



**International Test and Evaluation  
Program  
for Humanitarian Demining**

**Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire**

**AACEN 15044:2009 ; complément 4**

## **RETOUR D'EXPERIENCE**

**Test et évaluation de l'équipement de déminage mécanique  
dans le cadre de l'Accord d'Atelier du CEN (CWA 15044:2004,  
remplacé par AACEN 15044:2009)**

**4ème partie: méthodes statistiques employées pour calculer  
les performances des engins de déminage, les intervalles de  
confiance et les différences de performance**

**(dernière mise à jour en date du 15 juillet 2010)**

## Table des matières

1. Introduction.....	2
2. Principes statistiques.....	2
2.1 Terminologie et définitions.....	2
2.2 Calcul de l'intervalle de confiance.....	7
2.3 Vérification d'hypothèses.....	10
3. Références.....	16
4. Annexe 1: approches statistiques pour obtenir la part de neutralisation observée où les différences constatées sur les pourcentages de neutralisation deviennent statistiquement significatives .....	18
4.1 Graphiques.....	18
4.2 Application des différentes approches statistiques aux données des tests de performance du AACEN 15044.....	21

## 1. Introduction

La conception des tests de performance décrits dans l'AACEN 15044 (Test et évaluation de l'équipement de déminage mécanique), ainsi que les calculs qui en résultent pour déterminer les performances des engins mécaniques de dépollution, utilisent des méthodes et des principes de statistiques courants.

Ce document met en évidence les méthodes et principes statistiques utilisés dans le contexte de l'AACEN 15044, et fournit également une liste de références permettant des informations plus approfondies. Les procédures statistiques générales sont décrites point par point dans le corps du texte, puis converties pour l'application du AACEN 15044 dans les textes « encadrés » correspondants.

## 2. Principes statistiques

### *2.1 Terminologie et définitions*

La statistique est un domaine des mathématiques qui s'intéresse aux procédures et aux méthodes de recueil et de présentation de données (statistique descriptive) et qui interprète celles-ci, en tire des conclusions et fait des prédictions (statistique déductive, ou inférence statistique). De manière générale, dans le domaine de l'inférence statistique, des échantillons de données sont utilisés pour tirer des conclusions sur l'une ou plusieurs des populations d'où proviennent les échantillons. Une **population** consiste en une somme de sujets/objets partageant un facteur commun. Un **échantillon**, en revanche, est un ensemble de sujets/objets qui a été extrait d'une population [1] (les indications entre crochets renvoient à des documents fournis en liste de références, pages 18 et 19). Les objectifs fondamentaux des statistiques sont l'**estimation des paramètres d'une population** (les valeurs qui caractérisent une population donnée) et la **vérification d'hypothèses** concernant ces paramètres [2].

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Lorsque la capacité d'un engin de déminage/dépollution mécanique à neutraliser des mines est évaluée au cours d'un test de performance selon l'AACEN 15044, nous établissons la manière dont une population de mines antipersonnel va être traitée (neutralisée ou non) par l'engin. L'objectif principal du test est d'estimer la capacité de l'engin à dépolluer des mines AP et la probabilité de neutralisation de celles-ci. Pour cela, le pourcentage de mines AP neutralisées dans un couloir d'essai (l'échantillon) ainsi qu'un intervalle de confiance sont calculés. La configuration du test prescrite par l'AACEN 15044 permet également de vérifier un certain nombre d'hypothèses concernant la capacité de l'engin à dépolluer des mines AP. Les hypothèses suivantes peuvent par exemple être testées :

- La capacité de dépollution de l'engin est-elle la même sur les trois types de terrain ?
- La capacité de dépollution de l'engin sur un terrain sableux est-elle la même pour les trois profondeurs d'enfouissement des mines ?
- La capacité de dépollution dans les couches arables de deux machines différentes est-elle la même pour des mines enterrées à la limite de la surface du sol ?

Un échantillon doit être **représentatif** de la population, afin que les conclusions qui en résultent puissent être pertinentes pour la population dans sa totalité. L'échantillon idéal est un échantillon choisi au hasard, qui doit répondre aux critères suivants:

- Chaque sujet/objet de la population a les mêmes chances d'être sélectionné comme membre de l'échantillon ;
- La sélection de chaque sujet/objet est indépendante de la sélection de tout autre sujet/objet ;
- Pour une taille d'échantillon donnée, tous les échantillons potentiels issus d'une population ont les mêmes chances d'être utilisés [1].

Si les spécificités de préparation des couloirs de test de l'AACEN 15044 sont respectées, c'est-à-dire si les mines-cibles AP<sup>1</sup> sont enterrées de manière aléatoire dans les couloirs d'essai, chaque couloir représente alors un échantillon pris au hasard de la population de mines AP considérée. Cette population pourrait être constituée, par exemple, de mines AP enterrées à 10 cm de profondeur dans le gravier.

Notons qu'un plan de pose conçu par un opérateur n'est pas considéré comme aléatoire. Afin de disposer les mines-cibles de manière aléatoire, un générateur de nombres aléatoires (logiciel permettant de générer des nombres au hasard) tel que le RAND de Microsoft Excel, ou un générateur disponible gratuitement sur internet (voir, par exemple, [31] et [32]), peuvent être utilisés.

1 Les cibles-test utilisées sont représentatives d'un type générique de mines AP ou d'un type spécifique de mines AP, selon les objectifs des essais.

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Une **statistique** fait référence à une caractéristique d'un échantillon. C'est donc un nombre qui peut être calculé à partir des données recueillies sur un échantillon aléatoire. Un **paramètre**, en revanche, fait référence à une caractéristique d'une population. C'est donc un nombre décrivant une population [1] [6]. L'estimation des paramètres à partir des statistiques est un aspect essentiel de la statistique. Les statistiques recueillies à partir d'échantillons sont utilisées comme estimateurs des paramètres de la population correspondante [3].

Les statistiques définies dans l'AACEN 15044 représentent, sous forme de pourcentage, le nombre de mines-cibles AP neutralisées sur l'ensemble des mines-cibles AP enterrées dans le couloir d'essai. Ces données servent alors d'estimation de la capacité de l'engin à neutraliser les mines AP dans les conditions du couloir d'essai.

La **distribution d'échantillonnage** d'une statistique est la distribution<sup>2</sup> de cette statistique calculée à partir d'un nombre infini d'échantillons. Si une personne prend, à plusieurs reprises, des échantillons d'une taille  $n$  à partir d'une population et calcule, à chaque fois, une statistique spécifique pour cet échantillon, la distribution des valeurs obtenues pour la statistique s'appelle la distribution d'échantillonnage de cette statistique. Chaque statistique possède une distribution d'échantillonnage [3].

Imaginons que l'on utilise un engin de déminage sur un couloir de test comprenant 50 cibles-tests enterrées à 10 cm de profondeur dans le sable (taille de l'échantillon  $n=50$ ) et que le pourcentage de cibles neutralisées est déterminé. La machine est ensuite utilisée de nouveau sur un couloir de test ayant les mêmes caractéristiques (soit : 50 cibles-tests enterrées à 10 cm de profondeur dans le sable (taille de l'échantillon  $n=50$ )) et le pourcentage de cibles neutralisées est là aussi déterminé. Le second pourcentage obtenu ne sera pas nécessairement le même que le premier. Ainsi, quand le test est reproduit un nombre infini de fois, un nombre infini de pourcentages de neutralisation est obtenu. La distribution de ce nombre infini de pourcentages de neutralisation est appelée la distribution d'échantillonnage du pourcentage de neutralisation des mines.

Tout en gardant la distribution d'échantillonnage à l'esprit, il faut comprendre que, bien que la statistique obtenue à partir d'un échantillon suffisamment large s'approche très certainement du centre de distribution d'échantillonnage (puisqu'elle comprend la plupart des échantillons), il est aussi possible de tomber, par hasard, sur un des échantillons extrêmes. Si l'on prend la moyenne de distribution d'échantillonnage -la moyenne d'un nombre infini d'échantillons- on a plus de chance de s'approcher de la véritable moyenne de population - le paramètre qui nous intéresse. Ainsi, la moyenne de distribution d'échantillonnage est plus proche des paramètres de population [5]. L'étendue des statistiques pouvant être obtenue pour l'estimation des paramètres de population grâce à l'échantillonnage est nommée **l'erreur d'échantillonnage**. L'erreur d'échantillonnage donne une idée de la précision de l'estimation statistique. Une faible erreur d'échantillonnage signifie qu'il y a relativement moins de variabilité ou d'étendue dans la distribution d'échantillonnage et qu'il est donc plus probable que l'estimation obtenue soit plus proche de la valeur réelle d'intérêt de la population.

---

2 Définition et exemples de *distributions* sont disponibles dans [33] et [34]

Dans le cas du AACEN 15044, un test n'est effectué qu'une seule fois pour des conditions spécifiques (par exemple, des cibles à 10 cm de profondeur dans le sable). Ainsi, il nous faut prendre en compte que, même si le pourcentage de neutralisation obtenu s'approche du centre de distribution d'échantillonnage, c'est à dire des paramètres que nous recherchons (le pourcentage réel de neutralisation de l'engin de déminage dans les conditions données), la valeur obtenue peut être une valeur extrême de la distribution d'échantillonnage et, par conséquent, peut être relativement éloignée des paramètres que nous recherchons. Il est donc important de connaître l'erreur d'échantillonnage afin d'estimer le degré de certitude de notre estimation par rapport au pourcentage de neutralisation réel.

En pratique, l'erreur d'échantillonnage est indiquée par un **intervalle de confiance** calculé à l'aide de **l'écart-type** de la distribution d'échantillonnage<sup>3</sup>. L'écart-type est la mesure la plus communément utilisée pour l'étendue de la distribution. L'intervalle de confiance fournit une gamme des valeurs susceptible de contenir les paramètres de population qui nous concernent.

Les intervalles de confiance sont construits pour un **degré de confiance** sélectionné par l'utilisateur. Le degré de confiance permet de savoir avec quel degré de certitude les paramètres de la population sont compris dans l'étendue des valeurs fournies par l'intervalle de confiance autour de l'estimation. Le degré de confiance est exprimé en pourcentage. Un degré de confiance de 95% signifie que l'on peut être sûr à 95%. Un degré de confiance de 99% signifie que l'on peut être sûr à 99%. La plupart des chercheurs utilisent le degré de confiance de 95%. Le degré de confiance de 95% signifie que, si l'on extrait plusieurs échantillons d'une même population et qu'à chaque fois des estimations sur l'intervalle de confiance sont faites, les intervalles qui en découlent regrouperaient les véritables paramètres de la population dans 95% des cas. La probabilité que les véritables paramètres de la population se trouvent dans l'intervalle de confiance augmente avec le degré de confiance [8] [9].

Afin de calculer l'erreur d'échantillonnage et l'intervalle de confiance pour une estimation, il faut connaître la distribution d'échantillonnage. Dans le cadre du AACEN 15044, on suppose que la distribution d'échantillonnage suit la **loi binomiale de distribution** (voir encadré page suivante). La loi binomiale décrit le comportement d'une variable  $X$  si les conditions suivantes sont remplies:

- le nombre d'observations  $n$  est fixe ;
- chaque observation est indépendante ;
- chaque observation représente un de ces deux résultats (« succès » ou « échec ») ;
- la « probabilité de succès »  $p$  est la même pour chaque résultat.

Si ces conditions sont remplies, alors la variable  $X$  suit une loi binomiale avec des paramètres  $n$  et  $p$ . La loi binomiale indique la probabilité d'obtenir  $X$  résultats concordants lorsque l'on prend comme échantillon  $n$  objets/sujets dans une population avec un taux de succès théorique  $p$  [7] [10].

---

<sup>3</sup> Plus d'informations sur l'écart-type d'une distribution et les formules pour le calculer sont disponibles, par exemple voir référence [35]

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Les observations pour les tests de performance du AACEN 15044 sont des « valeurs comptées », c'est-à-dire que l'on compte le nombre de cibles neutralisées. Le nombre d'observations  $n$  a été fixé à 50 (50 mines AP cibles par couloir d'essai) et chaque observation ne peut avoir que deux résultats : mine-cible AP neutralisée (succès) ou non neutralisée (échec). La probabilité  $p$  de succès est la capacité réelle de neutralisation de l'engin dans les conditions spécifiques du couloir d'essai (nature du sol, profondeur d'enfouissement des mines) et est donc une valeur fixe. La neutralisation d'une mine AP dans le couloir d'essai est indépendante de la neutralisation d'une autre mine AP. En résumé, le nombre de mines-test neutralisées dans le couloir d'essai suit une loi binomiale de distribution.

La figure 1 montre un exemple de loi binomiale pour un couloir d'essai comportant 20 cibles-tests ( $n=20$ ) avec un engin d'une capacité théorique de dépollution de 80% ( $p=0.8$ ). Le graphique montre que si l'engin travaille sur un couloir d'essai comportant 20 mines-cibles, le taux de dépollution sera probablement de 16 mines sur 20 (dans 21,8% des tests) : la capacité de dépollution réelle de la machine. Mais il existe aussi de bonnes chances d'enregistrer un taux de dépollution de 14 mines sur 20 (dans 10,9% des tests) ou de 18 mines sur 20 (dans 13,7% des tests). Même des taux de dépollution extrêmes de 10 mines sur 20 (dans 0,2% des tests) ou de 20 mines sur 20 (dans 1,2% des tests) peuvent être enregistrés.

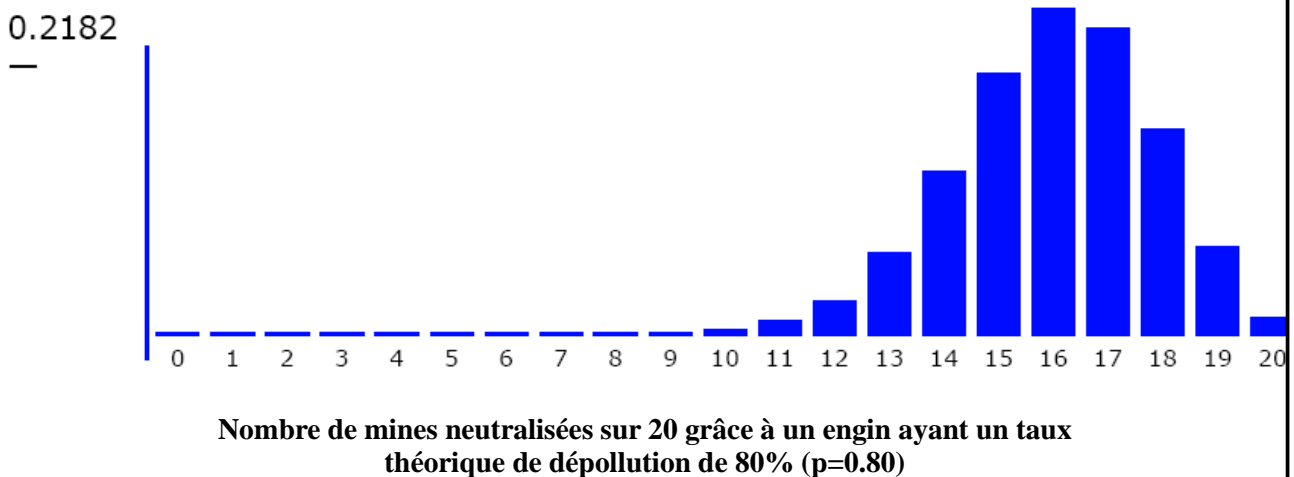


Figure 1: loi binomiale pour  $n=20$  et  $p=80$

### Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Notons que selon la théorie statistique, la distribution d'échantillonnage d'une variable-compte n'est bien décrite par la loi binomiale que dans les cas où la taille de la population est largement supérieure à la taille de l'échantillon. De manière générale, la loi binomiale ne devrait pas être appliquée aux observations sur un échantillon pris au hasard, sauf si la taille de la population est au moins 10 fois supérieure à celle de l'échantillon [7]. Dans le cas du AACEN 15044, les échantillons comportent 50 mines-test et la taille de la population (toutes les mines AP enterrées dans le sol à une profondeur spécifiée) est infinie.

La loi binomiale est une fonction mathématique à deux variables  $n$  et  $p$ . Les probabilités résultant d'une loi binomiale peuvent être directement calculées en utilisant la formule appropriée (voir [7]). Toutefois, comme ces calculs sont fastidieux, d'autres méthodes plus rapides sont disponibles, comme par exemple en déduisant les valeurs à partir de tableaux [36], en utilisant une calculatrice de loi binomiale sur internet [10] ou en utilisant une feuille de calculs contenant des fonctions statistiques (exemple : Microsoft Excel, BINOM DIST).

## 2.2 Calcul de l'intervalle de confiance

L'intervalle de confiance d'une statistique distribuée de manière binomiale peut être calculé grâce à différentes méthodes (formules), chacune d'elles fournissant des estimations d'intervalles légèrement différentes pour un même degré de confiance. Les méthodes les plus utilisées/citées sont : l'Approximation Normale, la Méthode de Wald, la Méthode de Wald ajustée, la Méthode Clopper-Pearson ou méthode Exacte, la Méthode dite des Intervalles (Wilson). Les formules correspondantes pour les intervalles de confiance, ainsi que les avantages et inconvénients de ces méthodes sont décrits en [11] et [12]. Il est facile de calculer les intervalles de confiance à l'aide de calculatrices au code source libre, comme celles présentées en [13] et [14].

Les facteurs importants permettant de déterminer la largeur de l'intervalle de confiance sont :

- La taille de l'échantillon. La probabilité que la statistique de l'échantillon reflète les véritables paramètres de la population augmente avec la taille de l'échantillon. Pour un degré de confiance donné, plus la taille de l'échantillon est importante, plus l'intervalle de confiance est petit. Cependant, la relation n'est pas linéaire (doubler la taille de l'échantillon ne diminue pas de moitié l'intervalle de confiance).
- La proportion/le pourcentage indiqué par l'échantillon. Par exemple, si 99% des cibles tests sont neutralisées et que 1% reste intact, l'incertitude du résultat est plus faible que si 51% sont neutralisées et 49% sont intactes, quelle que soit la taille de l'échantillon. Pour un degré de confiance donné, l'intervalle de confiance est plus petit avec un résultat extrême qu'avec un résultat dans la moyenne.

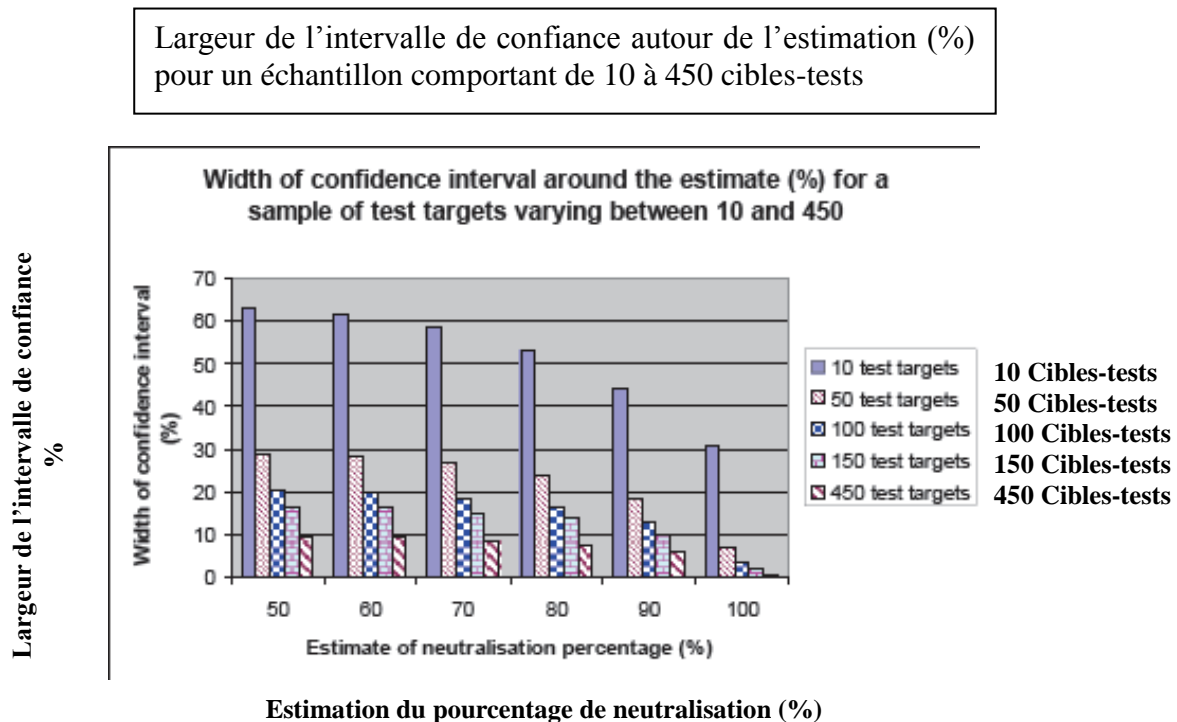
Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

La figure 2 illustre les facteurs précédents dans le cas du AACEN 15044. Plus l'échantillon est important (nombre de cibles-tests dans le couloir d'essai) et/ou plus le pourcentage de neutralisation estimé est important, plus l'intervalle de confiance autour de l'estimation est faible et donc plus l'estimation est sûre. Les intervalles de confiance ont été calculés grâce à la Méthode de Clopper-Pearson.

Pour les détails des calculs voir

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

Figure 2 : Largeur de l'intervalle de confiance (en pourcentage de l'estimation), au niveau de confiance 95%, pour des échantillons comportant entre 10 et 450 cibles-tests, et une capacité estimée de neutralisation de l'engin variant entre 50 et 100%.



Notons que les calculs de l'intervalle de confiance supposent que l'échantillon a été prélevé au hasard dans une population significative. Si l'échantillon n'est pas totalement aléatoire, l'intervalle de confiance n'est pas sûr [9].



Le AACEN 15044 recommande d'utiliser le niveau de confiance de 95% et de calculer les intervalles de confiance pour le pourcentage de neutralisation selon la Méthode de Clopper-Pearson (ou méthode exacte). Le groupe de travail à l'origine du AACEN 15044 a ensuite établi qu'un total de 50 cibles-test fournirait un intervalle de confiance satisfaisant autour de l'estimation obtenue, pour un budget d'essai acceptable.

Le graphique de la figure 3 montre que quand seulement 10 cibles-test sont utilisées, et qu'un pourcentage de neutralisation de 80% est obtenu, il existe une probabilité de 95% que la capacité réelle de l'engin dans les conditions du test (nature du sol, profondeur d'enfouissement des mines) soit entre 44% et 97%. Avec 50 cibles-test, il existe 95% de chances pour que les capacités réelles de l'engin dans les conditions du test soient entre 66% et 90%.

Pour le détail des calculs:

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

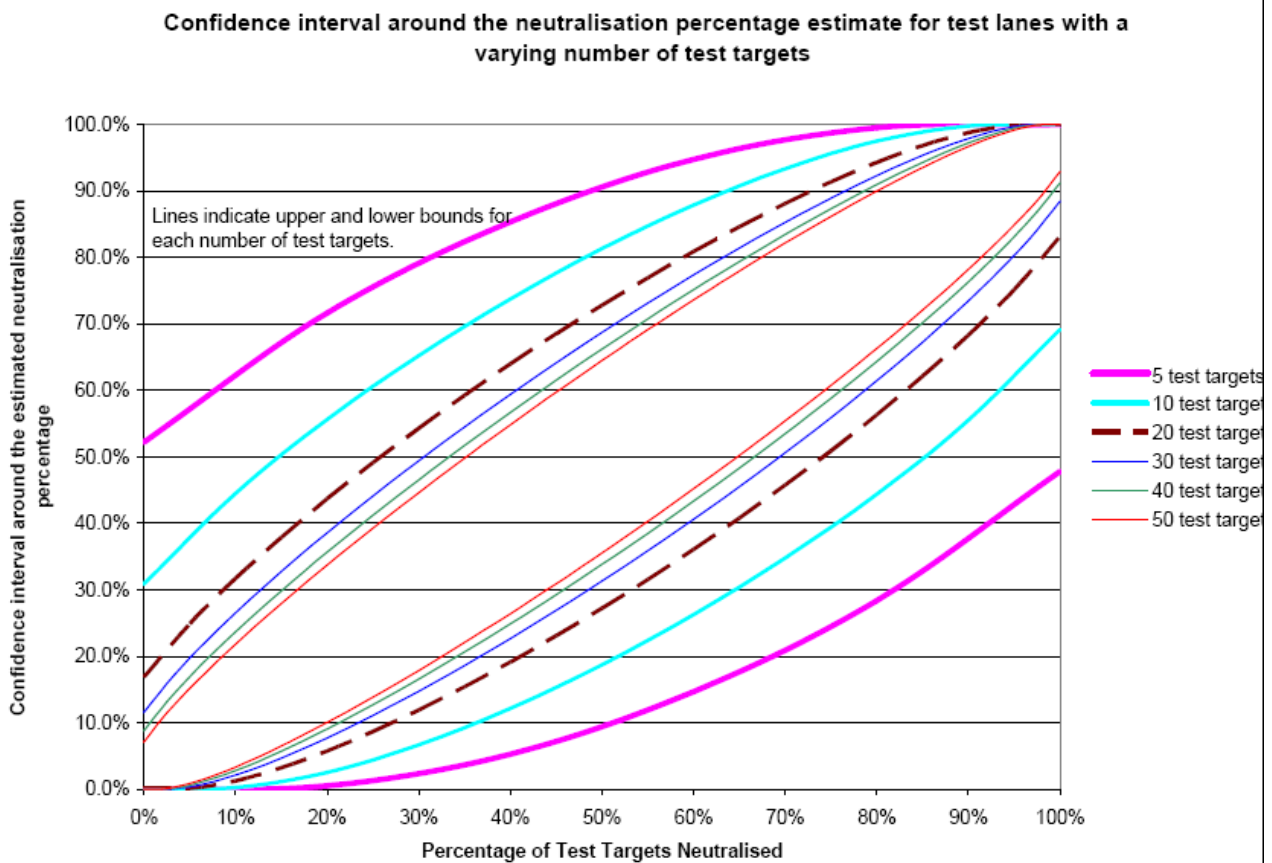


Figure 3: largeur de l'intervalle de confiance, au degré de confiance de 95%, pour des couloirs d'essai comportant un nombre variable de cibles-test.

Notons que, dans le cas hypothétique où les pourcentages de neutralisation obtenus pour tous les types de conditions testées (3 natures de sol x 3 profondeurs d'enfouissement des cibles-test) ne sont pas significativement différents, les données de neutralisation peuvent être mises en commun, ce qui donne un échantillon de 450 cibles-tests (9 tests avec 50 cibles chacun) et donc une estimation plus sûre de la capacité de neutralisation de l'engin testé. Dans le cas de l'exemple précédent, cela signifie qu'on peut être sûr à 95% que la capacité réelle de l'engin est comprise entre 76% et 84%.

### 2.3 Vérifications d'hypothèses

□

Un des aspects importants de l'utilisation des méthodes statistiques pour estimer des paramètres de populations est le fait qu'il existe des procédures statistiques standards afin de vérifier les hypothèses concernant les paramètres de populations en cours d'estimation. Lors des **vérifications d'hypothèses**, on doit décider si les données montrent un « véritable » effet ou si elles résultent du hasard, c'est à dire qu'on ne doit pas ces informations qu'à une erreur d'échantillonnage [15].

Une **hypothèse statistique** est une supposition à propos d'un paramètre de population. Cette supposition peut être avérée ou non. Si les données de l'échantillon (les observations) sont cohérentes avec l'hypothèse statistique, l'hypothèse est acceptée, sinon elle est rejetée. Il existe deux types d'hypothèses statistiques [15,16]:

- **L'Hypothèse Nulle, H0.** L'hypothèse nulle est généralement celle qui estime que les résultats collectés à partir d'échantillons ont été obtenus totalement par hasard. Il n'y a donc pas d'effet réel sur la population et les effets sur les échantillons ne sont dus qu'à l'erreur d'échantillonnage (hasard).
- **L'Hypothèse Alternative, H1.** L'hypothèse alternative est celle qui estime que les résultats d'échantillons sont influencés par une cause non aléatoire, c'est-à-dire par une variable. Alors, les différences observées dans les statistiques des échantillons reflètent des différences existantes entre les populations étudiées.

Dans le cas du AACEN 15044, une H0 serait que l'engin est aussi efficace dans le sable que dans la couche arable pour les mines enterrées à la limite de la surface du sol ; alors que l'hypothèse alternative H1 correspondante serait que les performances de l'engin sur les mines enterrées à la limite de la surface du sol sont différentes dans le sable et dans la couche arable. Une autre H0 possible serait que l'engin A est aussi efficace que l'engin B dans le sable, alors que l'H1 correspondante serait que les deux engins obtiennent des résultats différents.

Lors des vérifications d'hypothèses, l'approche suivante est utilisée:

- Énoncer les hypothèses. Cela implique d'énoncer l'hypothèse nulle (H0) et l'hypothèse alternative (H1). Les hypothèses sont énoncées de manière à ce qu'elles soient mutuellement exclusives. Ainsi, si l'une est vraie, l'autre doit être fausse ;
- Supposer que H0 est vraie ;
- Calculer la probabilité d'obtenir les résultats à partir des données si H0 était vraie ;
- Si cette probabilité est faible (par exemple moins de 5%), rejeter H0 ;
- Si H0 est rejetée, il ne reste plus que H1.

Lors des vérifications d'hypothèses, pour calculer la probabilité que l'effet observé dans les données se produirait par hasard si H0 était vérifiée, une seule **statistique-test** est calculée et évaluée. Dans le cas d'une vérification d'hypothèse pour une statistique distribuée de manière binomiale, comme une proportion (pourcentage), on utilise le **test du  $\chi^2$**  (aussi appelé chi-deux ou chi-carré) [15]. Le **test du  $\chi^2$**  est fondé sur la **statistique de test du  $\chi^2$**  et permet de savoir si les proportions de certaines catégories sont différentes dans des groupes différents [18]. Le test du  $\chi^2$  est soit un test du  $\chi^2$  simple (test **d'homogénéité du  $\chi^2$** ) soit un test du  $\chi^2$  double (**test d'indépendance du  $\chi^2$** ). Lors d'un test d'homogénéité, les observations sont comparées à une population dont les caractéristiques sont connues, alors que pour un test d'indépendance, on compare les fréquences d'occurrence de

**Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire**

deux catégories ou plus entre deux groupes ou plus. Afin de calculer la statistique du  $\chi^2$ , les données observées sont présentées dans un tableau, appelé **tableau de contingence**.

Lors des vérifications d'hypothèses du AACEN 15044, on utilise un **test d'indépendance du  $\chi^2$**  car les proportions de mines non-explosées et de mines neutralisées sont comparées pour deux groupes ou plus. Ces groupes peuvent-être, par exemple :

- a/ deux engins de déminage différents (M1, M2) traitant des mines-cibles enterrées dans le sable à la limite de la surface du sol,
- b/ un seul engin (M1) traitant des mines enterrées dans le sable, et dans le gravier, à la limite de la surface du sol,
- c/ un seul engin (M1) traitant des cibles enterrées dans le sable, à trois profondeurs différentes,

Les tableaux de contingence pouvant être établis à partir de ces exemples sont les suivants :

(a)

Nombre de mines	Mines enterrées à la limite de la surface du sable		Nombre total de mines
	M1	M2	
Neutralisées	48	40	88
Non-explosées	2	10	12
Total	50	50	100

(b)

Nombre de mines	M1, mines enterrées à la limite de la surface du sol		Nombre total de mines
	sable	gravier	
Neutralisées	48	46	94
Non-explosées	2	4	6
Total	50	50	100

(c)

Nombre de mines	M1, sable			Nombre total de mines
	Cibles enterrées à 0cm	Cibles enterrées à 10cm	Cibles enterrées à 20cm	
Neutralisées	48	47	39	133
Non-explosées	2	3	11	17
Total	50	50	50	150

Le **calcul de la statistique du test du  $\chi^2$**  repose sur la comparaison de valeurs observées et de résultats théoriques. Lors d'un test d'indépendance, la distribution des données pour les différentes catégories est inconnue. Les valeurs théoriques sont calculées à partir des valeurs observées (total des lignes du tableau, total des colonnes, total du tableau), et la somme des différences entre les valeurs théoriques et les valeurs observées est utilisée pour calculer la statistique du test du  $\chi^2$ .

Pour de plus amples informations sur le test du  $\chi^2$  et les calculs correspondants, voir [17], [21], [22], [23], [24]. Les paragraphes ci-dessous illustrent les calculs pour les exemples du AACEN 15044 cités précédemment.

**Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire**

La première étape du calcul de  $\chi^2$  consiste à générer la valeur théorique (E) pour chaque cellule du tableau contenant les données observées. La valeur théorique pour chaque cellule du tableau ( $E_{ij}$  : indices i pour les colonnes ; indices j pour les lignes) est alors calculée à l'aide de la formule suivante:

$$\text{Total de la ligne X Total de la colonne} / \text{Total du tableau}$$

Les valeurs théoriques de chaque cellule des tableaux fournis en exemple sont indiqués entre parenthèses et en italiques.

(a)

Nombre de mines	mines enterrées à la limite de la surface du sol dans le sable		Nombre de mines
	M1	M2	
Neutralisées	48 <i>(44)</i>	40 <i>(44)</i>	88
Non-explosées	2 <i>(6)</i>	10 <i>(6)</i>	12
Total	50	50	100

(b)

Nombre de mines	M1, mines enterrées à la limite de la surface du sol		Nombre de mines
	sable	gravier	
Neutralisées	48 <i>(47)</i>	46 <i>(47)</i>	94
Non-explosées	2 <i>(3)</i>	4 <i>(3)</i>	6
Total	50	50	100

©Nombre de mines	M1, sable			Nombre de mines
	Cibles enterrées à 0cm	Cibles enterrées à 10cm	Cibles enterrées à 20cm	
Neutralisées	48 <i>(45)</i>	47 <i>(45)</i>	39 <i>(45)</i>	133
Non-explosées	2 <i>(5)</i>	3 <i>(5)</i>	11 <i>(5)</i>	17
Total	50	50	50	150

L'étape suivante est de calculer la statistique du test du  $\chi^2$ , qui inclut les différences entre la valeur théorique ( $E_{ij}$ ) et la valeur observée ( $O_{ij}$ ) de chaque cellule, selon la formule suivante:

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

Tableau (a) :  $\chi^2 = 6,06$  ; Tableau (b) :  $\chi^2 = 0,71$  ; Tableau (c) :  $\chi^2 = 10,89$

Explication pour le tableau (a) :

$$\chi^2 = (48-44)^2/44 + (40-44)^2/44 + (2-6)^2/6 + (10-6)^2/6 = 16/44 + 16/44 + 16/6 + 16/6 = 32/44 + 32/6 = 6,06$$

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Notons que d'après la théorie du  $\chi^2$ , l'application de la **correction de Yates** est recommandée lors du calcul de la statistique  $\chi^2$  pour des tableaux à double lignes/colonnes comportant une ou plusieurs cellules inférieures à 5. Certains appliquent la correction pour tous les tableaux à double entrée. La **correction de Yates** est un ajustement [24] réducteur et arbitraire, dont l'application rend plus difficile d'établir des différences, c'est-à-dire que les différences observées devront être plus importantes pour que le test indique que l'hypothèse nulle doit être rejetée. Dans le test  $\chi^2$  corrigé, on ôte 0,5 à chaque valeur absolue (O-E). Les autres calculs restent identiques.

Lors des vérifications d'hypothèses du AACEN 15044, les tableaux de contingence à double entrée comportent des cellules inférieures à cinq et, par conséquent, la correction de Yates doit être appliquée.

Tableau (a) :  $\chi^2$  corrigé = 4,64 ; Tableau (b) :  $\chi^2$  corrigé = 0,18

**De même, explication pour le tableau (a) :**

$$\chi^2 \text{ corrigé} = 3,5^2/44 + 3,5^2/44 + 3,5^2/6 + 3,5^2/6 = 4,64$$

Lors des tests de vérifications, la valeur de la statistique-test calculée est ensuite comparée à une valeur seuil, appelée aussi **valeur critique**. Dans toute vérification d'hypothèses, la valeur critique dépend du **niveau de signification ( $\alpha$ )** (ou niveau statistiquement significatif) du test, et du caractère **unilatéral** ou **bilatéral** du test.

Le test de  $\chi^2$  évalue si la valeur  $\chi^2$  obtenue à partir des données observées est une valeur plausible pour une variable qui suit la distribution  $\chi^2$ , à un niveau statistiquement significatif déterminé ; elle est comparée à une valeur seuil qui détermine la limite entre 1) l'étendue des valeurs plausibles (plage d'acceptation), en supposant que l'hypothèse nulle est vérifiée et 2) l'étendue des valeurs peu plausibles (plage de rejet), en supposant que l'hypothèse nulle est vérifiée. Si les statistiques-tests tombent dans la plage d'acceptation, l'hypothèse nulle est acceptée. En revanche, si la valeur  $\chi^2$  obtenue à partir des données observées tombe dans la plage de rejet, l'hypothèse nulle  $H_0$  est rejetée et l'hypothèse alternative  $H_1$  est acceptée. La plage d'acceptation (et donc la valeur seuil) est déterminée de façon que les chances de rejet de l'hypothèse nulle quand celle-ci est vérifiée soient égales au niveau statistiquement significatif  $\alpha$ . Le niveau statistiquement significatif  $\alpha$  est lié au niveau de confiance choisi par l'utilisateur par la formule suivante : niveau de confiance (%) =  $(1-\alpha) \times 100$ . Ainsi, un niveau de confiance de 95% choisi par l'utilisateur implique l'utilisation d'un niveau significatif de 0,05 pour la vérification des hypothèses.

Dans le cas du AACEN 15044, la valeur seuil de la distribution de  $\chi^2$  est déterminée afin que la probabilité d'accepter l'hypothèse nulle quand celle-ci est vraie soit de 95% (niveau de confiance = 95%) et que les chances de rejeter l'hypothèse nulle quand celle-ci est vérifiée soient de 5% (niveau statistiquement significatif  $\alpha = 0.05$ ). Les exemples cités précédemment montrent que si l'on rejette l'hypothèse nulle et que, par conséquent, on conclut que : a) les deux engins testés n'ont pas les mêmes performances sur des mines enterrées dans le sable à la limite de la surface du sol , b) l'engin n'a pas les mêmes performances sur des mines enterrées dans le sable et le gravier à la limite de la surface du sol ou c) l'engin n'a pas les mêmes performances sur les minées enterrées dans le sable à des profondeurs différentes ; alors il y a 5% de chances que les conclusions tirées soient erronées.

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

La distribution de  $\chi^2$  est une distribution mathématique dont la forme est déterminée par le nombre de **degrés de liberté dl** [25]. Ainsi, la valeur seuil à laquelle est comparée la valeur calculée des tests dépend également des degrés de liberté de la distribution de  $\chi^2$ . Les degrés de liberté (dl) s'élèvent au nombre des informations indépendantes qui forment l'estimation d'un paramètre. De manière générale, les degrés de liberté d'une estimation sont égaux au nombre de résultats indépendants qui entrent dans l'estimation, moins le nombre de paramètres estimés comme étapes intermédiaires dans l'estimation du paramètre lui-même [26]. L'une des règles simples, lorsque l'on compare les fréquences d'occurrence dans deux catégories ou plus, entre deux groupes ou plus, est que les degrés de liberté sont égaux à : (nombre de colonnes moins une) x (nombre de lignes moins une), sans compter les lignes et les colonnes des totaux.

Quand les degrés de liberté sont connus, les valeurs seuils de la distribution de  $\chi^2$  pour des niveaux statistiquement significatifs différents peuvent être déterminées à partir des tableaux de valeurs seuils disponibles sur internet [29] ou de la fonction du tableur correspondante (pour EXCEL : CHIINV( $\alpha$ , dl)).

Lors des vérifications d'hypothèses, l'hypothèse nulle usuelle est qu'il n'existe pas de différence entre les populations d'où proviennent les données. Si l'hypothèse nulle n'est pas vérifiée, l'hypothèse alternative doit être vraie, et il doit donc exister une différence entre les populations. Puisque ni l'hypothèse nulle ni l'hypothèse alternative ne donnent d'indication sur la différence, le test est considéré comme bilatéral. Lors d'un test unilatéral, l'hypothèse alternative donne une indication : en médecine, par exemple, un traitement actif est plus efficace qu'un placebo [19]. Dans un test bilatéral, la plage d'acceptation est délimitée par deux valeurs seuils : quand la valeur obtenue est plus importante que la plus grande valeur seuil, ou plus faible que la plus petite valeur seuil, l'hypothèse nulle est rejetée. Lors d'un test unilatéral, la plage d'acceptation n'est délimitée que par une seule valeur seuil.

Il est généralement recommandé d'effectuer des tests bilatéraux, sauf contre-indication démontrée. Si l'on doit effectuer des tests unilatéraux, les spécificités du test doivent être précisées au préalable, avant de collecter les données [19], [20]. Il convient de faire un test unilatéral quand on peut affirmer avec certitude (et avant la collecte des données) qu'il n'y aura soit aucune différence, ou que l'éventuelle différence aura été anticipée. S'il est impossible de préciser la nature de la différence avant la collecte des données, il convient de faire un test bilatéral [20].

Dans le cadre des tests de vérification du AACEN 15044, il est préférable d'effectuer un test bilatéral car il est impossible, avant la collecte des données, d'affirmer avec certitude qu'un engin, par exemple, sera plus performant qu'un autre, ou que tel engin sera plus performant dans le sable qu'en couche arable, etc. Un test unilatéral sera plus approprié, par exemple, quand un engin amélioré est comparé à la version précédente et que l'on peut certifier que l'engin ne sera pas moins performant mais d'une efficacité identique ou supérieure à sa version précédente.

Quand un test bilatéral  $\chi^2$  est effectué à un niveau significatif de 5% ( $\alpha = 0,05$ ), la valeur obtenue de  $\chi^2$  est comparée aux deux valeurs seuils, soit à  $\alpha/2 = 0,025$  et  $1 - \alpha/2 = 0,975$ . Quand la valeur statistique du test est supérieure à la plus grande valeur seuil (soit  $\alpha/2 = 0,025$ ) ou inférieure à la plus petite valeur seuil (soit  $1 - \alpha/2 = 0,975$ ), l'hypothèse nulle est rejetée.

Quand un test unilatéral  $\chi^2$  est effectué à un niveau significatif de 5% ( $\alpha = 0,05$ ), la valeur obtenue de  $\chi^2$  est comparée à la plus grande (en réalité, la seule) valeur seuil soit  $\alpha = 0,05$ . Quand la valeur statistique du test est supérieure à cette valeur, l'hypothèse nulle est rejetée [29].

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Dans le cadre des exemples du AACEN 15044 cités précédemment, on suppose:

- un test bilatéral : l'hypothèse nulle est que les performances du/des engin(s) sont les mêmes alors que l'hypothèse alternative est qu'elles sont différentes ;
- un niveau significatif de 0.05. Les valeurs seuil sont donc (voir [29]) :

(a)  
dl=1  
Valeur seuil maximale  $\alpha/2=0,025=5,024$   
Valeur seuil minimale  $1-\alpha/2=0,975=0,001$

(b)  
dl=1  
valeur seuil maximale  $\alpha/2=0,025=5,024$   
valeur seuil minimale  $1-\alpha/2=0,975=0,001$

(c)  
dl=2 : (3 colonnes - 1)x(2 lignes - 1)  
Valeur seuil maximale  $\alpha/2=0,025=7,378$   
Valeur seuil minimale  $1-\alpha/2=0,975=0,051$

Les résultats des tests peuvent être interprétés ainsi:

(a)  
 $\chi^2$  corrigé=4,64<5,024 et >0,001  
H0 acceptée

(b)  
 $\chi^2$  corrigé=0,18<5,024 et >0,001  
H0 acceptée

Les engins testés affichent les mêmes performances sur des mines enterrées dans le sable à la limite de la surface du sol

L'engin est aussi efficace sur des mines enterrées dans le sable et le gravier à la limite de la surface du sol

(c)  
 $\chi^2=10,89>7,378$   
H0 rejetée et H1 acceptée

Les performances de la machine dépendent de la profondeur d'enfouissement des cibles.

Les calculs ci-dessus, illustrés par un ensemble de données du AACEN 15044, peuvent être effectués avec n'importe quel ensemble de données, du moment que la taille de l'échantillon est suffisamment importante. Il n'existe aucune limite reconnue pour cette taille. Pour certains experts, la taille minimale d'un échantillon doit être de 50 alors que pour d'autres il peut ne comporter que 20 sujets. Notons que le  $\chi^2$  doit être calculé sur des données réelles (et non sur des pourcentages) et que l'on présuppose **des tailles de cellules adéquates**. Certains experts exigent des tailles de cellules de 5 (sujets) ou plus et d'autres de 10 ou plus. Cependant, il ne convient pas d'utiliser le test du  $\chi^2$  si la fréquence est inférieure à 1 pour une quelconque cellule, ou inférieure à 5 dans plus de 20% des cellules. Dans le cas du test d'indépendance du  $\chi^2$  à double entrée, les fréquences théoriques inférieures à 5 sont généralement acceptées si la correction de Yates est utilisée [24]. La calculatrice utilisée [24] permet un calcul interactif de  $\chi^2$  et indique si la statistique de  $\chi^2$  convient pour cet ensemble de données.

La courbe de la figure 5 du AACEN 15044 (partie principale ; voir aussi l'annexe 1 du présent document) résume les calculs significatifs sur les différences entre deux échantillons, lorsque chaque échantillon contient 50 cibles-tests. La courbe indique le point-limite auquel la différence de pourcentage de neutralisation observée durant les tests peut être perçue comme une différence significative d'un point de vue statistique. L'annexe 1 fournit de plus amples détails sur les hypothèses utilisées afin d'obtenir la courbe de l'image 4 du AACEN 15044 et donc sur les calculs qui pourraient être utilisés. Elle illustre également l'effet des différentes hypothèses sur les conclusions des essais à l'aide d'exemples pratiques.

Notons que d'un point de vue purement statistique, la figure 4 du AACEN 15044 permet de conclure que les performances mécaniques des engins sont différentes, mais qu'il est impossible d'affirmer quel engin est plus performant que l'autre.

### 3. Références

- [1] Handbook of Parametric and Nonparametric Statistical Procedures, Third Edition, D. J. Sheskin, Chapman & Hall/CRC, 2004
- [2] Elementary Statistical Methods for Foresters, Agricultural Handbook 317, F. Freese, U.S. Department of Agriculture, Forest Service, 1980
- [3] Introductory Statistics: Concepts, Models, and Applications. The Sampling Distribution, <http://www.psychstat.missouristate.edu/Introbook/sbk19.htm>
- [4] Hyperstat Online, Sampling Distribution, <http://davidmlane.com/hyperstat/A11150.html>
- [5] Web Center for Social Research Methods, Statistical Terms in Sampling, <http://www.socialresearchmethods.net/kb/sampstat.php>
- [6] Yale University Department of Statistics, Sampling in Statistical Inference, <http://stat.yale.edu/Courses/1997-98/101/sampinf.htm>
- [7] Yale University Department of Statistics, The Binomial Distribution, <http://stat.yale.edu/Courses/1997-98/101/binom.htm>
- [8] Engineering Statistics Handbook, What are confidence intervals? <http://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm>
- [9] University of Connecticut, Confidence Interval, <http://www.gifted.uconn.edu/siegle/research/Samples/ConfidenceInterval.htm>
- [10] Statistical Tutorial: Binomial Distribution, StatTrek Teach Yourself Statistics, <http://stattrek.com/Lesson2/Binomial.aspx?Tutorial=Stat>
- [11] Estimating Completion Rates from Small Samples using Binomial Confidence Intervals: Comparisons and Recommendations, J. Sauro, J.R. Lewis, Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 2005 <http://www.measuringusability.com/papers/sauro-lewisHFES.pdf>
- [12] Understanding Binomial Confidence Intervals, P. Mayfield, SigmaZone.com, 2008 [http://www.sigmazone.com/binomial\\_confidence\\_interval.htm](http://www.sigmazone.com/binomial_confidence_interval.htm)
- [13] Southwest Oncology Group Statistical Center, Binomial confidence interval calculator, [http://www.swogstat.org/stat/public/binomial\\_conf.htm](http://www.swogstat.org/stat/public/binomial_conf.htm)
- [14] Measuring Usability, Confidence Interval Calculator for a Completion Rate, J. Sauro, 2005, <http://www.measuringusability.com/wald.htm>
- [15] DePaul University, Hypothesis testing and Comparing Two Proportions, <http://condor.depaul.edu/~dallbrit/extra/psy241/psy241-lec10a-chi-square.ppt#257>
- [16] Statistical Tutorial: Hypothesis Tests, StatTrek Teach Yourself Statistics, <http://stattrek.com/Lesson5/HypothesisTesting.aspx>
- [17] Research Design in Occupational Education, Module S7 – Chi Square, J.P. Key, 1997,



**Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire**

<http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm>

[18] MicrobiologyBytes: Maths & Computers for Biologists: Inferential Statistics – Comparing Groups II, chi-squared test, <http://www.microbiologybytes.com/math/1011-21.html>

[19] Statistics Notes: One and Two Sided Tests of Significance, J.M. Bland, D.G. Bland, Medical Statistics Laboratory, Imperial Cancer Research Fund, London, 1994, <http://www.bmj.com/cgi/content/short/309/6949/248>

[20] Intuitive Biostatistics: Choosing a statistical test, GraphPad.com, <http://www.graphpad.com/www/Book/Choose.htm>

[21] QMSS e-Lessons, The Chi-Square Test [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] Connections, What is the chi-square statistic? M. Mamahlodi, 2006 <http://cnx.org/content/m13487/latest/>

[23] Department of Mathematics and Computer Science, Hobart and William Smith Colleges, The Chi Square Statistic, <http://math.hws.edu/javamath/ryan/ChiSquare.html>

[24] An interactive calculation tool for chi-square tests of goodness of fit and independence, K. J. Preacher, 2001, Available from <http://www.quantpsy.org>, <http://people.ku.edu/~preacher/chisq/chisq.htm>

[25] MathWorld – A Wolfram Web Resource, Chi-Squared Distribution, E. W. Weisstein, <http://mathworld.wolfram.com/Chi-SquaredDistribution.html>

[26] Hyperstat Online, Degrees of Freedom, <http://davidmlane.com/hyperstat/A42408.html>

[29] Engineering Statistics Handbook, Critical Values of the Chi-Square Distribution, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm>

[30] NC State University, CHASS, Chi-Square Significance Tests, <http://faculty.chass.ncsu.edu/garson/PA765/chisq.htm>

[31] What's this fuss about true random? Random generator, <http://www.random.org/>

[32] Random Number Generator and Checker, <http://www.psychicscience.org/random.aspx>

[33] Concepts & Applications of Inferential Statistic, L. Lowry, <http://faculty.vassar.edu/lowry/webtext.html>

[34] Engineering Statistics Handbook, Gallery of Distributions, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm>

[35] Hyperstat Online, Standard Deviation and Variance, <http://davidmlane.com/hyperstat/A16252.html>

[36] Pearson Education, Binomial Probabilities Table, <http://wps.aw.com/wps/media/objects/384/394213/Binomial%20Table.pdf>

#### 4. Annexe 1: approches statistiques afin d'obtenir la part de neutralisation observée à partir de laquelle les différences constatées sur les pourcentages de neutralisation deviennent statistiquement significatives

Le graphique de la figure 4 du AACEN 15044, indiquant à quel moment les différences observées lors des tests de performance sont considérées comme statistiquement différentes, repose sur des principes statistiques expliqués dans le corps principal de ce document. Les formules utilisées sont celles du test unilatéral de  $\chi^2$ , sans correction de Yates.

Si la théorie statistique est correctement appliquée aux tests de performances du AACEN 15044, il est préférable d'utiliser un test de  $\chi^2$  bilatéral, et la correction de Yates pourrait être appliquée (se référer au corps principal du texte pour plus d'explications).

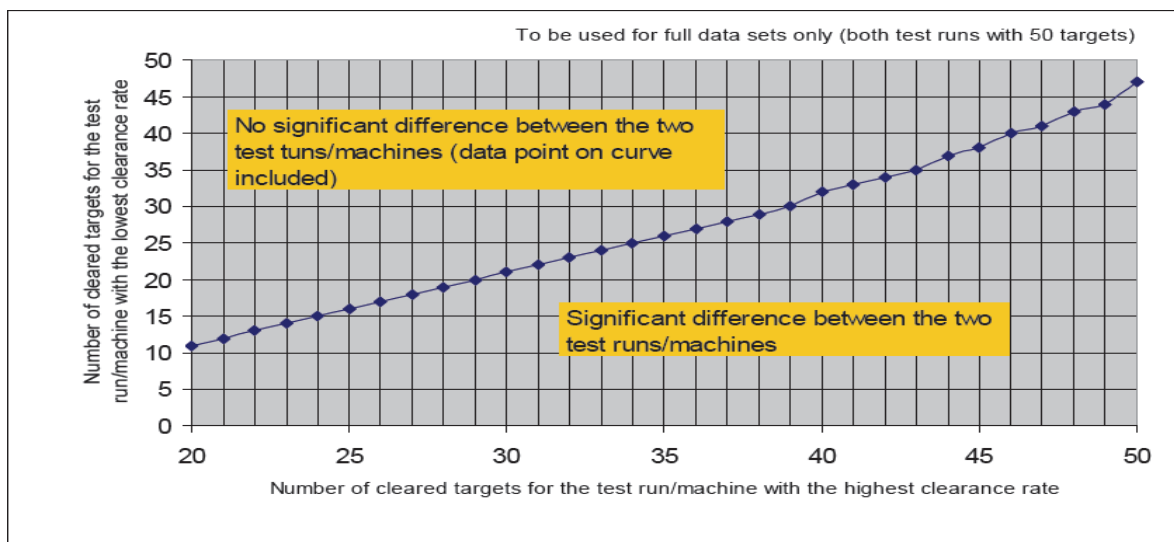
Cette annexe expose les différentes conclusions susceptibles d'être tirées en utilisant quatre approches différentes (test de  $\chi^2$  unilatéral sans correction de Yates, test de  $\chi^2$  bilatéral sans correction de Yates, test de  $\chi^2$  unilatéral avec correction de Yates, test de  $\chi^2$  bilatéral avec correction de Yates). A partir des graphiques suivants, on peut conclure que le graphique qui figure dans l'AACEN 15044 est le moins conservateur de tous, c'est-à-dire que des différences moindres dans les pourcentages de neutralisation seront considérées comme statistiquement différentes. A partir des graphiques et des explications de ce document, les utilisateurs devraient être en mesure de choisir la meilleure approche en fonction de leurs objectifs.

##### 4.1 Graphiques

Les calculs nécessaires aux graphiques suivants sont disponibles à cette adresse:

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

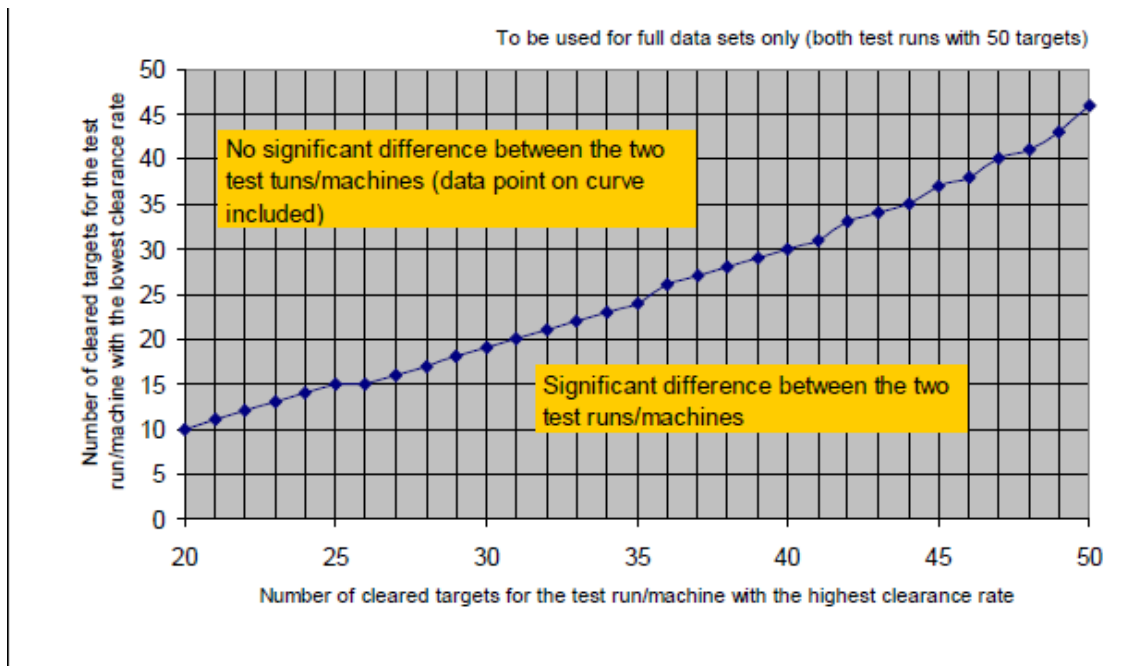
##### Approche 1, test de $\chi^2$ unilatéral sans correction de Yates



Exemple: deux tests sont effectués, chacun sur un couloir d'essais comportant 50 cibles-tests. Si l'un des tests montre que les 50 cibles ont été neutralisées, l'autre test doit comporter plus de 3 cibles non-explosées (soit moins de 47 neutralisées) afin que la différence observée soit statistiquement différente.

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

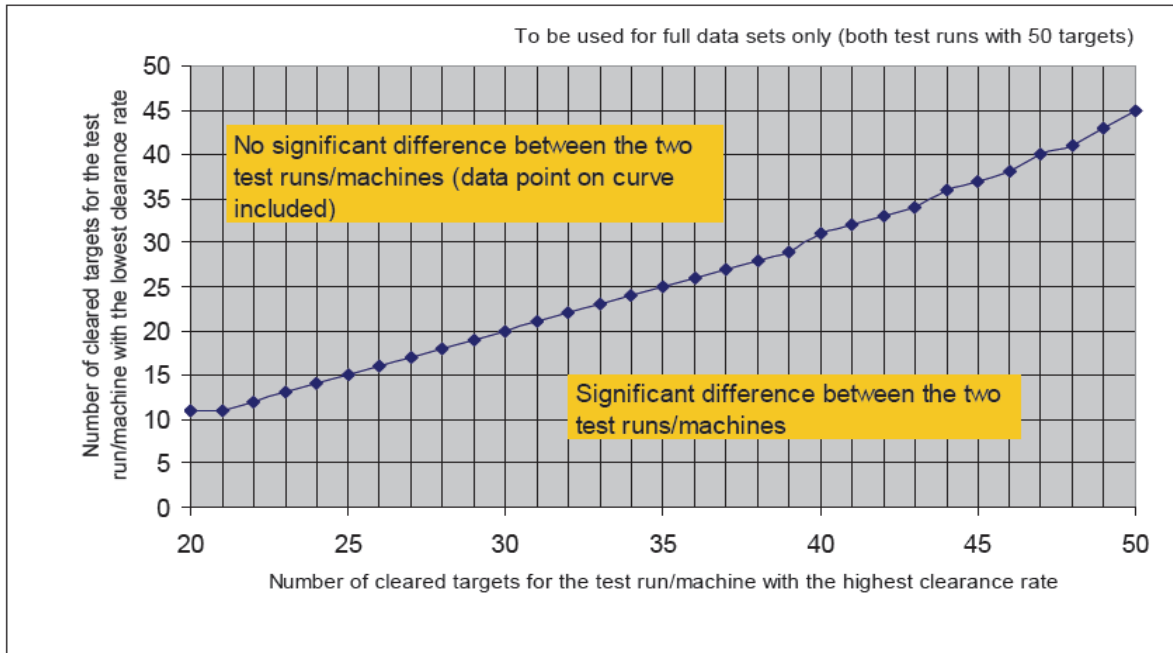
Approche 2: test de  $\chi^2$  bilatéral sans correction de Yates



Exemple: deux tests sont effectués, chacun sur un couloir d'essais comportant 50 cibles-tests. Si l'un des tests montre que les 50 cibles ont été neutralisées, l'autre test doit comporter plus de 4 cibles non-explosées (soit moins de 46 neutralisées) afin que la différence observée soit statistiquement différente.

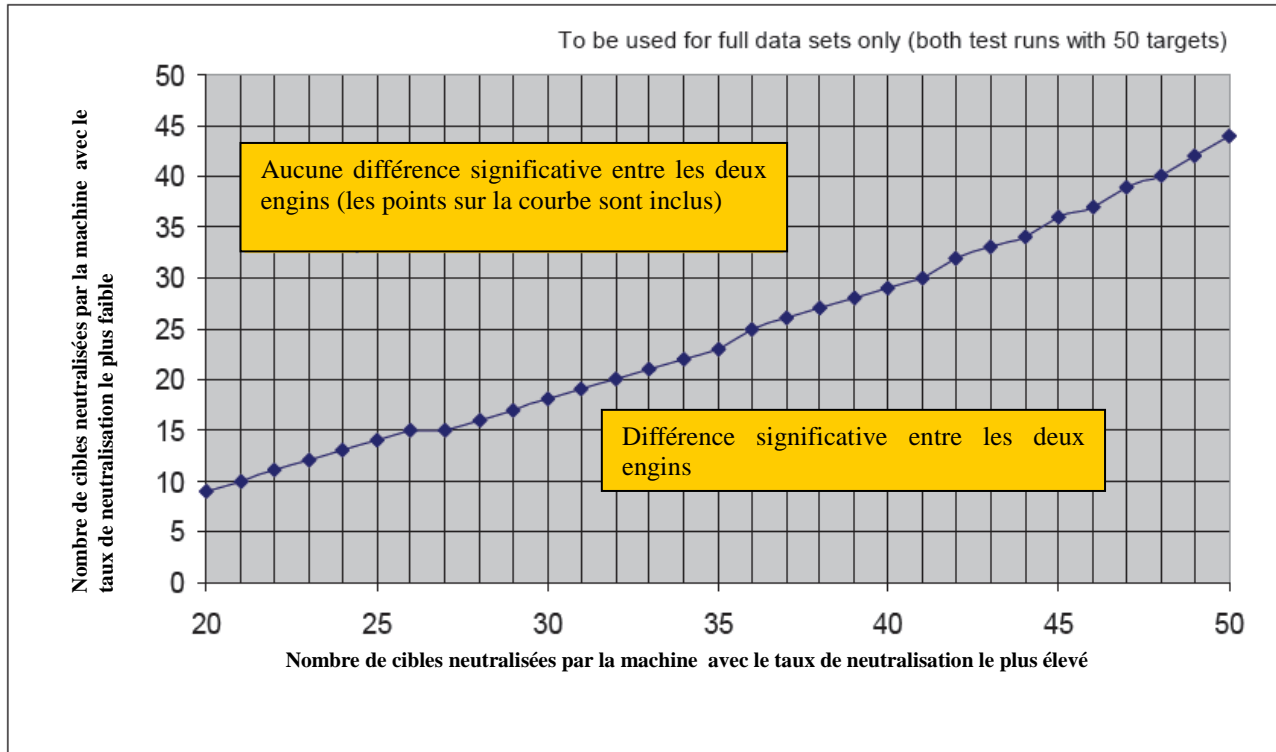
Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

Approche 3: test de  $\chi^2$  unilatéral avec correction de Yates



Exemple: deux tests sont effectués, chacun sur un couloir d'essais comportant 50 cibles-tests. Si l'un des tests montre que les 50 cibles ont été neutralisées, l'autre test doit comporter plus de 5 cibles non-explosées (soit moins de 45 neutralisées) afin que la différence observée soit statistiquement différente.

Approche 4: test de  $\chi^2$  bilatéral avec correction de Yates



Exemple: deux tests sont effectués, chacun sur un couloir d'essais comportant 50 cibles-tests. Si l'un des tests montre que les 50 cibles ont été neutralisées, l'autre test doit comporter plus de 6 cibles non-explosées (soit moins de 44 neutralisées) afin que la différence observée soit statistiquement différente.

## 4.2 Application des différentes approches statistiques aux données des tests de performances du AACEN 15044

### 4.2.1 Résultats des tests du AACEN 15044 sur le fléau Mini-MineWolf et les fraiseuses

Données extraites du rapport de test (ITEP Projet 3.2.44, 2007)

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

	fléau Mini-MineWolf			fraiseuse Mini-MineWolf		
	sable	gravier	couche arable	sable	gravier	couche arable
limite de la surface du sol	50/50	50/50	49/50	47/50	50/50	50/50
10cm	50/50	49/50	50/50	48/50	49/50	49/50
15cm	50/50	50/50	50/50	49/50	50/50	50/50

**Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire**

A l'aide du graphique correspondant à l'approche 1 et aux données du tableau ci-dessus, il est possible de conclure qu'il n'existe pas de différence significative, en ce qui concerne les performances du fléau, en fonction de la nature du sol ni de la profondeur d'enfouissement des mines ; le fléau est aussi efficace dans le sable, le gravier ou la couche arable, et sur des mines enterrées à plusieurs niveaux de profondeur, jusqu'à 15cm. Il en est de même pour les performances de la fraiseuse. De plus, les données montrent également qu'il n'y a pas de différence significative dans les performances de neutralisation des cibles-test entre le fléau et la fraiseuse, quelles que soient les conditions des tests.

Ces conclusions impliquent que les performances estimées de neutralisation des cibles AP, aussi bien pour le fléau que la fraiseuse, peuvent être calculées à l'aide d'un échantillon de 450 cibles-tests, au lieu d'un échantillon de 50 cibles, ce qui produit un intervalle de confiance plus petit et donc une meilleure estimation du pourcentage de neutralisation.

	Mini fléau MineWolf	Mini fraiseuse MineWolf
Performance de neutralisation	448/450 (= 99,6%) Sûr à 95% que le pourcentage de neutralisation réel du mini fléau minewolf se situe de 98,4% à 99,9%	442/450 (= 98,2%) Sûr à 95% que le pourcentage de neutralisation réel de la mini fraiseuse minewolf se situe de 96,5% à 99,2%

(Pour des détails sur les limites de l'intervalle de confiance, voir le corps principal du document)

#### 4.2.2 Résultats des tests de performance de l'AACEN 15044 pour le fléau Bozena-4 et le fléau Bozena-5

Données extraites des rapports de test du fléau Bozena-4 (ITEP Projet 3.2.22, 2004)

<http://www.itep.ws/pdf/Bozena4TestReport.pdf>

et du fléau Bozena-5 (ITEP Project 3.2.33, 2006)

[http://www.itep.ws/pdf/Bozena5\\_DRDC\\_2007.pdf](http://www.itep.ws/pdf/Bozena5_DRDC_2007.pdf)

	Fléau Bozena-4			Fléau Bozena-5 <sup>5</sup>		
	sable	gravier	couche arable	sable	gravier	couche arable
limite de la surface du sol	47/50	46/50	49/50	49/49	47/50	45/47
10cm	48/50	47/50	48/50	50/50	48/50	49/50
15cm	46/50	46/50	47/50	50/50	49/50	42/46

A l'aide du graphique correspondant à l'approche 1<sup>6</sup> et aux données du tableau ci-dessus, il est possible de conclure que les fléaux Bozena-4 et Bozena-5 sont aussi efficaces dans le sable, le gravier ou la couche arable pour toutes les mines enterrées à plusieurs niveaux de profondeur, jusqu'à 15cm. Si l'on s'intéresse aux différences de performance entre les deux fléaux, la seule différence observée, qui soit statistiquement significative selon l'approche 1, concerne les cibles enterrées à 15cm de profondeur dans le sable (surlignage en vert). Toutes les autres différences ne sont pas statistiquement significatives.

5 Notons qu'une ou plusieurs cibles n'ont pas été retrouvées après les tests. Il est donc impossible de savoir si elles ont été neutralisées. Elles ont donc été éliminées de l'ensemble des données dans ce cas précis. Dans le rapport de test, en revanche, les cibles manquantes ont été listées comme non-explosées.

6 Le graphique a été utilisé pour le fléau Bozena-4 ; mais pour le fléau Bozena-5, il a été fait usage de la calculatrice  $\chi^2$  (à l'adresse <http://people.ku.edu/~preacher/chisq/chisq.htm>) puisque l'ensemble des données était incomplet (le nombre total de cibles-tests était inférieur à 50 pour certains des tests effectués).

	Fléau Bozena-4	Fléau Bozena-5
Performance de neutralisation	424/450 Sûr à 95% que le pourcentage de neutralisation réel du fléau Bozena-4 se situe de 91,6% à 96,2%	429/450 Sûr à 95% que le pourcentage de neutralisation réel du fléau Bozena-5 se situe de 95% à 98,4%
	Fléau Bozena-4, sable, mines à 15cm de profondeur	Fléau Bozena-5, sable, mines à 15cm de profondeur
	46/50 Sûr à 95% que le pourcentage de neutralisation réel du fléau Bozena-4 se situe de 80,8% à 97,8%	50/50 Sûr à 95% que le pourcentage de neutralisation réel du fléau Bozena-5 se situe de 92,9% à 100%

(Pour des détails sur les limites de l'intervalle de confiance, voir le corps principal du texte)

Toutefois, si on utilise les autres approches, la différence observée pour les mines à 15 cm dans le sable n'est pas statistiquement différente. La conclusion serait alors que les fléaux Bozena-4 et -5 ont des performances identiques pour toutes les conditions testées.

#### 4.2.3 Comparaison des résultats de tests de performance du AACEN 15044 pour les fléaux Bozena et le fléau Mini Mine Wolf

	Fléau Mini Mine Wolf			Fléau Bozena-4		
	sable	gravier	couche arable	sable	gravier	couche arable
limite de la surface du sol	50/50	50/50	49/50	47/50	46/50	49/50
10cm	50/50	49/50	50/50	48/50	47/50	48/50
15cm	50/50	50/50	50/50	46/50	46/50	47/50

A l'aide du graphique correspondant à l'approche 1 et des données ci-dessus, il est possible de conclure que les différences de performance observées sont statistiquement significatives sur des cibles dans le gravier à la limite de la surface du sol (surlignage en vert), et sur les cibles enterrées à 15cm de profondeur dans le sable comme dans le gravier (surlignage respectivement en bleu et en rose). Cependant, quand les autres approches sont utilisées, ces différences ne sont pas considérées comme significatives.





International Test and Evaluation  
Program  
for Humanitarian Demining

Programme International d'Essais et d'Évaluations pour le Déminage Humanitaire

	Fléau Mini MineWolf			Fléau Bozena-5 <sup>i</sup>		
	sable	gravier	couche arable	sable	gravier	couche arable
limite de la surface du sol	50/50	50/50	49/50	49/49	47/50	45/47
10cm	50/50	49/50	50/50	50/50	48/50	49/50
15cm	50/50	50/50	50/50	50/50	49/50	42/46

A l'aide du graphique correspondant à l'approche 1 et des données ci-dessus, il est possible de conclure que la seule différence de performance observée qui soit statistiquement significative concerne les cibles enterrées à 15 cm de profondeur dans la couche arable (surlignage en vert). En revanche, elle n'est plus significative lorsque les autres approches sont utilisées.

<sup>i</sup> Notons qu'une ou plusieurs cibles-test n'ont pas été retrouvées après le test. Il est donc impossible de savoir si elles ont été neutralisées. Elles ont été éliminées de l'ensemble des données dans cet exemple. Dans le rapport de test, cependant, les cibles manquantes ont été répertoriées comme non explosées