

# SYS865 Inférence statistique avec programmation R

Ornwipa Thamsuwan

10 janvier 2024

# Brise-Glace

## Réponses anonymes

Go to wooclap.com

Enter the event code DOGLGJ



**Figure 1:** Lien à l'activité sur Wooclap

## Introduction individuelle

1. Quel est votre nom ? Comment devrions-nous vous appeler ?
2. Quel est la concentration de votre étude à l'ÉTS ?  
(c-à-d. votre programme au cycle supérieure, avec ou sans mémoire/thèse)
3. Quel est votre formation antérieure ? au niveau du baccalauréat ? et votre expérience professionnelle ?
4. Pourquoi est-ce que vous poursuivez ce cours ?
5. Quels sont vos objectifs principaux après avoir terminé votre étude à l'ÉTS ?

# Introduction au cours

- ▶ Travaux pratiques basés sur des projets en équipe
  - ▶ Variables aléatoires et distribution de probabilités
  - ▶ Échantillonnage et estimation des paramètres
  - ▶ Intervalles de confiance
  - ▶ Tests d'hypothèses des moyennes et des variances
  - ▶ Correlations et associations
  - ▶ Régression linéaire
  - ▶ Régression logistique
- ▶ Programmation R

- ▶ **Projet 1 : Tests d'hypothèses (14 février 2024) 40%**
  - ▶ Analyse exploratoire des données
  - ▶ Tests des suppositions pour les variables paramétriques
  - ▶ Tests d'hypothèses des moyennes et des variances pour un ou plusieurs échantillons
- ▶ **Projet 2 : Association des variables (21 mars 2024) 40%**
  - ▶ Régression linéaire ou logistique
  - ▶ Techniques de l'ajustement pour les facteurs confondants ou de la stratification si nécessaire
- ▶ **Projet 3 : Tous ensembles (10 avril 2024) 20%**

- ▶ **Projet 1 et 2**
  - ▶ 25 pts. pour les travaux (rapports écrits)
  - ▶ 10 pts. pour la présentation + les Q-et-R
  - ▶ 5 pts. pour l'évaluation par les pairs
- ▶ **Projet 3**
  - ▶ 15 pts. pour la présentation du progrès pendant deux périodes de l'accompagnement du projet
  - ▶ 5 pts. pour l'évaluation par les pairs

Grille d'autoévaluation et d'évaluation par les pairs (en p.j.)



- 10 janvier : introduction, variables aléatoires
- 17 janvier : espérance, variance, indépendance
- 24 janvier : échantillonnage, intervalle de confiance
- 31 janvier : test d'hypothèse
- 07 février : accompagnement du projet
- 14 février : **présentation du projet 1**
- 21 février : corrélation, régression linéaire
- 28 février : facteurs confondants, régression logistique
- 07 mars : *semaine de relache, il n'y a pas du cours*
- 14 mars : accompagnement du projet
- 21 mars : **présentation du projet 2**
- 28 mars : accompagnement du projet
- 03 avril : accompagnement du projet
- 10 avril : **présentation du projet 3**

**Brise-Glace**

**Introduction au  
cours**

**Information  
supplémentaire**

**Variables  
aléatoires**

**Recap**

# **Information supplémentaire**

## Pima Indians Diabetes Database

Lien vers la base de données sur Kaggle

## Référence

Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Symposium on Computer Applications and Medical Care* (pp. 261–265). IEEE Computer Society Press.

Merci de les télécharger avant le prochain cours.

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap



Lien vers l'instruction dans la vidéo sur YouTube

Merci de compléter toutes les installations avant le prochain cours.

# Variables aléatoires

Variables aléatoires discrètes et continues

→ Lien vers des exemples sur Wooclap

Fonction de masse et de densité de probabilité

Fonction de répartition

Lancer une pièce de monnaie,  
quelle est la probabilité d'obtenir une pile ? et une face ?

Lancer un dé équilibré,  
quelle est la probabilité d'obtenir "1" ? "2"... "6" ?

Lancer une pièce de monnaie,  
quelle est la probabilité d'obtenir une pile ? et une face ?

Lancer un dé équilibré,  
quelle est la probabilité d'obtenir "1" ? "2"... "6" ?

Quelle est la probabilité qu'un élève de la classe ait 23 ans ?



## Procès Bernoulli

Dans un essai Bernoulli, il n'y a que deux résultats possibles. Pour une pièce équilibrée, la probabilité d'obtenir face (succès) est de 0,5, et la probabilité d'obtenir pile (échec) est également de 0,5.

La fonction de masse de probabilité (FMP) pour une variable aléatoire Bernoulli est définie comme suit :

- ▶  $P(X = 1) = p$  (probabilité de succès)
- ▶  $P(X = 0) = 1 - p$  (probabilité d'échec)

où  $X$  est la variable aléatoire représentant le résultat (1 pour face, 0 pour pile) et  $p$  est la probabilité d'obtenir une face.

# Fonction de masse de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

## R code

```
pmf_coin_flip <- function(p) {  
  if (p < 0 || p > 1) { stop() }  
  pmf <- c(Pile = p, Face = 1 - p)  
  return(pmf)  
}
```

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

# Fonction de masse de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

## R code

```
pmf_coin_flip <- function(p) {  
  if (p < 0 || p > 1) { stop() }  
  pmf <- c(Pile = p, Face = 1 - p)  
  return(pmf)  
}
```

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

Un vecteur des probabilités d'obtenir une pile et une face

```
pmf_coin_flip(0.5)
```

```
## Pile Face  
## 0.5 0.5
```

## R code

```
simulate_coin_flips <- function(num_flips, p) {  
  pmf <- pmf_coin_flip(p)  
  outcomes <- names(pmf)  
  probabilities <- as.numeric(pmf)  
  flips <- sample(outcomes, size = num_flips,  
                  replace = TRUE,  
                  prob = probabilities)  
  return(flips)  
}
```

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

# Fonction de masse de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

## R code

```
simulate_coin_flips <- function(num_flips, p) {  
  pmf <- pmf_coin_flip(p)  
  outcomes <- names(pmf)  
  probabilities <- as.numeric(pmf)  
  flips <- sample(outcomes, size = num_flips,  
                  replace = TRUE,  
                  prob = probabilities)  
  return(flips)  
}
```

Simulation de lancer une pièce de monnaie 7 fois

```
simulate_coin_flips(7, 0.5)
```

```
## [1] "Pile" "Face" "Face" "Pile" "Pile" "Face" "Pile"
```

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

# Fonction de masse de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Pour un dé équilibré à six faces, la fonction de masse de probabilité (FMP) attribue une probabilité de  $\frac{1}{6}$  à chaque face (1 à 6). Chaque résultat a donc une chance égale d'apparaître lors d'un lancer, représentée par la formule  $P(X = x) = \frac{1}{6}$ , où  $X$  est le résultat du dé et  $x$  est une valeur spécifique entre 1 et 6.



**Figure 2:** Un dé

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

# Fonction de masse de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

## R code

```
simulate_dice_rolls <- function(num_rolls) {  
  
  # Résultats possibles : 1, 2, 3, 4, 5, 6  
  outcomes <- 1:6  
  
  # Chaque résultat a une probabilité égale  
  probabilities <- rep(1/6, 6)  
  
  # Simulation de lancer un dé  
  rolls <- sample(outcomes, size = num_rolls,  
                  replace = TRUE,  
                  prob = probabilities)  
  return(rolls)  
}
```

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

# Fonction de masse de probabilité (suite)

Simulation de lancer un dé 20 fois

```
simulate_dice_rolls(20)
```

```
## [1] 6 4 1 6 2 6 3 4 2 4 1 3 5 5 6 6 1 3 3 2
```



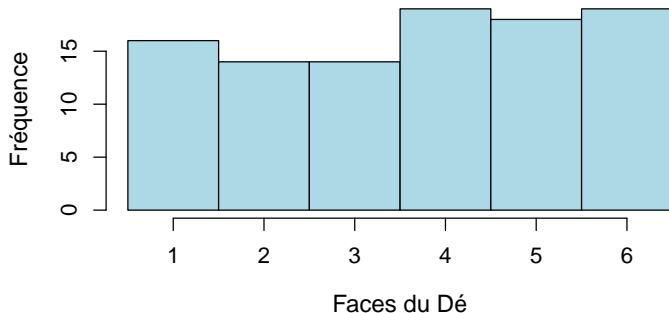
# Fonction de masse de probabilité (suite)

Simulation de lancer un dé 20 fois

```
simulate_dice_rolls(20)
```

```
## [1] 6 4 1 6 2 6 3 4 2 4 1 3 5 5 6 6 1 3 3 2
```

**Histogramme des Lancers de Dé 100 fois**



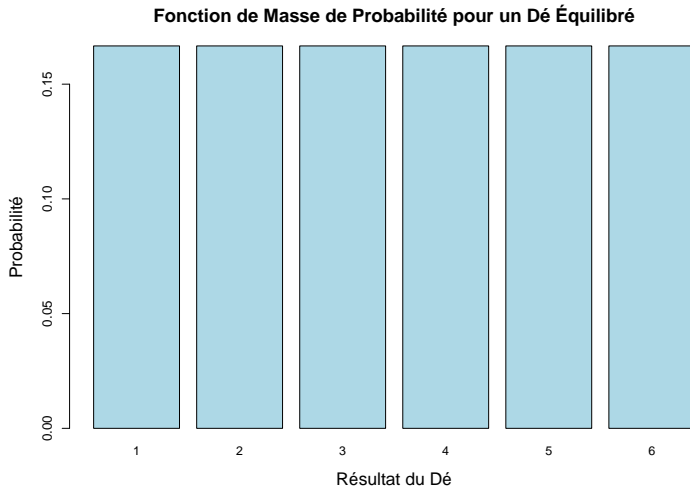
Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap



Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

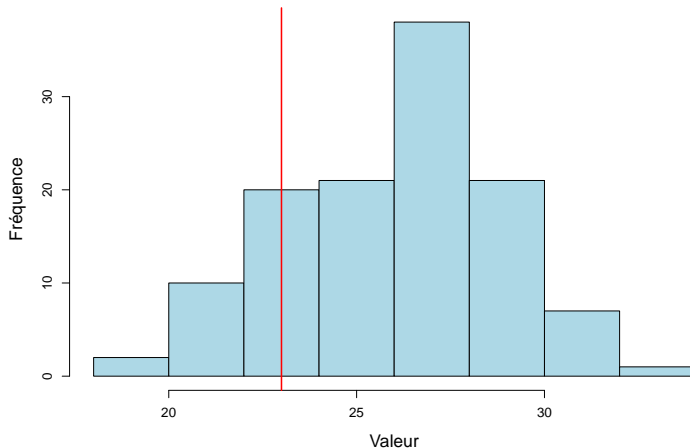
Variables  
aléatoires

Recap

Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

```
## 26 27 25 23 28 22 27 26 27 27 25 22 28 23 23
## 29 27 24 23 26 28 29 27 29 24 25 28 31 29 24
## 24 22 25 27 28 26 21 30 26 25 24 22 30 30 26
## 27 22 25 32 28 22 27 20 23 21 23 28 28 25 27
## 21 27 30 25 26 25 27 28 26 24 24 20 24 27 30
## 30 30 25 26 26 26 23 27 29 28 28 29 29 27 20
## 28 26 23 26 23 27 25 29 29 26 25 25 22 23 28
## 22 28 32 27 28 31 26 24 23 22 29 24 27 28 29
```

Histogramme des Échantillons Normalement Distribués



# Fonction de densité de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Brise-Glace

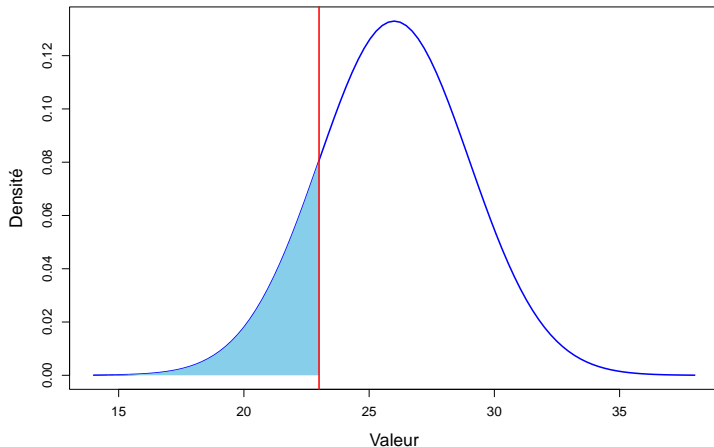
Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

Fonction de densité de probabilité



# Fonction de densité de probabilité (suite)

SYS865 Inférence  
statistique avec  
programmation R

Ornwipa  
Thamsuwan

Brise-Glace

Introduction au  
cours

Information  
supplémentaire

Variables  
aléatoires

Recap

La fonction de densité de probabilité (FDP) représente la distribution des probabilités pour une variable aléatoire continue.

Dans l'exemple, la variable suit une distribution normale (ou gaussienne) avec une moyenne ( $\mu$ ) de 26 et un écart-type ( $\sigma$ ) de 3.

Chaque point sur la courbe de la FDP indique la densité de probabilité pour une valeur donnée. La moyenne,  $\mu = 26$ , est le pic de cette courbe, et l'écart-type,  $\sigma = 3$ , indique la dispersion des valeurs autour de la moyenne.

Une ligne verticale à  $x = 23$  sur cette courbe montre la position de 23 par rapport à la distribution normale. Dans une telle distribution, les valeurs se trouvent généralement dans l'intervalle  $\mu \pm 3\sigma$ .

# Recap



- ▶ Fonction de masse de probabilité : Variables aléatoires discrètes
  - ▶ Lancer une pièce de monnaie, quelle est la probabilité d'obtenir une pile ? et une face ?
  - ▶ Lancer un dé équilibré, quelle est la probabilité d'obtenir "1"... "6" ?
- ▶ Fonction de densité de probabilité : Variables aléatoires continues
  - ▶ Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

- ▶ Fonction de masse de probabilité : Variables aléatoires discrètes
  - ▶ Lancer une pièce de monnaie, quelle est la probabilité d'obtenir une pile ? et une face ?
  - ▶ Lancer un dé équilibré, quelle est la probabilité d'obtenir "1"... "6" ?
- ▶ Fonction de densité de probabilité : Variables aléatoires continues
  - ▶ Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

Le prochain cours. . .

- ▶ Fonction de répartition
  - ▶ Lancer un dé équilibré, quelle est la probabilité d'obtenir un numéro moins de 2 ?
  - ▶ Quelle est la probabilité qu'un élève de la classe ait plus de 23 ans ?