# دراسة مقارنة بين اختبار t واختبار ولكوكسن باستخدام أسلوب المحاكاة

**د. عبدالحليم مولود الصويعي**
**قسم الإحصاء – كلية العلوم – الزاوية**
**جامعة الزاوية**

## مقدمة:

عند دراسة المقارنة بين مجموعتين مستقلتين من البيانات خصوصا عندما يكون حجم العينة صغير، يكون التساؤل أيهما أفضل استخدام اختبار معلمي أو اختبار غير معلمي؟ العديد من الدراسات تمت في هذا المجال لمحاولة تحديد الاختبار المناسب والأفضل. المقارنة بين اختبار t لعينتين مستقلتين واختبار ولكوكسن للرتب ويعرف أيضا باختبار مان وتني ( Weber and Sawilowsky, 2009) أحد هذه المقارنات التي تمت فيها العديد من الدراسات في حالات مختلفة. حيث يعتبر اختبار t اختبار معلمي بينما اختبار ولكوكسن لامعلمي.

لاستخدام اختبار t كاختبار معلمي، هناك فروض يجب توفرها في البيانات. أحد هـــذه الفروض هو أن البيانات يجب أن تكون مسحوبة عشوائيا من مجتمع له توزيع طبيعـــي. فـــي بعض الحالات هذا الفرض للبيانات لا يتحقق، خصوصا في الحياة العملية. فهـــل نســـتمر فـــي استخدام اختبار t أو الأفضل إيجاد اختبار آخر بديل استخدامه لايكون مرتبط بتوزيع البيانـــات؟ الاختبارات اللامعلمية تعتبر البديل المناسب في هذه الحالة. من هذه الاختبـــارات اللامعلميـــة، اختبار ولكوكسن للرتب.

عند المقارنة بين اختبارين مختلفين لنفس الفرضية، تكون المقارنة عادة من حيث الخطأ من النوع الأول حيث يتم دراسة مدى ثبات الخطأ من النوع الأول لكل اختبار حول قيمة محددة معطاة (عادة 0.05)، وكذلك من حيث قوة الاختبار. فإذا ما كان حجم الخطأ من النـــوع الأول لكل اختبار قريبين من بعض، فإن الاختبار الأفضل يكون الاختبار الذي له قوة اختبار أكبر.

عند عدم توفر فرض التوزيع الطبيعي للبيانات، وجد أن حجم الخطأ من النـــوع الأول عند استخدام اختبار t ثابت ومستقر عندما يكون حجم العينة كبير (n=30) وتوزيع البيانات غير ملتو بشكل كبير (Sawilowsky and Blair, 1992). كذلك نفس الباحثين وجد أنه عندما يكوم حجم العينة صغير (n<20) وتوزيع البيانات ملتو بدرجة متوسطة أو كبيرة، فإن حجم الخطأ من النوع الأول لاختبار t يكون غير ثابت.

اختبار ولكوكسن للرتب (Wilcoxon, 1945) هو الاختبار اللامعلمي البديل لاختبار t التقليدي الذي يوصى باستخدامه عند عدم توفر فرض التوزيع الطبيعي للبيانات وحجـــم العينـــة يكون صغير (Sawilowsky and Blair, 1999) . بشكل عام، اختبار ولكوكسن للرتب يعتبر

له قوة اختبار أكبر من اختبار t لعينتين مستقلتين عندما تكون البيانات مسحوبة من مجتمعات لها توزيع غير طبيعي (Blair and Higgins, 1980). اختبار ولكوكسن له قوة اختبار تصل إلى أربعة أضعاف أكبر من اختبار t عندما تكون البيانات مسحوبة من توزيع أسي (Sawilowsky and Blair, 1992). عليه يمكن القول في حالة التوزيعات غير الطبيعية، خاصة شديدة الالتواء مثل التوزيع الأسي والتوزيع اللوغاريتمي أو التوزيعات التي بها قيم شاذة، فإن اختبار ولكوكسن يكون له قوة اختبار أكبر من اختبار t (MacDonald, 1999). أيضا حجم الخطأ من النوع الأول يكون بعيد عن قيمة مستوى المعنوية المحددة عندما يتم استخدام التباين التجميعي في حالة عدم تساوي التباينين وحجم العينتين غير متساوي ( Winter and Dodou, 2012 ). السؤال هنا يبقى ماذا لو كانت البيانات مسحوبة من مجتمع غير طبيعي ولكنه متماثل (Symmetric)؟ هل أن اختبار ولكوكسن يعتبر مازال أفضل أم أن الإجابة ستختلف؟

الغرض من هذه البحث هو التحقق من السؤال أعلاه وذلك بدراسة مقارنة بين اختبار t واختبار ولكوكسن باستخدام أسلوب المحاكاة لبيانات مسحوبة من توزيع غير طبيعي آخر وهو التوزيع المنتظم. ومن المعروف أن التوزيع المنتظم هو توزيع غير طبيعي ولكنه متماثل (Symmetric).

## اختبار t لعينتين مستقلتين:

اختبار t لعينتين مستقلتين عادة يستخدم عندما تكون الرغبة في الحصول على استدلال حول مجتمعين مستقلين من خلال المقارنة بين متوسطي عينتين مسحوبتين من هذين المجتمعين

(Hinkle, Wiersma and Jurs, 2003). اختبار t لاختبار تساوي عينتين مستقلتين لهما حجم $n_1$ , $n_2$ على التوالي، تعرف كالتالي:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

حيث أن:

$\bar{X}_2, \bar{X}_1$ يمثلان متوسط ي العينة الأول والعينة الثانية على التوالي.

$\mu_2, \mu_1$ يمثلان متوسطي المجتمعين المأخوذ منهما العينتين.

$n_1$ , $n_2$ يمثلان حجمي العينتين الأولى والثانية على التوالي.

$S_P^2$ يمثل التباين التجميعي للعينتين تحت فرض تجانس تبـــايني المجتمعـــين، ويعـــرف كالآتي:

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

حيث أن:

$S_1^2$ و ${}_2^2$ يمثلان تبايني العينتين الأولى والثانية على التوالي.

وبما أن الاختبار يكون حول تساوي متوسطي المجتمعين، فإنه يمكن إعادة كتابة اختبار $t$ بالصيغة

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{S_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

أصبح موثقا جيدا أن استخدام اختبار $t$ لعينتين مستقلتين ربما يكون غير ملائم عنـــدما تكون أحد فروضه غير متوفرة. خصوصا عندما يكون توزيع البيانات غير طبيعي لأن له قـــوة

اختبار أقل مـــن الاختبـــار اللامعلمـــي البـــديل ( Bridge and Sawilowsky, 1999 and MacDonald, 1999).

## اختبار ولكوكسن للرتب:

اختبار ولكوكسن للرتب (Wilcoxon, 1945) هو الاختبار اللامعلمي البديل لاختبار t التقليدي الذي يوصى باستخدامه عند عدم توفر فرض التوزيع الطبيعي للبيانات وحجـــم العينـــة يكون صغير (Blair and Sawilowsky, 1999). هذا الاختبار يستخدم الرتب للبيانات بـــدلا من البيانات نفسها. حيث أن البيانات الأصلية يتم تحويلها إلى رتب بعد دمج المجموعتين. يـــتم مقارنة مجموع الرتب الأصغر من بين المجموعتين بالقيمة الحرجة لاختبار ولكوكسن. إذا كان حجم العينتين أكبر من 10، فإن توزيع المعاينة لاختبار ولكوكسن للرتب يقترب مـــن التوزيـــع الطبيعي، واختبار z يستخدم في هـــذه الحالـــة (Ott and Longnecker, 2010). الصـــيغة المستخدمة لاختبار تساوي توزيع المجتمعين عرف كالآتي:

$$z = \frac{T - \mu_T}{\sigma_T}$$

حيث أن :

T: تمثل مجموع رتب المشاهدات

$\mu_T$: متوسط توزيع $T$

$\sigma_T$ : الانحراف المعياري لتوزيع $T$

وأن :

$$\mu_T = \frac{\mathrm{n}_1(\mathrm{n}_1 + \mathrm{n}_2)}{2}$$

$$\sigma_T = \sqrt{\frac{n_1 n_2}{12}(n_1 + n_2 + 1)}$$

مع ملاحظة أنه لاستخدام اختبار ولكوكسن، يجب أن تكـــون العينتـــين مســـتقلتين وأن المجتمعين المسحوب منهما العينتين معرفين (Ott & Longnecker, 2010).

## استخدام أسلوب المحاكاة للمقارنة بين اختبار t واختبار ولكوكسن:

في عملية المحاكاة هذه، تتم المقارنة من حيث حجـــم الاختبـــار (α) وقـــوة الاختبـــار $(1-\beta)$ لهذين الاختبارين في حالتين مختلفتين، الحالة الأولى تتم بأخذ العينات مـــن مجتمـــع طبيعي التوزيع وفي الحالة الثانية تتم بأخذ العينات من مجتمع له توزيع منتظم، حيث أن التوزيع المنتظم هو توزيع ليس طبيعي ولكنه متماثل.

لدراسة التـــأثير بوضـــوح تـــم الحصـــول علـــى خمـــس عينـــات مختلفـــة بأحجـــام n=8,15,20,25,30 حيث هذه العينات تمثل عينة صغيرة جدا، عينة صغيرة، عينة متوســـطة، عينة كبيرة. عدد مرات المحاكاة في كل مرة هو r=10000.

في كل حالة يتم اختبار تساوي أو عدم تساوي متوسطي المجتمعين المســـحوب منهمـــا العينتين. نتوقع النتائج لحجم الاختبار وقوة الاختبار لكلا الاختبارين ستتغير باختلاف الحالتين، وبمقارنة النتائج سيتضح لنا أي اختبار أكثر أمناً من حيث الاستخدام.

## الحالة الأولى: الاختبارات في حالة مجتمع طبيعي التوزيع:

يتم توليد مجموعتين من البيانات كلاهما من توزيع طبيعي حيث إن مستوى المعنويـــة هو 0.05.

**1) في حالة تساوي متوسطي المجتمعين:**

يتم تقييم حجم الاختبار ($\alpha$) للاختبارين الإحصائيين (اختبار t واختبار ولكوكسن)، ولهذا الغرض تم أخذ متوسطين متساويين $\mu_1 = \mu_2 = 1$ وتباين العينتين مختلفين $\sigma_1^2 = \frac{1}{3}, \sigma_2^2 = \frac{1}{12}$ وبالتالي فإن الفرضية الصفرية هي: $H_0: \mu_1 = \mu_2$

في هذه الدراسة، تم استخدام البرنامج الإحصائي R لتوليد العينتين، بعد ذلك وبافتراض صحة الفرضية الصفرية، تم حساب نسبة رفض الفرضية الصفرية وذلك من خلال تحديد نسبة P-value<0.05 حيث أنه تم استخدام $\alpha = 0.05$ ، فإذا كانت النسبة تحقق مستوى المعنوية المعطى فإننا نحصل على الحجم المناسب للاختبار.

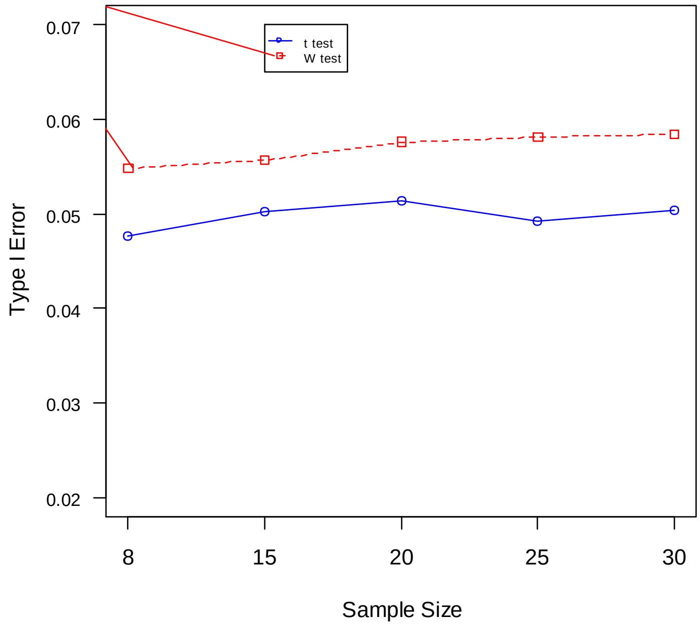
للحصول على حجم الاختبار لاختبار t واختبار ولكوكسن نقوم بإعادة العملية 10000 مره ومن ثم تحديد عدد مرات رفض الفرضية الصفرية (P_value<0.05 ) و بالقسمة على 10000 نحصل على حجم الاختبار الإحصائي. العملية السابقة تم تكرارها لأحجام العينات 8، 15، 20، 25، 30. البيانات المتحصل عليها من عملية المحاكاة تم عرضها في جدول (1)

**جدول رقم (1) يوضح حجم العينة، حجم الاختبار لاختبار t وحجم الاختبار لاختبار ولكوكسن.**

| حجم العينة | حجم اختبار t | حجم اختبار ولكوكسن |
|---|---|---|
| 8 | 0.0477 | 0.0548 |
| 15 | 0.0505 | 0.0557 |
| 20 | 0.0513 | 0.0576 |
| 25 | 0.0492 | 0.0581 |
| 30 | 0.0503 | 0.0584 |

نلاحظ من الجدول (1) إن قيم حجم الاختبار لاختبار t قريبة جدا من مستوى المعنوية المحددة وهي 0.05 ولا تتأثر بزيادة حجم العينة. بينما حجم الاختبار لاختبار ولكوكسن أكبر من مستوى المعنوية المحددة وتزداد مع زيادة حجم العينة.

شكل رقم (1) يوضح تقارب حجم الاختبار لاختبار t من 0.05 أكثر من حجم الاختبار لاختبار ولكوكسن مما يعطي الأفضلية لاختبار t.




**شكل (1) يوضح حجم الاختبار لاختبار t وحجم الاختبار لاختبار ولكوكسن**

من الشكل (1) نلاحظ مدى تقارب حجم الاختبار لاختبار t من مستوى المعنوية المحددة (0.05) أكثر من اختبار ولكوكسن

**2) في حالة عدم تساوي متوسطي المجتمعين:**

يتم إيجاد قوة الاختبار للاختبارين تحت الفرضية البديلة: $H_1: \mu_1 \neq \mu_2$

لتوليد بيانات العينة الأولى يتم استخدام $\mu_1=1,\ \sigma_1^2=\frac{1}{3}$

لتوليد بيانات العينة الثانية يتم استخدام $\mu_2=2,\ \sigma_2^3=\frac{3}{4}$

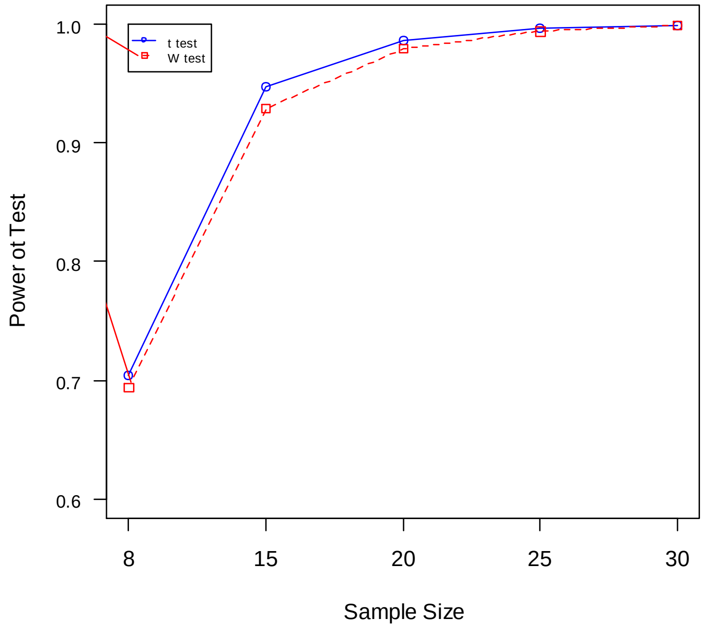
للحصول على قوة الاختبار لاختبار t واختبار ولكوكسن يتم إعادة العملية 10000 مره ومن ثم تحديد عدد مرات رفض الفرضية الصفرية ($P\_value<0.05$ ) و بالقسمة على 10000 نحصل على قوة الاختبار الإحصائي. العملية السابقة تم تكرارها لأحجام العينات 8، 15، 20، 25، 30. البيانات المتحصل عليها من عملية المحاكاة تم عرضها في جدول (2).

**جدول رقم (2) يوضح حجم العينة، قوة الاختبار لاختبار t وقوة الاختبار لاختبار ولكوكسن.**

| حجم العينة | قوة اختبار t | قوة اختبار ولكوكسن |
|---|---|---|
| 8 | 0.7037 | 0.6937 |
| 15 | 0.9469 | 0.9292 |
| 20 | 0.9859 | 0.9794 |
| 25 | 0.9963 | 0.994 |
| 30 | 0.9996 | 0.9989 |

نلاحظ من الجدول (2) أن قيم قوة الاختبار لاختبار t قريبة جدا من قوة الاختبار لاختبار ولكوكسن مع أفضلية بسيطة جدا لاختبار t مع ملاحظة زيادة قوة الاختبار لكلا الاختبارين مع زيادة حجم العينة.

شكل رقم (2) يوضح تقارب قوة الاختبار لاختبار t مع قوة الاختبار لاختبار ولكوكسن.




**شكل (2) يوضح قوة الاختبار لاختبار t وقوة الاختبار لاختبار ولكوكسن**

من الشكل (2) نلاحظ مدى تقارب قوة الاختبار لاختبار t مع قوة الاختبار لاختبار ولكوكسن مع أفضلية بسيطة لاختبار t .

## الحالة الثانية: الاختبارات في حالة مجتمعات منتظمة التوزيع:

يتم توليد مجموعتين من البيانات كلاهما من توزيع منتظم، حيث إن مستوى المعنوية هو 0.05.

### 1) في حالة تساوي متوسطي المجتمعين:

يتم تقييم حجم الاختبار للاختبارين الإحصائيين اختبار t واختبار ولكوكسن، ولهذا الغرض تم استخدام متوسطين متساويين $\mu_1 = \mu_2 = 1$

- ليكن المجتمع الأول يتبع التوزيع المنتظم (max=2, min=0)

- ليكن المجتمع الثاني يتبع التوزيع المنتظم (max=1.5 min=0.5).

يمكن حساب تباين المجتمعين باستخدام صيغة التباين للتوزيـــع المنـــتظم عليـــه فـــإن $\sigma_1^2 = \frac{1}{3}, \sigma_2^3 = \frac{1}{12}$.

وبالتالي الفرضية الصفرية هي: $H_0: \mu_1 = \mu_2$

لقد تم استخدام البرنامج الإحصائي R لتوليد العينتين U(0.2) ,U(0.5,1.5) . بعد ذلك وبافتراض صحة الفرضية الصفرية، تم حساب نسبة رفض الفرضية الصفرية وذلك بتحديد نسبة P-value<0.05 حيث أننا استخدمنا $\alpha = 0.05$.

حيث أنه للحصول على حجم الاختبار لاختبار t واختبار ولكوكسن تم إعـــادة العمليـــة 10000 مره ومن ثم تحديد عدد مرات رفـــض الفرضـــية الصـــفرية (P_value<0.05 ) و بالقسمة على 10000 نحصل على حجم الاختبار الإحصائي. العملية السابقة تم تكرارها لأحجام العينات 8، 15، 20، 25، 30.البيانات المتحصل عليها من عملية المحاكاة تم عرضـــها فـــي جدول (3)

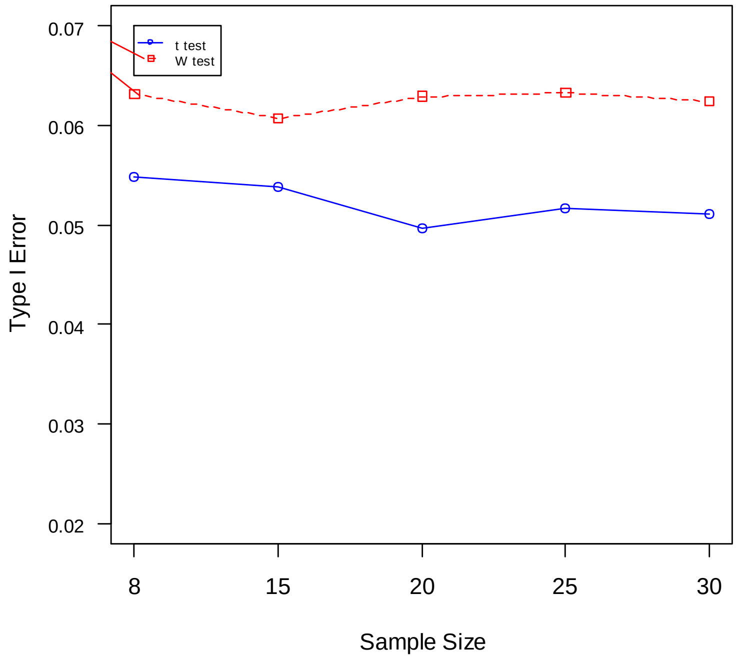
**جدول رقم (3) يوضح حجم العينة، حجم الاختبار لاختبار t وحجم الاختبار لاختبار ولكوكسن.**

| حجم العينة | حجم اختبار t | حجم اختبار ولكوكسن |
|---|---|---|
| 8 | 0.0548 | 0.0633 |
| 15 | 0.0538 | 0.0607 |
| 20 | 0.0496 | 0.0629 |
| 25 | 0.0516 | 0.0634 |
| 30 | 0.0511 | 0.0624 |

نلاحظ من الجدول (3) أن قيم حجم الاختبار لاختبار t أكبر قليلا من مستوى المعنويـــة المحددة (0.05) في حالة حجم العينة صغير (8،15) بينما تكون قريبة جدا من مستوى المعنوية المحددة 0.05 في حالة حجم العينة تساوي 20 فأكثر. بينما حجم الاختبار لاختبار ولكوكســـن

أكبر من مستوى المعنوية المحددة ولا تتأثر بزيادة حجم العينة. عليه يمكن القول بـــإن حجـــم الاختبار لاختبار t يعطينا قيم قريبة من 0.05 أفضل بكثير من اختبار ولكوكسن خاصـــة عنـــد زيادة حجم العينة.

شكل رقم (3) يوضح تقارب حجم الاختبار لاختبار t من 0.05 أكثر من حجم الاختبار لاختبار ولكوكسن مما يعطي الأفضلية لاختبار t.




**شكل رقم (3) يوضح حجم الاختبار لاختبار t وحجم الاختبار لاختبار ولكوكسن.**

**2) في حالة عدم تساوي متوسطي المجتمعين:**

يتم إيجاد قوة الاختبار للاختبارين (t، ولكوكسن) تحت الفرضية البديلـــة: $H_1: \mu_1 \neq \mu_2$

- تم توليد بيانات العينة الأولى من u(0,2).

- ثم توليد بيانات العينة الثانية من $u(0.5, 3.5)$.

حيث أن :

$\mu_2=1$ ، $\mu_1=2$

وبالتالي فإن:

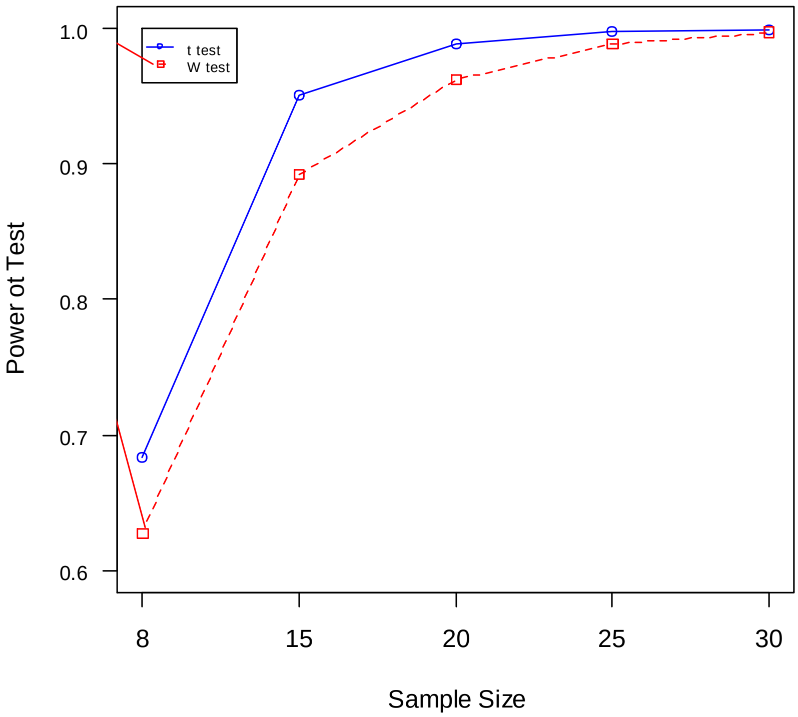
$$\sigma_2^2 = \frac{3}{4} , \sigma_1^2 = \frac{1}{3}$$

للحصول على قوة الاختبار لاختبار t واختبار ولكوكسن نقوم بإعادة العمليــة 10000 مره ومن ثم تحديد عدد مرات رفض الفرضية الصفرية ($P\_value<0.05$ ) و بالقسمة علــى 10000 نحصل على قوة الاختبار الإحصائي. العملية السابقة تم تكرارها لأحجام العينــات 8، 15، 20، 25، 30. البيانات المتحصل عليها من عملية المحاكاة تم عرضها في جدول (4)

**جدول رقم (4) يوضح حجم العينة، قوة الاختبار لاختبار t وقوة الاختبار لاختبار ولكوكسن.**

| حجم العينة | قوة اختبار t | قوة اختبار ولكوكسن |
|---|---|---|
| 8 | 0.6833 | 0.6268 |
| 15 | 0.9513 | 0.8922 |
| 20 | 0.9889 | 0.9626 |
| 25 | 0.9979 | 009884 |
| 30 | 0.9995 | 0.9971 |

نلاحظ من الجدول (4) أن قيم قوة الاختبار لاختبار t أكبر من قوة الاختبــار لاختبــار ولكوكسن في حالة العينات الصغيرة وتقترب قوة الاختبار لكلا الاختبارين من بعض مع زيــادة حجم العينة. والأفضلية تبقى لاختبار t .




**شكل (4) يوضح قوة الاختبار لاختبار t وقوة الاختبار لاختبار ولكوكسن**

من الشكل (4) نلاحظ أفضلية قوة الاختبار لاختبار t على قوة الاختبار لاختبار ولكوكسن عندما يكون حجم العينة صغير، وتزداد قوة الاختبار مع زيادة حجم العينة وشبه تطابق قوة الاختبار لكلا الاختبارين.

## الخلاصة:

1) في هذا البحث تم التأكيد على أفضلية استخدام اختبار t على اختبار ولكوكسن عندما تكون البيانات مسحوبة من مجتمع طبيعي التوزيع. حيث أن قوة الاختبار لاختبار t أكبر من قوة الاختبار لاختبار ولكوكسن خصوصا في حالة حجم العينة صغير. مع ملاحظة أن قوة

الاختبار لكلا الاختبارين تزداد طرديا مع زيادة حجم العينة وتصبح قـوة الاختبـار لكـلا الاختبارين شبه متطابقة كلما كان حجم العينة كبير.

2) معظم الدراسات السابقة والمشار إليها في هذا البحث، أشارت إلى أفضلية استخدام اختبـار ولكوكسن على اختبار t عندما تكون البيانات مسحوبة من مجتمع غير طبيعـي التوزيـع، خصوصا إذا كانت التوزيع شديد الالتواء أو ذو قيم شاذة. في هذا الدراسة لم تتوافـق مـع الدراسات السابقة إذا كانت البيانات مسحوبة من مجتمع غير طبيعي التوزيع ولكنه متماثـل وهو التوزيع المنتظم في هذا البحث. لقد وجد في حالة بيانات مسحوبة من توزيع منتظم أن استخدام اختبار t له أفضلية على اختبار ولكوكسن. حيث أن اختبار t له قوة اختبار أكبـر من اختبار ولكوكسن في حالة العينات صغيرة الحجم، وتزداد قوة الاختبار لكلا الاختبارين طرديا مع زيادة حجم العينة وتصبح قوة الاختبار لكلا الاختبارين شبه متطابقة كلمـا كـان حجم العينة كبير. مع ملاحظة أن قيمة حجم الخطأ من النوع الأول لاختبار t ثابت ومحافظ حول قيمة مستوى المعنوية المحددة وهي 0.05 ولا تتأثر بتغير حجم العينة. بينما قيمة حجم الخطأ من النوع الأول لاختبار ولكوكسن كانت كبيرة وبعيدة عن مستوى المعنوية المحـددة وهي 0.05 وتبقى بعيدة حتى مع زيادة حجم العينة.

**هوامش البحث :**


1) Blair, R. C., & Higgins, J. J. (1980). A comparison of the power of Wilcoxon's rank-sum statistic to that of student's t statistic under various nonnormal distributions. Journal of Educational Statistics, 5, 309-335.
2) Bridge, P. D., & Sawilowsky, S. S. (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: Comparative power of the t-test and Wilcoxon rank-sum test in small samples applied research. Journal of Clinical Epidemiology, 3, 229-235
3) de Winter, J. C. F. and D. Dodou (2012). Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon. Practical Assessment, Research & Evaluation, 15(11). Available online: http://pareonline.net/getvn.asp?v=15&n=11.
4) Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Applied statistics for the behavioral sciences (5th ed.). Boston, MA: Houghton Mifflin.
5) MacDonald, P. (1999). Power, type I, and type III error rates of parametric and nonparametric statistical tests. The Journal of Experimental Education, 67, 367-379.
6) Ott, R. L, & Longnecker, M. T. (2010). An introduction to statistical methods and data analysis. Belmont, CA: Brooks/Cole Cengage Learning.
7) Sawilowsky, S. (2005). Misconceptions leading to choosing the t test over the Wilcoxon Mann-Whitney U test for shift in location parameter. Journal of Modern Applied Statistical Methods, 4, 598–600.
8) Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. PsychologicalBulletin, 111, 352–360.

9) Weber, M. and Sawilowsky, S. (2009) "Comparative Power Of The Independent t, Permutation t, and WilcoxonTests," Journal of Modern Applied Statistical Methods: Vol. 8: Iss. 1, Article 3. Available at: http://digitalcommons.wayne.edu/jmasm/vol8/iss1/3

10) Wilcoxon, F. (1945). Individual comparisons by ranks methods. Biometrics Bulletin, 1, 88-83.