابى كمقياس للموضع Measure of Location واختبار "ت" t-Test كطريقة لاختبار وجود الفروق ، حيث يُعد المتوسط الحسابى من أكثر مقاييس الموضع

استخداماً بصورة متكررة فى مجال العلوم التربوية والسلوكية ، ويلاحظ أن الفروض المتصلة بالمتوسطات هى الأكثر استخداماً (Scheffe, 1970: 1501; Cohen, 1988: 19; Wilcox, 2002: 169; Ruxton, 2006; Ozdemir, 2013: 322; Rusticus & Lovato, 2011; 2014: 1; Kang, Harring, & Li, 2014: 1-2) .

فاختبار "ت" هو الاختبار البارامترى الكلاسيكى الذى يشيع استخدامه لاختبار الفرض المعتاد : تساوى بارامترى المتوسط لمجتمعين $\mu_1 = \mu_2$ ، لهذا يستخدم اختبار "ت" لمجموعتين مستقلتين بشكل متكرر عندما يرغب الباحثون فى عمل استدلالات عن مجتمعين مستقلين من خلال مقارنة متوسطى عينتين مشتقتين من المجتمعين (Kulkarni, 1993: 20; Hinkle, Wiersma, & Jurs, 2003: 238) .

وأكدت دراستا عبد الناصر السيد عامر (٢٠١٢) ؛ عبد العاطى أحمد الصياد وعبد الناصر السيد عامر (٢٠١٣ : ٧) أن التصميم البحثى الأكثر انتشاراً فى البحوث النفسية والتربوية هو الذى يتضمن مقارنة بين متوسطين على متغير تابع ، وقد بلغت نسبة استخدام اختبار "ت" فى عينة من الدراسات والبحوث العربية المنشورة فى الفترة (٢٠٠٠ – ٢٠١١) ٢٤،٧% وفى البيئة المصرية ٣٨،٢% . بينما أفادت دراسة Ruxton (2006) بعد مراجعة مائة وثلاثون بحثاً منشوراً فى دورية علم البيئة السلوكى Behavioral Ecology أن اختبار "ت" استخدم فى سبعة وستون موقفاً خلال ستة وعشرون دراسة بنسبة (١٩،٥٥ %) .

لكن الإجراءات الكلاسيكية لمقارنة مجموعتين مثل اختبار "ت" تكون عادة مقيدة بافتراضى الاعتدالية Normality وتجانس التباينات Homogeneity of Variance ، وعلى مر السنين ، قدمت العديد من الإجراءات لمعالجة انتهاك هذه الافتراضات ، مع ملاحظة أن الإجراءات اللابارامترية Nonparametric Procedures هى بدائل قابلة للتطبيق يمكن استخدامها عندما يكون التوزيع غير اعتدالى (Ahad, Abdullah, Heng, & Ali, 2012: 43)

ولافتراض تجانس التباين أهمية خاصة لأنه يوفر الأساس المنطقى لجمع مربعات الانحرافات للمجموعتين معاً لتشكيل تقدير ممزوج مشترك لتباين المجتمع ، ويُعد التباين الممزوج Pooled Variance تقديراً أكثر استقراراً لتباين المجتمع لأن خطأ المعاينة يميل إلى أن يكون صغيراً للتقدير الممزوج عما لو أخذت قيمة منفردة لكل عينة على حدة (Kulkarani, 1993: 3) .

ومع زيادة القدرة الحسابية لأجهزة الحاسب الآلى ، تتحسن الأساليب الإحصائية باستمرار ، ومن أشهر الأساليب الإحصائية الحديثة تلك المبنية على إعادة المعاينة Resampling ، ونتج عن

ذلك مجموعة من الإجراءات أو الطرق منها : اختبار العشوائية المحدد Randomization Exact Test ، والصدق التقاطعى Cross-Validation ، وطريقة Jackknife ، وإجراء Bootstrap (Akpanta & Okorie, 2015: 441-443) .

وقد حاز الاستدلال الإحصائى المبنى على إعادة معاينة البيانات Data Resampling على قدر كبير من الاهتمام فى السنوات الأخيرة مقارنة بالإحصاء الكلاسيكى الذى ينتمى له اختبار "ت"* والفكرة الأساسية حول هذه الطرق أنها لا تفترض الكثير عن توزيع المجتمع ، وبدلاً من ذلك تحاول الحصول على معلومات حول المجتمع من البيانات نفسها (Reddy, Boiroju, Yerukala, & Rao, 2011: 185) .

وأصبحت طرق إعادة المعاينة قابلة للتطبيق العملى أخذاً فى الاعتبار توافر الحاسبات رخيصة الثمن والحزم الإحصائية الجديدة . وهذه الطرق الحديثة أكثر بساطة ودقة مقارنة بالطرق المعيارية للاستدلال الإحصائى ، وتتطلب عدداً أقل من الافتراضات ولها إمكانية كبيرة للتعميم . وتوفر مزايا واضحة عندما لا تُستوفى افتراضات الاختبارات البارامترية التقليدية مع العينات الصغيرة المشتقة من توزيعات غير اعتدالية . بالإضافة إلى أن إعادة المعاينة تستطيع التعامل مع أسئلة لا يمكن الإجابة عنها باستخدام الطرق البارامترية واللابارامترية التقليدية مثل المقارنة بين النسب والوسائط الحسابية (Berger, 2007) .

ويُعد إجراء Bootstrap واحداً من الطرق المبنية على إعادة المعاينة ، ويستخدم لعمل أنواع معينة من الاستدلالات الإحصائية ، وجوهره فكرة مؤداها أنه فى غياب أى معرفة بالمجتمع ، يكون توزيع القيم فى عينة عشوائية حجمها n من المجتمع هو أفضل دليل للتوزيع فى المجتمع . وبالتالى يمكن استخدامه لاشتقاق تقديرات ذات منعة Robust Estimates للأخطاء المعيارية وفترات الثقة لإحصاءات مثل : المتوسط ، والوسيط ، والنسبة Proportion ، ونسبة الأرجحية Odds Ratio ، ومعامل الارتباط ، ومعاملات الانحدار . (Manly, 1997: 34; Wooldridge, 2013: 24) .

ولا يعتمد هذا الإجراء على توزيع المعاينة النظرى مثل نظرية النهاية المركزية Central Limits Theorem التى تتطلب عينات كبيرة الحجم كما فى حالة الاختبارات الكلاسيكية ومنها اختبار "ت" (Ahad et al., 2012: 43-44) .

---

* يُسمى بالنظرى لأنه يتم عمله بالاستعانة بمبادىء الاحتمالات وليس بالتجريب العملى ، وجميعها تشترك فى صفة واحدة ، وهى كونها نظرية تتحدد خصائصها من القياسات على عينة واحدة (أحمد سليمان عودة وخليل يوسف الخليلى ، ١٩٨٨ : ١٨٠) .

وقد بين Othman, Keselman, Padmanabhan, Wilcox, and Fradette (2003) المميزات العملية لاستخدام إجراء Bootstrap : أنه لا يتطلب المعرفة بتوزيع المعاينة للاختبار الإحصائى ، ولا يتطلب تقديرات للأخطاء المعيارية للمقدرات Estimators كالمتوسطات والتباينات ومعاملات الارتباط ، وهذا الشرط يجعل اختبار الفرضيات يتم بشكل مرن جداً .

كما قارن Krishnamoorthy, Lu, and Mathew (2007) بين إجراء Bootstrap واختبار "ف" والبديلين : Welch Test, James Test ، واقترح استخدام إجراء Bootstrap لأنه الإجراء الوحيد الذى كان أداؤه مرضياً بصرف النظر عن حجم العينة ، وقيم تباينات الخطأ ، وعدد المتوسطات التى تتم مقارنتها تحت شرط عدم تجانس التباينات .

وفى نفس السياق ، قام Higgins (2005) بتبنى وتبسيط عملية Bootstrapping ، من خلال استخدام مولدات أرقام عشوائية بطريقة Monte Carlo لتوليد عينات Bootstrap .بالإضافة إلى تنفيذ اختبار الفرضيات باستخدام الاختبار الإحصائى لتوزيع معاينة غير معلوم ، وبين أنه يمكن استخدام إجراء Bootstrap فى تقييم أداء الاختبار الإحصائى من حيث الخطأ من النوع الأول والقوة . علاوة على ذلك ، فإنه يُيسر أيضاً من إجراء تحليل الحساسية Sensitivity Analyses لأداء طرق بارامترية معروفة وطرق لابارامترية تحت شروط تجريبية معتادة وأخرى متطرفة .

ويسعى كل من إجراء Bootstrap والاستدلال البارامترى التقليدى لتحقيق الهدف نفسه باستخدام معلومات محدودة لتقدير توزيع المعاينة للمقدر المختار Chosen Estimator ، ويستخدم التقدير لعمل استدلالات حول بارامتر المجتمع ، والفرق الرئيس بين هذين النهجين الاستدلاليين هو كيفية الحصول على توزيع المعاينة حيث يستخدم الاستدلال البارامترى التقليدى افتراضات مسبقة حول شكل توزيع المقدر ، أما إجراء Bootstrap فهو إجراء متحرر من التوزيع Distribution–Free وهذا يعنى أنه لا يعتمد على فئة معينة من التوزيعات . وباستخدامه يتم تقدير توزيع المعاينة على أساس أن توزيع العينة هو تقدير جيد لتوزيع المجتمع (Reddy et al., 2011: 185) .

وتأتى الدراسة الحالية لتحليل أداء إجراء Bootstrap مقارنة باختبار "ت" ذو التباين الممزوج من حيث تقديرات الخطأ من النوع الأول وخصائص القوة الإحصائية فى توليفات مختلفة من شروط انتهاك/ عدم انتهاك افتراض تجانس التباين .

مشكلة الدراسة

يُعد اختبار "ت" t–Test الذى قدمه العالم الإنجليزى William Sealy Goset واحداً من أشهر الاختبارات الإحصائية الاستدلالية التقليدية ، وأكثرها استخداماً بجانب اختبار "ف" الذى قدمه

العلامة Fisher ، ويستخدم لاختبار الفروض الفارقة بين متوسطين سواء كانا مستقلين أم مرتبطين مقيداً بمجموعة من الافتراضات الصارمة على البيانات التى يتم جمعها.

ولدى الباحث سبيلان لاختبار الفروق بين متوسطين . يقوم الأول على تكرار التجربة على عينتين تخضع كل منهما لواحد من مستويات المتغير التجريبى ، حيث يخلص الباحث إلى انعدام أثر المتغير المستقل إذا نزلت نسبة زيادة المتوسط على المتوسط الآخر عن ٩٥ % ، لكن يبقى السؤال حول العدد المناسب لتكرار التجربة . لذلك ، يضطر الباحثون إلى سلوك السبيل الثانى ، وهو سبيل الاستدلال الإحصائى الذى يُمكن الباحث من تحديد الفرق بين المتوسطين من تجربة واحدة فقط (ميخائل أسعد ، ١٩٩٠ : ١٨٥) .

ويقوم الهدف الأساسى للقياس النفسى فى تحديد الصفة المدروسة لدى المجتمع الإحصائى الذى يحمل تلك الصفة بدرجة ما . وإن استطاع الباحث تحديد كم الصفة فى المجتمع كانت أحكامه قاطعة نهائية وثابتة بصدد متوسط الصفة وتباينها ، غير أنه من الصعب ، بل من المستحيل بلوغ مجتمع ما لتحديد الصفة المعينة . والمألوف أن يتناول الباحث عينة من المجتمع يدرسها ويحسب إحصاءاتها (Weinberg & Goldberg, 1990: 158) .

صحيح أنه بالإمكان جعل العينة عشوائية ، أى ممثلة للمجتمع محل الدراسة ، لكن العينة مهما بلغ حجم أفرادها ، ومهما احتيط لجعلها ممثلة للمجتمع المدروس ، تبقى عينة واحدة من أصل احتياطى كبير من عينات أخرى محتملة ، لكل منها متوسطها وتباينها . وفى مقدور الباحث إصدار أحكام تتعلق بكم الصفة المدروسة انطلاقاً من عينة ما ، لكن أحكامه تبقى احتمالية أى عرضة لدرجة ما من درجات الخطأ ، وعليه أن يقدر مدى الخطأ المحتمل لتصحيح الحكم (ميخائل أسعد ، ١٩٩٠ : ١٦٤-١٦٥) .

ومتى ما تم اختيار عينة ، يجب افتراض أن إحصاءات العينة (مثلاً المتوسط $\bar{X}$) لا تطابق تماماً بارامترات المجتمع (مثلاً المتوسط μ) إن أمكن قياسه ، وأى افتراض آخر يعد ضرباً من المجازفة ، ويكون خطأ المعاينة Sampling Error هو الفرق بين القياسين $(\bar{X} - \mu)$ ، والأمر الطبيعى هو توقع انحراف متوسط العينة عن متوسط المجتمع ، وخطأ المعاينة ليس خطأ فى حد ذاته ، ويجب أن يكون عشوائياً (Sprinthall, 1990: 118) .

ويذكر Maggio and Sawilowsky (2014) أن عملية اختيار الاختبار الإحصائى يمكن أن تكون معقدة وغامضة وفى بعض الأحيان مخيبة للآمال ، فعملية الاختيار يجب أن تخضع لاعتبارات مثل خصائص المنعة فيما يتصل بتقديرات الخطأ من النوع الأول لانحرافات اعتدالية

المجتمع والقوة الإحصائية . وعندما يتم انتهاك الافتراضات البارامترية يمكن البحث عن اختبار أو أكثر يتميز بدرجة أعلى من القوة تحت مجموعة معلومة من الشروط ، وبهذا يكون الاختيار غالباً مشوب بالحدس أو التخمين .

لذا ظلت وستظل مشكلة اختيار الاختبار المناسب قائمة لدى قطاع كبير من الباحثين وترتبط بشكل ما بضعف مستوى المهارات المكتسبة خلال برامج الإعداد فى مرحلة الدراسات العليا ، وتظهر الآثار اللاحقة لذلك فى استخدام اختبارات إحصائية دون أدنى اعتبار لانتهاك الافتراضات الأساسية التى تستند إليها .

وبعامة يعتمد كل اختبار للاستدلال الإحصائى ومنها بالطبع اختبار "ت" على مجموعة أساسية من الافتراضات ، عندما يتم استيفائها فإن الاختبار سوف يوظف كما هو مستهدف منه ومُعد له ، وعندما يتم انتهاك الافتراضات فإن الاختبار ربما يكون مضلل (Keselman et al., 1998).

وتتصف البيانات الحقيقية فى مجال علم النفس بثلاث خصائص هى : الإلتواء ، وغياب تجانس التباين **Heteroscedasticity** ، والدرجات المتطرفة **Outliers** ، وجميعها تؤثر على أداء اختبار "ت" والطرق الاستدلالية المعيارية الأخرى مثل اختبارى تحليل التباين وتحليل الانحدار من حيث القوة واحتمالية ارتكاب أخطاء فى اتخاذ القرارات الإحصائية . وكل خاصية من تلك الخصائص بصرف النظر عن الخاصيتين الأخريين يمكن أن تقلل إلى حد كبير فرص : (أ) تحرى الفروق الحقيقة بين المجموعات ، (ب) تحرى الارتباطات الحقيقية بين المتغيرات العشوائية ، (جـ) الحصول على فترات ثقة دقيقة لمعالم المجتمع المستهدف (Micceri, 1989; Wilcox, 1990, 2012; Harwell, 1988; Harwell & Serlin, 2001; Yuan & Hayashi, 2003: 93; Ahad et al., 2012: 43; Ozdemir, 2013: 322-323; Berge, 2015) .

وتشكل هذه الخصائص الثلاثة مجتمعة مصدر قلق خطير جداً ، فاختبار "ت" فى الواقع تحت الشروط العامة ليس تقاربياً بشكل صحيح **Asymptotically Correct** . ومنذ ثمانون عاماً وحتى الآن وجدت أدبيات مكثفة بشأن تأثيرات انتهاك افتراض تجانس التباين وعدم الاعتدالية (Boneau, 1960; Blair & Higgins, 1985; Zumbo & Jennings, 2002; Lumley, Diehr, Emerson, & Chen, 2002; Fradette, Keselman, Lix, Algina, & Wilcox, 2003; Wilcox, 1990, 2012; Ozdemir, 2013: 322) .

وعلى مدار السنوات ظهر مجموعة من الإجراءات لمعالجة انتهاك تجانس التباين تحت مُسمى إجراءات منيعة لاختبار الفروض **Robust Hypothesis Testing Procedures** مثل :

إجراء (James, 1951) ، وإجراء (Welch, 1951) ، وطريقة (Brown & Forsythe, 1974) ، وإجراء (Alexander & Govern, 1994) ، وطريقة (Odemir & Kurt, 2006) ، وطريقة (Keselman, et al., 2008) (In: Ahad et al., 2012: 43; Ozdemir, 2013: 323) .

وقد أدت هذه الطرق إلى تحسن السيطرة على احتمالية الخطأ من النوع الأول ، ولكن ظلت المشكلة كما هى ، فأى طريقة تعتمد على المتوسط يمكن أن تكون لها قوة نسبية منخفضة . كذلك أى انتهاك لافتراضات الاختبار الإحصائى البارامترى يُفسد توزيع الاختبار ويغير تقديرات الخطأ من النوع الأول والخطأ من النوع الثانى (Bradley, 1968: 25; Ozdemir, 2013: 322-323) .

والأمر المهم أن نسبة لا بأس بها من الباحثين فى المجال التربوى والنفسى يستخدمون اختبار "ت" بشكل روتينى دون فحص لافتراضاته ، ومعممين النتائج من العينة للمجتمع اعتماداً على مستويات الدلالة الإحصائية التى تقترب من ٠،٠٥ ، ويتفاخرون بنتائجهم إذا ما كان مستوى الدلالة الإحصائية يصل إلى ٠،٠١ أو ٠،٠٠١ ، وبهذا أصبحت هذه القيم مقدسة كعتبة حدية للقيم التى تحدد الدلالة الإحصائية .

ومن ناحية ثانية حاز العديد من الطرق الإحصائية الحديثة على إهتمام الباحثين السيكولوجيين والتربويين مثل : تحليل البيانات الاستكشافى ، وتصور البيانات ، والإجراءات المنيعة ، والطرق المبنية على إعادة المعاينة Resampling Methods ، ومع ذلك يميل العديد من الباحثين إلى تبنى الطرق الإحصائية التقليدية بدلاً من تجريب هذه الطرق الجديدة كممارسة محافظة (Efron, 1979a, b; Efron & Tibshirani, 1993: 13; Reddy et al., 2011: 185; Yu, 2003: 1) .

وتسهم ثلاثة عوامل فى هذه الممارسة المحافظة : الأول ، أن الطرق الجديدة غير متضمنة فى المقررات الإحصائية التى يدرسها الطلاب ، ونتيجة لذلك ، فإن المفاهيم المرتبطة بالطرق الجديدة تبدو غامضة للعديد من الباحثين ، والثانى ، أن معظم مطورى الحزم الإحصائية كرسوا جهدهم لتحليل البيانات باستخدام الاختبارات التقليدية وحتى لو كان الباحثون على علم بهذه الطرق الجديدة فإن الإتاحة المحدودة للحزم يعيقهم من استخدامها ، والثالث ، أنه بالرغم من الوعى بهذه المفاهيم وإتاحة الحزم الإحصائية يُنظر إلى الإجراءات التقليدية على أنها تقوم على تبرير نظرى متين (Yu, 2003: 1) .

يعتمد اختبار "ت" على مفهوم العينة العشوائية كسبيل لتعميم النتائج من العينة على المجتمع ، وهى معاينة بدون إحلال Sampling Without Replacement ، أما الطرق الحديثة

فتستخدم المعاينة مع الإحلال Sampling With Replacement فكل عنصر فى المجتمع يكون متاح للاختيار ضمن العينة فى كل مرحلة من مراحل اختيار أو إنتقاء العينة بغض النظر عما إذا كان قد تم اختياره من قبل ، ونتيجة للمعاينة مع الاحلال فإن العنصر الواحد يمكن إنتقاؤه أكثر من مرة (Weinberg & Goldberg, 1990: 240) .

ويذكر (2007) Berger أن أساليب إعادة المعاينة ومنها إجراء Bootstrap دخلت سريعاً إلى حقل تحليل البيانات ، ويعتقد بعض الإحصائيين أن إجراءات إعادة المعاينة سوف تحل فى القريب العاجل محل الإجراءات اللابارامترية ، وربما تحل محل معظم الإجراءات البارامترية أيضاً وقد بدأت البحوث فى مجال علم النفس استخدام إجراء Bootstrap ، ودعا أنصار هذا الإجراء إلى استخدامه بشكل خاص على عينات تتكون من (٢٠–٨٠) حالة ، وقد استجابت دورية علم النفس التطبيقى Journal of Applied Psychology لهذه الدعوة حيث استخدم الباحثون بشكل متزايد إجراء Bootstrap (Koopman, Howe, Hollenbeck, & Sin, 2015: 194)

وقد تباينت نتائج الدراسات والبحوث السابقة – التى أتيح للباحث الحالى الاطلاع عليها– والتى تناولت المقارنة بين اختبار "ت" وإجراء Bootstrap فيما يتصل بأداء الخطأ من النوع الأول والقوة الإحصائية والمنعة وحدود الثقة مثل : (Kulkarni, 1993; Lansing, 1999; Yin, Yusof, Yaacob, Othman, 2010; Ahad et al., 2012) ، ويعد ذلك المبرر الرئيس لإجراء الدراسة الحالية .

وعلى حد معرفة الباحث لم يستخدم إجراء Bootstrap فى دراسة عربية أو مصرية فى مجال التربية وعلم النفس كاستجابة سياقية فى إطار مواكبة التطورات المتلاحقة فى منحى معالجة البيانات القائم على طرق إحصائية حديثة مبنية على إعادة المعاينة Resampling ، وهذا هو المبرر الاضافى لإجراء الدراسة الحالية .

لذا تسعى الدراسة الحالية إلى الإجابة عن الأسئلة الآتية :

١. ما أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات الصغيرة ؟
٢. ما أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات المتوسطة ؟

٣. ما أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة ؟

٤. ما أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة جداً ؟

هدف الدراسة

هدفت الدراسة الحالية إلى تقصى أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات (الصغيرة ، المتوسطة ، الكبيرة ، الكبيرة جداً) .

أهمية الدراسة

أولاً : الأهمية النظرية

تكمن الأهمية النظرية للدراسة الحالية فى التالى :

١. تعرض لمفهوم إحصائى مهم هو إعادة المعاينة ، الذى يمثل فكرة جديدة حول التحليل الإحصائى الذى يختلف عن الإحصاء التقليدى (Schieber, 2013) .

٢. ارتباط موضوع الدراسة باستخلاص وتفسير النتائج التى تستخدم اختبار "ت" لدلالة الفروق بين متوسطين مستقلين .

٣. تعرض لبعض إجراءات إعادة المعاينة كبدائل للاختبارات الإحصائية الاستدلالية التقليدية معتادة الاستخدام مع التركيز على طريقة Bootstrap كأسلوب إحصائى غير معتاد فى البحوث والدراسات العربية والمصرية فى مجال التربية وعلم النفس ، حيث أوصت الدراسات بمزيد من البحوث لتحقيق فهم أفضل لإجراء Bootstrap (Lansing, 1999) .

ثانياً : الأهمية التطبيقية

تكمن الأهمية التطبيقية للدراسة الحالية فى التالى :

١. تزود بتفاصيل منهجية عن أثر مستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات (الصغيرة ، المتوسطة ، الكبيرة ، الكبيرة جداً) .

٢. توجيه انتباه الباحثين إلى الآثار اللاحقة لانتهاك أحد الافتراضات الأساسية التى يستند إليها الاختبار الإحصائى "ت" الأكثر استخداماً فى الدراسات والبحوث العربية والمصرية

حدود الدراسة

اقتصرت الدراسة الحالية فى حدودها على التالى :

١. عدد العينات المولدة عشوائياً التى ستتم دراستها ست وخمسون عينة .

٢. جميع المجتمعات التى ستسحب منها عينات الدراسة موزعة توزيعاً اعتدالياً .

٣. مقارنة أداء إجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة[®] بأداء اختبار "ت" ذو التباين الممزوج Pooled Variance t–Test فى حالة العينات (الصغيرة ، المتوسطة ، الكبيرة ، الكبيرة جداً) من حيث تقديرات الخطأ من النوع الأول والقوة تحت مستويات متباينة من أزواج التباينات .

٤. النقطة المرجعية Benchmark Point للقوة الإحصائية تساوى ٠،٨ .

٥. القيمة الحرجة لاختبار Kohr كمؤشر لدلالة الفروق بين مستوى الدلالة الإسمى ومستوى الدلالة الحقيقى (≥ 1.96) .

مصطلحات الدراسة

تجانس التباين Homogeneity of Variance

تباين المشاهدات فى المجتمع الأول لا يختلف عن تباين المشاهدات فى المجتمع الثانى (Best & Kahn, 2006: 416) .

الخطأ من النوع الأول A Type I Error

احتمال الرفض الخاطىء لفرض صفرى صحيح (Finch, Thompson, & Cumming, 2002) . وتتحدد قيمته التقديرية بمستوى الدلالة الحقيقى (α_a) Actual Segnificance Level (Best & Kahn, 2006: 410) .

قوة الاختبار الإحصائى Power of a Statistical Test

قدرة الاختبار الإحصائى على رفض الفرض الصفرى عندما يكون فى حقيقة الأمر خاطئاً ، أى احتمالية تجنب الخطأ من النوع الثانى وتتحدد قيمته التقديرية بالاحتمال المُكمل للخطأ من النوع الثانى $(p = 1 - \beta)$ (Pagano, 2010: 244) .

إجراء Bootstrap

يعرفه Schieber (2013) بأنه طريقة إحصائية لتوليد توزيع المعاينة لإحصاء عن طريق المعاينة مع الإحلال من عينة البيانات الأصلية . ويعرفه Strube (1988) بأنه طريقة تقوم على

[®] تضع بعض الحزم الإحصائية ومنها SPSS عدد التكرارات ١٠٠٠ عينة كخيار مفترض Default Option .

إعادة المعاينة المتتابعة Successive Resampling ، وتبعاً لها يتم توليد توزيع المعاينة لإحصاء ما خلال معاينات متتابعة من فئة بيانات مشاهدة .

المجتمع Population

يعرفه فؤاد أبو حطب وآمال صادق (١٩٩٦ : ٧٧ ؛ ٣٠٧) بأنه الكل أو الجميع الذى يتم التعميم إليه ، ويعرف أحمد سليمان عودة وخليل يوسف الخليلى (١٩٨٨ : ١٥ ؛ ١٧١) المجتمع الإحصائى بأنه أى مجموعة كلية محددة أو تجمع مُعرف من الأشياء أو الأشخاص أو الحوادث ، وهو المجموعة الشاملة التى يجرى اختيار العينات منها .

العينة Sample

يعرفها فؤاد أبو حطب وآمال صادق (١٩٩٦ : ٧٧) بأنها جزء من كل أو بعض من جميع ، ويعرفها أحمد سليمان عودة وخليل يوسف الخليلى (١٩٨٨ : ١٥ ؛ ١٧١) بأنها أى مجموعة جزئية من المجموعة الكلية أو المجتمع الإحصائى يتم جمع البيانات من خلالها بصورة مباشرة .

الاطار النظرى للدراسة

المعاينة والاستدلال الإحصائى Sampling and Statistical Inference

يقترن علم الاستدلال الإحصائى بعجز الباحث عن قياس الصفة فى المجتمع الإحصائى ، وتنعدم الحاجة إلى هذا العلم ، إن استطاع الباحث بلوغ كل المجتمع . لكن الباحث – أى باحث – يعجز عن قياس المجتمع أو بلوغه ، وسيبقى عمله مقتصراً على العينات ، وستبقى الحاجة إلى علم الاستدلال الإحصائى قائمة (ميخائل أسعد ، ١٩٩٠ : ١٦٥) .

ويوجد افتراضان أساسيان لمفهوم العينة حتى يمكن استخدام إحصاء العينة على نحو مقبول هما : الأول افتراض التمثيل Representation ، ويُقصد به أن تكون العينة ممثلة لجميع الوحدات التى يتألف منها الأصل ، وعادة ما يهتم الباحث اهتماماً كبيراً بحجم العينة أكثر من اهتمامه بمدى تمثيلها للأصل ، على الرغم من أن حجم العينة ليس محكاً كافياً للحكم على صلاحيتها للتعميم على الأصل . والثانى افتراض المصادفة Chance ، ويُقصد به أن اختيار العينة يتحدد بعدد كبير من العوامل المستقلة المعقدة التى لا يستطيع الباحث التحكم فيها أو توجيهها ، ويتيح هذا التعقد والتعدد فى عوامل الاختيار فرصاً متكافئة متساوية للوحدات التى يتألف منها الأصل فى أن تكون موضع الاختيار ، وهذا الافتراض يتضمن فى جوهره مفهوم العشوائية (فؤاد أبو حطب وآمال صادق ، ١٩٩٦ : ٧٧-٧٩) .

ويتقاسم العديد من المتخصصين Laypersons سوء الفهم بأن العينة الملائمة يجب أن

تكون صورة بالكربون أو لها خصائص مطابقة للمجتمع محل الدراسة ، إن ميزة الاختيار العشوائى تكمن فى أن إحصاءات العينة ستكون تقدير غير متحيز لبارامترات المجتمع ، لأن تباينات متوسطات العينة العشوائية معلومة ، ومن الممكن تقدير هذه التباينات على أساس الاحتمال (Best & Kahn, 2006: 402-403) .

والخطأ الذى لا يتحكم فيه الباحث يُعد حقيقة فى المعاينات العشوائية ، وانتقاء عينات عشوائية لا يضمن أنها سوف تكون ممثلة للمجتمع ، ولا توجد عينة سوف يكون تكوينها مطابقاً مطابقة دقيقة لتكوين المجتمع ، وإذا تم انتقاؤها بعناية وحجمها كبير بدرجة كافية ، فإنها ينبغى أن تمثل المجتمع بدرجة تقريبية (ل. ر. جاى وج. إ. ميلز وب. إيراسيان ، ٢٠١٢ ، ٢١٤-٢١٥) .

ويعزى التباين فى متوسطات العينة لما يعرف بخطأ المعاينة Sampling Error ، وهذا المصطلح لا يعنى أى خطأ فى عملية المعاينة Sampling Process لكنه يصف فقط تباينات الصدفة مستحيلة الحدوث عند حساب متوسطات عدد من العينات التى تم اختيارها عشوائياً . فتقدير بارامترات المجتمع أو الاستدلال عنها من إحصاءات عينة عشوائية ليست عملية محكمة ، حيث لوحظ أن المتوسطات المتعاقبة Successive Means لعينات مشتقة عشوائياً من المجتمع نفسه ليست متطابقة ، وعندئذ من المنطقى افتراض أن أى واحداً منها من المحتمل أن يختلف عن متوسط المجتمع ، وهذا بدوره يمثل معضلة للإحصائيين عند استخدام عينة واحدة كأساس للتعميم للمجتمع (Bartz, 1988: 240-241; Best & Kahn, 2006: 402) .

وتتضمن دائماً القرارات الإحصائية المبنية على دليل مشاهد فى عينة احتمالية الخطأ ، فالباحثون لا يتعاملون مع القرارات المبنية على الحتمية ، هم فقط يقدرون احتمالية أو عدم احتمالية وقوع الأحداث . والغرض من الإحصاء الاستدلالى هو عمل استدلالات بشأن النواتج بناء على عينة (Best & Kahn, 2006: 409) .

وعند استخدام الإحصاء الاستدلالى لعمل قرارات رفض أو قبول الفروض الصفرية ، فإنه توجد أربعة تجمعات أو توليفات ممكنة من النواتج (Bartz, 1988: 262) :

١. يقرر الباحث ، بناء على النتيجة الإحصائية (بارامتر الأصل ليس مساوياً لإحصاءة العينة) ، أن يرفض الفرض الصفرى عندما يكون خاطئاً (قرار صحيح) ، ويُعبر عن ذلك بقوة الاختبار الإحصائى .
٢. يقرر الباحث ، بناء على النتيجة الإحصائية (بارامتر الأصل مساوياً لإحصاءة العينة) ، أن يقبل الفرض الصفرى عندما يكون صحيحاً (قرار صحيح) .

٣. يقرر الباحث ، بناء على النتيجة الإحصائية (بارامتر الأصل مساوياً لإحصاءة العينة أى العينة مشتقة من هذا الأصل) ، أن يرفض الفرض الصفرى عندما يكون صحيحاً (قرار خاطىء) ، ويُعبر عن ذلك بالخطأ من النوع الأول .

٤. يقرر الباحث ، بناء على النتيجة الإحصائية (بارامتر الأصل ليس مساوياً لإحصاءة العينة أى العينة مشتقة من أصل مختلف) ، أن يقبل الفرض الصفرى عندما يكون خاطئاً (قرار خاطىء) ، ويُعبر عن ذلك بالخطأ من النوع الثانى .

ولهذا يحدث ، لسبب أو لآخر ، أن يكون الاستدلال الإحصائى المتعلق بالفرض المطروح خاطئاً . والخطأ نوعان : (١) الخطأ من النوع الثانى وهو قبول الفرض الخاطىء إنطلاقاً من منطقة ثقة مرتفعة أى (٠،٩٥) ومن مستوى دلالة منخفض أى (٠،٠٥) ، (٢) الخطأ من النوع الأول وهو رفض الفرض الصحيح إنطلاقاً من منطقة ثقة منخفضة دون (٠،٩٥) ومن مستوى دلالة مرتفع أكبر من (٠،٠٥) (ميخائيل أسعد ، ١٩٩٠ : ١٩٩) .

ويمكن تخفيض احتمال الخطأ من النوع الأول بضبط أو تخفيض مستوى الدلالة الإحصائية إلى مستوى أكثر تشدداً يكون عادة (٠،٠٠١) ، والمشكلة أن ذلك سوف يزيد احتمال الخطأ من النوع الثانى ، وهذا يعنى أن الخطأ من النوع الثانى يرتبط عكسياً بالخطأ من النوع الأول ، أى أن تجنب أحد الخطأين سوف يزيد فرصة ارتكاب الخطأ الآخر (Aron & Aron, 1994: 198-199; Best & Kahn, 2006: 410) .

والطريقة الفضلى لتقليل حجمى الخطأ من النوع الأول والخطأ من النوع الثانى α_a , β تتمثل فى زيادة حجم العينة ، حيث كلما زاد حجم العينة قلت قيمة الخطأ المعيارى للمتوسطات العينية ، وبالتالى زاد احتمال كون العينة ممثلة لمجتمعها ، إذا ما تم اختيارها حسب الأصول العلمية الخاصة بذلك (عبد الرحمن عدس ، ١٩٩٧ : ٣٥) .

ولا توجد طريقة سهلة لتحديد ما إذا كان الفرق المحسوب أو المشاهد خطأ معاينة أو فرق حقيقى ، لكن يمكن السؤال عما إذا كنا نفضل الخطأ من النوع الأول أو الخطأ من النوع الثانى ، وتتحدد المخاطرة بقبول أى نوع من الخطأ حسب موضوع الدراسة (Bartz, 1988: 263) .

وبهذا لا يمكن قبول أو رفض فرض صفرى بثقة تامة ، ولكن يمكن فقط بيان إن كان محتملاً أو غير محتمل ، إلى حد كبير ، وهناك احتمالان جرى العُرف على استخدامهما فى العلوم السلوكية وهما $(\alpha = .05)$ ، $(\alpha = .01)$ لقيم اختبار "ت" . وإذا رُفض الفرض الصفرى عندما تكون $(\alpha = .01)$ أو بأقل منها فإن الباحث لا يخاطر إلا قليلاً إذا رفض الفرض ولا يحتمل أن

يخطىء كثيراً فى هذا الرفض ، وعلى الأمد الطويل يقع الباحث فى خطأ رفض فرض بغير حق فيما لا يزيد عن (١%) من المرات وهذا معيارٌ على درجة كبيرة من التشدد (ج. ملتون سميث ، ١٩٧٨ : ٩٣) .

لذا ينبغى عند اتخاذ قرار إحصائى ، تذكر التمييز بين الدلالة الإحصائية ، وأهمية النتائج المتحصل عليها ، وإذا تبين من اختبار "ت" وجود نتائج ذات دلالة فى تجربة ما ، فلا يعنى هذا بالضرورة أن النتائج مهمة . فالأهمية كثيراً ما تعتمد على حجم الفرق المحتمل بين متوسطى المجتمع $(\mu_1 - \mu_2)$ ضمن أشياء أخرى . وجزء من عملية اتخاذ القرار ، تحديد ما الفرق الأدنى بين متوسطى المجتمع الذى سيعتبر مهماً (Weinberg & Goldberg, 1990: 291) .

إعادة المعاينة Resampling

يُعد بارامتر المجتمع رقم ثابت ، وللمتغير العشوائى توزيع معين أو كثافة احتمالية تصف احتمالية حدوث القيم المختلفة أثناء المعاينة العشوائية ، ويشار إلى دالة الكثافة **Density Function** بتوزيع المعاينة لإحصاءات عينة مثل المتوسط . ولكل إحصاءات العينة توزيعات معاينة ، وتُعد الافتراضات الخاصة بشكل توزيعات المعاينة من صميم الإجراءات المعيارية لاختبار الفروض أو اختبار الدلالة الإحصائية (Dwyer, 1983: 35) .

ولما كان تباين المجتمع ثابتاً فى كل الحالات ، وأن تباينات العينات المأخوذة منه تختلف باختلاف حجم العينة ونوعية العناصر الداخلة فيها ، فإنه يتوقع اختلافاً بين التوزيعات العينية لكل من هذين الإحصاءين ، ويرجع هذا الاختلاف فى جوهره إلى أن حوالى نصف العينات الممكن أخذها من مجتمع ما ينتظر أن تكون تبايناتها أكبر قيمة من تباين المجتمع ، بينما يحتمل أن يكون النصف الآخر أقل قيمة منه ، وهذا يدل على أن قيم تباينات العينات المختلفة من حجم معين ينتظر أن تتناثر حول تباين مجتمع الأصل (عبد الرحمن عدس ، ١٩٩٧ : ٥٤) .

وتقارن الاختبارات البارامترية الكلاسيكية ومنها اختبار "ت" الإحصاءات الملاحظة بتوزيعات معاينة نظرية ، وتُعد إعادة المعاينة منهجية ثورية لأنها تنحرف عن التوزيعات النظرية ، وبالأحرى يعتمد الاستدلال على معاينة متكررة داخل نفس العينة ، ولذلك تُسمى هذه المنهجية بإعادة المعاينة (Yu, 2003: 2) .

ومن مشكلات الإحصاء التقليدى أن معظم الحالات يتم فيها قبول افتراضات الإحصاء التقليدى كما لو كانت مستوفاة ، بالاضافة إلى أنه لا يمكن تطبيقها لبعض الإحصاءات مثل الوسيط والمنوال والمدى والنسبة ، فحين أن عمومية أساليب إعادة المعاينة تخلص من الصيغ الإحصائية المعقدة

. (Schieber, 2013)

وطريقة إعادة المعاينة ترتبط بالمحاكاة المسماة Monte Carlo وفيها يكون الباحثون بيانات ويشتقون استنتاجات اعتماداً على العديد من السيناريوهات المحتملة (Lunneborg, 2000)

الأساس المنطقى لدعم إعادة المعاينة Rationale of Supporting Resampling

أثار داعمو إعادة المعاينة مجموعة من الأسباب لتبرير استخدام أساليبها منها :

١. الامبريقى : تعتمد الإجراءات الكلاسيكية ومن أشهرها اختبار "ت" على التوزيعات النظرية التى تتطلب افتراضات قوية لكل من العينة والمجتمع . لكن الانتقال الاستدلالى المفاجىء من العينة إلى المجتمع ربما يكون مشكلاً وبخاصة إذا كان المجتمع معرفاً بطريقة سيئة ، وعند الشك فى جدوى استخدام التوزيعات النظرية فإن إعادة المعاينة المبنية على بيانات امبريقية بديلاً جيداً (Diaconis & Efron, 1983; Peterson, 1991)

٢. الوضوح : من الناحية المفاهيمية يُعد مفهوم إعادة المعاينة بسيطاً وواضحاً ، ولا يتطلب خلفية رياضية معقدة لاستيعاب إعادة المعاينة (Rudner & Shafer, 1992) .

٣. التوزيع : تتطلب الإجراءات الكلاسيكية افتراضات توزيعية ، والتى تستوفى عادة بالعينة كبيرة الحجم . وعندما يكون حجم العينة صغيراً ولا يتطابق مع الافتراضات البارامترية ، فإنه يُوصى بإعادة المعاينة كعلاج (Diaconis & Efron, 1983) .

٤. العينة غير العشوائية : تتطلب الإجراءات الكلاسيكية معاينة عشوائية لتحقيق صدق الاستدلال من العينة إلى المجتمع . وأكد Edgington (1995) أن إعادة المعاينة تتسم بالصدق لأى نوع من البيانات بما فيها البيانات العشوائية وغير العشوائية . واقترح Lunneborg (2000) أنه رغم استخدام العينات غير العشوائية فى إعادة المعاينة ربما لا يؤدى إلى استنتاجات استدلالية ، وعلى الأقل فإن المعاينة الفرعية للعينات غير العشوائية يمكن أن تعطى المزيد عن الوصف الموضعى Local Description للبيانات وإستقرار النتائج .

٥. العينة صغيرة الحجم : حتى لو استوفت بنية البيانات الافتراضات البارامترية ، فإن دراسة العينة صغيرة الحجم سيتعثر بواسطة مستوى القوة المنخفض ، وعملية Bootstrapping يمكن أن تعالج عينة صغيرة كمجتمع كبير لتوليد مزيد من المشاهدات (Yu, 2003: 12) .

٦. العينة كبيرة الحجم : تعالج إعادة المعاينة عادة العينة صغيرة الحجم ، ومع ذلك يمكن تطبيق الأسلوب نفسه أيضاً فى الموقف الذى يكون حجم العينة كبيراً (Yu, 2003: 12)

٧. الإعادات (المكررات) Replications : لا تخبر الإجراءات الكلاسيكية الباحثين بمقدار احتمال أن النتائج يمكن تكرارها (Thompson & Synder, 1997) . فالمعاينة العشوائية البسيطة

هى مثال لما يُسمى المعاينة بدون إحلال أو استبدال فبمجرد انتقاء عنصر من المجتمع ليكون ضمن عينة فإنه يتم حذفه من الاعتبار فى بقية المراحل المتبقية من عملية الاختيار (Weinberg & Goldberg, 1990: 240) .

أنواع إعادة المعاينة Types of Resampling

من طرق إعادة المعاينة شائعة الاستخدام لأغراض متباينة ما يلى :

أولاً : اختبار العشوائية المحدد Randomization Exact Test

والذى يُسمى أيضاً باختبار التباديل Permutation Test وطوره Fisher (1935/1960) ، وهذا الاختبار يحتاج إلى حسابات فى غاية التعقيد وبرمجة رفيعة المستوى (Yu, 2003: 2) . ويستخدم إعادة المعاينة بدون إحلال لاختبار الفرضية : لا يوجد تأثير No Effect (Berger, 2007) .

ثانياً : طريقة الصدق التقاطعى Cross-Validation

حيث اقترح Kurtz (1948) طريقة الصدق التقاطعى البسيطة Simple Cross-Validation كعلاج لاختبار الرورشاخ أحد أشهر اختبارات الشخصية الذى انتقده المتخصصون فى القياس النفسى لافتقاره للخصائص السيكومترية الشائعة مثل اعتدالية البيانات ، واعتماداً على هذه الطريقة طور Mosier (1951) طريقة الصدق التقاطعى المزدوجة Double Cross-Validation التى تم تمديدها فيما بعد إلى طريقة الصدق متعدد التقاطع Multicross-Validation عن طريق Krus and Fuller (1982) (In : Yu, 2003: 2)

والطريقة البسيطة هى الأسهل للتطبيق خلال عملية من ثلاث خطوات (Palomares, 1990: 7) :

١. يُقسم الباحث العينة الأصلية إلى مجموعتين فرعيتين منفصلتين بشكل عشوائى .
٢. يُجرى الباحث التحليل نفسه على كلا المجموعتين الفرعيتين بشكل منفرد .
٣. يقارن الباحث امبريقياً النتائج ، محاولاً أن يظهر إعادة النتائج لكلا المجموعتين الفرعيتين ، وهذا يزيد الثقة فى إعادة نتائج الدراسة .

ثالثاً : طريقة Jackknife

تعرف أيضاً بطريقة Quenouille-Tukey Jackknife وقد ابتكرها Quenouille (1949, 1956) وطورها فيما بعد Tukey (1958) ، وتختلف عن طريقة Bootstrap فى أن المجموعات الفرعية المختلفة أو مجموعات المفحوصين تُسحب بصورة متكررة من فئة البيانات الأصلية . ويتم حساب الإحصاءات موضع الاهتمام لكل فئة بيانات مستقطعة ثم يتم أخذ المتوسط للنتائج . ويتم

تنفيذ هذه الطريقة وفق الخطوات التالية :

١. إجراء التحليل التمييزى Discriminant Analysis على بيانات العينة الكاملة لينتج معاملات الدالة التمييزية ومعاملات البنية والنقاط الوسطى للمجموعة Group Centroids .

٢. تقسيم العينة الأصلية N إلى k من الفئات أو المجموعات الفرعية متساوية الحجم n ، وكل فئة فرعية يمكن أن تكون ذات حجم صغير يساوى ١ أو حجم كبير يساوى N .

٣. يحذف الباحث بصورة متكررة كل فئة فرعية من العينة الأصلية ويجرى التحليل التمييزى الوصفى على كل فئة بيانات مستقطعة ، وتفضل الفئات الفرعية الأصغر لأنها تنتج المزيد من الإعادات وهذا يجعل من السهل حذف الدرجات المتطرفة ويكون هناك المزيد من الثقة فى النتائج .

٤. تُحسب القيم الزائفة Pseudo Values من كل فئة بيانات مستقطعة باستخدام معاملات الدالة التمييزية الأصلية ومعاملات الدالة المستقطعة وعدد الفئات الفرعية k كالتالى :

٥. $$Pseudovalues = J_i\theta' = k(\theta) - (k-1)\theta$$

٦. يتم حساب متوسط القيم الزائفة للحصول على Jackknifed Coefficients كالتالى :

٧. $$Jackknifed\_Coefficient = J(\theta') = \frac{\sum J_i(\theta)}{k}$$

٨. تفسير المعاملات Jackknifed Coefficients .

ويمكن استخدام إحصاء "ت" لتقويم نتائج طريقة Jackknife ، ولأن Jackknifed Coefficients تتوزع اعتدالياً فإن قيمة "ت" يمكن حسابها من المعادلة التالية :

$$t = \frac{Jackknifed\_Coefficients}{standard\_error\_of\_the\_means\_of\_the\_pseudovalues}$$

بدرجة حرية $df = k - 1$.

رابعاً : إجراء Bootstrap

ابتكر هذه الطريقة (Efron (1979, 1981 ، كما طورت فيما بعد بواسطة Efron and Tibshirani (1993) ، وتفترض أن عينة واحدة متاحة تزاد إلى العديد من العينات عن طريق إعادة المعاينة . ويستخدم إعادة المعاينة بإحلال لتكوين حدود الثقة (In: Berger, 2007) .

وهى طريقة مبنية على الحاسب لتعيين قياسات لدقة التقديرات الإحصائية ، وتوفر بديلاً منافساً للاستدلال الإحصائى تحت انتهاك الشروط المعيارية المعتادة (Efron & Tibshirani, 1993: 10; Davison & Hinkley, 1997: 11) .

ويمكن النظر إليها باعتبارها مجموعة من الطرق المطورة لعمل أنواع معينة من الاستدلالات الإحصائية ، وقد طورت حديثاً لكونها تتطلب حاسبات حديثة تتمتع بالقدرة الحسابية لتبسيط الحسابات المعقدة فى النظرية الإحصائية التقليدية ، ونتج عن ذلك أن الفكرة الرئيسة للإحصاء تغيرت ولكن تطبيقاتها مازالت مستمرة ، وقد أسهمت الحاسبات الحديثة فى تطبيق هذه الأفكار بمرونة وسرعة وسهولة مع أقل عدد من الافتراضات الرياضية (Efron & Tibshirani, 1993: 1-2) .

وتُعد هذه الطرق من أقوى طرق تقويم إعادة النتائج حيث تعتمد الطريقة على نسخ البيانات الأصلية فوق بعضها عدداً كبيراً من المرات وذلك لتوليد ملف ضخم من البيانات ، ومن هذه المجموعة الضخمة ، تُختار المئات أو الآلاف من العينات العشوائية بحيث يكون حجم كل عينة هو نفس العدد n للعينة الأصلية ، ويجرى الاختبار المطلوب والمحدد فى الدراسة نفسها ، ثم تحسب نتائج تلك الاختبارات لكل عينة بشكل مستقل ويؤخذ المتوسط (Thompson, 1992: 19; Higgins, 2005: 57) .

اختبار "ت" t-Test

له توزيع احتمالى لمتغير متصل يُسمى t-Distribution ، ويشبه منحنى توزيع "ت" المنحنى الاعتدالى المعيارى فى أنه منتظم ، وجرسى الشكل ، ووحيد القمة ، وذو متوسط يساوى الصفر ، وهو متماثل ولكنه لا يمس المحور الأفقى كالتوزيع الاعتدالى ، ولكنه أكثر تفرطحاً من التوزيع الاعتدالى عند المتوسط وهذا يعنى أن توزيع "ت" أكثر تشتتاً من التوزيع الاعتدالى المعيارى ، وأن طرفى هذا التوزيع أكثر ارتفاعاً عن طرفى التوزيع الاعتدالى المعيارى عن المحور الأفقى لذا تكون قيمة "ت" أكبر من قيمة Z المناظرة للمساحة نفسها (Bartz, 1988: 248-249; 256) .

والحالة الأكثر عمومية هى التى يستخدم فيها اختبار "ت" لمقارنة مجموعتين لهما عدد مختلف من الأفراد ، فلا ضرورة أن يكون للمجموعتين الحجم نفسه ، وفى الكثير من التجارب يندر توافر أعداد متساوية من الأفراد . وإذا كانت n_1, n_2 هى أعداد المشاهدات فى المجموعتين ، تكون درجات الحرية المناظرة هى $n_1 - 1, n_2 - 1$ بإجمالى $(n_1 + n_2 - 2)$ ، ويختلف تباين متوسطى المجموعتين $\frac{S^2}{n_1}, \frac{S^2}{n_2}$ ، حيث S^2 هو التباين الممزوج* Pooled Variance ، بينما تباين الفرق بين المتوسطين ، مجموع التباينين كالتالى (Snedecor, 1946: 80-82) :

---

* يبرر افتراض تجانس التباين فى استخدام صيغة معادلة اختبار "ت" جمع تبايني المجموعتين للحصول على تقدير واحد لتباين مجتمع الأصل ، وكذلك استخدام درجات حرية المجموعتين معاً ، وبالطبع يصل تقدير التباين إلى أعلى درجات الدقة إذا صح افتراض أن $\sigma_1^2 = \sigma_2^2$ لمجتمعات الأصل ، لأن ذلك يعنى أن كلاً من S_1^2, S_2^2 كإحصاءين للعينات يتطابق مع بارامتر تباين مجتمع الأصل ، وهذا

$$\frac{S^2}{n_1}+\frac{S^2}{n_2}=S^2(\frac{1}{n_1}+\frac{1}{n_2})=S^2(\frac{n_1+n_2}{n_1n_2})$$

ويكون الانحراف المعيارى لمتوسط الفروق كالتالى :

$$S_{\bar{x}}=\sqrt{S^2(\frac{n_1+n_2}{n_1n_2})}=S\sqrt{(\frac{n_1+n_2}{n_1n_2})}$$

وتستخدم الصيغة التالية لحساب قيمة "ت" :

$$t=\bar{X}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{(n_1+n_2)S_{X^2}}}$$

حيث $\bar{X}$ هو الفرق بين متوسطى المجموعتين ، S_{X^2} مجموع المربعات الممزوجة The Pooled Sum of Squares .

واختبار "ت" كاختبار إحصائى بارامترى يستند إلى الافتراضات* الآتية :

١. افتراض الاستقلالية Independency : حيث يقتضى هذا الافتراض بأن العينتين مستقلتان ، أى لا توجد علاقة بين بيانات أو مشاهدات الأفراد فى العينة الأولى بالمقارنة مع العينة الثانية (Sprinthall, 1990: 186) ، أى أنه بموجب هذا الافتراض أن n_1 من المشاهدات قد تم الحصول عليها عشوائياً من المجتمع الأول بشكل مستقل عن n_2 من المشاهدات والتى تم الحصول عليها عشوائياً من المجتمع الثانى ، وهذا يعنى أن معامل الارتباط بين المتوسطين $\bar{X}_1,\bar{X}_2$ المحسوبين على عدد لانهائى من العينات يساوى صفراً (أحمد سليمان عودة وخليل يوسف الخليلى ، ١٩٨٨ : ٢٢٧) . وفشل الباحث فى تحقيق الضبط التجريبى لا يحله اللجوء إلى استخدام اختبار "ت" للمجموعات المرتبطة وإنما استخدام أسلوب تحليل التغاير (فؤاد أبو حطب وآمال صادق ، ١٩٩٦ : ٣٧٤) .

٢. البيانات فى المستوى الفترى أو النسبى Interval or Ratio Data : حيث يقتضى هذا الافتراض بأن يكون المتغير المقاس فى المستوى النسبى أو على الأقل فى المستوى الفترى

يعنى أن هاتين الإحصائيتين غير متحيزتين فى تقديرهما للبارامتر المشترك σ^2 ، وإذا كان الأمر كذلك فإنه من غير المنطقى عدم جمع المعلومات التى تتوافر للباحث منهما للحصول على تقدير أفضل وأكثر دقة للبارامتر σ^2 ، وحينئذ يتوافر أيضاً تقدير أكثر دقة للخطأ المعيارى للفروق بين المتوسطين (فؤاد أبو حطب وآمال صادق ، ١٩٩٦ : ٣٧٤) .

* فى حالة اختبار "ت" لعينتين مرتبطتين تكون الافتراضات الأساسية هنا هى نفس الافتراضات فى حالة اختبار "ت" لمتوسطى عينتين مستقلتين ماعدا افتراض الاستقلالية (صلاح أحمد مراد ، ٢٠٠٠ : ٢٤٩).

حتى يتسنى حساب المتوسط والانحراف المعيارى (Bartz, 1988: 255; Sprinthall, 1990: 186).

٣. افتراض العينات العشوائية : تؤكد النظرية الإحصائية التى تكمن وراء اختبار "ت" بشدة على افتراض المعاينة العشوائية وضرورة استيفائه (Bartz, 1988: 255) ، ويقتضى هذا الافتراض أن تكون البيانات من عينة عشوائية من مجتمع كبير . أى كلا العينتين عينات عشوائية بسيطة من المجتمعات المشتقة منها (Sprinthall, 1990: 186) .

٤. افتراض التوزيع الاعتدالى للمجتمعات Normally Distributed Populations : بموجب هذا الافتراض تتخذ المشاهدات X_i فى المجتمع الأول شكل التوزيع الطبيعى لمتوسط يساوى μ_1 ، وكذلك الأمر بالنسبة للمشاهدات فى المجتمع الثانى X_j يفترض فيها أن تتخذ شكل التوزيع الطبيعى لمتوسط يساوى μ_2 (أحمد سليمان عودة وخليل يوسف الخليلى ، ١٩٨٨ : ٢٢٦) ، وهو افتراض معقول فى معظم الأحوال ، نظراً لما يُعتقد بأن أكثرية الخصائص الفسيولوجية والسيكولوجية تتوزع اعتدالياً وهذا لا يكافىء أن العينات تتوزع اعتدالياً (Bartz, 1988: 256) ، والتوزيع الاعتدالى للأفراد أو الحالات فى مجتمع الأصل يؤدى إلى الحصول على توزيع اعتدالى للمتوسط وغيره من الإحصاءات التى تحسب للعينات ، وحتى لو كان توزيع الأصل بعيداً عن الاعتدالية فإن توزيع متوسطات العينات المشتقة منه يميل إلى الاعتدالية إلا إذا كانت العينات صغيرة جداً (فؤاد أبو حطب وآمال صادق ، ١٩٩٦ : ٣١١)

وتوجد فى الأدبيات عدة اختبارات لاختبار افتراض الاعتدالية منها : اختبار Shapiro–Wilk W ، واختبار Anderson–Darling ، واختبار Martinez–Iglewicz ، واختبار Kolmogorov–Simirnov ، واختبار D'Agostino–Pearson Omnibus ، واختبار Lilliefors ، وهذه الاختبارات تتميز بانخفاض القوة الإحصائية فى حالة $N \prec 10$ (Razali & Wah, 2011: 21–22; Saculinggan & Balase, 2013: 1–3) .

٥. افتراض تجانس التباين Variance Homogeneity : يقتضى بتساوى تباين المجتمعين ، وبموجب هذا الافتراض يكون لتباين المشاهدات فى كل من المجتمعين نفس القيمة σ^2 . وبذلك تكون القيمة المتوقعة للتباين فى كل من العينتين مساوية للمقدار σ^2 . أى يكون توقع $\sigma^2 = E(S_1^2)$ ، وكذلك توقع $\sigma^2 = E(S_2^2)$ ، حيث S_1^2 تباين قيمة المشاهدات فى العينة الأولى ، S_2^2 تباين قيمة المشاهدات فى العينة الثانية ، أى يكون كل من S_1^2, S_2^2 تقديراً مستقلاً لنفس المقدار σ^2 . ولذلك فإن الحصول على تقدير واحد لتباين المجتمع σ^2

من مزجهما أو خلطهما معاً يعطى تقديراً أفضل وأكثر كفاءة مما لو حصلنا على تقديرين مستقلين له من كل منهما على انفراد ، ويتم ذلك من خلال جمع مجموع المربعات فى البيانات من كل من العينتين والقسمة على درجات الحرية الاجمالية لكل من العينتين من خلال الصيغة :

$$S^2_{pooled} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

حيث S^2_{pooled} هو التباين الممزوج أو ما يسمى التباين داخل المجموعات (أحمد سليمان عودة وخليل يوسف الخليلى ، ١٩٨٨ : ٢٢٦-٢٢٧) .

واختبار افتراض تجانس التباين يمكن التعبير عنه كما يلى :

الفرض الصفرى : تساوى تباين العينة الأولى وتباين العينة الثانية $H_\circ : S_1^2 = S_2^2$

الفرض البديل : عدم تساوى تباين العينة الأولى وتباين العينة الثانية $H_a : S_1^2 \neq S_2^2$

ويتطلب استخدام اختبار "ف" الذى يعتمد على توزيع "ف" ويمكن حسابه بقسمة التباين الكبير على التباين الصغير كما بالمعادلة :

$$F_{calc} = \frac{S^2_{\max}}{S^2_{\min}}$$

ويتم الكشف عن دلالة القيمة المحسوبة F_{calc} بمقارنتها بالقيمة الجدولية F_{table} عند درجة حرية $df = N - 1$ لكل عينة ، وإذا كانت $F_{calc} \prec F_{table}$ فإن ذلك يعنى الفشل فى رفض الفرض الصفرى أى القرار قبول الفرض الصفرى والاستنتاج هو تجانس التباينات . واستخدام اختبار "ف" لاختبار تجانس التباين يتطلب أن تكون كلا العينتين مشتقتين من مجتمعات تتوزع اعتدالياً وهذا بدوره يتطلب اختبار افتراض الاعتدالية أولاً ، وإذا كانت كلا العينتين لا تتوزعان اعتدالياً لا يفضل استخدام اختبار النسبة الفائية ، وفى مثل هذه الحالة ينصح باستخدام Levene Test (Best & Kahn, 2006: 416-417) .

دراسات وبحوث سابقة

هدفت دراسة Kulkarni (1993) إلى مقارنة تقديرات الخطأ من النوع الأول لأربع طرق (اختبار "ت" ، Pooled Error Bootstrap Contrast ، Unpooled Error Bootstrap Contrast ، Robust Rank Order Test) تحت شرط انتهاك افتراض تجانس التباين فى حالة تساوى أو عدم تساوى حجم العينتين بنسب تباين مختلفة حيث بلغت أحجام العينات [(١٥ ، ٥) ، (٣ ، ٥) ، (٧ ، ٧)] ، ومن بين ما توصلت إليه نتائج الدراسة تشوه تقديرات الخطأ من النوع الأول عندما تشتق

العينات الصغيرة من المجتمعات الأكثر تبايناً ، وأن تقديرات الخطأ من النوع الأول تكون أقل من مستوى الدلالة الإحصائية عندما تشتق العينات الصغيرة من المجتمعات الأقل تبايناً ، وفشل زيادة حجم العينة فى تخفيض هذا التشوه ، كما أظهر اختبار "ت" وطريقة Pooled Error Bootstrap Contrast أداءً متقارباً ، وأيضاً تم التوصل إلى أن أداء اختبار "ت" أفضل من الطرق الثلاثة الأخرى فى المحافظة على تقديرات منخفضة للخطأ من النوع الأول فى حالة تساوى أحجام العينات ، وتبين أن طريقتا (Robust Rank Order Test ، Unpooled Error Bootstrap Contrast) قدمتا أداءً متميزاً فى حالة أحجام العينات الكبيرة من حيث السيطرة على تقديرات الخطأ من النوع الأول .

أما دراسة Lansing (1999) فقد هدفت إلى فحص تقديرات الخطأ من النوع الأول والقوة لأربع طرق (اختبار "ت" ، Pooled Error Bootstrap Contrast ، Unpooled Error Bootstrap Contrast ، Welch's Test) ، وتم توليد البيانات بالمحاكاة بواسطة طريقة Monte Carlo باستخدام الحزمة الإحصائية DATASIM حيث بلغت أحجام العينات [(٢ ، ٢) ، (٣ ، ٥) ، (٥ ، ٥) ، (٢ ، ٢٠) ، (٥ ، ٥٠) ، (٥٠ ، ٥٠) ، (٢٠ ، ١٠٠) ، (٢٠ ، ٢٠٠)] وبلغت نسب التباينات [(١ : ١) ، (١ : ١٠) ، (١ : ٥) ، (١٠ : ١) ، (٥ : ١)] ، ومن بين ما توصلت إليه نتائج الدراسة أن إجراءات Bootstrap أكثر قوة مقارنة ببقية الطرق فى حالة العينات كبيرة الحجم والعكس صحيح فى حالة العينات صغيرة الحجم .

فحين هدفت دراسة Yin et al., (2010) إلى فحص تقديرات الخطأ من النوع الأول لاختبار "ت" الممزوج لعينتين مستقلتين وإجراء Bootstrap وتم توليد توزيعات معاينة عشوائية اعتدالية وأخرى تتبع توزيع مربع كاى حيث بلغت أحجام العينات [(٥ ، ١٥) ، (١٥ ، ٢٥)] وبلغت نسب التباينات [(١ ، ١) ، (١ : ٩) ، (١ ، ٣٦)] ، ومن بين ما توصلت إليه نتائج الدراسة أن اختبار "ت" يتميز بالمنعة تحت حوالى ٢٠% ، ٣٣% من الشروط التى تمت دراستها للعينات على الترتيب ، بينما وجد أن إجراء Bootstrap يتميز بالمنعة تحت حوالى ٣٣% ، ٤٧% من الشروط التى تمت دراستها للعينات على الترتيب ، وأن اختبار "ت" تميز عن إجراء Bootstrap فى حالة عدم انتهاك افتراض تجانس التباين .

أما دراسة Li (2011) فقد هدفت إلى مقارنة خصائص القوة لثلاث اختبارات لمقارنة متوسطين مستقلين (اختبار "ت" ، لختبار العشوائية ، طريقة Bootstrap) ، وتم توليد البيانات بالمحاكاة بواسطة برنامج GAUSS ، ومن بين ما توصلت إليه نتائج الدراسة أن أداء القوة الإحصائية لطريقة Bootstrap أفضل من أداء القوة الإحصائية لاختبار "ت" فى حالة العينات

المتوسطة والكبيرة .

كما هدفت دراسة Ahad et al. (2012) إلى مقارنة القوة لاختبارى "ت" لعينتين مستقلتين وإجراء Bootstrap بتكرار ١٠٠٠ عينة ، وتم توليد بيانات المحاكاة بواسطة الحزمة RANDGEN ، وبلغ حجم العينتين (٥ ، ١٥) وتباينات العينتين [(١، ١) ، (١ ، ٩) ، (٩ ،١٠)] ، بافتراض أن العينة الأولى مسحوبة من مجتمع يتوزع اعتدالياً والثانية مسحوبة من مجتمع يتبع توزيع مربع كاى ، ومن بين ما توصلت إليه نتائج الدراسة أن إجراء Bootstrap تفوق بشكل طفيف على اختبار "ت" فيما يتصل بخصائص القوة الإحصائية عبر كل الشروط المستهدفة بالدراسة .

وأجرى Kang et al. (2014) دراسة هدفت إلى الكشف عن تأثير كل من انتهاك افتراض الاعتدالية وحجم التأثير وحجم العينة على القوة والمنعة لأربع طرق (اختبار "ت" ، اختبار ويلكوكسون لمجموع الرتب ، طريقة Bootstrap ، اختبار The Robust Welch–James) ، وتم توليد البيانات باستخدام المحاكاة بطريقة Monte Carlo ، ومن بين ما توصلت إليه نتائج الدراسة تمتع طريقة Bootstrap بميزة القوة عن اختبار "ت" لعينتين مستقلتين عندما تكون العينات كبيرة الحجم .

تعقيب على البحوث والدراسات السابقة

١. ندرة الدراسات – فى حدود معرفة الباحث– فى البيئتين المصرية والعربية على حد سواء فيما يتصل بسلوك اختبار "ت" فى مواقف القياس التى تنتهك واحداً أو أكثر من افتراضاته .
٢. اعتمدت جميع الدراسات على توليد البيانات عن طريق المحاكاة لتوفير شرط العشوائية فى اختيار العينة الذى يصعب توافره فى البيانات الحقيقية حيث يلجأ العديد من الباحثين إلى العينات غير العشوائية مثل العينات المقصودة التى تفتقر إلى درجة مقبولة من تمثيل المجتمع .
٣. تباينت الدراسات فى توليفات الشروط المستخدمة لاختبار أداء الخطأ من النوع الأول والقوة الإحصائية لإجراء Bootstrap واختبار "ت" ذو التباين الممزوج .
٤. تختلف الدراسة الحالية عن الدراسات السابقة فى عدد وأحجام العينات المستخدمة وعدد ومستويات انتهاك/ عدم انتهاك افتراض تجانس التباين وعدد مرات تكرار عينة إجراء Bootstrap .

فروض الدراسة

تسعى الدراسة الحالية إلى اختبار الفروض الآتية :–

١. لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة

لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين المزوج فى حالة العينات الصغيرة .

٢. لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين المزوج فى حالة العينات المتوسطة .

٣. لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين المزوج فى حالة العينات الكبيرة .

٤. لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين المزوج فى حالة العينات الكبيرة جداً .

إجراءات الدراسة

تشمل منهجية توليد البيانات والمعالجة الإحصائية :-

أولاً : توليد البيانات

استخدمت الحزمة الإحصائية Minitab لتوليد بيانات الدراسة الحالية بغرض فحص أثر انتهاك/ عدم انتهاك افتراض تجانس التباين على تقديرات الخطأ من النوع الأول والقوة الإحصائية لإجراء Bootstrap واختبار "ت" ذو التباين المزوج تحت توليفة الشروط التالية:

١. أحجام العينات : استخدمت أربع أزواج من العينات [(٢٣ ، ١٦) ؛ (٦٨ ، ٥١) ؛ (١٣٤ ، ١١٩) ؛ (٢٨٩ ، ٢٦٠)] لتمثل العينات الصغيرة والمتوسطة والكبيرة والكبيرة جداً على الترتيب ، وبلغ عدد العينات المولدة ست وخمسون عينة .

٢. التوزيع : مجتمعى الأصل موزعين توزيعاً اعتدالياً .

٣. عملية Bootstrapping : تم توليد (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap .

٤. التباينات : استخدمت أزواج التباينات [(١ ، ١)] ؛ [(٤ ، ١) ؛ (١ ، ٤)] ؛ [(٩ ، ١) ؛ (١ ، ٩)] ؛ [(٢٥ ، ١) ؛ (١ ، ٢٥)] .

ويمكن تلخيص توصيف تصميم الدراسة على النحو المبين بالجدول التالى :

جدول (١) توصيف تصميم الدراسة

الشروط	مستوى الانتهاك	بارامترات المجتمع σ_1^2, σ_2^2	بارامترات المجتمع μ_1, μ_2	حجم العينتين	مستوى الدلالة الإسمى α_n
الأول	لا يوجد	(١ ، ١)	(٤١،٧ ، ٤٢،٩)	(١٦ ، ٢٣)	٠،٠٥
الثانى	بسيط	(٤ ، ١) ؛ (١ ، ٤)		(٥١ ، ٦٨)	
الثالث	متوسط	(٩ ، ١) ؛ (١ ، ٩)		(١١٩ ، ١٣٤)	
الرابع	حاد	(٢٥ ، ١) ؛ (١ ، ٢٥)		(٢٦٠ ، ٢٨٩)	

وقد تم الرجوع إلى بعض الدراسات لتحديد مستويات انتهاك افتراض تجانس التباين منها : (Hess, Olejnik, & Huberty, 2001); (Yin et al., Othman, 2010); (Li, 2011); (Ahad et al., 2012); (Kang et al., 2014) .

ثانياً : المعالجة الإحصائية

(أ) تم فى الدراسة الحالية استخدام الأساليب الإحصائية التالية :

١. اختبار "ت" ذو التباين الممزوج .

٢. إجراء Bootstrap .

٣. اختبار Levene L Test للكشف عن تجانس التباين .

٤. اختبار Kohr للمقارنة بين تقدير مستوى الدلالة الإسمى ومستوى الدلالة الحقيقى ، ويتخذ الصيغة التالية :

$$Z_{Kohr} = \frac{(\alpha_a - \alpha_n)}{\sqrt{\frac{\alpha_n(1-\alpha_n)}{N^{*}}}}$$

بدلالة القيم الحدية (± ١،٩٦ ، ± ٢،٥٨) لمستوى دلالة (٠،٠٥ ، ٠،٠١ ، ٠،٠٠١) على الترتيب .

(ب) تم فى الدراسة الحالية استخدام الحزم الإحصائية التالية :

١. الحزمة الإحصائية Minitab لتوليد بيانات المحاكاة Simulated Data .

٢. الحزمة الإحصائية SPSS لتنفيذ عمليات الـــ Bootstrapping .

٣. الحزمة الإحصائية G*Power لحساب القوة الإحصائية (Soper, 2015) .

نتائج الدراسة وتفسيرها

نتائج اختبار الفرض الأول

ينص الفرض الأول على : لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap.

واختبار "ت" ذو التباين الممزوج فى حالة العينات الصغيرة ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات الخطأ من النوع الأول والقوة الإحصائية على النحو المبين بالجدولين التاليين :

جدول (٢) تقديرات الخطأ من النوع الأول والقوة

لاختبار "ت" ذو التباين الممزوج فى حالة العينات الصغيرة

بارامترات المجتمع		المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2	n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١٫٧ ، ٤٢٫٩)	(١ ، ١)	٢٣	٤٢٫٧٩٥	٠٫٧٠٣	١٦	٤٢٫٠٨٣	٠٫٩٤٤	لا يوجد	٠٫٠٥	٠٫٠١	٠٫٣١٣
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٤)	٢٣	٤٢٫٢٥	٢٫٢٤	١٦	٤١٫٦٧٣	٠٫٩٠٣	بسيط	٠٫٠٥	*٠٫٣١١	٠٫٥٨٤
(٤١٫٧ ، ٤٢٫٩)	(٤ ، ١)	٢٣	٤٣٫٠٣٦	٠٫٨٨٢	١٦	٤١٫٩٢	٢٫٠١	بسيط	٠٫٠٥	٠٫٠٢٤	٠٫١٣٥
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٩)	٢٣	٤٢٫٨٦	٢٫٧٩	١٦	٤١٫٣١	٠٫٨٩٥	متوسط	٠٫٠٥	٠٫٠٢٩	٠٫٥٠١
(٤١٫٧ ، ٤٢٫٩)	(٩ ، ١)	٢٣	٤٢٫٧٩	١٫١٧	١٦	٤٠٫٨٧	٢٫٧٥	متوسط	٠٫٠٥	٠٫٠٠٥	٠٫٣٧٩
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٢٥)	٢٣	٤٣٫٠٢	٥٫٣٢	١٦	٤١٫٦٨٩	٠٫٦٠٩	حاد	٠٫٠٥	*٠٫٣٢٧	٠٫٥٧٦
(٤١٫٧ ، ٤٢٫٩)	(٢٥ ، ١)	٢٣	٤٢٫٧٧	١٫٢١	١٦	٤٠٫٤٥	٢٫٩٠	حاد	٠٫٠٥	٠٫٠١١	٠٫٢٥٧

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠٫٨ .

ويتضح من الجدول (٢) السابق أن التباين المصحوب بالتزاوج الموجب Positive Pairing (الذى يحدث عندما تكون العينة الكبيرة الحجم منتمية للمجتمع ذى التباين الكبير والعينة صغيرة الحجم منتمية للمجتمع ذى التباين الصغير) ينتج أفضل تقدير للقوة الإحصائية فى حالة العينات الصغيرة .

كما يتضح أن التباين المصحوب بالتزاوج السالب Negative Pairing (الذى يحدث عندما تكون العينة الكبيرة الحجم منتمية للمجتمع ذى التباين الصغير والعينة صغيرة الحجم منتمية للمجتمع ذى التباين الكبير) ينتج أفضل تقدير للخطأ من النوع الأول فى حالة العينات الصغيرة وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتيجة دراسة (Ahad et al., 2012) .

ومن الجدول (٢) يتضح أيضاً تضخم تقدير الخطأ من النوع الأول بحسب اختبار Z_{Kohr} فى حالة انتهاك افتراض تجانس التباين فى المستويين البسيط والحاد المصحوب بالتزاوج الموجب .

وأن اختبار "ت" ذو التباين الممزوج لا يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠٫٨فى حالة توافر تجانس التباين وفى حالة انتهاك افتراض تجانس التباين فى جميع المستويات (بسيط ، متوسط ، حاد) .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين يؤدى إلى زيادة

تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الصغيرة ، ومع ما ورد فى الأدبيات مثل (Ruxton, 2006: 688) بأن اختبار "ت" يؤدى بشكل غير جيد Performs Badly من حيث الخطأ من النوع الأول فى حالة عدم تساوى أزواج التباينات ، وأيضاً تتفق جزئياً مع نتائج دراسة (Li, 2011) بأن اختبار "ت" لا يؤدى بشكل جيد عند التعامل مع العينات الصغيرة .

أما بشأن أداء الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الصغيرة ومرات تكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap فهو على النحو المبين بالجدول التالى :

جدول (٣)

تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الصغيرة

بارامترات المجتمع		عدد العينات	المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2		n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١٫٧ ، ٤٢٫٩)	(١ ، ١)	١٠٠٠	٢٣	٤٢٫٧٩٥٣	٠٫٧٠٢٤	١٦	٤٢٫٠٨٢٠	٠٫٩٤٣٦	لا يوجد	٠٫٠٥	٠٫٠١٥	٠٫٢٣٠
		٢٠٠٠	٢٣	٤٢٫٧٩٥٣	٠٫٧٠٢٤	١٦	٤٢٫٠٨٢٠	٠٫٩٤٣٦	لا يوجد	٠٫٠٥	٠٫٠١٤	٠٫٣٦٠
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٤)	١٠٠٠	٢٣	٤٢٫٢٤٥٦	٢٫٢٤٣٢	١٦	٤١٫٦٧٣٤	٠٫٩٠٢٨	بسيط	٠٫٠٥	٠٫٢٨٠*	٠٫٥٢٣
		٢٠٠٠	٢٣	٤٢٫٢٤٥٦	٢٫٢٤٣٢	١٦	٤١٫٦٧٣٤	٠٫٩٠٢٨	بسيط	٠٫٠٥	٠٫٢٧٩*	٠٫٥٢٢
(٤١٫٧ ، ٤٢٫٩)	(٤ ، ١)	١٠٠٠	٢٣	٤٣٫٠٣٥٨	٠٫٨٨٢٤	١٦	٤١٫٩٢٣٩	٢٫٠١٢٦	بسيط	٠٫٠٥	٠٫٠٣١	٠٫٦٢١
		٢٠٠٠	٢٣	٤٣٫٠٣٥٨	٠٫٨٨٢٤	١٦	٤١٫٩٢٣٩	٢٫٠٢١٦	بسيط	٠٫٠٥	٠٫٠٥٨	٠٫٥٨٤
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٩)	١٠٠٠	٢٣	٤٢٫٨٦٢٠	٢٫٧٩	١٦	٤١٫٣١٠١	٠٫٨٩٥٤	متوسط	٠٫٠٥	٠٫٠٣٤	٠٫٦٠٢
		٢٠٠٠	٢٣	٤٢٫٨٦٢٠	٢٫٧٩	١٦	٤١٫٣١٠١	٠٫٨٩٥٤	متوسط	٠٫٠٥	٠٫٠٣٢	٠٫٥٩٢
(٤١٫٧ ، ٤٢٫٩)	(٩ ، ١)	١٠٠٠	٢٣	٤٢٫٧٨٦٧	١٫٤٣٢٦	١٦	٤٠٫٨٣٣٦	٢٫٧٥١٢	متوسط	٠٫٠٥	٠٫٠١٨	٠٫٦٧٥
		٢٠٠٠	٢٣	٤٢٫٧٨٦٧	١٫٦٣٢٦	١٦	٤٠٫٨٣٣٦	٢٫٧٥١٢	متوسط	٠٫٠٥	٠٫٠٢٠	٠٫٦٩١
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٢٥)	١٠٠٠	٢٣	٤٣٫٠٢١٩	٥٫٣٢٠٩	١٦	٤١٫٦٨٨٨	٠٫٦٠٩٦	حاد	٠٫٠٥	٠٫٢٤٧*	٠٫٤٩٣
		٢٠٠٠	٢٣	٤٣٫٠٢١٩	٥٫٣٢٠٩	١٦	٤١٫٦٨٨٨	٠٫٦٠٩٦	حاد	٠٫٠٥	٠٫٢٦٥*	٠٫٥١٧
(٤١٫٧ ، ٤٢٫٩)	(٢٥ ، ١)	١٠٠٠	٢٣	٤٢٫٧٧٢٦	١٫٣١٠٤	١٦	٤٠٫٤٤٩١	٣٫٨٩٩٣	حاد	٠٫٠٥	٠٫٠٣٩	٠٫١٦٦
		٢٠٠٠	٢٣	٤٢٫٧٧٢٦	١٫٣١٠٤	١٦	٤٠٫٤٤٩١	٣٫٨٩٩٣	حاد	٠٫٠٥	٠٫٠٤٢	٠٫١٧٧

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .
* الفرق دال إحصائيا باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠٫٨ .

ويتضح من الجدول (٣) السابق أن إجراء Bootstrap حافظ على تقديرات الخطأ من النوع الأول فى مستوى أقل من مستوى الدلالة الإحصائية الإسمى (٠٫٠٥) فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج . وأن إجراء Bootstrap تميز بتقديرات مرتفعة للخطأ من النوع الأول بحسب اختبار Z_{Kohr} فى حالة انتهاك افتراض تجانس التباين فى المستوين البسيط والحاد المصحوبين بالتزاوج الموجب .

ويتضح أيضاً أن إجراء Bootstrap لا يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠٫٨ فى حالة توافر تجانس التباين وفى حالة انتهاك افتراض تجانس التباين فى جميع المستويات (بسيط ، متوسط ، حاد) .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتيجة دراسة (Ahad et (al., 2012 ، كما تتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الصغيرة ، وأيضاً تتفق جزئياً مع نتائج دراسة (Li, 2011) بأن إجراء Bootstrap ليس فعالاً بدرجة كافية عند التعامل مع العينات الصغيرة .

وبمراجعة نتائج الجدولين (٢) ، (٣) يُلاحظ أن إجراء Bootstrap لا يُظهر خصائص متميزة ملحوظة من حيث أداء القوة حال توافر افتراض تجانس التباين لزوج تباينات مجتمعى الأصل فى حالة العينات الصغيرة مقارنة باختبار "ت" ذو التباين الممزوج ، ولا تتفق هذه النتيجة التى تم التوصل إليها مع ما توصلت إليه نتائج دراسة (Ahad et al., 2012) أن إجراء Bootstrap تفوق بشكل طفيف على اختبار "ت" فى حالة العينات الصغيرة ، وتتفق أيضاً جزئياً مع ما توصلت إليه نتائج دراسة Yin et al. (2010) بأن إجراء Bootstrap لا يُظهر خصائص متميزة ملحوظة من حيث أداء الخطأ من النوع الأول حال توافر افتراض تجانس التباين مقارنة باختبار "ت" ذو التباين الممزوج

نتائج اختبار الفرض الثانى

ينص الفرض الثانى على : لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات المتوسطة ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات الخطأ من النوع الأول والقوة الإحصائية على النحو المبين بالجدولين التاليين :

جدول (٤) تقديرات الخطأ من النوع الأول والقوة لاختبار "ت" ذو التباين الممزوج فى حالة العينات المتوسطة

بارامترات المجتمع		المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2	n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١٫٧ ، ٤٢٫٩)	(١ ، ١)	٦٨	٤٢٫٩٩٦	٠٫٩٥٥	٥١	٤١٫٦٥٦	٠٫٩٦٦	لا يوجد	٠٫٠٥	*٠٫٠٠٠	**٠٫٩٩٩
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٤)	٦٨	٤٢٫٥٠	٢٫١٣	٥١	٤١٫٧٧	١٫٠٧	بسيط	٠٫٠٥	٠٫٠٢٦	٠٫٥٩٥
(٤١٫٧ ، ٤٢٫٩)	(٤ ، ١)	٦٨	٤٢٫٨١١	٠٫٩٦٥	٥١	٤٢٫٢٨	٢٫٢٠	بسيط	٠٫٠٥	٠٫٠٧٦	٠٫٥١١
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٩)	٦٨	٤٢٫٨٤	٣٫٣٣	٥١	٤١٫٨٥	١٫١٤	متوسط	٠٫٠٥	٠٫٠٤٣	٠٫٦٣٥
(٤١٫٧ ، ٤٢٫٩)	(٩ ، ١)	٦٨	٤٢٫٦٤٢	٠٫٩٦٦	٥١	٤١٫٣٩	٣٫١٧	متوسط	٠٫٠٥	*٠٫٠٠٣	٠٫٤٩٩
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٢٥)	٦٨	٤٣٫٠٠	٤٫٩٩	٥١	٤١٫٦٤٦	٠٫٨٨٥	حاد	٠٫٠٥	٠٫٠٥٨	٠٫٥٤٦
(٤١٫٧ ، ٤٢٫٩)	(٢٥ ، ١)	٦٨	٤٢٫٩٠٨	٠٫٨٥٤	٥١	٤١٫٦٥	٤٫٧٥	حاد	٠٫٠٥	٠٫٠٣٥	٠٫٤٠٤

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠،٨ .

ويتضح من الجدول (٤) السابق أن التزاوج الموجب بين التباين الأكبر وحجم العينة الأكبر ينتج تقديرات قوة أعلى مقارنة بالتزاوج السالب بين التباين الأصغر وحجم العينة الأكبر الذى ينتج تقديرات قوة أقل فى حالة انتهاك افتراض تجانس التباين فى جميع المستويات (بسيط ، متوسط ، حاد) فى حالة العينات المتوسطة ، وتتفق هذه النتيجة التى تم التوصل إليها مع ما توصلت إليه نتيجة دراسة (Ahad et al., 2012) .

كما يتضح أن التزاوج الموجب ينتج تقديرات أقل للخطأ من النوع الأول مقارنة بالتزاوج السالب الذى ينتج تقديرات أعلى للخطأ من النوع الأول فى حالة انتهاك تجانس التباين فى المستوى البسيط ، بينما ينتج التزاوج الموجب تقديرات أعلى للخطأ من النوع الأول مقارنة بالتزاوج السالب الذى ينتج تقديرات أقل للخطأ من النوع الأول فى حالة انتهاك تجانس التباين فى المستويين المتوسط والحاد

ومن الجدول (٤) يتضح أيضاً أن اختبار "ت" ذو التباين الممزوج تميز بتقديرات منخفضة للخطأ من النوع الأول بحسب اختبار Z_{Kohr} فى حالة تجانس التباين وفى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط المصحوب بالتزاوج السالب . وأن اختبار "ت" ذو التباين الممزوج يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة تجانس التباين فقط . وتميز الاختبار بتقديرات منخفضة للقوة الإحصائية فى حالة انتهاك افتراض تجانس التباين فى جميع المستويات (بسيط ، متوسط ، حاد) .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات المتوسطة ، ومع ما ورد فى الأدبيات مثل (Ruxton, 2006: 688) بأن اختبار "ت" يؤدى بشكل غير جيد من حيث الخطأ من النوع الأول فى حالة عدم تساوى أزواج التباينات .

أما بشأن أداء الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات المتوسطة ومرات تكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap فهو على النحو المبين بالجدول التالى

جدول (٥)

تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات المتوسطة

بارامترات المجتمع		حجم العينات	المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2		n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١،٧ ، ٤٢،٩)	(١ ، ١)	١٠٠٠	٦٨	٤٣،٠٠٢٩٩	١،٠٠٠٧٤	٥١	٤١،٩٥٥٥	٠،٩٦٦٥	لا يوجد	٠،٠٥	*٠،٠٠٠١	**٠،٩٩٩
		٢٠٠٠	٦٨	٤٣،٠٠٢٩٩	١،٠٠٠٧٤	٥١	٤١،٩٥٥٥	٠،٩٦٦٥	لا يوجد	٠،٠٥	*٠،٠٠٠٠	**٠،٩٩٩
(٤١،٧ ، ٤٢،٩)	(١ ، ٤)	١٠٠٠	٦٨	٤٣،٥٠٤١	٢،١٢٨٨	٥١	٤١،٧٦٩١	١،٠٦٧٥	بسيط	٠،٠٥	٠،٠١٢	٠،٥٨٩
		٢٠٠٠	٦٨	٤٣،٥٠٤١	٢،١٢٨٨	٥١	٤١،٧٦٩١	١،٠٦٧٥	بسيط	٠،٠٥	٠،٠١٥	٠،٦٢٣
(٤١،٧ ، ٤٢،٩)	(٤ ، ١)	١٠٠٠	٦٨	٤٢،٨١٠٧	٠،٩٦٤٦	٥١	٤٣،٢٧٥٤	٢،٢٠٠٥	بسيط	٠،٠٥	*٠،١٠٢	٠،٧٠٩
		٢٠٠٠	٦٨	٤٢،٨١٠٧	٠،٩٦٤٦	٥١	٤٣،٢٧٥٤	٢،٢٠٠٥	بسيط	٠،٠٥	*٠،١٢٢	٠،٧٢٨
(٤١،٧ ، ٤٢،٩)	(١ ، ٩)	١٠٠٠	٦٨	٤٢،٨٤٤٣	٣،٣٣٥٩	٥١	٤١،٨٤٧٨	١،١٣٨٤	متوسط	٠،٠٥	٠،٠٤٢	٠،٧٤٦
		٢٠٠٠	٦٨	٤٢،٨٤٤٣	٣،٣٣٥٩	٥١	٤١،٨٤٧٨	١،١٣٨٤	متوسط	٠،٠٥	٠،٠٣٥	٠،٧١٨
(٤١،٧ ، ٤٢،٩)	(٩ ، ١)	١٠٠٠	٦٨	٤٢،٦٤١٩	٠،٩٦٦٣	٥١	٤١،٣٩٠٦	٣،١٦٥١	متوسط	٠،٠٥	٠،٠٣٦	٠،٧٦٨
		٢٠٠٠	٦٨	٤٢،٦٤١٩	٠،٩٦٦٣	٥١	٤١،٣٩٠٦	٣،١٦٥١	متوسط	٠،٠٥	*٠،٠٠٨	٠،٧١٧
(٤١،٧ ، ٤٢،٩)	(١ ، ٢٥)	١٠٠٠	٦٨	٤٣،٠٠١٧	٤،٩٩٤٥	٥١	٤١،٦٤٥٧	٠،٨٨٥٣	حاد	٠،٠٥	٠،٠٣٦	٠،٥٦٥
		٢٠٠٠	٦٨	٤٣،٠٠١٧	٤،٩٩٤٥	٥١	٤١،٦٤٥٧	٠،٨٨٥٣	حاد	٠،٠٥	٠،٠٤١	٠،٦٠٩
(٤١،٧ ، ٤٢،٩)	(٢٥ ، ١)	١٠٠٠	٦٨	٤٢،٩٠٨١	٠،٨٥٣٦	٥١	٤١،٦٥٣٨	٤،٧٥٠٨	حاد	٠،٠٥	٠،٠٦٦	٠،٦٤٢
		٢٠٠٠	٦٨	٤٢،٩٠٨١	٠،٨٥٣٦	٥١	٤١،٦٥٣٨	٤،٧٥٠٨	حاد	٠،٠٥	٠،٠٧٤	٠،٦٦٤

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠،٨ .

يتضح من الجدول (٥) أن إجراء Bootstrap تميز بتقديرات منخفضة للخطأ من النوع الأول فى مستوى أقل من مستوى الدلالة الإحصائية الإسمى (٠،٠٥) بحسب اختبار Z_{Kohr} فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى البسيط المصحوب بالتزاوج الموجب ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج الموجب . وتميز بتقديرات مرتفعة للخطأ من النوع الأول أعلى من مستوى الدلالة الإحصائية الإسمى (٠،٠٥) بحسب اختبار Z_{Kohr} فى حالة انتهاك افتراض تجانس التباين فى المستوى البسيط المصحوب بالتزاوج السالب .

وأن إجراء Bootstrap يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة تجانس التباين فقط ، ولا يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة انتهاك افتراض تجانس التباين فى جميع المستويات (بسيط ، متوسط ، حاد) .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات المتوسطة عند استخدام إجراء Bootstrap .

وبمراجعة نتائج الجدولين (٤) ، (٥) يُلاحظ أن إجراء **Bootstrap** يُظهر خصائص ملحوظة من حيث أداء القوة حال انتهاك افتراض تجانس التباين لزوج تباينات مجتمعى الأصل فى حالة العينات المتوسطة مقارنة باختبار "ت" ذو التباين الممزوج مما يمنحه الأفضلية .

نتائج اختبار الفرض الثالث

ينص الفرض الثالث على : لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء **Bootstrap** بتكرار (١٠٠٠ ، ٢٠٠٠) عينة **Bootstrap** واختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات الخطأ من النوع الأول والقوة الإحصائية على النحو المبين بالجدولين التاليين :

جدول (٦) تقديرات الخطأ من النوع الأول والقوة لاختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة

بارامترات المجتمع		المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2	n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١٫٧ ، ٤٢٫٩)	(١ ، ١)	١٢٤	٤٢٫٨٥	١٫٠٠٩	١١٩	٤١٫٨٣	١٫٠٠٢	لا يوجد	٠٫٠٥	*٠٫٠٠٠٠	**٠٫٩٩٩
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٤)	١٢٤	٤٢٫٨٣	٢٫٠٠٩	١١٩	٤١٫٦٥٦	٠٫٩٢١	بسيط	٠٫٠٥	*٠٫٠٠٠٠	**٠٫٩٧٢
(٤١٫٧ ، ٤٢٫٩)	(٤ ، ١)	١٢٤	٤٢٫٧٦٦	٠٫٨٧٢	١١٩	٤١٫٤٢	١٫٩٥	بسيط	٠٫٠٥	*٠٫٠٠٠٠	**٠٫٩٩٧
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٩)	١٢٤	٤٢٫٩٥	٢٫٩٣	١١٩	٤١٫٦٦	١٫٠٥	متوسط	٠٫٠٥	*٠٫٠٠٠٠	٠٫٧٢٣
(٤١٫٧ ، ٤٢٫٩)	(٩ ، ١)	١٢٤	٤٢٫٨٣	١٫٠٢	١١٩	٤١٫٦٤	٢٫٨٥	متوسط	٠٫٠٥	*٠٫٠٠٠٠	٠٫٦٠٧
(٤١٫٧ ، ٤٢٫٩)	(١ ، ٢٥)	١٢٤	٤٢٫٣٥	٥٫٢٣	١١٩	٤١٫٦٩٣	٠٫٩٩٦	حاد	٠٫٠٥	*٠٫٠١٧٦	٠٫٥٣٩
(٤١٫٧ ، ٤٢٫٩)	(٢٥ ، ١)	١٢٤	٤٢٫٠٢٥	٠٫٩٢٠	١١٩	٤٢٫٥٧	٤٫٤٣	حاد	٠٫٠٥	*٠٫٠٢٤٧	٠٫٥٧٤

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Koh} ، ** القوة تتجاوز النقطة المرجعية ٠٫٨ .

ويتضح من الجدول (٦) السابق أن اختبار "ت" ذو التباين الممزوج قد حافظ على تقديرات منخفضة جداً للخطأ من النوع الأول بحسب اختبار Z_{Kohr} أقل من مستوى الدلالة الإسمى (٠٫٠٥) فى حالة تجانس التباين وفى حالة انتهاك افتراض تجانس التباين فى المستويين البسيط والمتوسط ، بينما تميز بتقديرات متضخمة للخطأ من النوع الأول فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد . وأن اختبار "ت" ذو التباين الممزوج يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠٫٨ فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى البسيط وبخاصة المصحوب بالتزاوج السالب . وتميز الاختبار بتقديرات منخفضة إلى حد ما للقوة فى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط ، وتميز بتقديرات منخفضة بشكل ملحوظ فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد . وتتفق هذه النتيجة التى تم

التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين فى المستوى الحاد يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الكبيرة لاختبار "ت" ذو التباين الممزوج ، ومع ما ورد فى الأدبيات مثل (Ruxton, 2006: 688) بأن اختبار "ت" يؤدى بشكل غير جيد من حيث الخطأ من النوع الأول فى حالة عدم تساوى أزواج التباينات .

أما بشأن أداء الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الكبيرة ومرات تكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap فهو على النحو المبين بالجدول التالى :

جدول (٧)

تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الكبيرة

بارامترات المجتمع		حجم العينات	المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2		n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤٢،٩ ، ٤١،٧)	(١ ، ١)	١٠٠٠	١٢٤	٤٢،٨٥٤١	١٠،٩٣٩	١١٩	٤١،٨٣٢٣	١٠،١٦٩	لا يوجد	٠،٠٥	*٠،٠٠٠١	**٠،٩٩٩
		٢٠٠٠	١٢٤	٤٢،٨٥٤١	١٠،٩٣٩	١١٩	٤١،٨٣٢٣	١٠،١٦٩	لا يوجد	٠،٠٥	*٠،٠٠٠٠	**٠،٩٩٩
(٤٢،٩ ، ٤١،٧)	(١ ، ٤)	١٠٠٠	١٢٤	٤٢،٨٣١٥	٢،٠٨٨٠	١١٩	٤١،٦٥٥٧	٠،٩٢١٢	بسيط	٠،٠٥	*٠،٠٠٠١	**٠،٩٩٥
		٢٠٠٠	١٢٤	٤٢،٨٣١٥	٢،٠٨٨٠	١١٩	٤١،٦٥٥٧	٠،٩٢١٢	بسيط	٠،٠٥	*٠،٠٠٠٠	**٠،٩٧٣
(٤٢،٩ ، ٤١،٧)	(٤ ، ١)	١٠٠٠	١٢٤	٤٢،٧٦٦٠	٠،٨٧٢١	١١٩	٤١،٤٢٢٣	١،٩٤٩٢	بسيط	٠،٠٥	*٠،٠٠٠١	**٠،٩٩٩
		٢٠٠٠	١٢٤	٤٢،٧٦٦٠	٠،٨٧٢١	١١٩	٤١،٤٢٢٣	١،٩٤٩٢	بسيط	٠،٠٥	*٠،٠٠٠٠	**٠،٩٩٧
(٤٢،٩ ، ٤١،٧)	(١ ، ٩)	١٠٠٠	١٢٤	٤٢،٩٤٦٧	٢،٩٢٧٠	١١٩	٤١،٦٦٢٦	١٠،٠٥٤٧	متوسط	٠،٠٥	*٠،٠٠٠١	**٠،٩٢٣
		٢٠٠٠	١٢٤	٤٢،٩٤٦٧	٢،٩٢٧٠	١١٩	٤١،٦٦٢٦	١٠،٠٥٤٧	متوسط	٠،٠٥	*٠،٠٠٠٠	٠،٧٨١
(٤٢،٩ ، ٤١،٧)	(٩ ، ١)	١٠٠٠	١٢٤	٤٢،٨٢٦	١،٠٢١	١١٩	٤١،٦٤١٨	٢،٨٥٢٩	متوسط	٠،٠٥	*٠،٠٠٠١	**٠،٨٥٩
		٢٠٠٠	١٢٤	٤٢،٨٢٦	١،٠٢١	١١٩	٤١،٦٤١٨	٢،٨٥٢٩	متوسط	٠،٠٥	*٠،٠٠٠٠	٠،٦٦٥
(٤٢،٩ ، ٤١،٧)	(١ ، ٢٥)	١٠٠٠	١٢٤	٤٢،٣٥٣١	٥،٢٢٨٦	١١٩	٤١،٦٩٢٦	٠،٩٩٦٤	حاد	٠،٠٥	*٠،١٤٦	٠،٥٠٦
		٢٠٠٠	١٢٤	٤٢،٣٥٣١	٥،٢٢٨٦	١١٩	٤١،٦٩٢٦	٠،٩٩٦٤	حاد	٠،٠٥	*٠،١٥٩	٠،٥١٩
(٤٢،٩ ، ٤١،٧)	(٢٥ ، ١)	١٠٠٠	١٢٤	٤٣،٠٢٥٠	٠،٩٩٩٩	١١٩	٤٢،٥٢٠٨	٤،٤٢٨٩	حاد	٠،٠٥	*٠،٢٧٢	٠،٤٧٨
		٢٠٠٠	١٢٤	٤٣،٠٢٥٠	٠،٩٩٩٩	١١٩	٤٢،٥٢٠٨	٤،٤٢٨٩	حاد	٠،٠٥	*٠،٢٨٧	٠،٤٨٨

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠،٨ .

ويتضح من الجدول (٧) السابق أن إجراء Bootstrap حافظ على تقديرات الخطأ من النوع الأول (٠،٠٠٠٠ ، ٠،٠٠٠١) بحسب اختبار Z_{Kohr} فى مستوى أقل من مستوى الدلالة الإحصائية الإسمى (٠،٠٥) فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستويين البسيط والمتوسط ، بينما تميز بتقديرات مرتفعة للخطأ من النوع الأول فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج الموجب ، وتميز بتقديرات متضخمة للخطأ من النوع الأول فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج السالب . وأن إجراء Bootstrap يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى البسيط ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط المصحوب بالتزاوج الموجب ومرات تكرار (١٠٠٠) عينة

Bootstrap ، وفى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط المصحوب بالتزاوج السالب ومرات تكرار (١٠٠٠) عينة Bootstrap . وتميز بتقديرات قوة منخفضة فى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط المصحوب بالتزاوج الموجب ومرات تكرار (٢٠٠٠) عينة Bootstrap ، وتميز بتقديرات قوة منخفضة فى حالة انتهاك افتراض تجانس التباين فى المستوى المتوسط المصحوب بالتزاوج السالب ومرات تكرار (٢٠٠٠) عينة Bootstrap ، وتميز بتقديرات قوة منخفضة جداً فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد وبخاصة المصحوب بالتزاوج السالب . وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين فى المستوى الحاد يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الكبيرة لطريقة Bootstrap ، كما تتفق هذه النتيجة مع ما توصلت إليه نتائج دراسة Lansing (1999) بأن إجراء Bootstrap يتمتع بخصائص قوة متميزة فى حالة العينات الكبيرة .

نتائج اختبار الفرض الرابع

ينص الفرض الرابع على : لا يوجد أثر دال لمستويات متباينة من أزواج التباينات على تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap بتكرار (١٠٠٠ ، ٢٠٠٠) عينة Bootstrap واختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة جداً ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات الخطأ من النوع الأول والقوة الإحصائية على النحو المبين بالجدولين التاليين :

جدول (٨)

تقديرات الخطأ من النوع الأول والقوة لاختبار "ت" ذو التباين الممزوج فى حالة العينات الكبيرة جداً

بارامترات المجتمع		المجموعة الأولى			المجموعة الثانية			مستوى الانتهاك	α_n	α_a	P
μ	σ^2	n_1	$\bar{x}_1$	S_1	n_2	$\bar{x}_2$	S_2				
(٤١،٧ ، ٤٢،٩)	(١ ، ١)	٢٨٩	٤٢،٨٤	١،٠٠١	٢٦٠	٤١،٦٧٥	٠،٩٣١	لا يوجد	٠،٠٠٥	٠،٠٠٠٠*	١،٠٠٠٠**
(٤١،٧ ، ٤٢،٩)	(١ ، ٤)	٢٨٩	٤٢،٩٠	٢،٠٠٤	٢٦٠	٤١،٧٤٣	٠،٩٧٨	بسيط	٠،٠٠٥	٠،٠٠٠٠*	٠،٩٩٩**
(٤١،٧ ، ٤٢،٩)	(٤ ، ١)	٢٨٩	٤٢،٩١	١،٠٠٣	٢٦٠	٤١،٧٨	٢،٠٠٧	بسيط	٠،٠٠٥	٠،٠٠٠٠*	٠،٩٩٩**
(٤١،٧ ، ٤٢،٩)	(١ ، ٩)	٢٨٩	٤٢،٨٩	٣،١٥	٢٦٠	٤١،٦٨	١،٠٠٢	متوسط	٠،٠٠٥	٠،٠٠٠٠*	٠،٩٩٦**
(٤١،٧ ، ٤٢،٩)	(٩ ، ١)	٢٨٩	٤٢،٩٣٨	٠،٩٩٩	٢٦٠	٤١،٥٨	٢،٩٦	متوسط	٠،٠٠٥	٠،٠٠٠٠*	٠،٩٩٩**
(٤١،٧ ، ٤٢،٩)	(١ ، ٢٥)	٢٨٩	٤٣،٠٨	٥،٠٠	٢٦٠	٤١،٨٢	١،٠١	حاد	٠،٠٠٥	٠،٠٠٠٠*	٠،٥٦٧
(٤١،٧ ، ٤٢،٩)	(٢٥ ، ١)	٢٨٩	٤٢،٨٧	١،٠٠٥	٢٦٠	٤٢،٣٣	٥،٥٢	حاد	٠،٠٠٥	٠،٠١٠٤*	٠،٥٥٠

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kohr} ، ** القوة تتجاوز النقطة المرجعية ٠،٨ .

يتضح من الجدول (٨) السابق أن اختبار "ت" ذو التباين الممزوج تميز بتقديرات منخفضة جداً للخطأ من النوع الأول أقل من مستوى الدلالة الإسمى (٠،٠٠٥) بحسب اختبار Z_{Kohr} فى حالة

تجانس التباين ، وفى حالة انتهاك إفتراض تجانس التباين فى المستويين البسيط والمتوسط ، وفى حالة انتهاك تجانس التباين فى المستوى الحاد المصحوب بالتزاوج الموجب ، بينما تميز بتقديرات مرتفعة فى حالة انتهاك افترض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج السالب .

وأن اختبار "ت" ذو التباين الممزوج يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستويين البسيط والمتوسط . وتميز الاختبار بتقديرات قوة منخفضة فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين فى المستوى الحاد يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الكبيرة جداً لاختبار "ت" ذو التباين الممزوج ، ومع ما ورد فى الأدبيات مثل (Ruxton, 2006: 688) بأن اختبار "ت" يؤدى بشكل غير جيد من حيث الخطأ من النوع الأول فى حالة عدم تساوى أزواج التباينات .

أما بشأن أداء الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الكبيرة جداً ومرات تكرار (١٠٠٠ ، ٢٠٠٠) عينة فهو على النحو المبين بالجدول التالى :

جدول (٩)

تقديرات الخطأ من النوع الأول والقوة لإجراء Bootstrap فى حالة العينات الكبيرة جداً

بارامترات المجتمع μ	بارامترات المجتمع σ^2	عدد العينات	المجموعة الأولى n_1	المجموعة الأولى $\bar{x}_1$	المجموعة الأولى S_1	المجموعة الثانية n_2	المجموعة الثانية $\bar{x}_2$	المجموعة الثانية S_2	مستوى الانتهاك	α_n	α_a	P
(٤٢،٩ ، ٤١،٧)	(١ ، ١)	١٠٠٠	٢٨٩	٤٢،٨٣٥٨	١٠،٠١٢٨	٢٦٠	٤١،٦٧٤٩	٠،٩٣١١	لا يوجد	٠،٠٥	*٠،٠٠١	**١،٠٠٠
		٢٠٠٠	٢٨٩	٤٢،٨٣٥٨	١٠،٠١٢٨	٢٦٠	٤١،٦٧٤٩	٠،٩٣١١	لا يوجد	٠،٠٥	*٠،٠٠٠	**١،٠٠٠
(٤٢،٩ ، ٤١،٧)	(١ ، ٤)	١٠٠٠	٢٨٩	٤٢،٨٩٧٩	٢٠،٠٣٧٦	٢٦٠	٤١،٧٤٣٣	٠،٩٧٨٠١	بسيط	٠،٠٥	*٠،٠٠١	**٠،٩٩٩
		٢٠٠٠	٢٨٩	٤٢،٨٩٧٩	٢٠،٠٣٧٦	٢٦٠	٤١،٧٤٣٣	٠،٩٧٨٠١	بسيط	٠،٠٥	*٠،٠٠٠	**٠،٩٩٩
(٤٢،٩ ، ٤١،٧)	(٤ ، ١)	١٠٠٠	٢٨٩	٤٢،٩٠٨٣	١٠،٠٣٣٩	٢٦٠	٤١،٧٨١١	٢٠،٠٧٠٣	بسيط	٠،٠٥	*٠،٠٠١	**٠،٩٩٩
		٢٠٠٠	٢٨٩	٤٢،٩٠٨٣	١٠،٠٣٣٩	٢٦٠	٤١،٧٨١١	٢٠،٠٧٠٣	بسيط	٠،٠٥	*٠،٠٠٠	**٠،٩٩٩
(٤٢،٩ ، ٤١،٧)	(١ ، ٩)	١٠٠٠	٢٨٩	٤٢،٨٨٨٣	٣٠،١٤٨٨	٢٦٠	٤١،٦٧٥٤	١٠،٠١٦٨	متوسط	٠،٠٥	*٠،٠٠١	**٠،٩٩٩
		٢٠٠٠	٢٨٩	٤٢،٨٨٨٣	٣٠،١٤٨٨	٢٦٠	٤١،٦٧٥٤	١٠،٠١٦٨	متوسط	٠،٠٥	*٠،٠٠٠	**٠،٩٩٥
(٤٢،٩ ، ٤١،٧)	(٩ ، ١)	١٠٠٠	٢٨٩	٤٢،٩٣٧٥	٠،٩٩٩٤	٢٦٠	٤١،٥٨٠٢	٢،٩٥٩٥	متوسط	٠،٠٥	*٠،٠٠١	**٠،٩٩٩
		٢٠٠٠	٢٨٩	٤٢،٩٣٧٥	٠،٩٩٩٤	٢٦٠	٤١،٥٨٠٢	٢،٩٥٩٥	متوسط	٠،٠٥	*٠،٠٠٠	**٠،٩٩٩
(٤٢،٩ ، ٤١،٧)	(٢٥ ، ١)	١٠٠٠	٢٨٩	٤٣،٠٧٧٥	٥،٠٠٢١	٢٦٠	٤١،٨١٧٢	١٠،٠١١٤	حاد	٠،٠٥	*٠،٠٠١	٠،٧٨٢
		٢٠٠٠	٢٨٩	٤٣،٠٧٧٥	٥،٠٠٢١	٢٦٠	٤١،٨١٧٢	١٠،٠١١٤	حاد	٠،٠٥	*٠،٠٠٠	٠،٥٦٨
(٤٢،٩ ، ٤١،٧)	(١ ، ٢٥)	١٠٠٠	٢٨٩	٤٢،٨٧١٩	١٠،٠٥٢١	٢٦٠	٤٢،٣٣٢٥	٥،٠٥٢٠٩	حاد	٠،٠٥	*٠،١١٥	٠،٥٦٩
		٢٠٠٠	٢٨٩	٤٢،٨٧١٩	١٠،٠٥٢١	٢٦٠	٤٢،٣٣٢٥	٥،٠٥٢٠٩	حاد	٠،٠٥	*٠،١٢٥	٠،٥٨٥

ملحوظة : α_n مستوى الدلالة الإسمى ، α_a مستوى الدلالة الفعلى ، P قوة الاختبار .

* الفرق دال إحصائياً باستخدام Z_{Kalc} ، ** القوة تتجاوز النقطة المرجعية ٠،٨ .

يتضح من الجدول (٩) السابق أن إجراء Bootstrap تميز بتقديرات منخفضة جداً للخطأ من النوع الأول فى مستوى أقل من مستوى الدلالة الإحصائية الإسمى (٠،٠٥) بحسب اختبار Z_{Kohr} فى حالة تجانس التباين ، وفى حالات انتهاك افتراض تجانس التباين فى المستويين البسيط والمتوسط ، بينما تميز بتقديرات مرتفعة إلى حد ما للخطأ من النوع الأول فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد المصحوب بالتزاوج السالب .

وأن إجراء Bootstrap يتمتع بخصائص قوة مرتفعة تتجاوز النقطة المرجعية ٠،٨ فى حالة تجانس التباين ، وفى حالة انتهاك افتراض تجانس التباين فى المستويين البسيط والمتوسط . وتميز بتقديرات قوة منخفضة فى حالة انتهاك افتراض تجانس التباين فى المستوى الحاد .

وتتفق هذه النتيجة التى تم التوصل إليها جزئياً مع ما توصلت إليه نتائج دراستى (Kang, Harring, 2012; Kang et al., 2014) بأن انتهاك افتراض تجانس التباين فى المستوى الحاد يؤدى إلى زيادة تقدير الخطأ من النوع الأول وانخفاض القوة الإحصائية فى حالة العينات الكبيرة جداً لطريقة Bootstrap ، كما تتفق هذه النتيجة مع ما توصلت إليه نتائج دراسة Lansing (1999) بأن إجراء Bootstrap يتمتع بعامة خصائص قوة متميزة فى حالة العينات الكبيرة جداً .

توصيات الدراسة

فى ضوء النتائج التى توصلت إليها الدراسة الحالية ، يورد الباحث التوصيات الآتية :

١. ضرورة اختبار افتراض تجانس التباين باستخدام أحد الاختبارات المناسبة ومن أمثلتها اختبار Levene L Test قبل البدء فى اختبار الفروض باستخدام اختبار "ت" لعينتين مستقلتين .
٢. يُعد الحجم الكبير للعينة أحد أهم ضمانات الاستدلال الإحصائى الجيد فى حالة استخدام اختبار "ت" ذو التباين الممزوج .
٣. يفضل التزاوج الموجب (الذى يحدث عندما تكون العينة الكبيرة الحجم منتمية للمجتمع ذى التباين الكبير والعينة صغيرة الحجم منتمية للمجتمع ذى التباين الصغير) عن التزاوج السالب عند استخدام اختبار "ت" ذو التباين الممزوج أو طريقة Bootstrap مع العينات الكبيرة والكبيرة بدرجة كافية .
٤. العدد (١٠٠٠) عينة Bootstrap كاف لتنفيذ إجراء Bootstrap متميزاً بالقوة الإحصائية فى حالة العينات المتوسطة فأكثر .
٥. التوجه لاستخدام أحد الصور المعدلة لاختبار "ت" لتناسب الموقف المتكرر الخاص بعدم تساوى أو تكافؤ التباينات وهو اختبار "ت" ذو التباينات غير المتساوية t-Test for

Unequal Variances فى حالة تعذر استخدام إجراء Bootstrap أو أحد البدائل الأخرى المبنية على إعادة المعاينة مثل اختبار العشوائية أو اختبار التباديل

٦. تطوير محتوى مقرر الإحصاء النفسى والتربوى لطلاب مرحلة الدراسات العليا بما يمكنهم من إكتساب مهارات التعامل مع الاختبارات والحزم الإحصائية .

المراجع

أحمد سليمان عودة ، خليل يوسف الخليلى (١٩٨٨) . الإحصاء للباحث فى التربية والعلوم الإنسانية . عمان : دار الفكر للطباعة والنشر والتوزيع .

ج. ملتون سميث (١٩٧٨) . الدليل إلى الإحصاء فى التربية وعلم النفس (ترجمة : إبراهيم بسيونى عميرة) . القاهرة : دار المعارف .

عبد الرحمن عدس (١٩٩٧) . مبادىء الإحصاء فى التربية وعلم النفس (الجزء الثانى : مبادىء الإحصاء التحليلى) (ط٢) . عمان : دار الفكر للطباعة والنشر والتوزيع .

عبد الناصر السيد عامر (٢٠١٢) . النماذج والاختبارات الإحصائية المستخدمة فى البحث النفسى المصرى والعربى . المجلة المصرية للدراسات النفسية ، ٢٢(٧٤) ، ٣٥٥-٣٧١ .

عبد العاطى أحمد الصياد ، عبد الناصر السيد عامر (٢٠١٣) . نحو معيار موحد لتقدير حجم التأثير لاختبار "ت" لعينتين مستقلتين فى البحث التربوى النفسى العربى . المجلة المصرية للدراسات النفسية ، ٢٣(٨٠) ، ٥-٢١ .

فؤاد أبو حطب ، آمال صادق (١٩٩٦) . مناهج البحث وطرق التحليل الإحصائى فى العلوم النفسية والتربوية والاجتماعية (ط٢) . القاهرة : مكتبة الأنجلو المصرية .

ل. ر. جاى ، ج. إ. ميلز ، ب. إيراسيان (٢٠١٢) . البحث التربوى : كفايات للتحليل والتطبيقات (ترجمة : صلاح الدين محمود علام) .عمان : دار الفكر للطباعة والنشر والتوزيع .

ميخائل أسعد (١٩٩٠) . الإحصاء النفسى وقياس القدرات الإنسانية . بيروت : دار الأفاق الجديدة .

Ahad, N. A., Abdullah, S., Heng, L. C., & Ali, N. M. (2012). Relative power performance of t-test and bootstrap procedure for two-sample. *Pertanika Journal of Science & Technology, 20*(1), 43-52.

Akpanta, A., & Okorie, I. (2015). Investigating the significance of a correlation coefficient using Jackknife estimates. *International Journal of Sciences: Basic and Applied Research, 22*(2), 441-448.

Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational*

Statistics, 19, 91-101.

Aron, A., & Aron, E. N. (1994). *Statistics for psychology*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Bartz, A. E. (1988). *Basic statistical concepts* (3rd ed.). New York: Macmillan Publishing Company.

Berger, D. E. (2007). *A gentle introduction to resampling. Demonstration* Presented at the Annual Meeting of the American Evaluation Association, Baltimore, Maryland.

Best, J. W., & Kahn, J. V. (2006). *Research in education* (8th ed.). Boston: Pearson Education Inc.

Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples t test to that of Wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin, 97*(1), 119-128.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t-test. *Psychological Bulletin, 57*(1), 49-64.

Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test equality of several means. *Technometrics, 16*, 129–132.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. New York: Cambridge University Press.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American, 248*(5), 116-130.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford University Press.

Edgington, E. S. (1995). *Randomization tests*. New York: M. Dekker.

Efron, B. (1979a). Bootstrap Methods: Another look at the Jackknife. *Annals of Statistics, 7*, 1-26.

Efron, B. (1979b). Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review, 21*(4), 460-480.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.

Finch, S., Thomason, N., & Cumming, G. (2002). Past and future American Psychological Association guidelines for statistical practice. *Theory and Psychology, 12*, 825–853.

Fisher, R. A. (1935/1960). *The design of experiments* (7th ed.). New York: Hafner Publishers.

Fradette, K., Keselman, H. J., Lix, L., Algina, J., & Wilcox, R. R. (2003). Conventional and robust paired and independent-samples t tests: Type I error and power rates, *Journal of Modern Applied Statistical Methods, 2*(2), 481-496.

Harwell, M. R. (1988). Choosing between parametric and non-parametric tests. *Journal of Counseling and Development, 67*, 35-38.

Harwell, M. R., & Serlin, R. C. (2001, April). *Review of non-parametric tests for complex experimental designs in educational research.* Paper Presented at the Annual of the American Educational Research Association (WA : Seattle).

Hess, B., Olejnik, S., & Huberty, C. J. (2001). The efficacy of two improvement-over-chance effect sizes for two-group univariate comparisons under variance heterogeneity and nonnormality. *Educational and Psychological Measurement, 61*, 909-936.

Higgins, G. E. (2005). Statistical significance testing: The bootstrapping method and an application to self-control theory. *The Southwest Journal of Criminal Justice, 2*(1), 54-75.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika, 38*, 324-329.

Kang, Y., & Harring, J. R. (2012). *Investigating the impact of non-normality, effect size, and sample size on two-group comparison procedures: An empirical study.* Paper Presented at the Annual Meeting of the American Educational Research Association.

Kang, Y., Harring, J. R., & Li, M. (2014). Reexamining the impact of nonnormality in two-group comparison procedures. The *Journal of Experimental Education, 83*(2), 1–28.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analysis. *Review of Educational Research, 68*(3), 350-386.

Kohr, R. L. A (1970). *Comparison of statistical procedures for testing* $\mu_1 = \mu_2$ *unequal n's and variances* (Unpublished doctoral dissertation). Pennsylvania State University.

Koopman, J., Howe, H., Hollenbeck, J., & Sin, H. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology, 100*(1), 194-202.

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis, 51*(12), 5731-5742.

Kulkarni, S. (1993). *A comparison of type I error rates for the bootstrap contrast with the t test and the robust rank order test for various samples sizes and variances* (Unpublished master's thesis). Lehigh University, Bethlehem, Pennsylvania.

Kurtz, A. K. (1948). A research test of Rorschach test. *Personnel Psychology, 1*, 41-53.

Lansing, L. L. (1999). Bootstrapping versus the student's t: the problems of type I error and power (Unpublished master thesis). Lehigh University, Bethlehem, Pennsylvania.

Li, H. (2011). *Power analysis for alternative tests for the equality of means* (Unpublished master thesis). East Tennessee State University.

Little, R. J. A. (1989). Testing the equality of two independent binomial proportions. *The American Statistician, 43*(4), 283-288.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health, 23*, 151-169.

Lunneborg, C. E. (2000). *Data analysis by resampling: Concepts and applications*. Pacific Grove, CA: Duxbury.

Maggio, S., & Sawilowsky, S. S. (2014). A new maximum test via the dependent samples t-test and the Wilcoxon gigned-ranks test. *Applied Mathematics, 5*, 110-114.

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (2nd ed.). London: Chapman & Hall.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Mosier, C. I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement, 11*, 5-11.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. (2003). *An improved robust Welch-James test statistic*. In the proceeding of the Regional Conference on Integrating Technology in the Mathematical Sciences, 2003. Universiti Sains Malaysia, Pulau Pinang, Malaysia.

Ozdemir, A. F. (2013). Comparing two independent groups: A test based on a one-step M-estimator and bootstrap-t. *Britich Journal of Mathematical and Statistical Psychology, 66*, 322-337.

Ozdemir, A. F., & Kurt S. (2006). One way fixed effect analysis of variance under variance heterogeneity and a solution proposal. *Selc,uk Journal of Applied Mathematics, 7*, 81–91.

Pagano, R. R. (2010). *Understanding statistics in the behavioral sciences* (9th ed.). Belmont: Wadsworth, Cengage Learning.

Palomares, R. S. (1990, November). *Alternatives to statistical significance testing*. Paper Presented at the Annual Meeting of the Mid-South Educational Research Association (LA: New Orleans).

Peterson, I. (1991). Pick a sample. *Science News, 140*, 56-58.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Soc. Series B, 11*, 18-84.

Quenouille, M. (1956). Notes on bias in estimation. *Biometrika, 43*, 353-360.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Reddy, M. K., Boiroju, N. K., Yerukala, R., & Rao, M. V. (2011). Bootstrap graphical test for equality of variances. *Electronic Journal of Applied Statistical Analysis, 4*(2), 184-188.

Rudner, L. M., & Shafer, M. M. (1992). *Resampling: a marriage of computers and statistics*. Practical Assessment, Research & Evaluation, 3(5). Available online: http://PAREonline.net/getvn.asp?v=3&n=5

Rusticus, S. A., & Lovato, C. Y. (2011). *Applying tests of equivalence for multiple group comparisons: Demonstration of the confidence interval approach*. Practical Assessment, Research & Evaluation, 16(7). Available online: http://pareonline.net/getvn.asp?v=16&n=7

Rusticus, S. A., & Lovato, C. Y. (2014). *Impact of sample size and variability on the power and type I error rates of equivalence tests: A simulation study*. Practical Assessment, Research & Evaluation, 19(11). Available online: http://pareonline.net/getvn.asp?v=19&n=11.

Ruxton, G. D. (2006). The unequal variance t-test in an underused alternative to student's t-test and the Mann-Whitney U test. *Behavioral Ecology, 17*(4), 688-690.

Saculinggan, M., & Balase, E. A. (2013). Empirical power comparison of goodness of fit tests for normality in the presence of outliers. *Journal of Physics: Conference Series, 435*(1), 1-11.

Scheffe, H. (1970). Practical solutions of the Behrens-Fisher problem. *Journal of the American Statistical Association, 65*(332), 1501-1508.

Schieber, F. (2013). *Resampling: The new statistics*. Heimstra Labs Colloquium.

Snedecor, G. W. (1946). *Statistical methods* (5th ed.). Ames, Iowa: The Iowa State College Press.

Soper, D.S. (2015). *Post-hoc statistical power calculator for a student t-test* [Software]. Available from http://www.danielsoper.com/statcalc

Sprinthall, R. C. (1990). *Basic statistical analysis* (3rd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Strube, M. J. (1988). Bootstrap type I error rates for the correlation coefficient: An Examination of procedures. *Psychological Bulletin, 104*(2), 290-292

Thompson, B. (1992, April). *The use of statistical significance tests in research: Some criticisms and alternatives*. Paper Presented at the Annual Meeting of the American Educational Research Association. (CA : San Francisco).

Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the Journal of Experimental Education. *Journal of Experimental Education, 66*, 75-83.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics, 29*, 614.

Weinberg, S., & Goldberg, K. (1990). *Statistics for the behavioral sciences*. New York: Camridge University Press.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38*, 330-336.

Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrical Journal, 32*, 771–780.

Wilcox, R. R. (2002). Comparing the variances of two independent groups. *Britich Journal of Mathematical and Statistical Psychology, 55*, 169-175.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Waltham, MA: Academic Press.

Wooldridge, J. (2013). *Introductory econometrics: Modern approach*. New York: Content Technologies, Inc.

Yin, T. S., Yusof, Z. M., Yaacob, C. R., & Othman, A. (2010). Performance of the traditional pooled variance t-test against the bootstrap procedure of difference between sample means. *Malaysian Journal of Mathematical Sciences, 4*(1), 85-94.

Yu, C. H. (2003). *Resampling methods: concepts, applications, and justification*. Practical Assessment, Research & Evaluation, 8(19). Retrieved July 31, 2015 from http://PAREonline.net/getvn.asp?v=8&n=19.

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology, 56*, 93-110.

Zumbo, B. D., & Jennings, M. J. (2002). The robustness of validity and efficiency of the related samples t-test in the presence of outliers. *Psicológica, 23*, 415-450.

The Impact of Different Pairs of Variances on Type I Error Rates and Power of the Bootstrap Procedure and the Pooled Variance t-Test: A Simulation Study

By
Mahsoub Abdelkader Aldowy Hassan
Associated Professor of Educational Psychology
Dept. of Educational Psychology
Faculty of Education at Qena
South Valley University

The purpose of this study was to test impact of different pairs of variances on type I error rates and power of the bootstrap procedure and the pooled variance t-test. This study was based on simulated data, data were generated using Minitab to draw (56) samples from normally distributed populations.

The study conditions that considered were as follows:
Group sample sizes: four different group samples (23, 16), (68, 51), (134, 119), (289, 260) were proposed relating to small, medium, large and sufficiently (very) large samples sizes.
Distribution: normal.
Bootstrapping: based on (1000, 2000) bootstrap samples
Variances: pairs variances [(1, 1), (4, 1), (1, 4), (9, 1), (1, 9), (25, 1), (1, 25)].

Results of the study showed:-

- Heteroscedasticity can have a pronounced effect on the actual significant level (α_a) compared to the nominal significant level (α_n), the (α_a) of the pooled variance t-test exceeds when smaller samples are drawn from the variable populations.
- The bootstrap is just as powerful as the pooled variance t-test when sample sizes are large but does't perform well when sample sizes are small.
- Increasing the sample size increases statistical power for both the bootstrap procedure and the pooled varience t-test.
- The pooled variance t-test and the bootstrap procedure controlled the type I error rates when sample size were large or very large. However, the type I error rates were lightly inflated when sample sizes were small or medium.
- There was no need to use 2000 bootstrap samples.

The study recommended that researchers check the assumption of homogeneity of variance when they try to compare mean difference between two groups. Researchers might use the bootstrap procedure as an alternate method to pooled variance t-test in the case of the existence of Heteroscedasticity.