



Faculté
de Médecine

Aix-Marseille Université



Sciences Economiques et Sociales de la
Santé & Traitement de l'Information Médicale

Inserm / IRD / Aix-Marseille Université

Tests Non Paramétriques

Plan

1. Paramétriques ou non?
2. Test d'une distribution de probabilité
3. Comparaison de moyennes
4. Comparaison de pourcentages
5. Corrélation

I. Paramétriques ou Non?

1. Paramétrique?

2. Distribution
3. Moyennes
4. Pourcentages
5. Corrélation

■ Tests paramétriques:

Construits en supposant que les V.A. ont des distributions de probabilité ayant des propriétés précises

Ex: Test de Student: distribution Normale des moyennes

Soit parce que la VA est Normale dans chaque groupe,

Soit parce que n_1 et $n_2 > 30 \Rightarrow$ approximation Normale
(théorème central limite)

Si les conditions d'application des tests paramétriques ne sont pas respectés ?

- Tests paramétriques **robustes**
aux faibles écarts aux conditions d'application
- Tests **non paramétriques**

■ Tests non paramétriques

- Ne nécessitent aucune hypothèse de distribution

- Inconvénients

- Puissance ↓

- Calculs ↑

II. Test d'une distribution de probabilité

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Test de Kolmogorov-Smirnov

Principe:

Soit un échantillon de n valeurs x_1, x_2, \dots, x_n d'une V.A. X

H_0 : $X \sim$ distribution de probabilité donnée

H_1 : $X \sim$ autre distribution

Exemple: Effet du Tabac~TAS,

Echantillon de 32 sujets

■ Importer le fichier de données *Tabac.xls*

```
data<-read.csv2("D:\\EISIS_DATA\\Tabac.csv", header=TRUE)
```

```
attach(data)
```

17 fumeurs $m_f = ?$ mmHg
 $s_f^2 = ?$ mmHg²
 histogramme

```
mean(TAS[Tabac==1])
```

```
var(TAS[Tabac==1])
```

```
hist(TAS[Tabac==1], col='red')
```

15 non fumeurs

$m_{nf} = ?$ mmHg

$s_f^2 = ?$ mmHg²

histogramme

```
mean(TAS[Tabac==0])
```

```
var(TAS[Tabac==0])
```

```
hist(TAS[Tabac==0], col='blue')
```

Exemple: Effet du Tabac~TAS,

Echantillon de 32 sujets

1. Paramétrique?

2. Distribution

3. Moyennes

4. Pourcentages

5. Corrélation

■ Importer le fichier de données *Tabac.xls*

```
data<-read.csv2("D:\\EISIS_DATA\\Tabac.csv", header=TRUE)
```

```
attach(data)
```

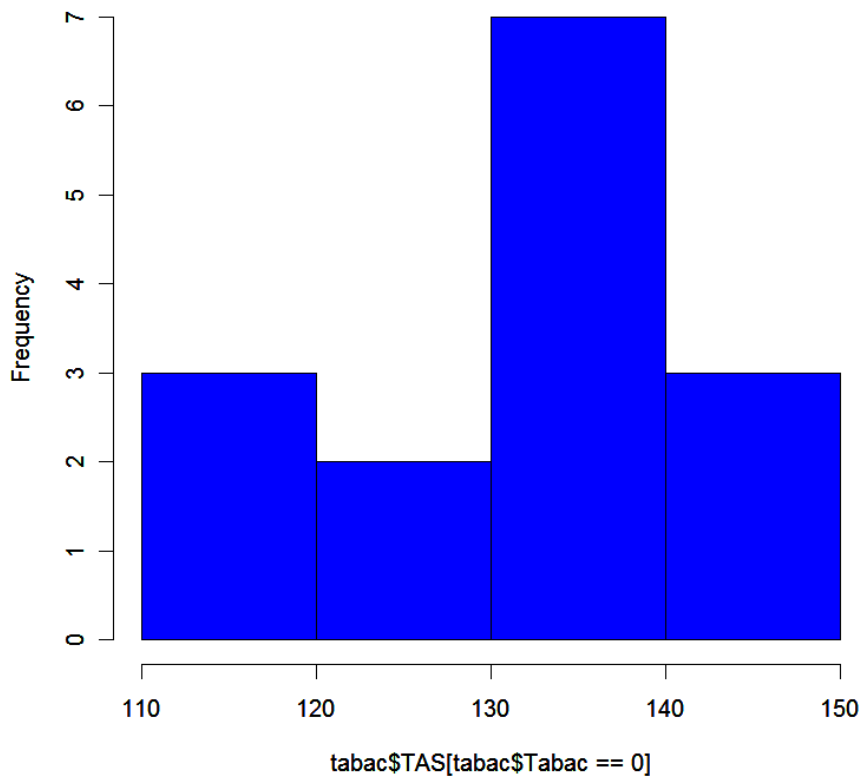
17 fumeurs $m_f = 149,6$ mmHg *mean(TAS[Tabac==1])*
 $s_f^2 = 212,1$ mmHg² *var(TAS[Tabac==1])*
histogramme *hist(TAS[Tabac==1], col='red')*

15 non fumeurs

$m_{nf} = 130,9$ mmHg *mean(TAS[Tabac==0])*
 $s_f^2 = 118,1$ mmHg² *var(TAS[Tabac==0])*
histogramme *hist(TAS[Tabac==0], col='blue')*

15 non fumeurs

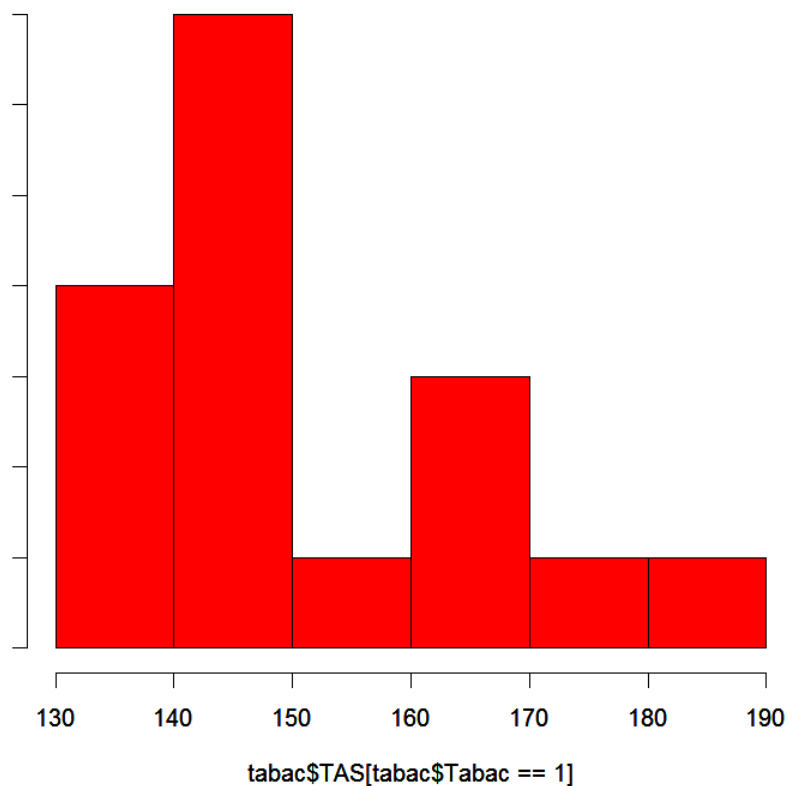
Histogram of `tabac$TAS[tabac$Tabac == 0]`



$$m_{nf} = 130,9 \text{ mmHg}$$
$$s_f^2 = 118,1$$

17 fumeurs

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation



$$m_f = 149,6 \text{ mmHg}$$
$$s_f^2 = 212,1$$

- On ordonne les valeurs observées:

$$(x_1 < x_2 < x_3 < \dots < x_n)$$

Fumeurs

$$130 < 133 < 139 < 141 < \dots < 163 < 167 < 171 < 181$$

- On calcul la fonction de répartition $F_o(x_i)$ observée:

$$F_o(x_1) = 1/n; F_o(x_2) = 2/n; F_o(x_3) = 3/n; \dots; F_o(x_n) = n/n = 1$$

$$p(\text{TAS} \leq 130) = 1/17 = 0,06; \quad p(\text{TAS} \leq 133) = 2/17 = 0,12; \dots$$

$$p(\text{TAS} \leq 181) = 17/17 = 1$$

- On compare $F_o(x_i)$ à la fonction de répartition théorique $F(x_i)$ connue.

$$X \sim N(\mu_f = 149,6; \sigma_f^2 = 212,1)$$

$$p(X \leq 130) = 0,089; \quad p(X \leq 133) = 0,13; \quad \dots; \quad p(X \leq 181) = 0,98$$

- On ordonne les valeurs observées:

$$(x_1 < x_2 < x_3 < \dots < x_n)$$

Fumeurs

$$130 < 133 < 139 < 141 < \dots < 163 < 167 < 171 < 181$$

- On calcul la fonction de répartition $F_o(x_i)$ observée:

$$F_o(x_1) = 1/n; F_o(x_2) = 2/n; F_o(x_3) = 3/n; \dots; F_o(x_n) = n/n = 1$$

$$p(\text{TAS} \leq 130) = 1/17 = 0,06; \quad p(\text{TAS} \leq 133) = 2/17 = 0,12; \dots$$

$$p(\text{TAS} \leq 181) = 17/17 = 1$$

- On compare $F_o(x_i)$ à la fonction de répartition théorique $F(x_i)$ connue.

$$X \sim N(\mu_f = 149,6; \sigma_f^2 = 212,1)$$

$$p(X \leq 130) = 0,089; \quad p(X \leq 133) = 0,13; \quad \dots; \quad p(X \leq 181) = 0,98$$

- On ordonne les valeurs observées:

$$(x_1 < x_2 < x_3 < \dots < x_n)$$

Fumeurs

$$130 < 133 < 139 < 141 < \dots < 163 < 167 < 171 < 181$$

- On calcul la fonction de répartition $F_o(x_i)$ observée:

$$F_o(x_1) = 1/n; F_o(x_2) = 2/n; F_o(x_3) = 3/n; \dots; F_o(x_n) = n/n = 1$$

$$p(\text{TAS} \leq 130) = 1/17 = 0,06; \quad p(\text{TAS} \leq 133) = 2/17 = 0,12; \dots$$

$$p(\text{TAS} \leq 181) = 17/17 = 1$$

- On compare $F_o(x_i)$ à la fonction de répartition théorique $F(x_i)$ connue.

$$X \sim N(\mu_f = 149,6; \sigma_f^2 = 212,1)$$

$$p(X \leq 130) = 0,089; \quad p(X \leq 133) = 0,13; \quad \dots; \quad p(X \leq 181) = 0,98$$

- On ordonne les valeurs observées:

$$(x_1 < x_2 < x_3 < \dots < x_n)$$

Fumeurs

$$130 < 133 < 139 < 141 < \dots < 163 < 167 < 171 < 181$$

- On calcul la fonction de répartition $F_o(x_i)$ observée:

$$F_o(x_1) = 1/n; F_o(x_2) = 2/n; F_o(x_3) = 3/n; \dots; F_o(x_n) = n/n = 1$$

$$p(\text{TAS} \leq 130) = 1/17 = 0,06; \quad p(\text{TAS} \leq 133) = 2/17 = 0,12; \dots$$

$$p(\text{TAS} \leq 181) = 17/17 = 1$$

- On compare $F_o(x_i)$ à la fonction de répartition théorique $F(x_i)$ connue.

$$X \sim N(\mu_f = 149,6; \sigma_f^2 = 212,1)$$

$$p(X \leq 130) = 0,089; \quad p(X \leq 133) = 0,13; \quad \dots; \quad p(X \leq 181) = 0,98$$

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

■ **Conditions d'application**

- ❑ Indépendance des individus
- ❑ Pas (ou peu) d'ex aequo

Exemple

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

Test de Kolmogorov-Smirnov

`ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))`

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

Test de Kolmogorov-Smirnov



```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

variable: TAS chez les fumeurs

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

Test de Kolmogorov-Smirnov



Distribution sous H0



```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

variable: TAS chez les
fumeurs



1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

Test de Kolmogorov-Smirnov



Distribution sous H0



```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

variable: TAS chez les fumeurs



moyenne de la distribution sous H0, estimée par le moyenne de la TAS chez les fumeurs



1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

Test de Kolmogorov-Smirnov



Distribution sous H0



```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

variable: TAS chez les fumeurs



moyenne de la distribution sous H0, estimée par le moyenne de la TAS chez les fumeurs



écart-type de la distribution sous H0, estimé par l'écart type de la TAS chez les fumeurs



1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélations

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

One-sample Kolmogorov-Smirnov test

```
data: TAS[Tabac == 1]  
D = 0.1673, p-value = 0.7283  
alternative hypothesis: two.sided
```

Warning message:

```
impossible de calculer les p-values correctes avec des ex-aequos in:  
ks.test(TAS[Tabac == 1], "pnorm", mean(TAS[Tabac == 1]), sd(TAS[Tabac  
== 1]))
```

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

One-sample Kolmogorov-Smirnov test

```
data: TAS[Tabac == 1]
D = 0.1673, p-value = 0.7283
alternative hypothesis: two.sided
```

Test

Warning message:
impossible de calculer les p-values correctes avec des ex-aequos in:
ks.test(TAS[Tabac == 1], "pnorm", mean(TAS[Tabac == 1]), sd(TAS[Tabac == 1]))

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

One-sample Kolmogorov-Smirnov test

data: TAS[Tabac == 1]

Données

D = 0.1673, p-value = 0.7283

alternative hypothesis: two.sided

Warning message:

impossible de calculer les p-values correctes avec des ex-aequos in:

```
ks.test(TAS[Tabac == 1], "pnorm", mean(TAS[Tabac == 1]), sd(TAS[Tabac == 1]))
```


1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélations

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

One-sample Kolmogorov-Smirnov test

```
data: TAS[Tabac == 1]  
D = 0.1673, p-value = 0.7283  
alternative hypothesis: two.sided
```

« petit p »

Warning message:

```
impossible de calculer les p-values correctes avec des ex-aequos in:  
ks.test(TAS[Tabac == 1], "pnorm", mean(TAS[Tabac == 1]), sd(TAS[Tabac  
== 1]))
```

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

Exemple

```
ks.test(TAS[Tabac==1], "pnorm", mean(TAS[Tabac==1]), sd(TAS[Tabac==1]))
```

Interprétation

- $p > 0,05$
- Test Non Significatif
- Non rejet de H_0 , au risque β
- On ne met pas en évidence de différence significative entre la distribution de la TAS chez les **fumeurs** et la distribution normale de moyenne **149,6** et de variance **212,1**

1. Paramétrique?

2. Distribution

3. Moyennes

4. Pourcentages

5. Corrélation

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélations

1. Hypothèses

H0: Loi \sim $N(\mu_{nf}=130,9; \sigma_{nf}^2=118,1)$

H1: Loi \sim autre loi

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

1. Hypothèses

H0: Loi \sim $N(\mu_{nf}=130,9; \sigma_{nf}^2=118,1)$

H1: Loi \sim autre loi

2. Prédiction sous H0 $110 < 112 < 119 < \dots < 141 < 143 < 147$

$p(X \leq 110) = 0.027; p(X \leq 112) = 0,041; \dots; p(X \leq 147) = 0,93$

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

1. Hypothèses

H0: Loi \sim $N(\mu_{nf}=130,9; \sigma_{nf}^2=118,1)$

H1: Loi \sim autre loi

2. Prédiction sous H0 $110 < 112 < 119 < \dots < 141 < 143 < 147$

$p(X \leq 110) = 0.027; p(X \leq 112) = 0,041; \dots; p(X \leq 147) = 0,93$

3. Confrontation

ks.test(TAS[Tabac==0], "pnorm", mean(TAS[Tabac==0]), sd(TAS[Tabac==0]))

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

1. Hypothèses
2. Prédiction sous H0
3. Confrontation

1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

```
ks.test(TAS[Tabac==0], "pnorm", mean(TAS[Tabac==0]), sd(TAS[Tabac==0]))
```

One-sample Kolmogorov-Smirnov test

```
data: TAS[Tabac == 0]  
D = 0.2134, p-value = 0.5015  
alternative hypothesis: two.sided
```

Warning message:

```
impossible de calculer les p-values correctes avec des ex-aequos in:  
ks.test(TAS[Tabac == 0], "pnorm", mean(TAS[Tabac == 0]), sd(TAS[Tabac  
== 0]))
```

Exercice : La distribution de la TAS chez les non-fumeurs est-elle Normale?

1. Hypothèses

2. Prédiction sous H_0

3. Confrontation

4. Interprétation

- $p > 0,05$
- Test Non Significatif
- Non rejet de H_0 , au risque β
- On ne met pas en évidence de différence significative entre la distribution de la TAS chez les non fumeurs et la distribution normale de moyenne 130,9 et de variance 118,1

1. Paramétrique?

2. Distribution

3. Moyennes

4. Pourcentages

5. Corrélation

■ Principale utilisation:

Comparaison d'une distribution observée à une
distribution Normale,

préalable à certains tests paramétriques:

- test de Student
- Anova

$$H_0: X \sim N(\mu; \sigma^2)$$

$$H_1: X \sim \text{autre distribution}$$

Attention:



1. Paramétrique?
- 2. Distribution**
3. Moyennes
4. Pourcentages
5. Corrélation

1. on veut conserver $H_0 \Rightarrow$ test basé sur la **puissance**

2. μ et σ^2 sont estimées sur l'échantillon:

Si X suit une loi Normale de moyenne **différente**

\Rightarrow rejet de H_0

Exemple:

$H_0: \text{TAS}[\text{tabac}==0] \sim N(125;118)$

$p=0,018$

1. Paramétrique?
2. Distribution
3. Moyennes
4. Pourcentages
5. Corrélation

■ Références

- Jean Bouyer: *Méthodes statistiques, Médecine-Biologie*, éditions INSERM
- Coll. (CIMES): *Biostatistiques*, éditions Omnisciences

■ Contact

jean.gaudart@univ-amu.fr

<http://sesstim.univ-amu.fr>

Faculté de Médecine de Marseille