

# **SYS865 Inférence statistique avec programmation R**

Ornwipa Thamsuwan

24 janvier 2024

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

# Plan de la séance

- ▶ Récap: variables aléatoires
  - ▶ Espérance
  - ▶ Variance et covariance
  - ▶ Indépendance
- ▶ Échantillonnage
  - ▶ Méthodes d'échantillonnage
  - ▶ Taille d'échantillon
- ▶ Début de l'inférence statistique
  - ▶ Intervalle de confiance

# Récap et matière à réflexion

Lire des données

```
data <- read.csv("diabetes.csv")
```

Espérance

```
expectations <- sapply(data, mean)
```

Variance et covariance

```
covariances <- var(data)
```

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Comment savoir si deux variables sont indépendantes l'une de l'autre ?

Comment savoir si deux variables sont indépendantes l'une de l'autre ?

- Inspection visuelle par graphiques de dispersion (“Scatter plot” en anglais)

Comment savoir si deux variables sont indépendantes l'une de l'autre ?

- ▶ Inspection visuelle par graphiques de dispersion (“Scatter plot” en anglais)
- ▶ Test de hypothèse
  - ▶ Test  $\chi^2$
  - ▶ Test de corrélation
  - ▶ Regression linéaire
  - ▶ Regression logistique



# Graphique de dispersion

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

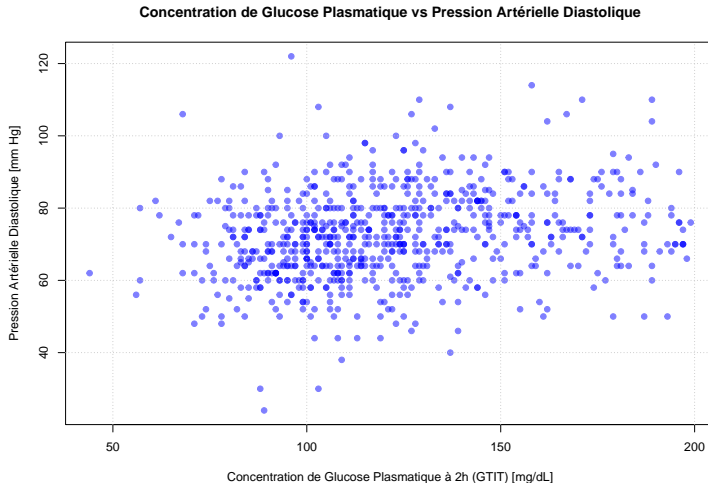
Récap et matière  
à réflexion

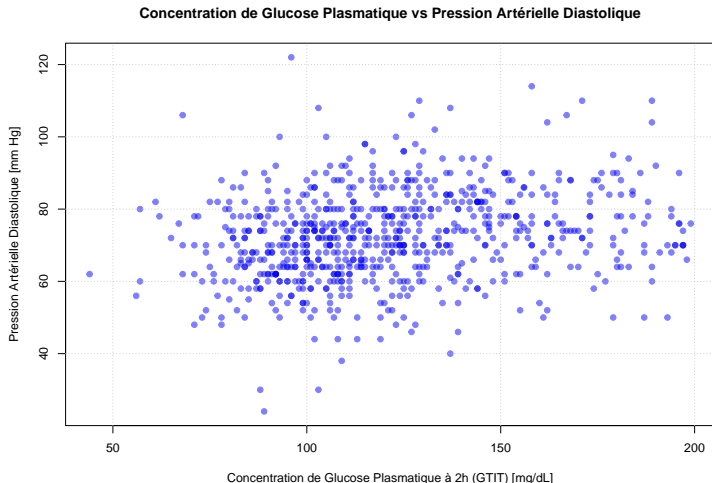
Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques



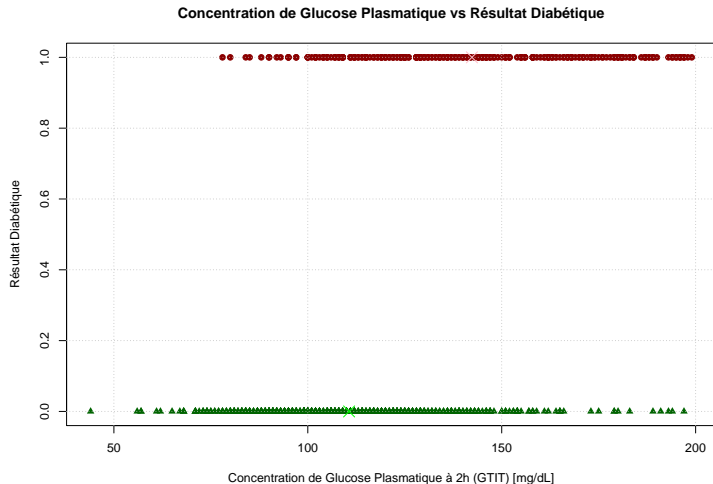


À votre avis, la covariance entre le glucose et la pression artérielle est positive, négative ou proche de zéro ?

# Graphique de dispersion

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan



Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

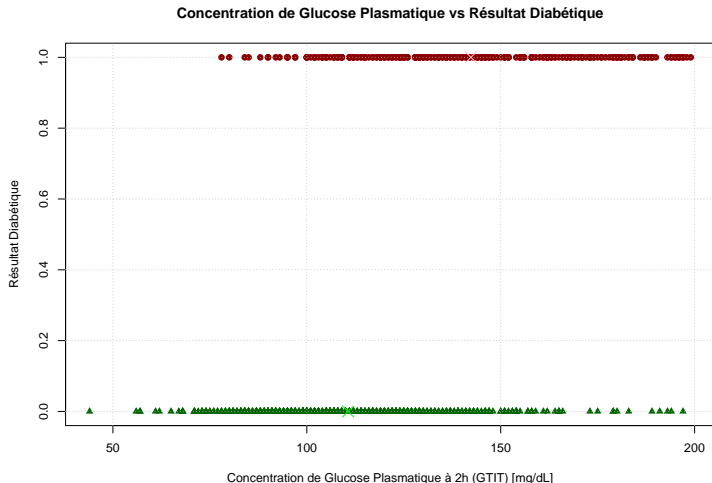
Intervalle de  
confiance

Travaux pratiques

# Graphique de dispersion

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan



Remarquez la différence dans la moyenne et dans la plage en comparant le cas des diabétiques et des non-diabétiques ?

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

# Récap et matière à réflexion (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

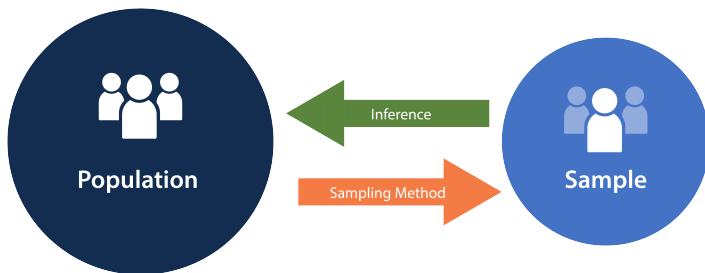
Travaux pratiques

Pouvons-nous utiliser les données fournies pour répondre à ces questions ?

Les données sont-elles représentatives de la population ?

Pouvons-nous utiliser les données fournies pour répondre à ces questions ?

Les données sont-elles représentatives de la population ?



**Figure 1:** Relation entre population et échantillon

Plan de la séance

Récap et matière  
à réflexion

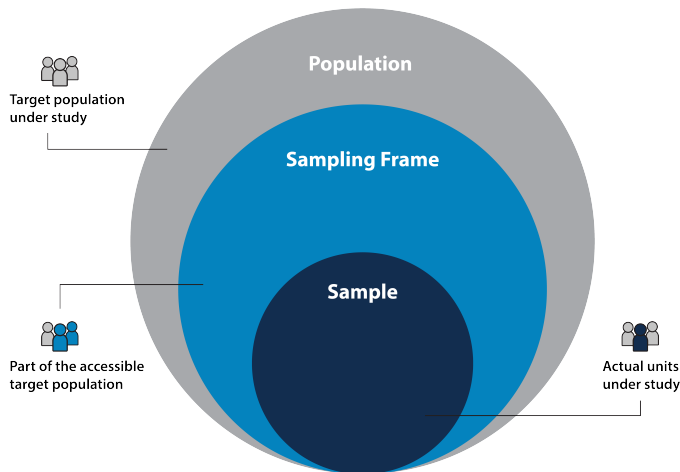
Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

# Méthodes d'échantillonnage



**Figure 2:** Échantillonnage

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques



*Recherche sur la complexité des problèmes de statistiques dispensés par les professeurs aux différentes spécialisations à l'ÉTS :*

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

*Recherche sur la complexité des problèmes de statistiques dispensés par les professeurs aux différentes spécialisations à l'ÉTS :*

**Population** : Un ensemble complet d'éléments (personnes, objets ou sujets) ayant des caractéristiques spécifiques que vous souhaitez étudier et sur lesquelles vous souhaitez faire des inférences.

- *Tous les enseignants de l'ÉTS qui dispensent des cours en statistiques aux différentes spécialisations.*

*Recherche sur la complexité des problèmes de statistiques dispensés par les professeurs aux différentes spécialisations à l'ÉTS :*

**Population** : Un ensemble complet d'éléments (personnes, objets ou sujets) ayant des caractéristiques spécifiques que vous souhaitez étudier et sur lesquelles vous souhaitez faire des inférences.

- *Tous les enseignants de l'ÉTS qui dispensent des cours en statistiques aux différentes spécialisations.*

**Cadre d'échantillonnage** : le matériel source ou la liste complète à partir de laquelle un échantillon est tiré. C'est une compilation exhaustive de tous les éléments de votre population.

- *Le registre de l'ÉTS qui liste tous les enseignants des cours en statistiques.*

**Échantillon** : Un sous-ensemble d'une population. Il s'agit de l'ensemble spécifique d'éléments à partir desquels vous collecterez des données.

- *Sous-ensemble des enseignants de l'ÉTS que vous sélectionnez pour votre étude.*

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillon** : Un sous-ensemble d'une population. Il s'agit de l'ensemble spécifique d'éléments à partir desquels vous collecterez des données.

- *Sous-ensemble des enseignants de l'ÉTS que vous sélectionnez pour votre étude.*

**Taille de l'échantillon** : le nombre de membres de la population enquêtés, mesurés ou observés.

La taille de l'échantillon détermine la quantité de données, ce qui influence davantage la précision de votre étude et la fiabilité de vos résultats.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillonnage aléatoire simple** : Chaque membre de la population a une chance égale d'être sélectionné.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillonnage aléatoire simple** : Chaque membre de la population a une chance égale d'être sélectionné.

**Échantillonnage systématique** : On sélectionne des membres de la population à intervalles réguliers, par exemple, choisir chaque 10ème personne de la liste.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillonnage aléatoire simple** : Chaque membre de la population a une chance égale d'être sélectionné.

**Échantillonnage systématique** : On sélectionne des membres de la population à intervalles réguliers, par exemple, choisir chaque 10ème personne de la liste.

**Échantillonnage stratifié** : La population est divisée en sous-groupes (strates) qui partagent des caractéristiques similaires. Un échantillon aléatoire est ensuite prélevé dans chacune de ces strates. Cette méthode garantit une représentation de chaque sous-groupe.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques



**Échantillonnage aléatoire simple** : Chaque membre de la population a une chance égale d'être sélectionné.

**Échantillonnage systématique** : On sélectionne des membres de la population à intervalles réguliers, par exemple, choisir chaque 10ème personne de la liste.

**Échantillonnage stratifié** : La population est divisée en sous-groupes (strates) qui partagent des caractéristiques similaires. Un échantillon aléatoire est ensuite prélevé dans chacune de ces strates. Cette méthode garantit une représentation de chaque sous-groupe.

**Échantillonnage par grappes** : La population est divisée en grappes, généralement basées sur des zones géographiques, et un échantillon aléatoire de ces grappes est choisi. Tous les individus des grappes sélectionnées sont dans l'échantillon.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

# Échantillonnage probabiliste (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

Récap et matière  
à réflexion

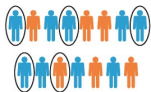
Méthodes  
d'échantillonnage

Taille de  
l'échantillon

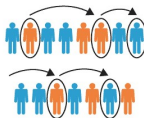
Intervalle de  
confiance

Travaux pratiques

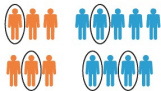
Simple random sample



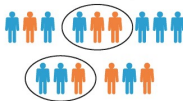
Systematic sample



Stratified sample



Cluster sample



**Figure 3:** Échantillonnage probabiliste

**Échantillonnage de convenance** : Les participants choisis sont les plus faciles à atteindre. Ce n'est pas un échantillon aléatoire et est souvent utilisé pour les tests pilotes.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillonnage de convenance** : Les participants choisis sont les plus faciles à atteindre. Ce n'est pas un échantillon aléatoire et est souvent utilisé pour les tests pilotes.

**Échantillonnage intentionnel** : Les participants sont sélectionnés en fonction du but de l'étude et du jugement du chercheur.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

**Échantillonnage de convenance** : Les participants choisis sont les plus faciles à atteindre. Ce n'est pas un échantillon aléatoire et est souvent utilisé pour les tests pilotes.

**Échantillonnage intentionnel** : Les participants sont sélectionnés en fonction du but de l'étude et du jugement du chercheur.

**Échantillonnage à réponse volontaire** : C'est les sujets qui choisissent de participer, souvent en réponse à une invitation générale.

**Échantillonnage de convenance** : Les participants choisis sont les plus faciles à atteindre. Ce n'est pas un échantillon aléatoire et est souvent utilisé pour les tests pilotes.

**Échantillonnage intentionnel** : Les participants sont sélectionnés en fonction du but de l'étude et du jugement du chercheur.

**Échantillonnage à réponse volontaire** : C'est les sujets qui choisissent de participer, souvent en réponse à une invitation générale.

**Échantillonnage boule de neige** : Les sujets actuels recrutent de futurs sujets parmi leurs connaissances. Cela est particulièrement utile pour atteindre des populations difficiles d'accès.

# Échantillonnage non-probabiliste (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

Récap et matière  
à réflexion

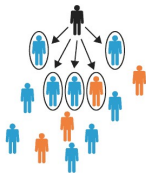
Méthodes  
d'échantillonnage

Taille de  
l'échantillon

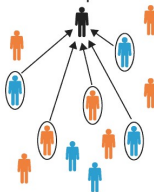
Intervalle de  
confiance

Travaux pratiques

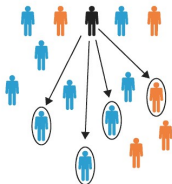
Convenience sample



Voluntary response sample



Purposive sample



Snowball sample

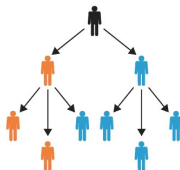


Figure 4: Échantillonnage non-probabiliste

En groupe de 3-4, discutez de quelle serait la situation dans laquelle chacune des méthodes d'échantillonnage serait utilisée ?



Pour une étude sur une maladie rare, un chercheur choisi délibérément les patients connus pour souffrir de cette maladie à partir des dossiers médicaux ou des établissements de santé spécialisés.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Pour une étude sur une maladie rare, un chercheur choisi délibérément les patients connus pour souffrir de cette maladie à partir des dossiers médicaux ou des établissements de santé spécialisés.

- Échantillonnage intentionnel

Pour une étude sur une maladie rare, un chercheur choisi délibérément les patients connus pour souffrir de cette maladie à partir des dossiers médicaux ou des établissements de santé spécialisés.

- Échantillonnage intentionnel

Pour évaluer l'efficacité d'une nouvelle campagne de santé, une organisation sélectionne au hasard cinq quartiers d'une ville. Ils enquêtent ensuite sur chaque foyer de ces quartiers.

Pour une étude sur une maladie rare, un chercheur choisi délibérément les patients connus pour souffrir de cette maladie à partir des dossiers médicaux ou des établissements de santé spécialisés.

- Échantillonnage intentionnel

Pour évaluer l'efficacité d'une nouvelle campagne de santé, une organisation sélectionne au hasard cinq quartiers d'une ville. Ils enquêtent ensuite sur chaque foyer de ces quartiers.

- Échantillonnage par grappes

Dans une étude sur la santé mentale des personnes sans abri, les premiers participants sans abri recommandent d'autres personnes sans abri qu'ils connaissent.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Dans une étude sur la santé mentale des personnes sans abri, les premiers participants sans abri recommandent d'autres personnes sans abri qu'ils connaissent.

- Échantillonnage boule de neige

Dans une étude sur la santé mentale des personnes sans abri, les premiers participants sans abri recommandent d'autres personnes sans abri qu'ils connaissent.

► Échantillonnage boule de neige

Une équipe de recherche sur la prévalence de l'hypertension divise la population en catégories ethniques, puis sélectionne au hasard un nombre proportionné d'individus dans chaque groupe pour garantir que tous soient représentés.

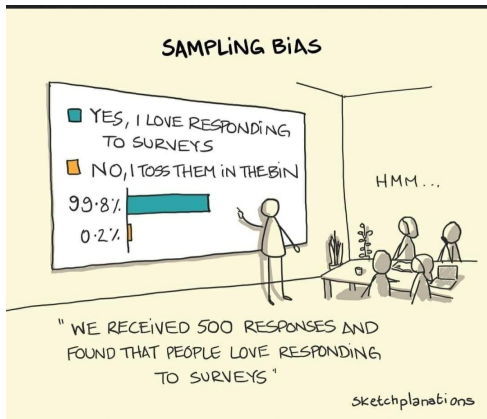
Dans une étude sur la santé mentale des personnes sans abri, les premiers participants sans abri recommandent d'autres personnes sans abri qu'ils connaissent.

- Échantillonnage boule de neige

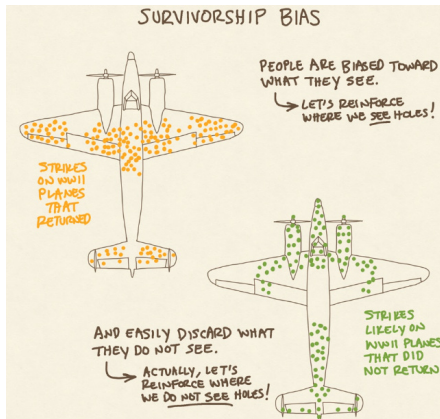
Une équipe de recherche sur la prévalence de l'hypertension divise la population en catégories ethniques, puis sélectionne au hasard un nombre proportionné d'individus dans chaque groupe pour garantir que tous soient représentés.

- Échantillonnage stratifié





**Figure 5:** Biais de réponse



**Figure 6:** Biais de survie

# Taille de l'échantillon

Le Théorème Central Limite (TCL) stipule que, pour une taille d'échantillon suffisamment grande, la distribution des moyennes d'échantillons se rapprochera d'une distribution normale, indépendamment de la distribution originale de la population.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Le Théorème Central Limite (TCL) stipule que, pour une taille d'échantillon suffisamment grande, la distribution des moyennes d'échantillons se rapprochera d'une distribution normale, indépendamment de la distribution originale de la population.

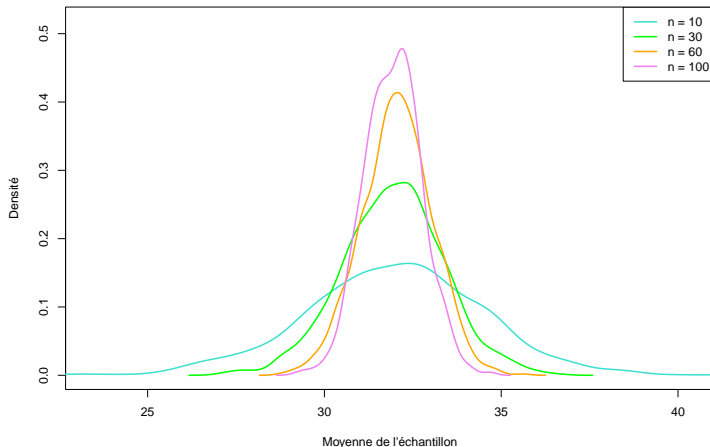
## Définition

Si  $X_1, X_2, X_3, \dots, X_n$  sont des échantillons aléatoires pris d'une population avec une moyenne générale  $\mu$  et une variance finie  $\sigma^2$ , la moyenne de l'échantillon  $\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \dots + X_n)$  sera approximativement distribuée normalement avec une moyenne  $\mu$  et une variance  $\frac{\sigma^2}{n}$ , à mesure que  $n$  devient grand.

La distribution normale est notée  $N(\mu, \frac{\sigma^2}{n})$ .

À partir de la base de données de “Pima Indian Diabetes Dataset”, les IMC (ou “BMI” en anglais) sont échantillonnées.

Distribution des moyennes d'échantillons pour différentes tailles d'échantillons



Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Le TCL justifie l'utilisation de la distribution normale dans l'inférence statistique et les tests d'hypothèses, même lorsque la population sous-jacente n'est pas normalement distribuée.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Le TCL justifie l'utilisation de la distribution normale dans l'inférence statistique et les tests d'hypothèses, même lorsque la population sous-jacente n'est pas normalement distribuée.

À mesure que  $n$  (la taille de l'échantillon) augmente, la forme de la distribution de la moyenne de l'échantillon  $\bar{X}$  devient de plus en plus en cloche ("bell-shaped" en anglais) ou normale.



L'**erreur standard**, qui mesure la variabilité des moyennes d'échantillons  $\bar{X}$ , est donnée par  $SE = \frac{\sigma}{\sqrt{n}}$ .

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

L'**erreur standard**, qui mesure la variabilité des moyennes d'échantillons  $\bar{X}$ , est donnée par  $SE = \frac{\sigma}{\sqrt{n}}$ .

À mesure que la taille de l'échantillon  $n$  augmente, l'erreur standard  $SE$  diminue. Cela indique que des échantillons plus grands fournissent des estimations plus précises de la moyenne de la population, réduisant ainsi le risque d'erreur d'échantillonnage.

# À noter que . . .

La taille d'échantillon "suffisamment grande" pour le TCL est généralement considérée comme étant 30 ou plus, mais cela peut varier en fonction de la population.

# Intervalle de confiance

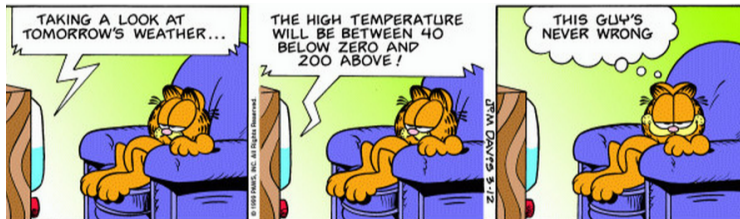


Figure 7: Intervalle de confiance ?

# Définition de l'Intervalle de Confiance

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

Récap et matière  
à réflexion

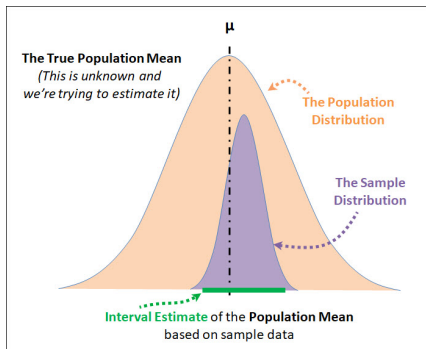
Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

Un intervalle de confiance est une plage de valeurs estimée à partir des données d'un échantillon, destinée à contenir un paramètre inconnu de la population (par exemple, moyenne de la population  $\mu$ ).



**Figure 8:** Estimation de l'IC

# Composants d'un Intervalle de Confiance

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

- ▶ **Estimation Ponctuelle** est généralement la moyenne de l'échantillon ( $\bar{x}$ ).
- ▶ **Niveau de Confiance**, exprimé en pourcentage (90 %, 95 %, 99 %, etc.), indique la **probabilité** que cet intervalle contienne le paramètre de la population si l'expérience est répétée plusieurs fois.
- ▶ **Marge d'Erreur (ME)** reflète l'incertitude autour de l'estimation ponctuelle et dépend de l'écart-type de la population  $\sigma$  et de la taille de l'échantillon  $n$ .

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

# Composants d'un Intervalle de Confiance

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

- **Estimation Ponctuelle** est généralement la moyenne de l'échantillon ( $\bar{x}$ ).
- **Niveau de Confiance**, exprimé en pourcentage (90 %, 95 %, 99 %, etc.), indique la **probabilité** que cet intervalle contienne le paramètre de la population si l'expérience est répétée plusieurs fois.
- **Marge d'Erreur (ME)** reflète l'incertitude autour de l'estimation ponctuelle et dépend de l'écart-type de la population  $\sigma$  et de la taille de l'échantillon  $n$ .

$$\mu = \bar{x} \pm Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

**Figure 9:** Composants d'un IC



## Lorsque $\sigma$ est Connue

- ▶ Formule :  $CI = \bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$
- ▶  $z$  : Score Z de la distribution normale, correspondant au niveau de confiance souhaité.

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

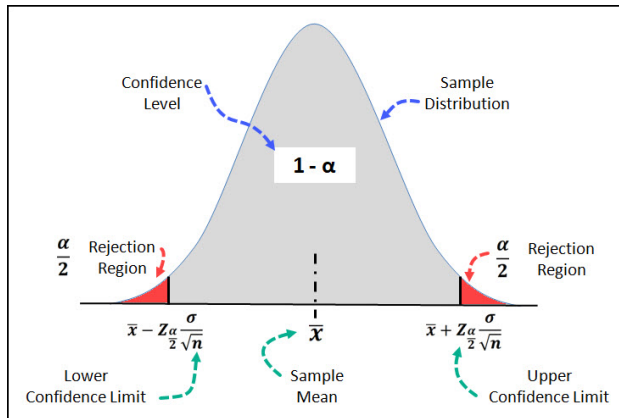
Travaux pratiques

## Lorsque $\sigma$ est Connue

- ▶ Formule :  $CI = \bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$
- ▶  $z$  : Score Z de la distribution normale, correspondant au niveau de confiance souhaité.

## Lorsque $\sigma$ est Inconnue

- ▶ Formule :  $CI = \bar{x} \pm t \times \frac{s}{\sqrt{n}}$
- ▶  $t$  : Score t de la distribution t, variant selon la taille de l'échantillon.



**Figure 10:** Niveau de Confiance ou  $1 - \alpha$

- L'erreur  $\alpha$  sera expliqué dans la prochaine séance sur les tests d'hypothèse.

Un IC de 95 % signifie que si de nombreux échantillons sont pris et qu'un IC est construit à partir de chacun, environ 95 % de ces intervalles contiendront la vraie moyenne de la population. Cela ne signifie pas qu'il y a 95 % de probabilité que l'intervalle donné contienne la moyenne populationnelle.

IC du paramètre “Glucose” dans la base de données “Pima Indian Diabetes”

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

IC du paramètre “Glucose” dans la base de données “Pima Indian Diabetes”

- Calculer la moyenne et l'écart type

```
data <- subset(data, Glucose > 0)
mean_glucose <- mean(data$Glucose, na.rm = TRUE)
sd_glucose <- sd(data$Glucose, na.rm = TRUE)
```

Plan de la séance

Récap et matière  
à réflexion

Méthodes  
d'échantillonnage

Taille de  
l'échantillon

Intervalle de  
confiance

Travaux pratiques

IC du paramètre "Glucose" dans la base de données "Pima Indian Diabetes"

- Calculer la moyenne et l'écart type

```
data <- subset(data, Glucose > 0)
mean_glucose <- mean(data$Glucose, na.rm = TRUE)
sd_glucose <- sd(data$Glucose, na.rm = TRUE)
```

- Déterminer la taille de l'échantillon et l'erreur standard

```
n <- sum(!is.na(data$Glucose))
se <- sd_glucose / sqrt(n)
```

IC du paramètre "Glucose" dans la base de données "Pima Indian Diabetes"

- Calculer la moyenne et l'écart type

```
data <- subset(data, Glucose > 0)
mean_glucose <- mean(data$Glucose, na.rm = TRUE)
sd_glucose <- sd(data$Glucose, na.rm = TRUE)
```

- Déterminer la taille de l'échantillon et l'erreur standard

```
n <- sum(!is.na(data$Glucose))
se <- sd_glucose / sqrt(n)
```

- Préciser le niveau de confiance de 95 % avec le niveau de signification de 0,05

```
alpha <- 0.05
```



# Exemples avec R (suite)

- Calculer la valeur critique à partir de la distribution t (t-score)

```
t_critical <- qt(1 - alpha/2, df = n - 1)
```

# Exemples avec R (suite)

- Calculer la valeur critique à partir de la distribution t (t-score)

```
t_critical <- qt(1 - alpha/2, df = n - 1)
```

- Calculer la marge d'erreur et le IC

```
margin_error <- t_critical * se  
ci_lower <- mean_glucose - margin_error  
ci_upper <- mean_glucose + margin_error
```

# Exemples avec R (suite)

- Calculer la valeur critique à partir de la distribution t (t-score)

```
t_critical <- qt(1 - alpha/2, df = n - 1)
```

- Calculer la marge d'erreur et le IC

```
margin_error <- t_critical * se  
ci_lower <- mean_glucose - margin_error  
ci_upper <- mean_glucose + margin_error
```

L'intervalle de confiance à 95 % pour le taux de glucose plasmatique moyen se situe entre ... mg/dL

## 119.5166 et 123.8569

# Exemples avec R (suite)

Et les ICs à 95 % du taux de glucose plasmatique des personnes non diabétiques (Outcome = 0) vs diabétiques (Outcome = 1) ?

# Exemples avec R (suite)

Et les ICs à 95 % du taux de glucose plasmatique des personnes non diabétiques (Outcome = 0) vs diabétiques (Outcome = 1) ?

► Couper des données

```
data_outcome_0 <- subset(data, Outcome == 0)
data_outcome_1 <- subset(data, Outcome == 1)
```

# Exemples avec R (suite)

Et les ICs à 95 % du taux de glucose plasmatique des personnes non diabétiques (Outcome = 0) vs diabétiques (Outcome = 1) ?

- Couper des données

```
data_outcome_0 <- subset(data, Outcome == 0)
data_outcome_1 <- subset(data, Outcome == 1)
```

- Calculer les ICs

```
## Outcome 0: [ 108.4602 , 112.8275 ]
```

```
## Outcome 1: [ 138.7462 , 145.8929 ]
```

## Exemples avec R (suite)

Et les ICs à 95 % du taux de glucose plasmatique des personnes non diabétiques (Outcome = 0) vs diabétiques (Outcome = 1) ?

- Couper des données

```
data_outcome_0 <- subset(data, Outcome == 0)
data_outcome_1 <- subset(data, Outcome == 1)
```

- Calculer les ICs

```
## Outcome 0: [ 108.4602 , 112.8275 ]
```

```
## Outcome 1: [ 138.7462 , 145.8929 ]
```

Observez que ces deux intervalles de confiance ne se chevauchent pas.

**Plan de la séance**

**Récap et matière  
à réflexion**

**Méthodes  
d'échantillonnage**

**Taille de  
l'échantillon**

**Intervalle de  
confiance**

**Travaux pratiques**

# Travaux pratiques



Pour chacun des neuf paramètres dans la base de données sur les diabètes . . .

1. Calculer l'intervalle de confiance
2. En séparant selon le paramètre "Outcome" (0 ou 1), recalculer l'IC
3. Déterminer si les IC des deux groupes sont différents
4. (Bonus) Créer un graphique pour présenter la distribution de probabilité avec l'estimation ponctuelle et la marge d'erreur