

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

## الدروس المستفادة

إختبار وتقييم معدات أنشطة إزالة الألغام الميكانيكية وفقاً لإتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير (إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 : 2004)

الجزء الرابع: الأساليب الإحصائية المستخدمة في إحتساب الأداء آلة إزالة الألغام، وفترات ثقة الأداء واختلافات الأداء

**يرجى أخذ العلم أن هذا النص لا يزال قيد الدرس. وسيتم نشر النسخة النهائية في نهاية أبريل 2010.**

تمت ترجمة هذا البروتوكول إلى اللغة العربية بتمويل من الصندوق العربي للإنماء الاقتصادي والاجتماعي



فريق البرنامج الدولي للإختبار والتقييم العامل على إختبار وتقييم المساعدة الميكانيكية لمعدات التطهير (ITEP WGMAE)

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

## جدول المحتويات

1-	المقدمة-----	2-
2-	المبادئ الإحصائية-----	2-
2.1-	المصطلحات والتعاريف-----	7-
2.2-	إحتساب فترة الثقة-----	9-
2.3-	اختبار الفرضيات-----	17-
3-	المراجع-----	4-
4-	الملحق 1: نهج إحصائية للحصول على جزء التحييد الملاحظ حيث تصبح الاختلافات في نسبة التحييد ملحوظة إحصائياً-----	20-
4.1-	الرسوم البيانية-----	20-
4.2-	تطبيق نهج إحصائية مختلفة لبيانات إختبار أداء الخاص بـ CWA 15044-----	23-

### 1- المقدمة

يستخدم كل من تصميم تجربة الأداء المنصوص عنها في اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير حول اختبار وتقييم آلات إزالة الألغام (CWA 15044) وإحتساب نتيجة أداء معدّات إزالة الألغام الميكانيكية، المبادئ والأساليب الإحصائية المشتركة.

تسلط هذه الوثيقة الضوء على المبادئ والأساليب الإحصائية المستخدمة في سياق اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 وتدرج قوائم المراجع التي يمكنها إعطاء المعلومات الأكثر شمولاً. وكذلك جرى وصف الإجراءات الإحصائية العامة خطوة خطوة في صلب النصّ الرئيسي ومن ثم تحويلها إلى تطبيق اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 في خانات النصوص المناسبة.

### 2- المبادئ الإحصائية

#### 2.1- المصطلحات والتعاريف

تدخل الإحصائيات في مجال الرياضيات مع أساليب وإجراءات أخرى لجمع وتقديم وتلخيص البيانات (إحصاءات وصفية) وكذلك لاستخلاص الاستنتاجات أو تقديم التنبؤات (الإحصاء الاستدلالي). وفي الإحصاء الاستدلالي، يتم استخدام بيانات العينة بغية استخلاص استنتاجات حول المجموعة التي استمدت منها العينات. إذ تتكون المجموعة من المجموع الكلي للمواضيع / الأشياء التي تحتوي على شيء مشترك في ما بينها، أما العينة فهي عبارة عن مجموعة من المواضيع / الأشياء التي استمدت من المجموعة [1]. إذ تتمثل الأهداف الأساسية للإحصاءات بـ: (1) تقدير معايير المجموعة (القيم التي تميز مجموعات معينة) و (2) اختبار الفرضيات حول هذه المعايير [2].

عندما يتم تقييم قدرة المعدات الميكانيكية على تحييد الألغام في تجربة أداء اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، يتم بذلك تحديد كيفية معالجة الألغام المضادة للأفراد (AP) بواسطة المعدات (يتم تحييدها أم لا). والهدف الرئيسي من الإختبار هو تقدير قدرة الآلات على إزالة الألغام المضادة للأفراد كنسبة مئوية من الألغام المضادة للأفراد التي تم تحييدها في ممر الإختبار (عينة). وإن إجراء الإختبار المنصوص عنه في مبادئ الإختبار التوجيهية بموجب إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، يسمح كذلك باختبار سلسلة من الفرضيات حول قدرة الآلات على إزالة الألغام المضادة للأفراد. وكذلك يمكن اختبار الفرضيات الآتية:

- هل تبقى قدرة آلات إزالة الألغام نفسها في أنواع التربة الثلاثة ؟
- هل تكون آلات إزالة الألغام في التربة الرملية نفسها على أعماق طمر الألغام الثلاثة؟
- هل تكون قدرة آلتين مختلفتين لإزالة الألغام في التربة السطحية نفسها بالنسبة لمستوى سطح الألغام المطمورة؟
- إلخ

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

ينبغي أن تشمل العينة أكبر عدد من المجموعات التي أخذت منها العينة المذكورة، وذلك كي تكون النتائج فعالة ومفيدة لاستخلاص الاستنتاجات. وفي هذا السياق، تعتبر العينة العشوائية بمثابة العينة المثالية للعمل بها. ولأخذ العينات، لا بدّ من إتباع الخطوات الآتية:

- قد يشكّل أي موضوع /شيء في المجموعة احتمالاً بأن يكون من بين مكونات العينة،
- إن إختيار الموضوع /شيء ينبغي أن يكون مستقلاً عن المواضيع أو الأشياء الأخرى.
- لحجم عينة متساوٍ، يكون لكلّ عينة محتملة يمكن استخلاصها من المجموعة، احتمال الحدوث نفسه.

وفي حال التقيّد بمواصفات إنشاء ممر الإختبار المحددة بموجب إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، أي أن تكون أهداف الإختبار<sup>1</sup> مضمورة عشوائياً في الممرات كي يشكّل كلّ ممر إختبار بحدّ ذاته عينة عشوائية من الألغام. وكذلك، ينبغي أن تكون هذه الألغام مضادة للأفراد ومضمورة على عمق 10 سم في الحصى.

تجدر الإشارة إلى أنّ مخطط الوضعية الموضوع من قبل المشغل الذي يطرر أهداف الإختبار ليس عشوائياً.

ولتخصيص أهداف الإختبار بشكل عشوائي، من الممكن استخدام محرّك رقم عشوائي مثل RAND () في Microsoft Excel أو أي محرّك رقم عشوائي متوفر على الويب (انظر على سبيل المثال [31] و [32])

تشير **الإحصائيات** إلى سمة من سمات العينة، أي أنها رقم قد يتم إحتسابه من خلال البيانات المرصودة في العينة العشوائية.

وتشير المعايير من الناحية الأخرى إلى سمة من سمات الألغام أي أنّها الرقم الذي يصف الألغام [1] [6].

إذ يبقى الجزء الأهم من الإحصاءات متمثلاً بتقييم **المقاييس والإحصائيات**. إذ يتم استخدام الإحصاءات المستمدة من العينات كأداة تقدير لمعايير الألغام ذات الصلة [3].

إنّ الإحصاءات المحددة في إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، هي عبارة عن عدد أهداف إختبار الألغام المضادة للأفراد التي تمّ تحييدها، والذي يتمّ التعبير عنه كنسبة مئوية من إجمالي عدد أهداف إختبار الألغام المضادة للأفراد المضمورة في ممرّ الإختبار. وبالتالي تستخدم هذه الإحصائية كتقدير لقدرة الآلة على تحييد الألغام المضادة للأفراد في الظروف المشابهة لظروف ممرّ الإختبار.

**توزّع العينة** هو توزيع<sup>2</sup> الإحصائية المحتسبة من العينة. في حال جرى أخذ العينات بحجم (ن) أكثر من مرّة واحتماب إحصائية محددة لتلك العينة في كل مرّة، تُسمّى عملية توزيع القيم التي تم الحصول عليها للإحصاء، "توزّع العينة". لكلّ إحصاء، توزّع عينة خاص به [3].

لنفترض أنّه يتم تشغيل آلة إزالة الألغام على ممرّ إختبار يحتوي على 50 هدفاً على عمق 10 سم في الرمال (حجم العينة ن = 50) وأنّ نسبة الأهداف التي تمّ تحييدها محددة. وبعد ذلك، يتمّ تشغيل الآلة مرة أخرى على ممر إختبار مع الحفاظ على الخصائص نفسها، أي 50 هدف مضمور على عمق 10 سم في الرمال (حجم العينة ن = 50) وكذلك على تحديد نسبة الأهداف التي تمّ تحييدها أيضاً. فليس من الضروري الحصول على النتيجة نفسها في الإختبارين. وبالتالي، عند تكرار الإختبار لمرّات عديدة غير محددة، يمكن الحصول على عدد لا حصر له من نسب التحييد. ويطلق على توزيع هذا العدد غير المحصور لنسب تحييد الألغام، اسم **توزّع العينة** لنسبة تحييد الألغام.

## البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

<sup>1</sup> تمثّل أهداف الإختبار المستخدمة فئة عامة من الألغام المضادة للأفراد أو فئة معينة من الألغام المضادة للأفراد اعتماداً على أهداف التجربة.  
<sup>2</sup> يمكن الإطلاع على تعريف وأمثلة من التوزيعات في [33] و [34]

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

ومع الأخذ بعين الاعتبار لعملية توزُّع العينة ، لا بدّ من إدراك أنّ الإحصائية من عينة واحدة قريبة من مركز توزُّع العينة (لأن معظم العينات ستكون هناك) تعطي نتيجة متطرّفة عن طريق الحظ. أمّا في حال جرى اعتماد متوسط توزُّع العينات - متوسط عدد لا حصر له من العينات - سيكون الأمر أقرب بكثير إلى متوسط عدد الألغام الحقيقي - معايير الإهتمام. وبالتالي فإن متوسط توزُّع العينة يعادل في الأساس معايير المجموعات الأخرى [5]. ويُشار إلى مجموعة الإحصاءات التي يمكن الحصول عليها لتقدير المعايير من خلال أخذ العينات على أنّها **خطأ في أخذ العينات**. يعطي خطأ المعاينة فكرة عن دقة التقديرات الإحصائية. ويشير خطأ أخذ العينات المنخفض إلى قلة التباين نسبياً في توزُّع العينات وأنه لذلك من المرجح أن تكون التقديرات التي تم الحصول عليها على مقربة من قيمة الألغام الحقيقية.

في إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، يتم إجراء الاختبار مرة واحدة فقط في حالة معينة (على سبيل المثال أهداف الاختبار على عمق 10 سم في الرمال). ومن هنا، يجب أن ندرك أنه بالرغم من اقتراب نسبة التحييد التي نحصل عليها من مركز توزُّع العينة، أي قريبة من المعايير التي نسعى إليها (نسبة إبطال مفعول للألة في ظروف معينة)، يمكننا الحصول على القيمة القصوى لتوزُّع العينات، وبالتالي من الممكن أن نكون بعيدين نسبياً عن المعايير التي نسعى إليها. ولذلك فمن المهم أن نعرف الخطأ في أخذ العينات لمعرفة وتقييم مدى قرب تقييمنا من نسبة التحييد.

عملياً، يشار إلى خطأ المعاينة قبل إحتماب فترة الثقة باستخدام الانحراف المعياري لتوزيع أخذ العينات<sup>3</sup>. الانحراف المعياري هو المقياس الأكثر شيوعاً لانتشار التوزع. وتقدم فترة الثقة مجموعة من القيم التي من المرجح أن تحتوي على معيار الألغام موضوع الإهتمام. تتشكل فترات الثقة في مستوى الثقة المحدد من قبل المستخدم. إذ يصف مستوى الثقة مدى إرتباط المعايير ضمن نطاق القيم التي حصلنا عليها حول التقدير. ويُعبّر عن مستوى الثقة كنسبة مئوية. إذ كان مستوى الثقة 95% يعني من الممكن أن نكون متأكدين بنسبة 95%. وإن كان مستوى الثقة 99% يعني بالإمكان التأكيد بنسبة 99%. يستخدم معظم الباحثون مستوى الثقة 95%. مستوى الثقة 95% يعني أنه إذا تم أخذ عينات من المجموعة نفسها مرّات عديدة وتم إجراء تقديرات فاصل الثقة في كل من هذه الحالات، تكون الفترات الناجمة عنها دأمة لمعايير العينة الأصلية في حوالي 95% من الحالات. كلما ارتفع مستوى الثقة بالقبول، كلما يمكننا التأكيد أنّ معايير العينة الأصلية تندرج ضمن النطاق المعين من خلال فترة الثقة [8] [9].

من أجل حساب هامش الخطأ وفترة الثقة لتقدير ما، ينبغي أن يكون توزُّع العينات معروفاً. ولتطبيق إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، ينبغي أتباع التوزيع الثنائي للعينات (أنظر النص داخل المربع). يصف التوزيع الثنائي سلوك العدّ المتغير  $X$  في حال انطباق الشروط الآتية:

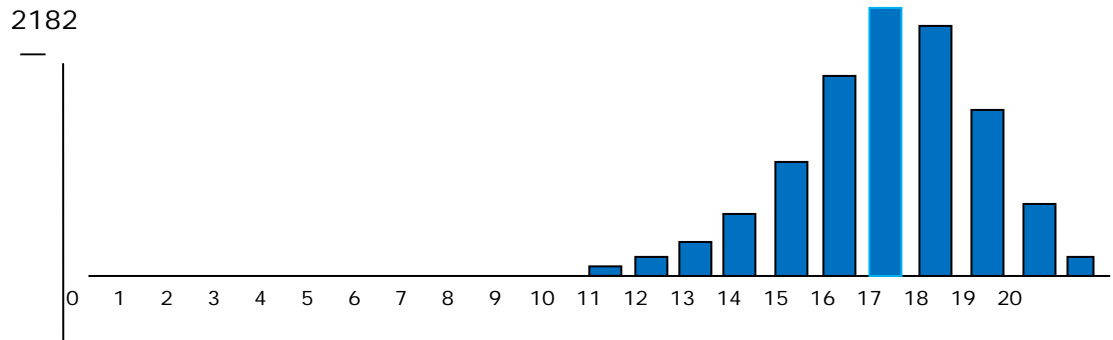
- يتم تحديد عدد من الملاحظات  $n$ ،
  - كل ملاحظة مستقلة عن الأخرى،
  - وتمثل كل ملاحظة واحدة من نتيجتين ("النجاح" أو "الفشل")، و
  - و"احتمال النجاح"  $p$  هو نفسه لكل نتيجة.
- وعند استيفاء هذه الشروط، يتبع العد المتغير  $X$  التوزيع الثنائي مع المعايير  $n$  و  $p$ . ويشير التوزيع الثنائي إلى احتمال الحصول على عدّ صحيح  $X$  عند اخذ العينات  $n$  مع معدّل نجاح نظري  $p$  [7] [10].

الملاحظات في اختبار الأداء بموجب إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، هي قيم العد، أي نحصى عدد الأهداف التي جرى تحييدها. تجدر الإشارة إلى تحديد عدد الملاحظات  $n$  بـ 50 (50 لغماً مستهدفاً في ممرّ الاختبار) ولكلّ ملاحظة نتيجتان فقط: ألغام مستهدفة جرى تحييدها (نجاح) أم لا (فشل). وإنّ احتمال النجاح  $p$  هو القدرة الحقيقية للألة على تحييد الألغام في ظلّ ظروف معينة في ممرّ الإختبار (نوع التربة وعمق طمر الألغام) وبالتالي، إنه قيمة ثابتة. وإنّ تحييد أي من الألغام المضادة للأفراد في الممرّ يكون مستقلاً عن تحييد الألغام الأخرى المضادة للأفراد الأخرى. وباختصار، يتبع عدد الألغام المستهدفة المضادة للأفراد التي تمّ تحييدها في ممرّ الإختبار ، التوزيع الثنائي.

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

3 للمزيد من المعلومات حول الانحراف المعياري للتوزيع وصيغ إحتسابه في [35].

يظهر المستند الأول مثالا على التوزيع الثنائي لممر الاختبار مع 20 هدفاً ( $n = 20$ ) وآلة ذات قدرة نظرية على إزالة الألغام المضادة للأفراد بنسبة 80% ( $p = 0.8$ ). ويظهر الرسم البياني أنه إذا كنا نقوم بتشغيل الآلة المحددة على ممر اختبار يحتوي على 20 هدفاً، يكون من المرجح إزالة 16 لغماً من أصل 20 (21.8% من الاختبارات)، أي القدرة الحقيقية لآلات إزالة الالغام، وبالمقابل، هناك أيضا احتمال كبير نسبياً بإزالة 14 هدفاً فقط من أصل 20 (10.9% من الاختبارات) أو 18 من أصل 20 (13.7% من الاختبارات). وكذلك قد نحصل على نتائج تشير إلى قدرة المعدات على إزالة 10 ألغام من أصل 20 لغماً (0.2% من الاختبارات) أو 20 من أصل 20 لغماً (1.2% من الاختبارات).



عدد الألغام التي تم تحييدها من أصل 20 بواسطة آلات ذات قدرة نظرية على إزالة الألغام بنسبة 80% ( $p=0.8$ )

المستند 1: التوزيع الثنائي لـ  $n=20$  و  $p=0.8$

وتجدر الإشارة إلى أنه وفقاً للنظرية الإحصائية، يعتبر توزع العينات الثنائي الأكثر دقة في الحالات التي يكون فيها حجم المجموعة التي أخذت منها العينة أكبر بكثير من حجم العينة. كقاعدة عامة، يجب أن لا يتم تطبيق التوزيع الثنائي على ملاحظات من عينة عشوائية إلا إذا كان عدد المجموعة التي أخذت منها العينة أكبر بـ 10 مرات على الأقل من حجم العينة [7].

والتوزيع الثنائي هو دالة رياضية مع اثنتين من المتغيرات  $n$  و  $p$ . ومن الممكن إحتساب إحتتمالات التوزيع الثنائي مباشرة باستخدام الصيغة المناسبة (المتاحة في [7]). ومع ذلك، تنطوي هذه الحسابات على الكثير من الحسابات وأساليب أسرع متوفرة، كقراءة القيم في الجدول (متوفر في [36])، وذلك باستخدام آلة حاسبة ذات توزيع ثنائي متاحة على شبكة الإنترنت (على سبيل المثال في [10]) أو باستخدام جدول يحتوي على دالة إحصائية (على سبيل المثال: مايكروسوفت اكسل - (BINOMDIST).

## 2.2- حساب فترة الثقة

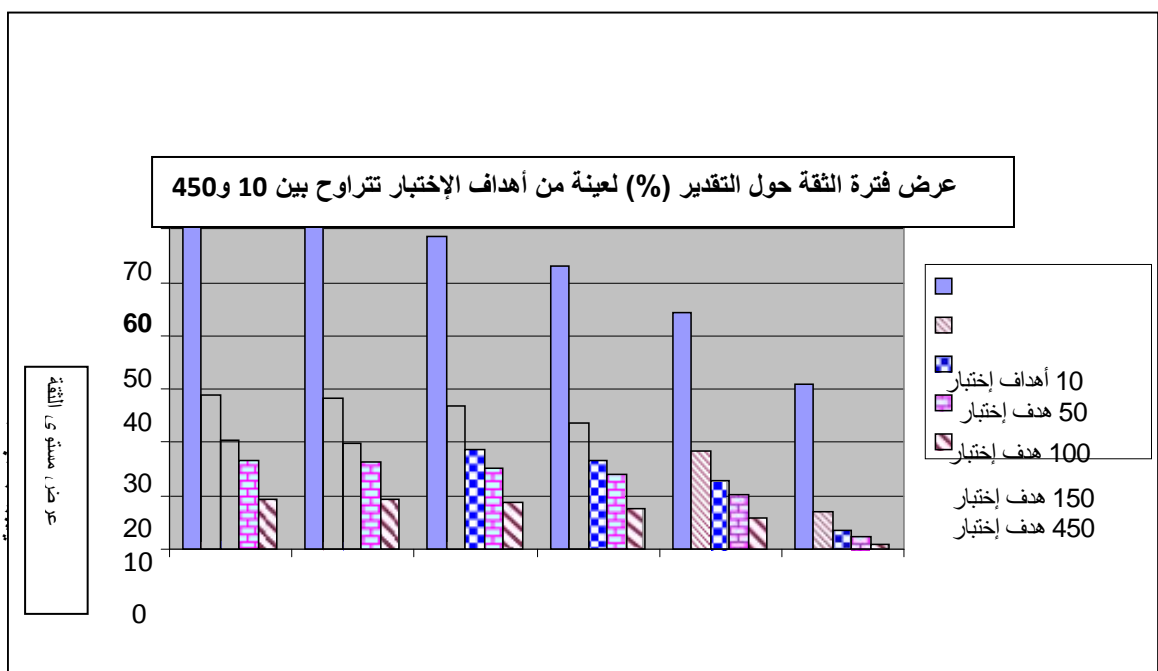
ويمكن حساب فترة الثقة من خلال إحصائية ثنائية التوزيع باستخدام أساليب مختلفة (صياغات)، بحيث تؤدي كل منها إلى تقديرات مختلفة قليلاً بالنسبة لمستوى الثقة. إذ أنّ الاستخدام الأكثر شيوعاً هو استخدام الأسلوب التقريبي العادي (Normal Approximation Method) وأسلوب وَاُلْد (Wald Method) وأسلوب وَاُلْد المعدل (Adjusted Wald Method) وأسلوب Clopper-Pearson والأسلوب الدقيق (Exact Method) وأسلوب الفاصل الزمني (Score Interval (Wilson)). وتجدون كلّ ما يتعلق بصياغات مستوى الثقة فضلاً عن المزايا والعيوب في الأساليب في [11] و [12]. وتحسب مستويات الثقة بسهولة باستخدام الآلات الحاسبة المفتوحة المصدر مثل تلك المبيّنة في [13] و [14].

## من العوامل الهامة لتحديد عرض مستوى الثقة:

- حجم العينة. كلما زاد حجم العينة أكثر، تستطيعون التأكد أكثر بأن العينة الإحصائية تعكس المعيار الحقيقي للمجموعة التي أخذت منها هذه العينة. هذا يشير إلى أنّه بالنسبة لمستوى معين من الثقة، كلما يكبر حجم العينة، كلما قلّ مستوى الثقة. ومع ذلك، فإن العلاقة ليست خطية (أي مضاعفة حجم العينة لا يخفض مستوى الثقة).
- النسبة المشار إليها من خلال العينة. على سبيل المثال، في حال جرى تحييد 99% من أهداف الإختبار وترك 1% منها، تكون فرص الخطأ أصغر من الحالة التي يتمّ بها تحييد 51% فقط وترك 49% من الألغام على حالها، بغض النظر عن حجم العينة. فالتأكد من النتائج النهائية أسهل من النتائج التي نحصل عليها في منتصف الطريق.

ويوضح المستند 2 العوامل المذكورة أعلاه في إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير. وكلّما كانت العينة كبيرة (عدد أهداف الإختبار في ممر اختبار) و / أو كلّما كانت نسبة التحييد أكبر، يكون مستوى الثقة أقلّ في ما يتعلق بالتقديرات وبالتالي، تكون النتيجة أكثر دقّة. يتمّ إحتساب مستويات الثقة باستخدام Clopper- Pearson أو الأسلوب الدقيق (Exact Method) وللمزيد من التفاصيل حول عملية الإحتساب أنظر:

([http://www.itep.ws/pdf/CWA15044/BinaryConfidenceIntervals\\_calc.xls](http://www.itep.ws/pdf/CWA15044/BinaryConfidenceIntervals_calc.xls)).



0 50 60 70 80 90 100

تقدير نسبة التحديد (%)

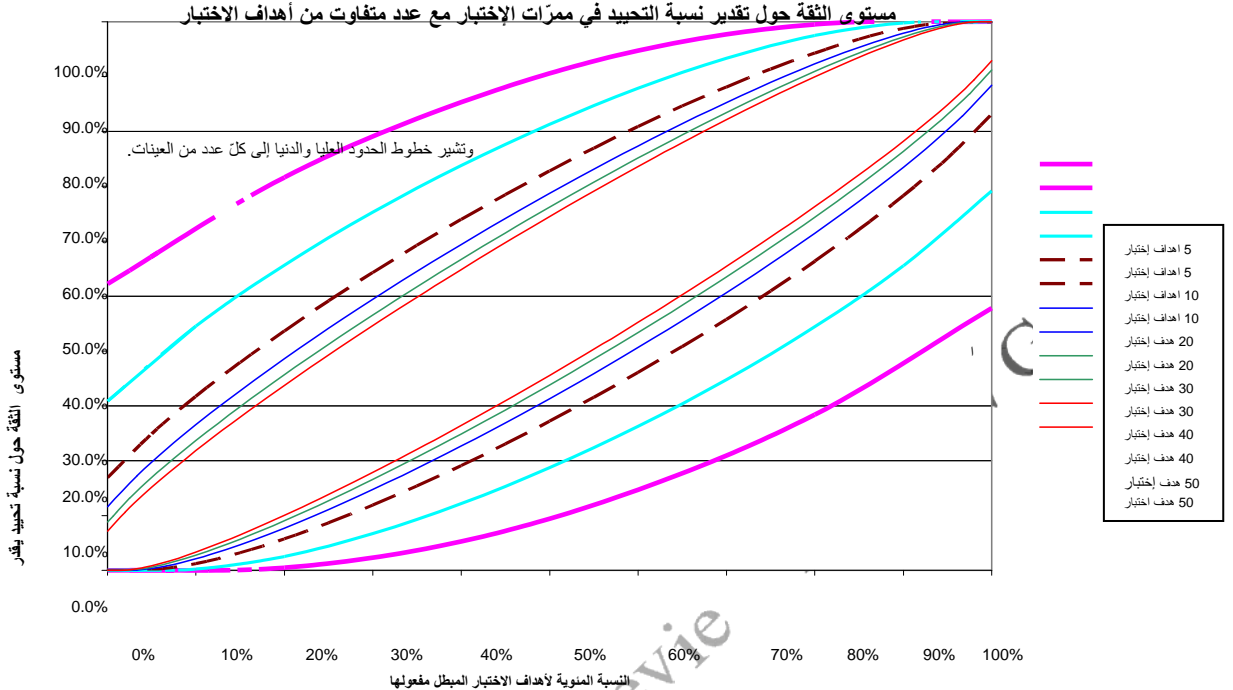
المستند 2: عرض مستوى الثقة (كنسبة مئوية للتقدير) عند مستوى الثقة 95% للعينات ذات عدد أهداف متراوح ما بين 10 و 450، وقدرة التحديد بين 50 و100%.

تجدر الإشارة إلى أن حسابات مستوى الثقة تفترض وجود عينة عشوائية حقيقية من المجموعة ذات الصلة. وإن لم تكن العينة عشوائية حقاً، إذاً لا يمكن الإعتماد على مستويات الثقة [9]

توصي إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 باستخدام مستوى الثقة 95% (مستوى الدلالة 5%) وبإحتساب مستويات الثقة لنسبة التحديد وفقاً لأسلوب Clopper-Pearson أو الأسلوب الدقيق (Exact Method). ووافقت ورشة العمل التي صاغت إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 أيضاً على عدد الأهداف المحدد بـ 50 باعتباره عدد كافٍ من شأنه أن يوفر مستوى ثقة مرضي حول التقدير الذي تم الحصول عليه بكلفة معقولة.

فالرسم البياني الظاهر في المستند (3) يبين أنه عندما يتم استخدام 10 أهداف ألغام ويتم الحصول على نسبة تحييد 80%، يمكننا التأكد بنسبة 95% من أن القدرة الفعلية للألة في ظل ظروف هذا الإختبار (نوع التربة وعمق طمر اللغم) تتراوح بين 44% و79%. ومع استخدام 50 هدفاً من الألغام يمكن التأكد بنسبة 95% من أن القدرة الفعلية للألة في ظل ظرف هذا الإختبار تتراوح بين 66% و 90% (للحصول على تفاصيل الحساب أنظر الرابط الآتي:

[http://www.itep.ws/pdf/CWA15044/BinaryConfidenceIntervals\\_calc.xls](http://www.itep.ws/pdf/CWA15044/BinaryConfidenceIntervals_calc.xls))





البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

### المستند (3): عرض مستوى الثقة عند مستوى 95% لممرات الإختبار مع عدد متفاوت من أهداف الإختبار.

تجدر الإشارة إلى أنه في الحالة الافتراضية التي تكون فيها نسب التحييد في جميع الظروف المختبرة غير مختلفة اختلافاً كبيراً (3 أنواع تربة x إختبار 3 أعماق مختلفة لطمر الهدف)، يمكن جمع البيانات التي من شأنها أن تؤدي إلى عينة من 450 هدف ، وبالتالي إلى نتيجة أكثر موثوقية لناحية تقدير قدرة الآلة على تحييد الألغام. أما بالنسبة للمثال السابق، فإن ذلك يعني أنه بالإمكان التأكد بنسبة 95% من أن القدرة الفعلية للآلة ستكون بين 76% و 84%.

### 2.3- إختبار الفرضيات

والجانب الأهم من جوانب استخدام الأساليب الإحصائية، هو وجود إجراءات إحصائية قياسية لإختبار الفرضيات المتعلقة بمعايير مجموعة العينات. ففي إختبار الفرضيات ، يقرّر الشخص ما إذا كانت البيانات تظهر تأثيراً "حقيقياً" أو يمكن أن تكون قد حدثت نتيجة تغير ما، أي هي مجرد نتيجة لخطأ في أخذ العينة [15].

الفرضية الإحصائية هي عبارة عن افتراض حول معايير مجموعة العينات. قد يكون أو لا يكون هذا الافتراض صحيحاً.

إذا كانت بيانات العينة (الملاحظات) تتفق مع الفرضية الإحصائية، يتم قبول الفرضية. وإن لم يكن كذلك، يتم دحض الفرضية. هناك نوعان من الفرضيات الإحصائية [15، 16]:

- فرضية العدم،  $H_0$ : فرضية العدم هي عادة الفرضية القائلة بأن الفرق الملاحظ في العينة ناتج عن الصدفة أي لا يوجد "تأثير حقيقي" إذ يكون التأثير في بيانات العينة هو مجرد نتيجة لخطأ في أخذ العينات (الصدفة).
- الفرضية البديلة،  $H_1$ : الفرضية البديلة هي الفرضية القائلة بأن ملاحظات العينة تتأثر ببعض الأسباب غير العشوائية، أي متغيرة. وبالتالي، إن الفرق الملاحظ في العينة يعكس الفرق الحقيقي في المجموعة.

وبالنسبة لحالة إتفاقيّة ورشة عمل اللّجنة الأوروبيّة لتوحيد المعايير 15044، إنّ فرضية العدم  $H_0$  هي عدم تغيير أداء الآلة سواء في الرمال أو التربة السطحية بالنسبة للألغام المظمورة في حين أنّ فرضية البديل  $H_1$  فهي تكمن عندما يتغير أداء الآلة بالنسبة للألغام إذ لا يكون نفسه في الرمال والتربة السطحية. وهناك فرضية عدم ( $H_0$ ) أخرى محتملة وهي أنّ الآلة (أ) تعمل تمامًا كآلة (ب) بالنسبة للألغام المظمورة في الرمال في حين الفرضية البديلة  $H_1$  تفترض أنّ الآلة تعمل بشكلٍ مختلف.

يتبع إختبار الفرضية اللوجيستيات الآتية:

- ذكر الفرضية. وذلك ينطوي على ذكر الفرضية العدم ( $H_0$ ) والفرضة البديلة ( $H_1$ ). والفرضيات ينبغي أن تكون متعارضة مع بعضها البعض. وبطبيعة الحال، عندما تكون إحداها صحيحة، تكون الأخرى خاطئة.
- إفتراض أنّ فرضية العدم  $H_0$  صحيحة.
- إحتساب احتمال الحصول على النتائج الملاحظة في البيانات الخاصة بك إذا كانت فرضية  $H_0$  صحيحة.
- إذا كان هذا الاحتمال منخفضاً (على سبيل المثال أقل من 5%)، يتم دحض هذه الفرضية.
- إذا تمّ دحض فرضية العدم، لا يتبقّى سوى الفرضية البديلة  $H_1$ .

في إختبار الفرضيات، لإحتساب احتمال إمكانية حدوث فرق ملحوظ في البيانات عن طريق الصدفة إذا كان  $H_0$  صحيح، يتم احتساب إختبار إحصائي واحد وتقييمه. لإختبار الفرضية الإحصائية الموزعة ثنائياً، مثل النسبة (نسبة مئوية)، يتم استخدام إختبار كاي المربع [15]. ويستند إختبار كاي المربع على إحصائية إختبار كاي المربع ( $\chi^2$ ) ويتحقّق ما إذا كانت نسب فئات معينة مختلفة في مجموعات مختلفة

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

[18]. إما يكون إختبار كاي المربع باتجاه واحد (يُطلق عليه أيضًا اسم: إختبار مربع كاي لجودة التوافق) أو باتجاهين (يُطلق عليه أيضًا اسم: إختبار الاحصائي كاي تربيع للاستقلالية). في إختبار كاي المربع باتجاه واحد ، يتم مقارنة الملاحظات للمجموعة المعروفة خصائصها في حين إختبار الاحصائي كاي تربيع باتجاهين، يتم مقارنة ترددات الحدوث في فئتين أو أكثر بين مجموعتين أو أكثر. ولغرض احتساب إختبار الإحصائية كاي تربيع، تُعرض بيانات النسبة الملحوظة في جدول يسمى **جدول التصادف**.

ولإختبار فرضيات إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير **15044**، يتم استخدام الإختبار الاحصائي كاي تربيع باتجاهين لتتم مقارنة نسب الألغام التي لا تزال فعالة والألغام التي تم تحييدها بين مجموعتين أو أكثر. يمكن أن تكون المجموعة الأخيرة على سبيل المثال كالآتي:

- لتان مختلفتان لمعالجة أهداف الإختبار المظمورة على سطح الأرض في الرمال، أو
- آلة واحدة لمعالجة أهداف الإختبار المظمورة على سطح الأرض في الرمال والحصى، أو
- آلة واحدة لمعالجة أهداف الإختبار المظمورة في الرمال على ثلاثة أعماق مختلفة،
- إلخ..

جداول التصادف التي يمكن إعدادها للأمثلة المذكورة هي على الشكل الآتي:

(ب)				(أ)		
	الآلة 1، الأهداف المظمورة على سطح الأرض			الأهداف المظمورة على سطح الأرض، في الرمال		
	الرمل	الحصى		الآلة 1	الآلة 2	
الأهداف التي تم تحييدها	48	46	94	48	40	88
الألغام التي لا تزال فعالة	2	4	6	2	10	12
	50	50	100	50	50	100

(ت)				
	الآلة 1، رمال،			
	الأهداف المظمورة على عمق 20 سم	الأهداف المظمورة على عمق 10 سم	الأهداف المظمورة على عمق 0 سم	
الأهداف التي تم تحييدها	39	47	48	133
الألغام التي لا تزال فعالة	11	3	2	17
	50	50	50	150

ويستند إحتساب الإختبار الاحصائي كاي تربيع ( $\chi^2$ ) على المقارنة بين القيم المرصودة والقيم المتوقعة. وفي إختبار كاي تربيع ذات الإتجاهين، من غير المعروف كيف ينبغي توزيع البيانات على فئات مختلفة. أو بالأحرى، يتم احتساب القيم المتوقعة استنادًا للقيم الملاحظة (مجموع الخانات الأفقية في الجدول ومجموع الخانات العمودية في الجدول ومجموع الجداول) ومجموع الفروق بين القيم المحسوبة المتوقعة والقيم المرصودة، ليستخدم بعد ذلك لإحتساب الإختبار الاحصائي كاي تربيع.

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

وللمزيد من المعلومات حول الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) وإحتسابه، أنظر [17]، [21]، [22]، [23]، [24]. تظهر الفقرات التالية احتساب لأمثلة إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 الواردة أعلاه.

الخطوة الأولى في حساب الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) هو إيجاد القيمة المتوقعة (E) لكل خانة من الجدول الذي يحتوي على البيانات المرصودة. ومن ثم يتم احتساب القيمة المتوقعة لكل خانة من الجدول ( $E_{ij}$ ) باستخدام المعادلة التالية:

مجموع الخانات الأفقية  $\times$  مجموع الخانات العمودية / مجموع الجدول

تجدون أدناه داخل قوسين وبخط مائل، القيم المتوقعة لكل خانة من جداول الأمثلة.

(ب)

	الألة 1، الأهداف المطمورة على سطح الأرض		
	الرمل	الحصى	
الأهداف التي تمّ تحييدها	48 (47)	46 (47)	94
الألغام التي لا تزال فعالة	2 (3)	4 (3)	6
	50	50	100

(أ)

	الأهداف المطمورة على سطح الأرض، في الرمال		
	الألة 1	الألة 2	
الأهداف التي تمّ تحييدها	48 (44)	40 (44)	88
الألغام التي لا تزال فعالة	2 (6)	10 (6)	12
	50	50	100

(ت)

	الألة 1، رمال،			
	الأهداف المطمورة على عمق 20 سم	الأهداف المطمورة على عمق 10 سم	الأهداف المطمورة على عمق 0 سم	
الأهداف التي تمّ تحييدها	39 (45)	47 (45)	48 (45)	133
الألغام التي لا تزال فعالة	11 (5)	3 (5)	2 (5)	17
	50	50	50	150

تجري الخطوة الثانية لإحتساب الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) الذي يضم الاختلافات بين القيمة المتوقعة ( $E_{ij}$ ) والقيمة المرصودة ( $O_{ij}$ ) في كل خانة وفقا للمعادلة التالية:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

الجدول (ب):  $\chi^2 = 0.71$

الجدول (أ):  $\chi^2 = 4.64$

الجدول (ت):  $\chi^2 = 10.21$

تجدر الإشارة إلى أنه وفقاً لنظرية كاي تربيع، من المستحسن تطبيق تصحيح بيتس لدى إحتماب إحصائية إختبار  $\chi^2$  لكلّ جدولين على حدى مع خانة واحدة أو أكثر ولكن أقل من خمسة. والبعض يطبق التصحيح لجدولّي القيمة المتوقّعة والقيمة المرصودة. إن تصحيح بيتس هو تعديل تقديري وتحفظي [24]. ويعني هذا الأخير أن التصحيح سوف يجعل الأمر أكثر صعوبة بالنسبة لوضع الفروقات، أي أن الفروقات الملحوظة يجب أن تكون أكبر في الإختبار للإشارة إلى أنه ينبغي دحض فرضية العدم. وفي إحصائية إختبار  $\chi^2$  المصححة، يتم إستنتاج القيمة المطلقة (للقيم المرصودة - القيم المتوقعة) بنسبة 0.5. يُذكر أنّ جميع الحسابات الأخرى تبقى كما هي.

وفي فرضية إتفاقيّة ورشة عمل اللّجنة الأوروبيّة لتوحيد المعايير 15044 يجري إختبار جدولّي التصادف اللذين يحتويان على الخانات الأقل من خمسة بغالبية الأحيان وبالتالي، ينبغي تطبيق تصحيح بيتس.

الجدول (ب):  $\chi^2$  المصححة: 0.18

الجدول (أ):  $\chi^2$  المصححة = 6.06

في إختبار الفرضية، يتم مقارنة قيمة إحصاءات الإختبار المحتسبة بالقيمة العتبية، وتسمى أيضا القيمة الحرجة. القيمة الحرجة لأي إختبار فرضية تعتمد على مستوى الدلالة ( $\alpha$ ) الذي يتم فيه إجراء الإختبار، وإذا كان الإختبار وحيداً أو مزدوجاً.

يقيم إختبار كاي تربيع إذا كانت قيمة  $\chi^2$  التي تم الحصول عليها من البيانات المرصودة هي القيمة المحتمل الحصول عليها بالنسبة للمتغير الذي يتبع توزيع كاي تربيع ولمستوى معين من الدلالة، أي أنه بالمقارنة مع القيمة الحرجة التي تحدد الحد الفاصل بين (1) مجموعة من القيم التي من المحتمل أن يتم الحصول عليها (منطقة القبول) على افتراض أن فرضية العدم هي الصحيحة و (2) مجموعة من القيم التي من غير المحتمل الحصول عليها (منطقة الدحض) على افتراض أن فرضية العدم هي الصحيحة. وفي حال جرى قبول إحصاء الإختبار، يتم قبول فرضية العدم. ومن ناحية أخرى، إذا كانت قيمة  $\chi^2$  التي تم الحصول عليها من البيانات المرصودة تقع ضمن نطاق رفض فرضية العدم، يتم دحض فرضية العدم  $H_0$  وقبول الفرضية البديلة  $H_1$ . يتم تحديد منطقة القبول (وبالتالي القيمة الحرجة) بحيث تكون فرصة رفض فرضية العدم عندما تكون صحيحة مساوية لمستوى الدلالة  $\alpha$ . ويرتبط مستوى الدلالة  $\alpha$  بمستوى الثقة الذي اختاره المستخدم وفقاً للمعادلة الآتية: مستوى الثقة (%) =  $(\alpha - 1) \times 100$ . وبالتالي، في حال إختار المستخدم مستوى ثقة 95%، ينبغي استخدام مستوى دلالة 0.05 في إختبار الفرضية.

وفي حالة إتفاقيّة ورشة عمل اللّجنة الأوروبيّة لتوحيد المعايير 15044، جرى تحديد القيمة الحرجة لتوزيع كاي تربيع وبذلك يكون إختبار قبول فرضية العدم عندما تكون صحيحة 95% (مستوى الثقة = 95%)، وفرصة دحضها عندما تكون صحيحة 5% (مستوى الدلالة  $\alpha = 0.05$ ). ومن خلال الأمثلة أعلاه، نستنتج الآتي عند دحض فرضية العدم: (أ) يتمكّن كلّ من الآلتين بأداء مختلف لدى إزالة الألغام المطمورة على سطح الأرض في الرمال، (ب) تعمل الآلة بشكل مختلف عند تطهير الألغام المطمورة على سطح الأرض في الرمال والحصى، (ت) تعمل الآلة بشكل مختلف لدى إزالة الألغام المطمورة على سطح الأرض في الرمال اعتماداً على عمقها، وتغييراً بنسبة 5% وهذا ما أدى إلى خطأ في النتيجة التي توصلنا لها.

توزيع كاي تربيع هو توزيع حسابي يتم تحديد شكله عن طريق عدد درجات الحرية [25]. وبالتالي، فإن القيمة الحرجة التي يتم مقارنة القيمة المحسوبة لإحصاء الإختبار بها، تعتمد أيضا على درجات حرية توزيع كاي تربيع. تصل درجات الحرية (DF) إلى عدد أقسام المعلومات المستقلة التي تدخل في تقدير معيار ما. بشكل عام، تكون درجات الحرية للتقدير مساوية لعدد الدرجات المستقلة التي تدخل في التقدير ناقص عدد المعلمات المقدر على نحو الخطوات الوسيطة في تقدير المعلمة نفسها [26]. وهناك قاعدة بسيطة لإختبار مقارنة ترددات الحدوث في فئتين أو أكثر بين مجموعتين أو أكثر وهي أنّ درجات الحرية تساوي (عدد الأعمدة ناقص واحد)  $\times$  (عدد الخانات الأفقية ناقص واحد) بإستثناء الصفوف أو الأعمدة للمجموع.

عندما تعرف درجات الحرية، يمكن تحديد القيم الحرجة لتوزيع إختبار كاي تربيع لمستويات الدلالة المختلفة من جداول القيم الحرجة المتاحة على شبكة الإنترنت (على سبيل المثال في [29]) أو من وظيفة دالة الجدول المناسب<sup>4</sup>.

وفي إختبار الفرضية، نقول فرضية العدم المعتادة أنّه لا يوجد فرق بين المجموعات التي تُستمدّ منها البيانات. إذا كانت فرضية العدم غير صحيحة، تكون الفرضية البديلة إذن هي الصحيحة وأنّ الفرق موجود. بما أنّ فرضية العدم والفرضية البديلة لم يحددا أي فرق، يعتبر الإختبار مزدوجاً. وفي إختبار التكافؤ الأحادي الجانب، تحدد الفرضية البديلة إتجاهاً على سبيل المثال في الطّب حيث العلاج الفعال خير من العلاج الوهمي [19]. إذا تم تنفيذ إختبار تكافؤ مزدوج، ينبغي تحديد منطقة القبول من قبل إثنين من القيم الحرجة. عندما تكون القيمة التي تم الحصول عليها من إحصاء إختبار أكبر من القيمة الحرجة العليا أو أقل من القيمة الحرجة الدنيا، يتم دحض فرضية العدم. وبالنسبة لإختبار التكافؤ الأحادي الجانب، يتم تحديد منطقة القبول من خلال قيمة حرجة واحدة فقط.

وينبغي استخدام إختبار التكافؤ المزدوج بإستثناء حالة وجود سبب وجيه لعدم القيام بذلك. وإذا كان هناك ضرورة لإستخدام إختبار التكافؤ الأحادي الجانب، ينبغي تحديد إتجاه الإختبار مسبقاً أي قبل جمع البيانات [19] [20]. يكون إختبار التكافؤ الأحادي الجانب مناسباً عندما يكون بإمكانك القول على وجه اليقين (وقبل جمع أي بيانات) أنّه لن يكون هناك فرقاً أو أنّ الفرق يسلك إتجاهاً يمكنك تحديده مسبقاً. إذا لم تتمكن من تحديد إتجاه أي فرق قبل جمع البيانات، إذن يكون الإختبار المزدوج هو الأنسب [20].

وبالنسبة لإختبار فرضية إتفاقيّة ورشة عمل اللّجنة الأوروبيّة لتوحيد المعايير 15044، يعتبر الإختبار المزدوج الخيار الأنسب كوننا لا نستطيع عادةً أن نذكر على وجه اليقين قبل جمع البيانات أنّه وعلى سبيل المثال، أنّ هذه الآلة أفضل من تلك في مجال تحييد أهداف الإختبار أو أنّ الآلة الخاصّة ستكون أفضل في عمليّة تحييد أهداف الإختبار في الرمال أكثر من سطح التربة، إلخ.

قد يكون إختبار التكافؤ الأحادي مناسباً، على سبيل المثال، عندما تتم مقارنة تطوّر الآلة بالنسبة لإصدارها السابق، حيث نستطيع القول على وجه اليقين أنّ الآلة لن تكون أسوأ لدى عملية إزالة أهداف الإختبار لا بل مثل أو أفضل من الإصدار السابق.

عندما يتم إجراء إختبار كاي التربيع المزدوج عند مستوى دلالة 5% ( $\alpha = 0.05$ )، تتم مقارنة القيمة التي تم الحصول عليها لإحصاء كاي تربيع باثنين من القيم الحرجة ذات الصلة  $\alpha / 2 = 0.025$  و  $1 - \alpha / 2 = 0.975$ . عندما تكون قيمة إحصائية الإختبار أكبر من القيمة الحرجة العليا ( $\alpha / 2 = 0.025$ ) أو أقل من القيمة الحرجة الدنيا ( $1 - \alpha / 2 = 0.975$ )، يتم دحض فرضية العدم. عندما يتم إجراء إختبار كاي التربيع الوحيد عند مستوى دلالة 5% ( $\alpha = 0.05$ )، تتم مقارنة القيمة التي تم الحصول عليها لإحصاء كاي تربيع بالقيمة الحرجة العليا ( $\alpha = 0.05$ ). عندما تكون قيمة الإختبار الإحصائية أكبر من هذه القيمة الحرجة العليا، يتم دحض فرضية العدم [29].

In Excel: CHIINV( $\alpha$ ,df) <sup>4</sup>

وبالنسبة لأمتلة إتفاقيّة ورشة عمل اللجنة الأوروبية لتوحيد المعايير **15044** الواردة أعلاه نعتبر: (1) أنه في الإختبار المزدوج لا يختلف أداء الآلة (الآلات) في فرضية العدم لا يختلف في حين يختلف الأداء في الفرضية البديلة و(2) يكون مستوى الدلالة 0.05. والقيم الحرجة كالاتي (أنظر الجداول [29])

(أ)

درجات الحرية=1

القيمة الحرجة العليا:  $5.024 = 0.025 = 2/\alpha$

القيمة الحرجة الدنيا:  $0.001 = 0.975 = 2/\alpha - 1$

(أ)

درجات الحرية= 1

القيمة الحرجة العليا:  $5.024 = 0.025 = 2/\alpha$

القيمة الحرجة الدنيا:  $0.001 = 0.975 = 2/\alpha - 1$

(ت)

درجات الحرية= 2

القيمة الحرجة العليا:  $7.378 = 0.025 = 2/\alpha$

القيمة الحرجة الدنيا:  $0.051 = 0.075 = 2/\alpha - 1$

من الممكن تفسير نتائج الاختبارات على النحو الآتي:

(ب)

الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) المصحح =  $5.024 > 0.18$

الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) المصحح =  $0.001 < 0.18$

قبول فرضية العدم  $H_0$

لا يختلف أداء آلات إختبار الألغام سواء في الرمل أو الحصى بالنسبة للألغام المظمورة على سطح الأرض.

(أ)

الاختبار الاحصائي كاي تربيع ( $\chi^2$ ) المصحح =  $5.024 < 6.06$

دحض فرضية العدم  $H_0$

قبول الفرضية البديلة  $H_1$

يختلف أداء آلات الإختبار بالنسبة للألغام المظمورة على سطح الأرض في الرمال.

(ت)

$7.378 < 10.21 = (\chi^2)$

دحض فرضية العدم  $H_0$

## قبول الفرضية البديلة $H_1$

يختلف أداء الآلات مع إختلاف عمق الهدف.

يمكن إجراء الحسابات المذكورة أعلاه من خلال مجموعة بيانات إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 شريطة أن يتم ضمان حجم عينة كبيرة بما فيه الكفاية. لا يوجد حد أدنى مقبول معتمد. إذ أن البعض يقول أن حجم العينة ينبغي أن يكون 50 في حين حددها البعض الآخر بـ 20. وتجدر الإشارة إلى أن إختبار كاي تربيع يجب أن يكون محسوبًا على بيانات العد الفعلي (وليس على النسب المستبدلة) ويفترض أيضا أحجام خانات كافية. فالبعض حدّد عدد الخانات بـ 5 أو أكثر في حين حددها آخرون بـ 10 أو أكثر. وإنّ إستخدام إختبار كاي تربيع ليس مناسبًا في حال كانت الخانات أقلّ من 1 أو إذا كانت أقل من 5 في أكثر من 20% منها. وفي جداول القيم المتوقعة والقيم المرصودة ذات الصلة بإختبار كاي تربيع للإستقلالية، عادة ما تُعتبر الخانات المتوقعة أقل من 5 مقبولة في حال تمّ إستخدام تصحيح بيتس [24]. فالآلة الحاسبة الظاهرة في [24] تسمح بإحتساب تفاعلي لإحصائية الإختبار وتشير ما إذا كانت إحصائية كاي تربيع مناسبة لمجموعة البيانات الواردة.

يلخص المنحنى الظاهر في المستند 5 من إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، الحسابات الذالة على الفروقات بين العينتين، حيث تحتوي كل عينة على 50 هدف إختبار. يشير المنحنى إلى نقطة ارتكاز حيث شكّل الفرق الملاحظ في نسبة التحييد في خلال الإختبار فرقًا لا يُستهان به من الناحية الإحصائية. ويقدم الملحق 1 تفاصيل إضافية حول الافتراضات المستخدمة لإعداد المنحنى المبيّن في المستند 5 من إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 والافتراضات البديلة ومن ثم الحسابات التي يمكن استخدامها. ويوضح كذلك تأثير الافتراضات المختلفة على استنتاجات التجربة مع بعض الأمثلة العملية.

ومن وجهة نظر إحصائية محض، يتيح المستند 5 من إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 استنتاج إختلاف أداء الآلة ولكن بالمقابل، لا يمكن القول أن الآلة تؤدي عملها أفضل من غيرها.



### 3- المراجع

- [1] كتيب الإجراءات الإحصائية البارامترية وغير البارامترية، النسخة الثالثة، D. J. Sheskin, Chapman & Hall/CRC، 2004.
- [2] الطرق الإحصائية الأولية للغابات، الدليل الزراعي 317، F. Freese، وزارة الزراعة الأمريكية، دائرة الغابات، 1980
- [3] الإحصاءات التمهيدية: المفاهيم، النماذج، والتطبيقات. توزع العينات، <http://www.psychstat.missouristate.edu/Introbook/sbk19.htm>
- [4] هيرستات أونلين، توزع العينات، <http://davidmlane.com/hyperstat/A11150.html>
- [5] مركز الويب لأساليب البحث الاجتماعي، الشروط الإحصائية في أخذ العينات، <http://www.socialresearchmethods.net/kb/sampstat.php>
- [6] جامعة بيل، قسم الإحصاء، أخذ العينات في الاستدلال الإحصائي، <http://stat.yale.edu/Courses/1997-98/101/sampinf.htm>
- [7] جامعة بيل، قسم الإحصاء، توزع العينات الثنائي، <http://stat.yale.edu/Courses/1997-98/101/binom.htm>
- [8] دليل الإحصاء الهندسية، ما هي فترات الثقة؟ <http://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm>
- [9] جامعة كونيتيكت، فترة الثقة <http://www.gifted.uconn.edu/siegle/research/Samples/ConfidenceInterval.htm>
- [10] دروس إحصائية: توزع العينات الثنائي، StatTrek Teach Yourself Statistics، <http://stattrek.com/Lesson2/Binomial.aspx?Tutorial=Stat>
- [11] تقدير معدلات الإنجاز من العينات الصغيرة باستخدام فترات الثقة الثنائية: مقارنات وتوصيات، J. Sauro, J.R. Lewis، وقائع الاجتماع السنوي التاسع والأربعون لجمعية العوامل الإنسانية وبيئة العمل، 2005 <http://www.measuringusability.com/papers/sauro-lewisHFES.pdf>
- [12] فهم فترات الثقة الثنائية، P. Mayfield، SigmaZone.com، 2008 [http://www.sigmazone.com/binomial\\_confidence\\_interval.htm](http://www.sigmazone.com/binomial_confidence_interval.htm)
- [13] المركز الإحصائي للمجموعة الجنوبية الغربية للأورام، احتساب فترة الثقة الثنائية، [http://www.swogstat.org/stat/public/binomial\\_conf.htm](http://www.swogstat.org/stat/public/binomial_conf.htm)
- [14] قياس قابلية الاستخدام، حاسبة فترة الثقة للحصول على معدل الإنجاز، J. Sauro، 2005 <http://www.measuringusability.com/wald.htm>
- [15] جامعة دي بول، اختبار الفرضية ومقارنة اثنين من النسب، <http://condor.depaul.edu/~dallbrit/extra/psy241/psy241-lec10a-chi-square.ppt#257>
- [16] دروس إحصائية: اختبارات الفرضية، StatTrek Teach Yourself Statistics، <http://stattrek.com/Lesson5/HypothesisTesting.aspx>
- [17] تصميم البحوث في التعليم المهني، الوحدة S7- كاي مربع، J.P. Key، 1997. <http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm>
- [18] MicrobiologyBytes: رياضيات ومعلوماتية للمتخصصين في علم الأحياء: إحصاءات استنتاجية- مقارنة المجموعات II- اختبار كاي المربع. <http://www.microbiologybytes.com/maths/1011-21.html>
- [19] ملاحظات الإحصاءات: اختبار تكافؤ أحادي الجانب واختبار تكافؤ مزدوج، J.M. Bland, D.G. Bland، مختبر الإحصائيات الطبية، الصندوق الملكي البريطاني لأبحاث السرطان، لندن، 1994 <http://www.bmj.com/cgi/content/short/309/6949/248>
- [20] الإحصائيات الحيوية البديهية: اختيار اختبار إحصائي، GraphPad.com، <http://www.graphpad.com/www/Book/Choose.htm>

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

- [21] الدروس الإلكترونية في الطريقة الكمية للعلوم الإجتماعية. إختبار كاي المربع  
[http://ccnmtl.columbia.edu/projects/qmss/the\\_chisquare\\_test/about\\_the\\_chisquare\\_test.html](http://ccnmtl.columbia.edu/projects/qmss/the_chisquare_test/about_the_chisquare_test.html)
- [22] روابط، ما هو احصاء كاي المربع؟ M. Mamahlodi، 2006  
<http://cnx.org/content/m13487/latest/>
- [23] قسم الرياضيات والمعلوماتية، كليات هوبارت وويليام سميثن، إحصاء كاي المربع  
<http://math.hws.edu/javamath/ryan/ChiSquare.html>
- [24] أداة حسابية تفاعلية لإختبارات كاي المربع حول حسنات الصلاحية والإستقلالية، K. J. Preacher، 2001،  
 متاح على: <http://www.quantpsy.org>, <http://people.ku.edu/~preacher/chisq/chisq.htm>
- [25] عالم الرياضيات - A Wolfram Web Resource، توزيع كاي المربع، E. W. Weisstein،  
<http://mathworld.wolfram.com/Chi-SquaredDistribution.html>
- [26] Hyperstat Online، درجات الحرية  
<http://davidmlane.com/hyperstat/A42408.html>
- [29] دليل الإحصاءات الهندسية، القيم الحرجة لتوزيع كاي المربع،  
<http://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm>
- [30] جامعة ولاية نورث كارولينا، كلية العلوم الإنسانية والاجتماعية، إختبارات كاي المربع للتكافؤ  
<http://faculty.chass.ncsu.edu/garson/PA765/chisq.htm>
- [31] ما هذه الضجة حول العشوائية الصحيحة؟ مولد عشوائي، <http://www.random.org/>
- [32] مولد رقم عشوائي ومدقق  
<http://www.psychicscience.org/random.aspx>
- [33] مفاهيم وتطبيقات إحصائية استنتاجية، L. Lowry،  
<http://faculty.vassar.edu/lowry/webtext.html>
- [34] دليل الإحصاءات الهندسية، معرض التوزيعات،  
<http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm>
- [35] Hyperstat Online، الانحراف المعياري والتباين،  
<http://davidmlane.com/hyperstat/A16252.html>
- [36] Pearson Education، جدول الاحتمالات ذات اتجاهين.  
<http://wps.aw.com/wps/media/objects/384/394213/Binomial%20Table.pdf>



البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

#### 4-الملحق 1: نهج إحصائية للحصول على جزء التحييد الذي أصبحت فيه الفروقات الملاحظة في نسبة التحييد ذات دلالة إحصائية.

ويستند الرسم البياني في المستند 5 من اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044، والذي يظهر اختلافاً ملحوظاً إحصائياً في اختبار أداء إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير، على المبادئ الإحصائية المفصلة في سياق الدروس المستفادة. وكانت الصيغ هي نفسها المستخدمة في اختبار كاي المربع ذات الإتجاه الواحد من دون تصحيح بيتس.

إذا تم تطبيق نظرية إحصائية بشكل صحيح على اختبار أداء إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 فيكون اختبار كاي المربع ذات الإتجاهين الخيار الأفضل على الأرجح ويمكن كذلك تطبيق تصحيح بيتس (انظر صلب النص الأساسي لمزيد من التوضيح).

يوضح هذا الملحق الاستنتاجات المختلفة التي يمكن للمرء أن يستخلصها عند استخدام النهج المختلفة (اختبار كاي المربع ذات الجانب الواحد مع تصحيح بيتس، واختبار كاي المربع ذات الإتجاهين من دون تصحيح بيتس، واختبار كاي المربع ذات الإتجاه الواحد مع تصحيح بيتس واختبار كاي المربع ذات الإتجاهين مع تصحيح بيتس). ويتبين أن الرسم البياني أدناه الوارد في إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 هو الأقل تحفظاً، أي أن الفروقات الصغرى مهما كانت في نسبة التحييد تعتبر مختلفة إحصائياً.

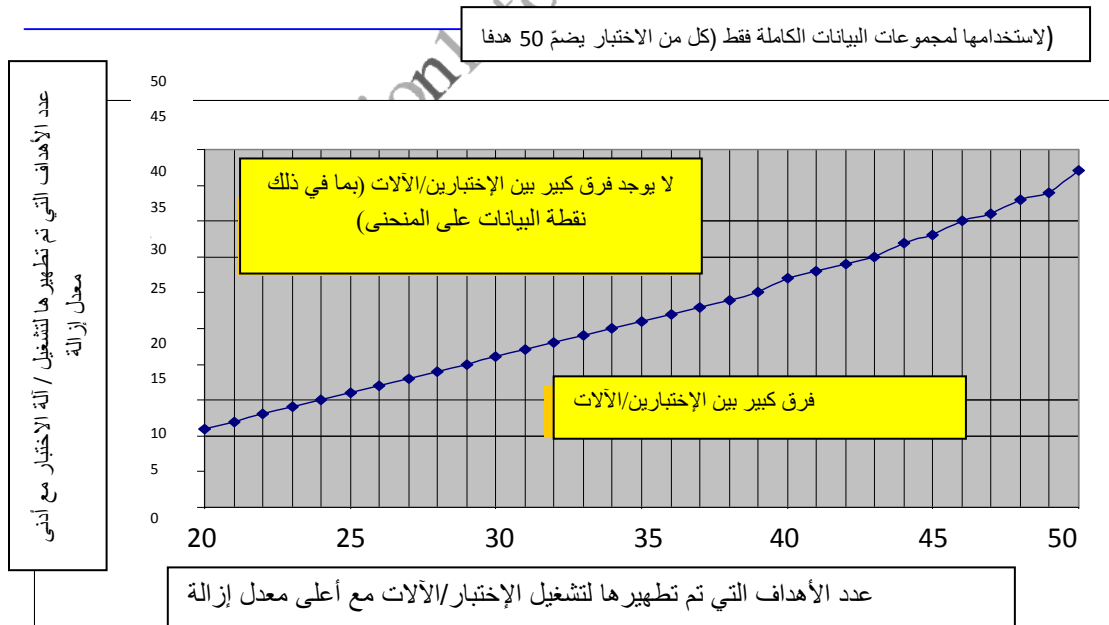
واستناداً إلى الرسوم البيانية والشروحات الواردة في وثيقة الدروس المستفادة، يعتبر من الضروري اختيار النهج الأفضل من قبل المستخدمين وذلك وفق الاغراض المنشودة.

#### 4.1- الرسوم البيانية

حسابات الرسوم البيانية أدناه متاحة على الرابط الآتي:

[http://www.itep.ws/pdf/CWA15044/Statistics\\_CWA15044.xls](http://www.itep.ws/pdf/CWA15044/Statistics_CWA15044.xls)

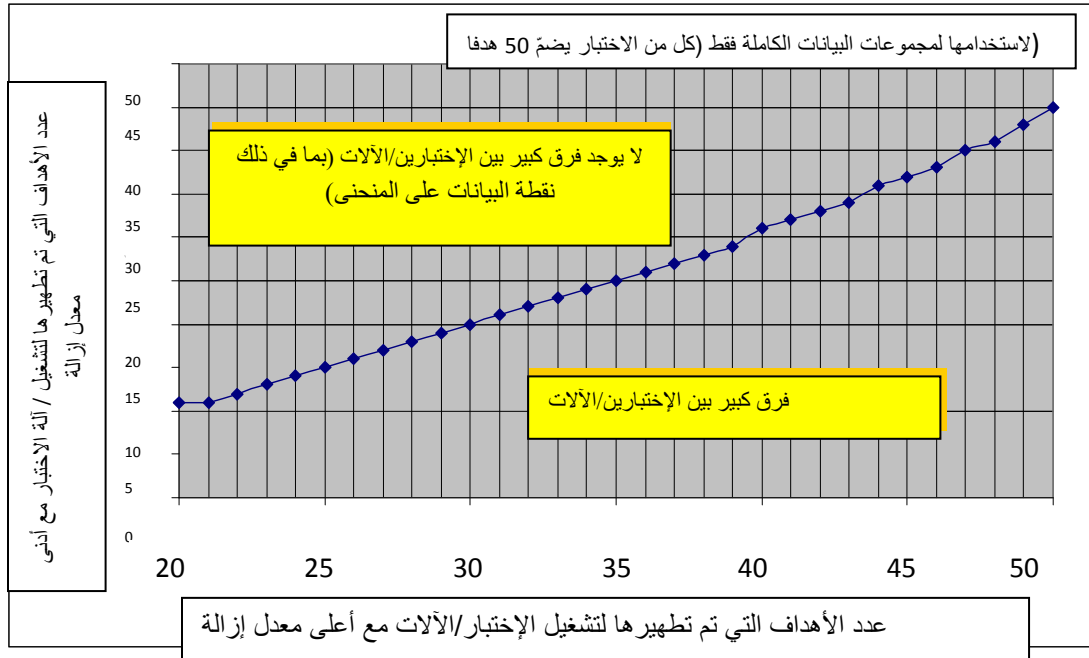
#### النهج 1، إختبار كاي المربع الأحادي الجانب من دون تصحيح بيتس



البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

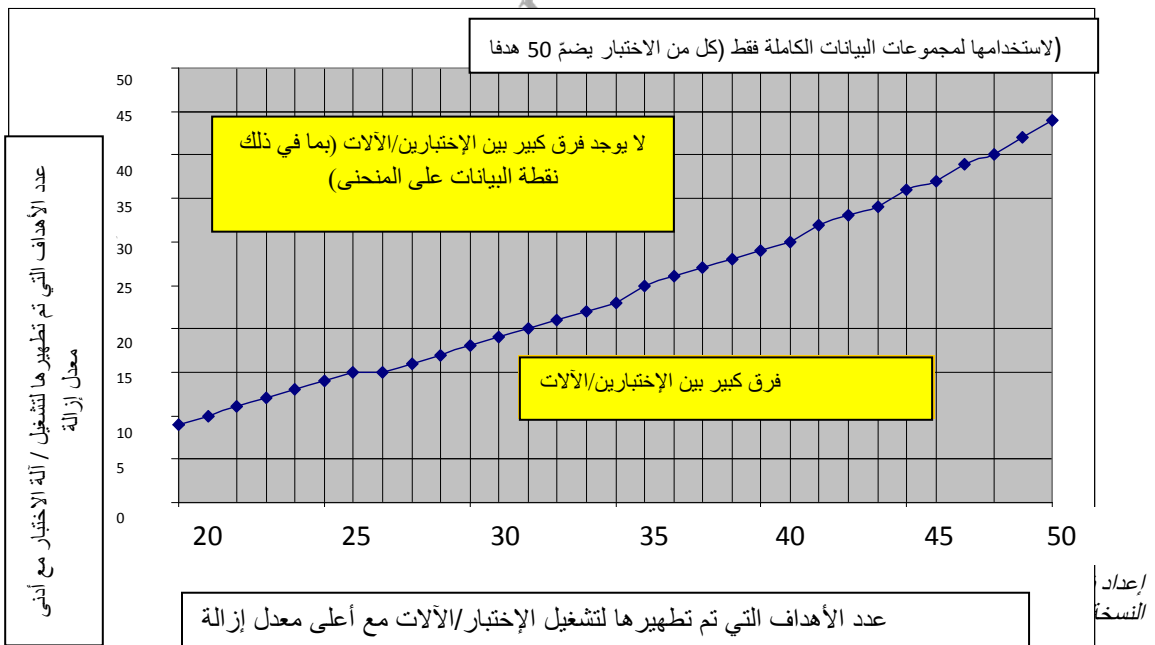
على سبيل المثال: تم تشغيل الإختبار مرتين، وكلّ منهما تضمّ 50 هدفًا. بالنسبة لمرّة الأولى التي تظهر تحييد 50 هدفًا بشكل تامّ، يحتاج الإختبار الثاني إلى ترك أكثر من 4 أهداف فعّالة (أي أقلّ من 46) للفرق الملحوظ ليكون مختلف إحصائيًا.

النهج 3: إختبار كاي المربع الأحادي الجانب مع تصحيح بيتس



على سبيل المثال: تم تشغيل الإختبار مرتين، وكلّ منهما تضمّ 50 هدفًا. بالنسبة لمرّة الأولى التي تظهر تحييد 50 هدفًا بشكل تامّ، يحتاج الإختبار الثاني إلى ترك أكثر من 5 أهداف فعّالة (أي أقلّ من 45) للفرق الملحوظ ليكون مختلف إحصائيًا.

النهج 4: إختبار كاي المربع الثاني مع تصحيح بيتس



البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

على سبيل المثال: تم تشغيل الإختبار مرتين، وكلّ منهما تضمّ 50 هدفًا. بالنسبة لمرّة الأولى التي تظهر تحييد 50 هدفًا بشكل تامّ، يحتاج الإختبار الثاني إلى ترك أكثر من 6 أهداف فعالة (أي أقلّ من 44) للفرق الملحوظ ليكون مختلف إحصائيًا.

## 4.2- تطبيق نهج إحصائية مختلفة على بيانات اختبار إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044

4.2.1 - نتائج اختبار أداء إتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 حول مدرسة ميني ميني ولف وأدوات الحراثة.

البيانات مأخوذة من تقرير الاختبار (مشروع البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية)

<http://www.itep.ws/pdf/FinalReportMiniMineWolf2007.pdf>

محرث ميني ماين ولف			مدرسة ميني ماين ولف			
سطح التربة	الحصى	الرمل	سطح التربة	الحصى	الرمل	
50/50	50/50	50/47	50/49	50/50	50/50	سطح التربة
50/49	50/49	50/48	50/50	50/49	50/50	على عمق 10 سم
50/50	50/50	50/49	50/50	50/50	50/50	على عمق 15 سم

وباستخدام الرسم البياني العائد للنهج 1 والبيانات الواردة في الجدول أعلاه يمكن الاستنتاج أنه لا توجد فروق ذات دلالة إحصائية في أداء المدرّسة على مختلف أنواع التربة وأعماق الطمر المختلفة، أي أنّ المدرسة تعمل أيضًا في الرمل والحصى وسطح التربة وكذلك بالنسبة للألغام المظمورة على أعماق مختلفة تصل إلى 15 سم. وكذلك الأمر بالنسبة للمحرثات. وعلاوة على ذلك، تظهر البيانات أيضًا أنه لا يوجد فرق كبير في أداء المدرسو والمحرثات في مجال تحييد الألغام المستهدفة وفي جميع ظروف التي جرى اختبارها.

ومن خلال الإستنتاج السابق، نستطيع احتساب أداء المدرسة والمحرثات أيضًا المقدر لتحييد الألغام المضادة للأفراد وذلك استخدام عينة من 450 هدفًا اختبار بدلا من عينة تحتوي على 50 هدفًا اختباريا، الأمر الذي يؤدي إلى إنتاج فاصل ثقة أصغر وبالتالي الحصول على تقدير أكثر موثوقية لنسبة التحييد.

محرث ميني ماين ولف	مدرسة ميني ماين ولف	أداء التحييد
450/442	450/448	
تأكد بنسبة 95% من أنّ نسبة التحييد بواسطة محرث ميني ماين ولف ستتراوح بين 96.5% و 99.2%.	تأكد بنسبة 95% من أنّ نسبة التحييد بواسطة مدرسة ميني ماين ولف ستتراوح بين 98.4% و 99.9%.	

(للمزيد من المعلومات حول حدود فترة الثقة، أنظر وثيقة الدروس المستفادة)

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

4.2.2 نتائج اختبار اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 بالنسبة لمدرسة **Bozena-4** ومدرسة **Bozena-5**.

البيانات مأخوذة من تقارير إختبار مدرسة Bozena-4 . (مشروع البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية (3.2.22.2004

[http://www.suffield.drdc-rddc.gc.ca/reports/English/DRDC\\_Suffield\\_TR\\_2005-](http://www.suffield.drdc-rddc.gc.ca/reports/English/DRDC_Suffield_TR_2005-138.pdf)

ومدرسة Bozena-5 . (مشروع البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية [138.pdf](http://www.itep.ws/pdf/Bozena5_DRDC_2007.pdf) (3.2.33.2006

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

مدرسة Bozena-4			مدرسة Bozena-5			
سطح التربة	الحصى	الرمل	سطح التربة	الحصى	الرمل	
50/45	50/47	49/49	50/49	50/46	50/47	سطح التربة
50/49	50/48	50/50	50/48	50/47	50/48	على عمق 10 سم
50/42	50/49	50/50	50/47	50/46	50/46	على عمق 15 سم

وباستخدام الرسم البياني العائد للنهج<sup>6</sup> والبيانات الواردة في الجدول أعلاه يمكن الاستنتاج أنه لا توجد فروق ذات دلالة إحصائية في أداء مدرستَي Bozena-4 و Bozena-5 في الرمل والحصى وسطح التربة وكذلك بالنسبة للألغام المظمورة على أعماق تصل إلى 15 سم. وبالنظر إلى الاختلافات في الأداء بين مدرستَي Bozena-4 و Bozena-5، فنلاحظ أن الاختلاف الوحيد الملاحظ أنه مختلف إحصائياً وفقاً للنهج 1 هو بالنسبة للأهداف المظمورة على عمق 115 سم في الرمل. أما الاختلافات الأخرى، لا دلالة إحصائية لها.

مدرسة Bozena-5	مدرسة Bozena-4	أداء التحديد
429/449 تأكد بنسبة 95% من أن نسبة التحديد بواسطة مدرسة Bozena-5 ستتراوح بين 95% و 98.4%	450/424 تأكد بنسبة 95% من أن نسبة التحديد بواسطة مدرسة Bozena-4 ستتراوح بين 91.6% و 96.2%	
مدرسة Bozena-5، رمال، ألغام مظمورة على عمق 15 سم	مدرسة Bozena-4، رمال، ألغام مظمورة على عمق 15 سم	
50/50 تأكد بنسبة 95% من أن نسبة التحديد بواسطة مدرسة Bozena-5 ستتراوح بين 92.9% و 100%	50/46 تأكد بنسبة 95% من أن نسبة التحديد بواسطة مدرسة Bozena-4 ستتراوح بين 80.8% و 97.8%	

(للمزيد من المعلومات حول حدود فترة الثقة، أنظر وثيقة الدروس المستفادة)

وباستخدام النهج الأخرى، فالإختلاف الملحوظ بالنسبة للألغام المظمورة عند عمق 15 سم في الرمال ليس إحصائياً. وبالتالي نستطيع الإستنتاج بأن أداء مدرستَي Bozena-4 و Bozena-5 لا يختلف الحالات التي جرى اختبارها.

### 4.2.3 مقارنة نتائج اختبار اداء اتفاقية ورشة عمل اللجنة الأوروبية لتوحيد المعايير 15044 بالنسبة لمدرسات Bozena ومدرسة ميني ماين ولف.

مدرسة Bozena-4			مدرسة ميني ماين ولف			
سطح التربة	الحصى	الرمل	سطح التربة	الحصى	الرمل	
50/49	50/46	50/47	50/49	50/50	50/50	سطح التربة
50/48	50/47	50/48	50/50	50/49	50/50	10 سم
50/47	50/46	50/46	50/50	50/50	50/50	15 سم

5 - تجدر الإشارة إلى إمكانية استرداد واحد أو أكثر من أهداف اختبار بعد الاختبارات. ولكن لا يكون من المعروف ما إذا تم تحييدها أم لا، وبالتالي تم حذفها من مجموعة البيانات في هذا المثال. غير أنه في تقرير الاختبار، أدرجت الأهداف المفقودة على أنها أهداف فعالة.

6 استخدم الرسم البياني لمدرسة Bozena-4 ولكن بالنسبة لمدرسة Bozena-5، فاستخدم حاسبة اختبار كاي المربع المتاحة على الرابط الآتي: <http://people.ku.edu/~preacher/chisq/chisq.htm> بمثابة مجموعة بيانات غير مكتملة بعد. أي أن العدد الإجمالي لأهداف الإختبار كان أقل من 50 في بعض دورات الإختبار.





البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية

وباستخدام الرسم البياني العائد للنهج 1 والبيانات الواردة في الجدول أعلاه يمكن الاستنتاج أنه لا توجد فروق ذات دلالة إحصائية في الأداء بالنسبة للألغام المستهدفة المظمورة على سطح التربة في الحصى وكذلك بالنسبة للألغام المظمورة على عمق يصل إلى 15 سم. وباستخدام النهج الأخرى، فالإختلاف الملحوظ بالنسبة للألغام المظمورة عند عمق 15 سم في الرمال ليس إحصائياً. ومع ذلك، عندما تستخدم النهج الأخرى لا تعتبر الاختلافات الأخيرة كبيرة.

مدرسة 5-Bozena			مدرسة ميني ماين ولف			
سطح التربة	الحصى	الرمل	سطح التربة	الحصى	الرمل	
47/45	50/47	49/49	50/49	50/50	50/50	سطح التربة
50/49	50/48	50/50	50/50	50/49	50/50	10 سم
46/42	50/49	50/50	50/50	50/50	50/50	15 سم

وباستخدام الرسم البياني العائد للنهج 1 والبيانات الواردة في الجدول أعلاه يمكن الاستنتاج أنّ الفرق الوحيد المُلاحظ في الأداء وذات دلالة إحصائية هو بالنسبة للأهداف المظمورة على عمق 15 سم على سطح التربة. ولا يكون الفرق مهماً إذا استخدمت النهج الأخرى.

<sup>7</sup> تجدر الإشارة إلى عدم إمكانية استرداد أي من الأهداف بعد الاختبارات. ومن ثم ليس من المعروف ما إذا كان قد تم تحييدها، وبالتالي تم حذفها من مجموعة البيانات في هذا المثال. غير أنه في تقرير الاختبار، أدرجت الأهداف المفقودة كأهداف فعّالة.



البرنامج الدولي لإختبار وتقييم أنشطة إزالة الألغام للأغراض الإنسانية