

Statistiques — niveau débutant

Hervé Seitz (email : herve.seitz@igh.cnrs.fr)

IGH (UMR 9002 CNRS et université de Montpellier)

8 mars 2024

Ce diaporama est accessible à :

<http://www.igh.cnrs.fr/equip/Seitz/Stats24.pdf>

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Pourquoi les statistiques ?

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

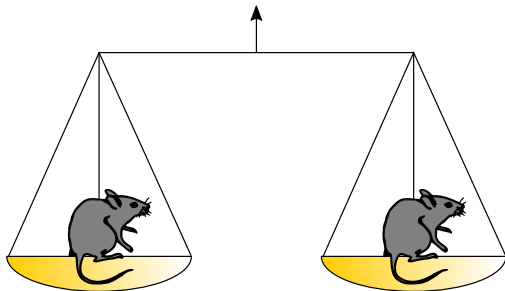
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Pourquoi les statistiques ?



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

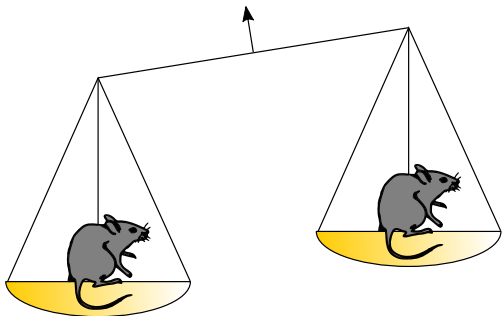
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Pourquoi les statistiques ?



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Variable continue :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Variable continue :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

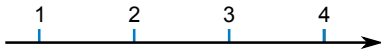
Variables continues et discontinues

Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Variable continue :



Variable discontinue (= « discrète ») :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

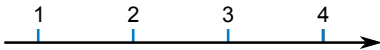
Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Exemple de variable discontinue (= « discrète ») : le nombre de souriceaux par portée (ne peut prendre que certaines valeurs bien précises).

Variable continue :



Variable discontinue (= « discrète ») :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

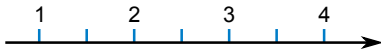
Exemple de variable continue : la masse corporelle d'une souris (peut prendre n'importe quelle valeur intermédiaire).

Exemple de variable discontinue (= « discrète ») : le nombre de souriceaux par portée (ne peut prendre que certaines valeurs bien précises).

Variable continue :



Variable discontinue (= « discrète ») :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Variable continue : deux mesures parfaitement précises ne donneront jamais exactement le même résultat, avec tous les chiffres après la virgule.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Variable continue : deux mesures parfaitement précises ne donneront jamais exactement le même résultat, avec tous les chiffres après la virgule.

Implication : si je fais une mesure « Condition 1 » et une mesure « Condition 2 », je sais d'avance qu'il y aura une différence !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

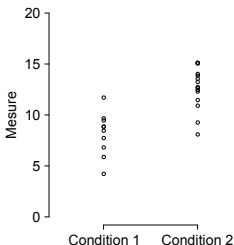
→ il faut mesurer plusieurs réplicats, et comparer la variabilité intra-groupe à la variabilité inter-groupe.

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

→ il faut mesurer plusieurs réplicats, et comparer la variabilité intra-groupe à la variabilité inter-groupe.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

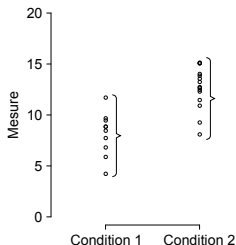
Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

→ il faut mesurer plusieurs réplicats, et comparer la variabilité intra-groupe à la variabilité inter-groupe.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

—> il faut mesurer plusieurs réplicats, et comparer la variabilité intra-groupe à la variabilité inter-groupe.

En biologie : « réplicats biologiques » (objets biologiques distincts), « réplicats techniques » (plusieurs mesures sur le même objet biologique).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

L'irreproductibilité expérimentale

Sources d'irreproductibilité expérimentale :

- ▶ Imperfection de l'appareil de mesure.
- ▶ Réelle différence entre les objets mesurés, à laquelle je choisis de ne pas accorder d'importance.

→ il faut mesurer plusieurs réplicats, et comparer la variabilité intra-groupe à la variabilité inter-groupe.

En biologie : « réplicats biologiques » (objets biologiques distincts), « réplicats techniques » (plusieurs mesures sur le même objet biologique ; ne capturent pas la variabilité biologique).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Variable continue : deux mesures parfaitement précises ne donneront jamais exactement le même résultat, avec une tous les chiffres après la virgule.

Implication : si je fais une mesure « Condition 1 » et une mesure « Condition 2 », je sais d'avance qu'il y aura une différence !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Variable continue : deux mesures parfaitement précises ne donneront jamais exactement le même résultat, avec une tous les chiffres après la virgule.

Implication : si je fais une mesure « Condition 1 » et une mesure « Condition 2 », je sais d'avance qu'il y aura une différence !

Variable discontinue : les résultats de mesure peuvent être rigoureusement égaux.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Variables continues et discontinues

Variable continue : deux mesures parfaitement précises ne donneront jamais exactement le même résultat, avec une tous les chiffres après la virgule.

Implication : si je fais une mesure « Condition 1 » et une mesure « Condition 2 », je sais d'avance qu'il y aura une différence !

Variable discontinue : les résultats de mesure peuvent être rigoureusement égaux.

... mais les sources d'irreproductibilité impliquent quand même de répliquer les mesures, et de comparer les variabilités intra- et inter-groupe.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations échantillonnées aient la même moyenne.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations échantillonnées aient la même moyenne.

Danger! Il est très facile de mal comprendre cette définition !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations **échantillonnées** aient la même moyenne.

Danger ! Il est très facile de mal comprendre cette définition !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Une autre façon de concevoir l'expérience :

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

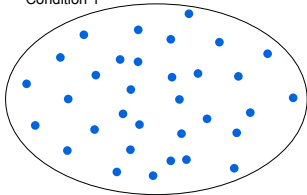
Conclusion

Suppléments

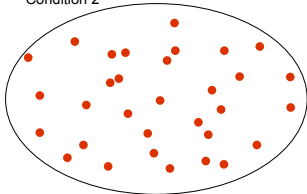
La p -value

Une autre façon de concevoir l'expérience :

Condition 1



Condition 2



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

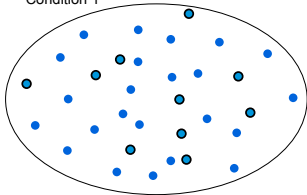
Conclusion

Suppléments

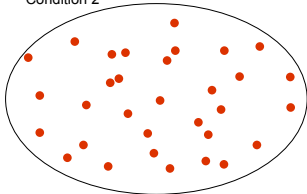
La p -value

Une autre façon de concevoir l'expérience :

Condition 1



Condition 2



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

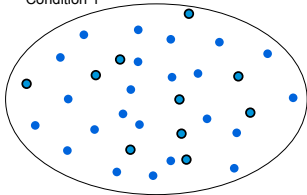
Conclusion

Suppléments

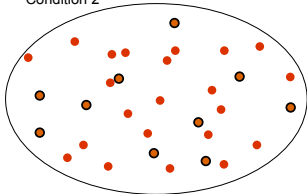
La p -value

Une autre façon de concevoir l'expérience :

Condition 1



Condition 2



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

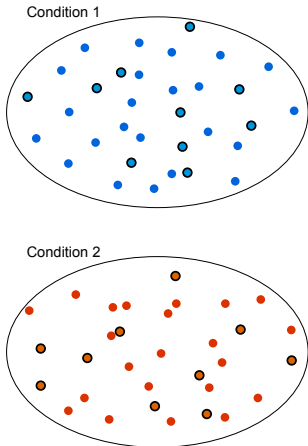
Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Une autre façon de concevoir l'expérience :



Un tirage aléatoire à partir d'une population d'effectif infini.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations **échantillonnées** aient la même moyenne.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations **échantillonnées** aient la même moyenne.

La « population échantillonnée » est cette population théorique, d'effectif infini, à partir de laquelle on tire aléatoirement des réplicats.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value

Le t-test compare deux jeux de données (exemple : une même mesure dans deux conditions expérimentales, avec réplicats).

Sa p -value est la probabilité que les deux populations **échantillonnées** aient la même moyenne.

La « population échantillonnée » est cette population théorique, d'effectif infini, à partir de laquelle on tire aléatoirement des réplicats.

→ le t-test permet de conclure sur l'égalité des moyennes de cohortes infinies (hors d'atteinte à l'expérience).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

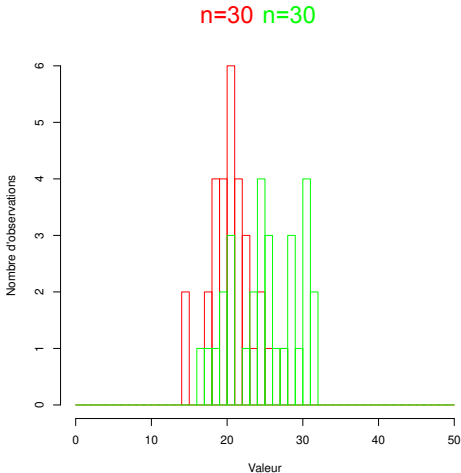
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value



Commandes R ayant généré ce graphe : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

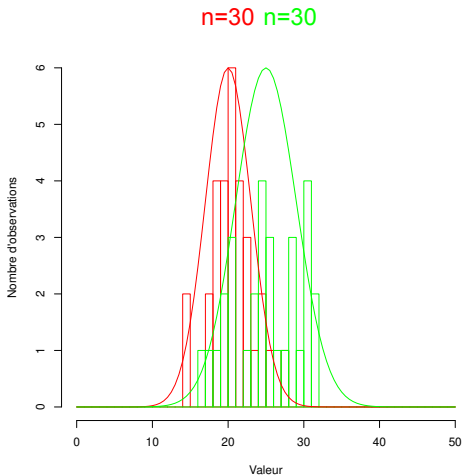
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

La p -value



Commandes R ayant généré ce graphe : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

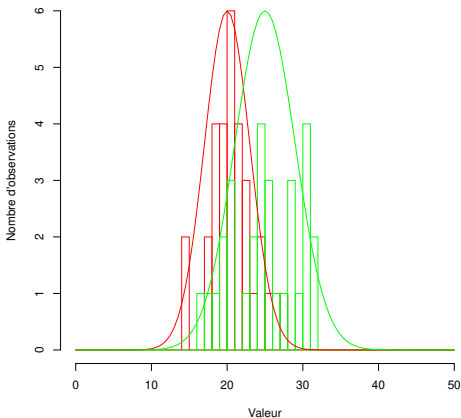
Conclusion

Suppléments

La p -value

$$p=5,945 \times 10^{-5}$$

$n=30$ $n=30$



Commandes R ayant généré ce graphique : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation


Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.


$$m = \frac{2,2+5,0+7,8+9,5+12,1}{5} = 7,32$$

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne.

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

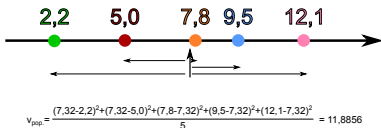
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne.



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

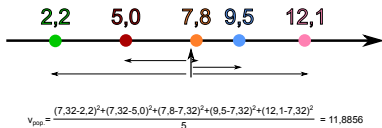
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne. Variance d'une population : $\frac{n}{n-1} \times$ variance d'un échantillon d'effectif n .



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne. Variance d'une population : $\frac{n}{n-1} \times$ variance d'un échantillon d'effectif n .



$$V_{\text{pop.}} = \frac{(7,32-2,2)^2 + (7,32-5,0)^2 + (7,8-7,32)^2 + (9,5-7,32)^2 + (12,1-7,32)^2}{5} = 11,8856$$

$$V_{\text{éch.}} = \frac{(7,32-2,2)^2 + (7,32-5,0)^2 + (7,8-7,32)^2 + (9,5-7,32)^2 + (12,1-7,32)^2}{4} = 14,857$$

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne. Variance d'une population : $\frac{n}{n-1} \times$ variance d'un échantillon d'effectif n .
- ▶ L'écart-type (*standard deviation*) : la racine carrée de la variance (donc : homogène à la moyenne).

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne (implicitement : « arithmétique ») (*mean*) : somme des valeurs, divisée par leur effectif.
- ▶ La variance (*variance*) : moyenne des carrés des écarts à la moyenne. Variance d'une population : $\frac{n}{n-1} \times$ variance d'un échantillon d'effectif n .
- ▶ L'écart-type (*standard deviation*) : la racine carrée de la variance (donc : homogène à la moyenne).



$$\sigma_{\text{pop.}} = \sqrt{v_{\text{pop.}}} = 3,44755$$

$$\sigma_{\text{éch.}} = \sqrt{v_{\text{éch.}}} = 3,854478$$

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

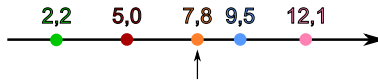
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).
- ▶ Le coefficient de variation (*coefficient of variation*) : écart-type divisé par la moyenne (permet de normaliser par l'ampleur des mesures).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

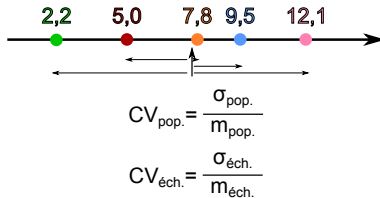
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).
- ▶ Le coefficient de variation (*coefficient of variation*) : écart-type divisé par la moyenne (permet de normaliser par l'ampleur des mesures).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).
- ▶ Le coefficient de variation (*coefficient of variation*) : écart-type divisé par la moyenne (permet de normaliser par l'ampleur des mesures).
- ▶ Le z-score (*z-score*) : pour chaque réplicat : son écart à la moyenne (mesuré en écarts-types).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

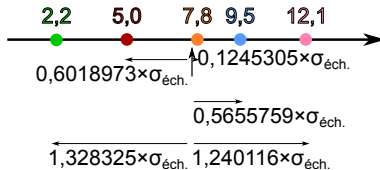
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La médiane (*median*) : dans une liste triée de manière croissante : valeur du milieu (leur moyenne si *ex-aequo*).
- ▶ Le coefficient de variation (*coefficient of variation*) : écart-type divisé par la moyenne (permet de normaliser par l'ampleur des mesures).
- ▶ Le z-score (*z-score*) : pour chaque réplicat : son écart à la moyenne (mesuré en écarts-types).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

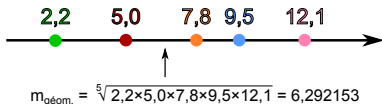
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.
- ▶ L'erreur-type (implicitement : « de la moyenne ») (*standard error [of the mean]*) : écart-type de la distribution d'une série d'évaluations de la moyenne sur des échantillons différents.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

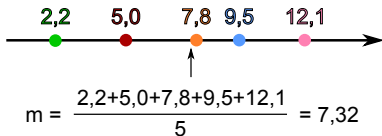
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.
- ▶ L'erreur-type (implicitement : « de la moyenne ») (*standard error [of the mean]*) : écart-type de la distribution d'une série d'évaluations de la moyenne sur des échantillons différents.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

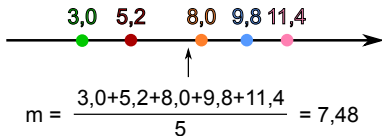
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.
- ▶ L'erreur-type (implicitement : « de la moyenne ») (*standard error [of the mean]*) : écart-type de la distribution d'une série d'évaluations de la moyenne sur des échantillons différents.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

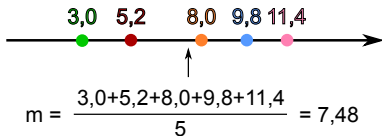
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Vocabulaire de statistiques descriptives

- ▶ La moyenne géométrique (*geometric mean*) : racine $n^{\text{ème}}$ du produit des n valeurs.
- ▶ L'erreur-type (implicitement : « de la moyenne ») (*standard error [of the mean]*) : écart-type de la distribution d'une série d'évaluations de la moyenne sur des échantillons différents. Si on note σ l'écart-type (théoriquement commun) de chacun de ces échantillons, l'erreur-type vaut : $\frac{\sigma}{\sqrt{n}}$.



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Représentation graphique

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

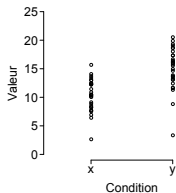
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Représentation graphique

Points individuels :



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

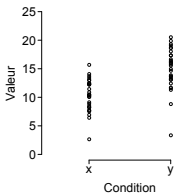
Designs expérimentaux plus élaborés

Conclusion

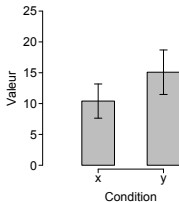
Suppléments

Représentation graphique

Points individuels :



Moyenne +/- écart-type :



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

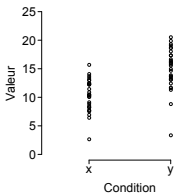
Designs expérimentaux plus élaborés

Conclusion

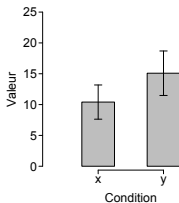
Suppléments

Représentation graphique

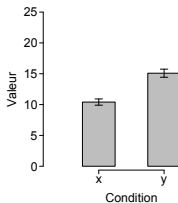
Points individuels :



Moyenne +/- écart-type :



Moyenne +/- erreur-type :



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

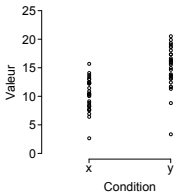
Designs expérimentaux plus élaborés

Conclusion

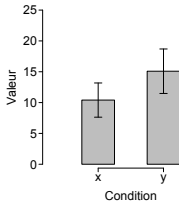
Suppléments

Représentation graphique

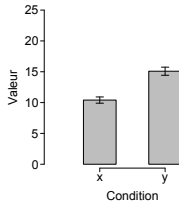
Points individuels :



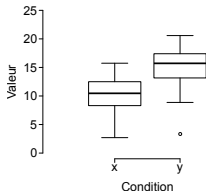
Moyenne +/- écart-type :



Moyenne +/- erreur-type :



Boxplot :



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

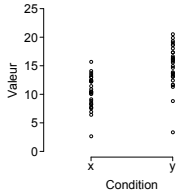
Designs expérimentaux plus élaborés

Conclusion

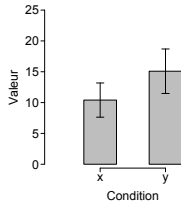
Suppléments

Représentation graphique

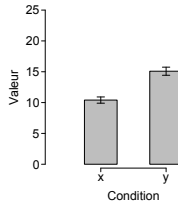
Points individuels :



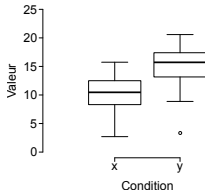
Moyenne +/- écart-type :



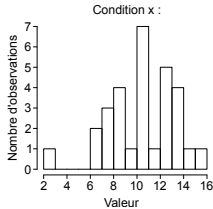
Moyenne +/- erreur-type :



Boxplot :



Histogramme :



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

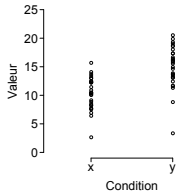
Designs expérimentaux plus élaborés

Conclusion

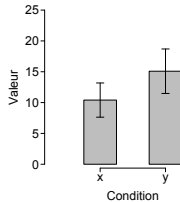
Suppléments

Représentation graphique

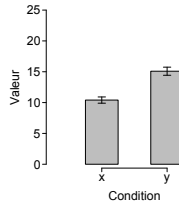
Points individuels :



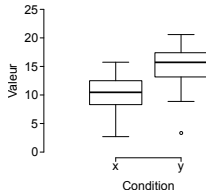
Moyenne +/- écart-type :



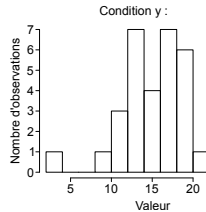
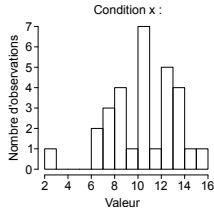
Moyenne +/- erreur-type :



Boxplot :



Histogramme :



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

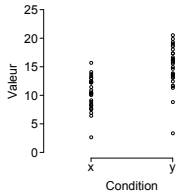
Designs expérimentaux plus élaborés

Conclusion

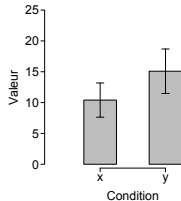
Suppléments

Représentation graphique

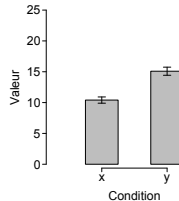
Points individuels :



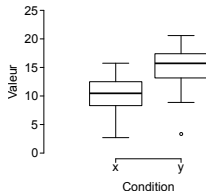
Moyenne +/- écart-type :



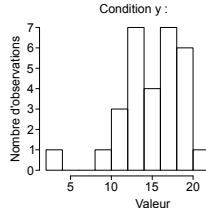
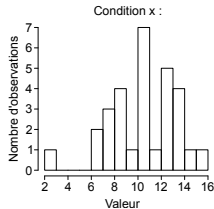
Moyenne +/- erreur-type :



Boxplot :



Histogramme :



Toujours décrire ce que représentent les barres d'erreur !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
écart des moyennes $\uparrow \implies p$ -value \downarrow

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

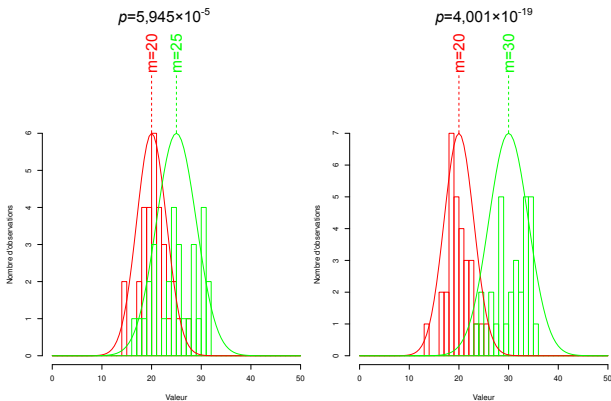
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
écart des moyennes $\uparrow \implies p$ -value \downarrow



Commandes R ayant généré ces graphes : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
écarts-types $\downarrow \implies p$ -value \downarrow

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

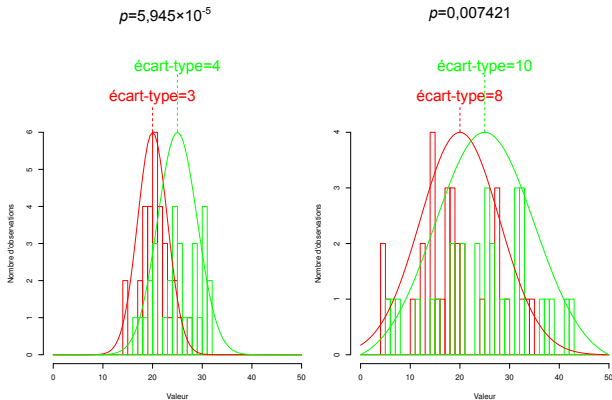
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
 écarts-types $\downarrow \implies p$ -value \downarrow



Commandes R ayant généré ces graphes : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
effectifs $\uparrow \implies p$ -value \downarrow

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

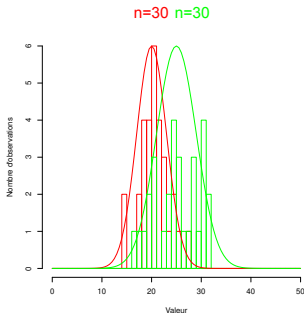
Conclusion

Suppléments

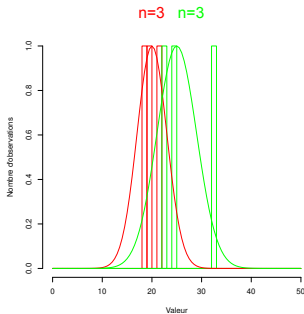
Déterminants de la p -value du t-test

Toutes choses égales par ailleurs :
effectifs $\uparrow \implies p$ -value \downarrow

$p=5,945 \times 10^{-5}$



$p=0,1447$



Commandes R ayant généré ces graphes : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Relativiser l'importance de la p -value !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Relativiser l'importance de la p -value ! Augmenter le nombre de réplicats tendra à la faire baisser.

Déterminants de la p -value du t-test

Relativiser l'importance de la p -value ! Augmenter le nombre de réplicats tendra à la faire baisser.

L'ampleur de l'effet (mesuré par le *fold-change*) et sa significativité (mesurée par la p -value) sont importantes toutes les deux.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Déterminants de la p -value du t-test

Relativiser l'importance de la p -value ! Augmenter le nombre de réplicats tendra à la faire baisser.

L'ampleur de l'effet (mesuré par le *fold-change*) et sa significativité (mesurée par la p -value) sont importantes toutes les deux.

L'effet se mesure par son ampleur ; la p -value mesure la confiance qu'on peut avoir dans cette ampleur.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations échantillonnées se distribue selon une loi normale (= « loi gaussienne »).

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations **échantillonnées** se distribue selon une loi normale (= « loi gaussienne »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations **échantillonnées** se distribue selon une loi normale (= « loi gaussienne »).

On regarde si les échantillons (= nos séries de réplicats) peuvent crédiblement être tirés d'une population normalement distribuée.

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations échantillonnées se distribue selon une loi normale (= « loi gaussienne »).

On regarde si les échantillons (= nos séries de réplicats) peuvent crédiblement être tirés d'une population normalement distribuée.

Le test de Shapiro-Wilk

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations échantillonnées se distribue selon une loi normale (= « loi gaussienne »).

On regarde si les échantillons (= nos séries de réplicats) peuvent crédiblement être tirés d'une population normalement distribuée.

Le test de Shapiro-Wilk (p -value : probabilité que la série de réplicats soit issue d'une population normalement distribuée).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

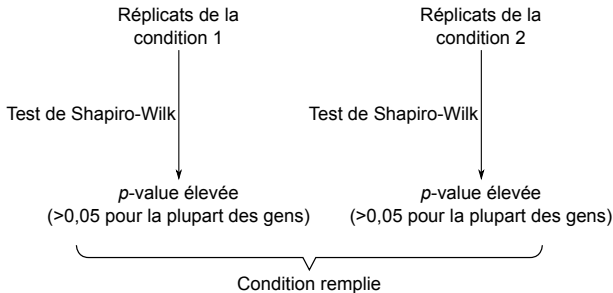
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

1^{ère} condition : chacune des deux populations échantillonnées se distribue selon une loi normale (= « loi gaussienne »).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

2^{ème} condition : les variances des deux conditions sont homogènes (\approx égales)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

2^{ème} condition : les variances des deux conditions sont homogènes (\approx égales) (condition d'« homoscédasticité »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

2^{ème} condition : les variances des deux conditions sont homogènes (\approx égales) (condition d'« homoscedasticité »).

Le test de Levene

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

2^{ème} condition : les variances des deux conditions sont homogènes (\approx égales) (condition d'« homoscedasticité »).

Le test de Levene (p -value : probabilité que les deux séries de réplicats soient issues de populations de même variance).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

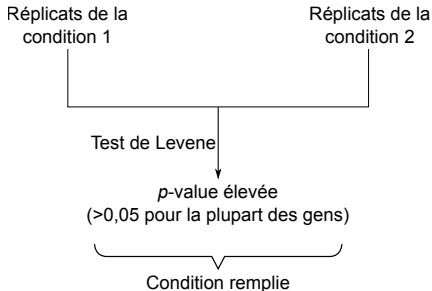
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

2^{ème} condition : les variances des deux conditions sont homogènes (\approx égales) (condition d'« homoscédasticité »).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Homoscédasticité : utiliser le t-test de Welch (= « t-test à variances hétérogènes ») au lieu du t-test de Student (= « t-test à variances homogènes »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Homoscédasticité : utiliser le t-test de Welch (= « t-test à variances hétérogènes ») au lieu du t-test de Student (= « t-test à variances homogènes »).

Si le t-test de Student est utilisable, il faut le préférer : il est plus puissant.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Homoscédasticité : utiliser le t-test de Welch (= « t-test à variances hétérogènes ») au lieu du t-test de Student (= « t-test à variances homogènes »).

Si le t-test de Student est utilisable, il faut le préférer : il est plus puissant (détecte la significativité avec plus de sensibilité).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Homoscédasticité : utiliser le t-test de Welch (= « t-test à variances hétérogènes ») au lieu du t-test de Student (= « t-test à variances homogènes »).

Si le t-test de Student est utilisable, il faut le préférer : il est plus puissant (détecte la significativité avec plus de sensibilité).

Normalité : utiliser une transformation mathématique (logarithme, racine carrée, ... ; de préférence si elle a un sens physique).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Et que faire si l'une au moins de ces conditions n'est pas respectée ?

Homoscédasticité : utiliser le t-test de Welch (= « t-test à variances hétérogènes ») au lieu du t-test de Student (= « t-test à variances homogènes »).

Si le t-test de Student est utilisable, il faut le préférer : il est plus puissant (détecte la significativité avec plus de sensibilité).

Normalité : utiliser une transformation mathématique (logarithme, racine carrée, ... ; de préférence si elle a un sens physique). Si le log des données est normalement distribué, faire le t-test sur le log des données.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Si même une transformation mathématique ne suffit pas :
utiliser le test de Wilcoxon (= test de Mann-Whitney).

Pourquoi les
statistiques ?

La p -value

Conditions
d'application du
t-test

Variantes du t-test

Lois de distribution

Généralisation : les
tests statistiques

Tests de
comparaison de
distributions

Tests d'hypothèses
multiples

Tests de
corrélation

Designs
expérimentaux plus
élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Si même une transformation mathématique ne suffit pas :
utiliser le test de Wilcoxon (= test de Mann-Whitney).

Travaille sur les rangs des données (robuste aux déviations à la normalité).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Si même une transformation mathématique ne suffit pas :
utiliser le test de Wilcoxon (= test de Mann-Whitney).

Travaille sur les rangs des données (robuste aux déviations à
la normalité).

Si les données sont normalement distribuées : préférer le
t-test, plus puissant.

Pourquoi les
statistiques ?

La p -value

Conditions
d'application du
t-test

Variantes du t-test

Lois de distribution

Généralisation : les
tests statistiques

Tests de
comparaison de
distributions

Tests d'hypothèses
multiples

Tests de
corrélation

Designs
expérimentaux plus
élaborés

Conclusion

Suppléments

Conditions d'application du t-test

Si même une transformation mathématique ne suffit pas : utiliser le test de Wilcoxon (= test de Mann-Whitney).

Travaille sur les rangs des données (robuste aux déviations à la normalité).

Si les données sont normalement distribuées : préférer le t-test, plus puissant.

Exemple : comparer deux séries de 3 réplicats chacune : par construction, la p -value du test de Wilcoxon ne pourra pas descendre en-dessous de 0,1 quelles que soient les distributions des deux séries !

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Appariement des données

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Appariement des données

Existe-t-il un lien particulier entre chaque réplikat d'une série et un (et un seul) réplikat de l'autre ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Appariement des données

Existe-t-il un lien particulier entre chaque réplikat d'une série et un (et un seul) réplikat de l'autre ?

Exemple : il s'agit du même individu analysé dans deux conditions (et chaque réplikat est un individu différent).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Appariement des données

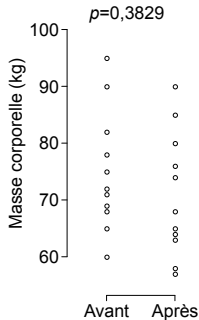
Existe-t-il un lien particulier entre chaque réplikat d'une série et un (et un seul) réplikat de l'autre ?

Exemple : il s'agit du même individu analysé dans deux conditions (et chaque réplikat est un individu différent).

Implication : les deux séries ont le même nombre de répliquats.

Appariement des données

Existe-t-il un lien particulier entre chaque réplikat d'une série et un (et un seul) réplikat de l'autre ?



Commandes R ayant généré ces graphes : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

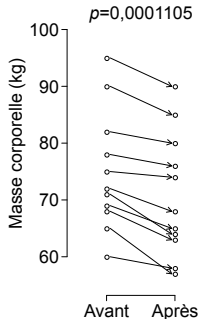
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Appariement des données

Existe-t-il un lien particulier entre chaque réplikat d'une série et un (et un seul) réplikat de l'autre ?



Commandes R ayant généré ces graphes : [\[lien\]](#)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Si je cherche une différence, quelle que soit son sens :
two-tailed t-test (= *two-sided* t-test).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Si je cherche une différence, quelle que soit son sens :
two-tailed t-test (= *two-sided* t-test).

Si je ne m'intéresse qu'à un sens de variation (ex. : les souris mutantes sont-elles plus légères que les souris sauvages ?) :
one-tailed t-test (= *one-sided* t-test).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Si je cherche une différence, quelle que soit son sens :
two-tailed t-test (= *two-sided* t-test).

Si je ne m'intéresse qu'à un sens de variation (ex. : les souris mutantes sont-elles plus légères que les souris sauvages ?) :
one-tailed t-test (= *one-sided* t-test).

Attention ! Il est rare de se désintéresser d'un des deux sens de variation, or le *one-tailed* t-test a tendance à donner des p -values plus basses.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Si je cherche une différence, quelle que soit son sens :
two-tailed t-test (= *two-sided* t-test).

Si je ne m'intéresse qu'à un sens de variation (ex. : les souris mutantes sont-elles plus légères que les souris sauvages ?) :
one-tailed t-test (= *one-sided* t-test).

Attention ! Il est rare de se désintéresser d'un des deux sens de variation, or le *one-tailed* t-test a tendance à donner des p -values plus basses.

⇒ possibilité d'un usage abusif, pour faire baisser la p -value de manière illégitime.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

t-test « *one-tailed* » ou « *two-tailed* »

Est-ce que je ne m'intéresse qu'à un sens de variation ?

Si je cherche une différence, quelle que soit son sens :
two-tailed t-test (= *two-sided* t-test).

Si je ne m'intéresse qu'à un sens de variation (ex. : les souris mutantes sont-elles plus légères que les souris sauvages ?) :
one-tailed t-test (= *one-sided* t-test).

Attention ! Il est rare de se désintéresser d'un des deux sens de variation, or le *one-tailed* t-test a tendance à donner des p -values plus basses.

⇒ possibilité d'un usage abusif, pour faire baisser la p -value de manière illégitime. Si on n'a pas de raison impérieuse de négliger un sens de variation, choisir le *two-tailed* t-test.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

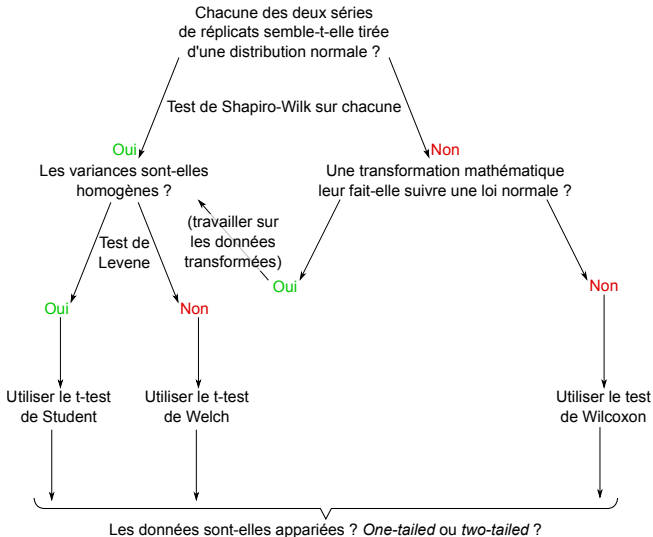
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Arbre décisionnel sur l'usage du t-test



Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

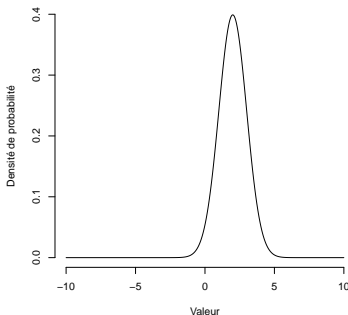
Conclusion

Suppléments

Lois de distribution

Quelques exemples courants :

- ▶ Loi normale (=« gaussienne ») : continue, symétrique, de $-\infty$ à $+\infty$.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

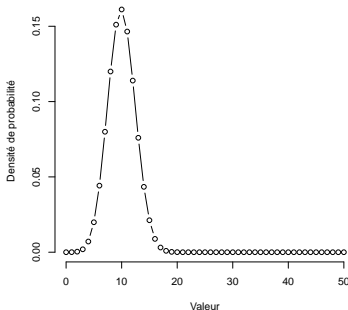
Conclusion

Suppléments

Lois de distribution

Quelques exemples courants :

- ▶ Loi normale (=« gaussienne »)
- ▶ Loi binomiale (probabilité d'obtenir un nombre donné de succès sur des tirages aléatoires indépendants) :
discontinue, définie seulement pour les valeurs positives.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

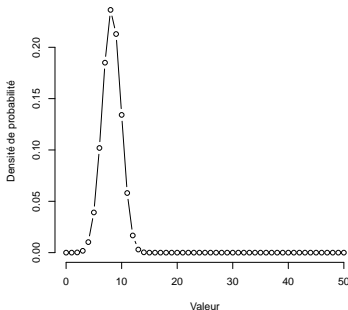
Conclusion

Suppléments

Lois de distribution

Quelques exemples courants :

- ▶ Loi normale (=« gaussienne »)
- ▶ Loi binomiale
- ▶ Loi hypergéométrique (idem mais pour des tirages sans remise).



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

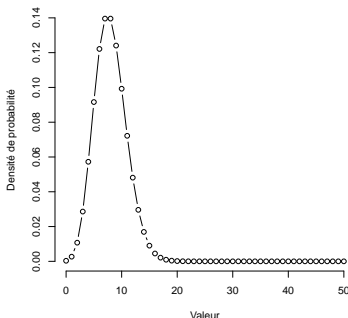
Conclusion

Suppléments

Lois de distribution

Quelques exemples courants :

- ▶ Loi normale (=« gaussienne »)
- ▶ Loi binomiale
- ▶ Loi hypergéométrique
- ▶ Loi de Poisson (probabilité d'obtenir k succès dans un temps donné, s'il y a en moyenne λ succès indépendants dans un tel intervalle) : discontinue, définie seulement pour les valeurs positives.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Le t-test évalue la probabilité d'une hypothèse (égalité des moyennes des populations échantillonnées).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Le t-test évalue la probabilité d'une hypothèse (égalité des moyennes des populations échantillonnées).

→ L'« hypothèse nulle ».

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Le t-test évalue la probabilité d'une hypothèse (égalité des moyennes des populations échantillonnées).

→ L'« hypothèse nulle ».

Hypothèse nulle du test de Shapiro-Wilk : la population échantillonnée suit une loi normale ; du test de Levene : les populations échantillonnées ont la même variance.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Le t-test évalue la probabilité d'une hypothèse (égalité des moyennes des populations échantillonnées).

→ L'« hypothèse nulle ».

Hypothèse nulle du test de Shapiro-Wilk : la population échantillonnée suit une loi normale ; du test de Levene : les populations échantillonnées ont la même variance.

Mesure de l'incertitude : par une probabilité (la p -value), ou par un intervalle (« la différence entre les moyennes des populations échantillonnées se trouve dans l'intervalle [...] avec une probabilité $\geq 0,95$ »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Le t-test évalue la probabilité d'une hypothèse (égalité des moyennes des populations échantillonnées).

→ L'« hypothèse nulle ».

Hypothèse nulle du test de Shapiro-Wilk : la population échantillonnée suit une loi normale ; du test de Levene : les populations échantillonnées ont la même variance.

Mesure de l'incertitude : par une probabilité (la p -value), ou par un intervalle (« la différence entre les moyennes des populations échantillonnées se trouve dans l'intervalle [... ; ...] avec une probabilité $\geq 0,95$ »).

Attention à la formulation ! Dire « je n'ai pas vu de différence significative », pas « il n'y a pas de différence » (pourrait ressortir avec davantage de réplicats).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Pour chaque test : le seuil de significativité détermine le pourcentage attendu de faux positifs.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Pour chaque test : le seuil de significativité détermine le pourcentage attendu de faux positifs. Exemple du t-test : seuil de p -value à 0,05 : même des échantillons tirés de populations de même moyenne apparaîtront significativement différents $\approx 5\%$ du temps.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Pour chaque test : le seuil de significativité détermine le pourcentage attendu de faux positifs. Exemple du t-test : seuil de p -value à 0,05 : même des échantillons tirés de populations de même moyenne apparaîtront significativement différents $\approx 5\%$ du temps.

Vocabulaire : un faux positif est également appelé « Erreur de type 1 ».

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Pour chaque test : le seuil de significativité détermine le pourcentage attendu de faux positifs. Exemple du t-test : seuil de p -value à 0,05 : même des échantillons tirés de populations de même moyenne apparaîtront significativement différents $\approx 5\%$ du temps.

Vocabulaire : un faux positif est également appelé « Erreur de type 1 ».

Faux négatif (« Erreur de type 2 ») : n'apparaît pas significatif alors que les populations échantillonnées ont des moyennes différentes.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Généralisation : les tests statistiques

Pour chaque test : le seuil de significativité détermine le pourcentage attendu de faux positifs. Exemple du t-test : seuil de p -value à 0,05 : même des échantillons tirés de populations de même moyenne apparaîtront significativement différents $\approx 5\%$ du temps.

Vocabulaire : un faux positif est également appelé « Erreur de type 1 ».

Faux négatif (« Erreur de type 2 ») : n'apparaît pas significatif alors que les populations échantillonnées ont des moyennes différentes. Sa fréquence dépend de l'ampleur de la différence, des variances, du nombre de réplicats, du seuil de significativité.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

	Cond. 1	Cond. 2
G1	19	5
S	4	8
G2	25	36
M	3	1

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

	Cond. 1	Cond. 2
G1	19	5
S	4	8
G2	25	36
M	3	1

Un t-test sur chaque catégorie ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

	Cond. 1	Cond. 2
G1	19	5
S	4	8
G2	25	36
M	3	1

Un t-test sur chaque catégorie ? Nécessite des réplicats du comptage (lui-même déjà répliqué).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

	Cond. 1	Cond. 2
G1	19	5
S	4	8
G2	25	36
M	3	1

Un t-test sur chaque catégorie ? Nécessite des réplicats du comptage (lui-même déjà répliqué).

Si une catégorie apparaît significativement différente mais pas les autres, que conclure ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°1 : comparer des comptages par catégorie entre plusieurs conditions expérimentales.

	Cond. 1	Cond. 2
G1	19	5
S	4	8
G2	25	36
M	3	1

Un t-test sur chaque catégorie ? Nécessite des réplicats du comptage (lui-même déjà répliqué).

Si une catégorie apparaît significativement différente mais pas les autres, que conclure ?

→ le t-test n'est pas adapté.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Pour comparer des tables de comptage : test du χ^2 , ou (plus précis) test exact de Fisher.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Pour comparer des tables de comptage : test du χ^2 , ou (plus précis) test exact de Fisher.

Ici : test du χ^2 : p -value=0.005921 ; test exact de Fisher : p -value=0.003375.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Pour comparer des tables de comptage : test du χ^2 , ou (plus précis) test exact de Fisher.

Ici : test du χ^2 : p -value=0.005921 ; test exact de Fisher : p -value=0.003375.

Test du χ^2 : imprécis pour de petits effectifs (moins de ≈ 10 observations dans au moins une catégorie).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Pour comparer des tables de comptage : test du χ^2 , ou (plus précis) test exact de Fisher.

Ici : test du χ^2 : p -value=0.005921 ; test exact de Fisher : p -value=0.003375.

Test du χ^2 : imprécis pour de petits effectifs (moins de ≈ 10 observations dans au moins une catégorie). Test exact de Fisher : les calculs peuvent être très longs si les effectifs sont grands (nombreuses permutations possibles).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Pour comparer des tables de comptage : test du χ^2 , ou (plus précis) test exact de Fisher.

Ici : test du χ^2 : p -value=0.005921 ; test exact de Fisher : p -value=0.003375.

Test du χ^2 : imprécis pour de petits effectifs (moins de ≈ 10 observations dans au moins une catégorie). Test exact de Fisher : les calculs peuvent être très longs si les effectifs sont grands (nombreuses permutations possibles).

Danger! Ces tests utilisent les données brutes de comptage (pas de normalisation!).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

3	2
6	6
5	6

$p\text{-value}=1$

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

3	2
6	6
5	6

$p\text{-value}=1$

30	20
60	60
50	60

$p\text{-value}=0,2413$

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

3	2
6	6
5	6

$p\text{-value}=1$

30	20
60	60
50	60

$p\text{-value}=0,2413$

300	200
600	600
500	600

$p\text{-value}=4,588 \times 10^{-7}$

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

3	2
6	6
5	6

$p\text{-value}=1$

30	20
60	60
50	60

$p\text{-value}=0,2413$

300	200
600	600
500	600

$p\text{-value}=4,588 \times 10^{-7}$

Une normalisation (ex. : pourcentage) perdrait l'information du nombre brut d'observations.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque). Version historique du test : réservée aux variables continues.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

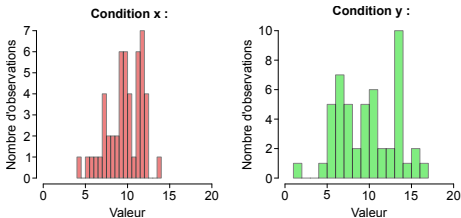
Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).



Commandes R ayant généré ces graphes : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

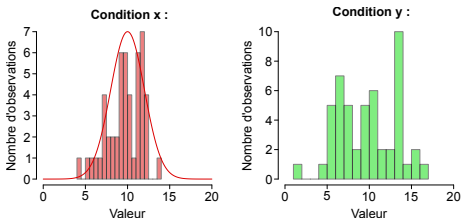
Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).



Commandes R ayant généré ces graphes : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

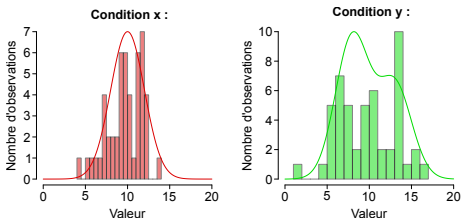
Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).



Commandes R ayant généré ces graphes : [\[lien\]](#).

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

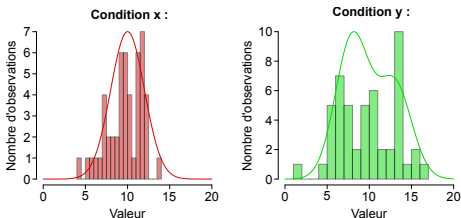
Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).



Commandes R ayant généré ces graphes : [\[lien\]](#). p -values :
t-test : 0.9005 ; test de Kolmogorov-Smirnov : 0.02171.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de comparaison de distributions

Contexte n°2 : comparer des distributions globales (pas simplement leurs moyennes).

Test de Kolmogorov-Smirnov : hypothèse nulle : les deux séries de réplicats sont échantillonnées d'une même distribution (inconnue, quelconque).

→ Plus sensible, mais plus difficile à interpréter (sauf connaissance mécanistique détaillée).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Seuil de p -value de 0,05 : il faut s'attendre à $\approx 5\%$ de faux positifs.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Seuil de p -value de 0,05 : il faut s'attendre à $\approx 5\%$ de faux positifs.

Si on fait de nombreux tests (y a-t-il une différence significative entre les conditions « x » et « y » au jour 1 ? Au jour 2 ? Au jour 3 ?...)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Seuil de p -value de 0,05 : il faut s'attendre à $\approx 5\%$ de faux positifs.

Si on fait de nombreux tests (y a-t-il une différence significative entre les conditions « x » et « y » au jour 1 ? Au jour 2 ? Au jour 3 ?...) le nombre de faux positifs sera de $\approx 0,05 \times$ nombre de tests.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Seuil de p -value de 0,05 : il faut s'attendre à $\approx 5\%$ de faux positifs.

Si on fait de nombreux tests (y a-t-il une différence significative entre les conditions « x » et « y » au jour 1 ? Au jour 2 ? Au jour 3 ?...) le nombre de faux positifs sera de $\approx 0,05 \times$ nombre de tests.

→ en faisant 20 000 tests, on observera $\approx 1 000$ faux positifs (une expérience de transcriptomique serait toujours fausse !).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Seuil de p -value de 0,05 : il faut s'attendre à $\approx 5\%$ de faux positifs.

Si on fait de nombreux tests (y a-t-il une différence significative entre les conditions « x » et « y » au jour 1 ? Au jour 2 ? Au jour 3 ?...) le nombre de faux positifs sera de $\approx 0,05 \times$ nombre de tests.

→ en faisant 20 000 tests, on observera $\approx 1 000$ faux positifs (une expérience de transcriptomique serait toujours fausse!).

Différence entre vrais et faux positifs : les vrais sont reproductibles (limitations pratiques à la reproductibilité de grosses expériences).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Correction de Bonferroni : si n est le nombre d'hypothèses testées, et α le seuil habituel, utiliser plutôt le seuil α/n .

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Correction de Bonferroni : si n est le nombre d'hypothèses testées, et α le seuil habituel, utiliser plutôt le seuil α/n .
Plus simple d'usage : multiplier les p -values par n (et ramener à 1 si elles sont > 1) plutôt que de diviser le seuil.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Correction de Bonferroni : si n est le nombre d'hypothèses testées, et α le seuil habituel, utiliser plutôt le seuil α/n .
Plus simple d'usage : multiplier les p -values par n (et ramener à 1 si elles sont > 1) plutôt que de diviser le seuil.

Correction de Benjamini-Hochberg : ne pas multiplier les n p -values par n , mais : par un facteur croissant (de 1 à n) dans la liste décroissante des p -values.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Correction de Bonferroni : si n est le nombre d'hypothèses testées, et α le seuil habituel, utiliser plutôt le seuil α/n .
Plus simple d'usage : multiplier les p -values par n (et ramener à 1 si elles sont > 1) plutôt que de diviser le seuil.

Correction de Benjamini-Hochberg : ne pas multiplier les n p -values par n , mais : par un facteur croissant (de 1 à n) dans la liste décroissante des p -values. Moins stringent, moins de faux négatifs. Principe *ad hoc*, mais très populaire en biologie moléculaire à haut débit.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests d'hypothèses multiples

Méthode empirique : rendre le seuil de significativité d'autant plus stringent que le nombre de tests est grand.

Correction de Bonferroni : si n est le nombre d'hypothèses testées, et α le seuil habituel, utiliser plutôt le seuil α/n .
Plus simple d'usage : multiplier les p -values par n (et ramener à 1 si elles sont > 1) plutôt que de diviser le seuil.

Correction de Benjamini-Hochberg : ne pas multiplier les n p -values par n , mais : par un facteur croissant (de 1 à n) dans la liste décroissante des p -values. Moins stringent, moins de faux négatifs. Principe *ad hoc*, mais très populaire en biologie moléculaire à haut débit.

Cas particulier de t-tests multiples contre une condition-contrôle commune : le test de Dunnett.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation

Principe : deux variables ont-elles tendance à co-varier, ou à varier indépendamment ?

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

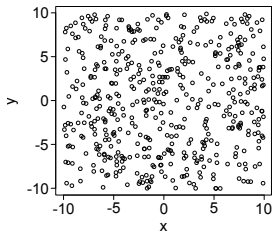
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

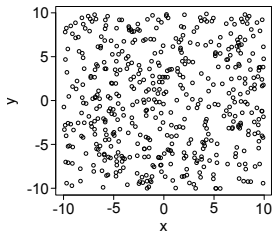
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Coefficient de Pearson : $r = 0.0352$ ($p\text{-value}=0.432$).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

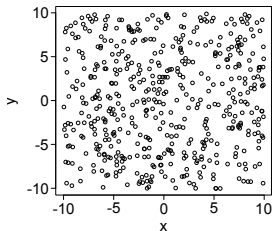
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ ($p\text{-value}=0.432$).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

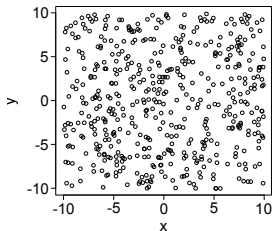
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ ($p\text{-value}=0.432$).

Vaut +1 pour une corrélation linéaire parfaite et croissante, -1 si elle est décroissante, des valeurs intermédiaires si la corrélation est imparfaite.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

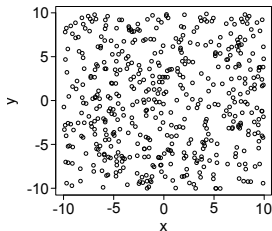
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ (p -value=0.432).

Coefficient de Kendall : $\tau = 0.0203$ (p -value=0.4968).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

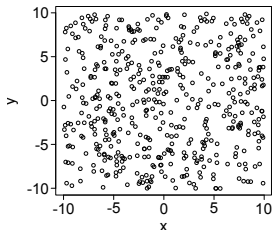
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ (p -value=0.432).

Coefficient de Kendall : $\tau = 0.0203$ (p -value=0.4968).

Vaut +1 si toutes les paires de points varient de façon concordante, -1 si elles varient toutes de manière discordante, des valeurs intermédiaires sinon.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

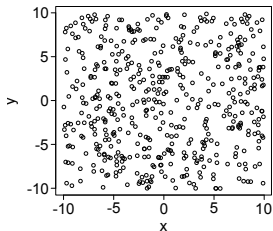
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ (p -value=0.432).

Coefficient de Kendall : $\tau = 0.0203$ (p -value=0.4968).

Coefficient de Spearman : $\rho = 0.0315$ (p -value=0.4827).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

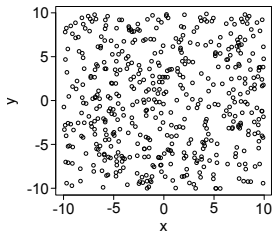
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Commandes R ayant généré ce graphe : [\[lien\]](#).

Hypothèse nulle : le coefficient de corrélation vaut 0.

Coefficient de Pearson : $r = 0.0352$ (p -value=0.432).

Coefficient de Kendall : $\tau = 0.0203$ (p -value=0.4968).

Coefficient de Spearman : $\rho = 0.0315$ (p -value=0.4827).

Coefficient de Pearson sur les rangs des valeurs (cherche une relation monotone, pas nécessairement linéaire).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

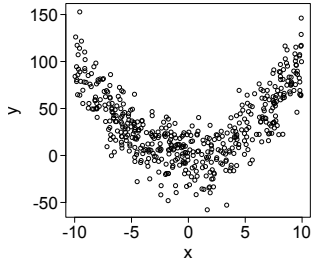
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

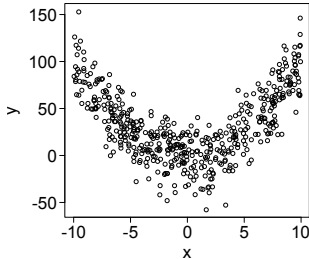
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Coefficient de Pearson : $r = 0.0846$ (p -value=0.05882)

Coefficient de Kendall : $\tau = 0.0126$ (p -value=0.6732)

Coefficient de Spearman : $\rho = 0.0435$ (p -value=0.3313)

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

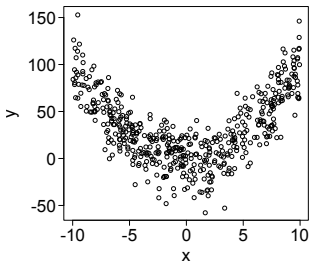
Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation



Coefficient de Pearson : $r = 0.0846$ (p -value=0.05882)

Coefficient de Kendall : $\tau = 0.0126$ (p -value=0.6732)

Coefficient de Spearman : $\rho = 0.0435$ (p -value=0.3313)

→ Disposer d'un modèle mathématique de la réponse y à x (« y a-t-il une corrélation entre y et x^2 ? »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation

Un piège classique : corrélation n'est pas causalité.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Tests de corrélation

Un piège classique : corrélation n'est pas causalité.

A est la cause de B, ou B est la cause de A ? A et B sont-elles deux conséquences d'une même cause C ? ...

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Comparer entre eux plus de deux groupes.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Comparer entre eux plus de deux groupes (e.g., « Entre les boulangers, les professeurs, les policiers, les infirmiers, y a-t-il une différence de temps passé à regarder la télévision ? »).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

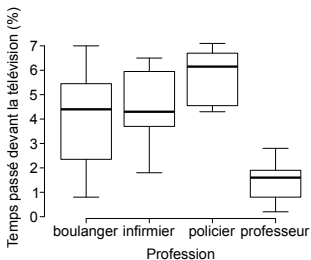
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Comparer entre eux plus de deux groupes.



Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

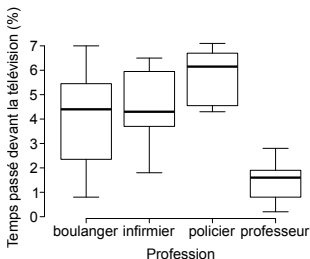
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Comparer entre eux plus de deux groupes.



Analysis of variance (ANOVA) : conditions : normalité des résidus (\implies normalité des observations à l'intérieur de chaque groupe), homogénéité des variances, et indépendance des observations.

Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

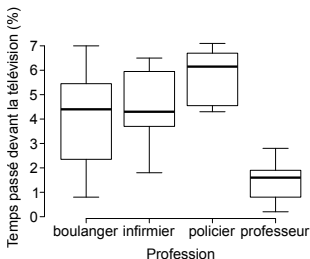
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Comparer entre eux plus de deux groupes.



Analysis of variance (ANOVA) : conditions : normalité des résidus (\implies normalité des observations à l'intérieur de chaque groupe), homogénéité des variances, et indépendance des observations.

ANOVA p -value = $7,39 \times 10^{-6}$ \longrightarrow un effet de la profession (sans détail!).

Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

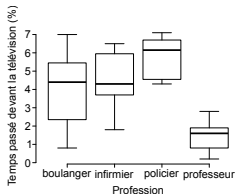
Tests « post-hoc » (ici : t-tests deux à deux) pour identifier les groupes significativement différents.

Designs expérimentaux plus élaborés

Tests « post-hoc » (ici : t-tests deux à deux) pour identifier les groupes significativement différents.

p -value des t-tests avec correction de Benjamini-Hochberg :

	boulangier	infirmier	policier
infirmier	0,47648	-	-
policier	0,03867	0,12304	-
professeur	0.00290	0.00036	5.2×10^{-6}



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Tests « post-hoc » (ici : t-tests deux à deux) pour identifier les groupes significativement différents.

p -value des t-tests avec correction de Benjamini-Hochberg :

	boulangier	infirmier	policier
infirmier	0,47648	-	-
policier	0,03867	0,12304	-
professeur	0.00290	0.00036	5.2×10^{-6}

Danger! Commencer par l'ANOVA avant de se lancer dans les tests 2 à 2 (sinon risque de faux positifs : test d'hypothèses multiples).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Plusieurs variables à la fois (ex. : effet de l'âge et de la souche de drosophiles sur une mesure physiologique).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

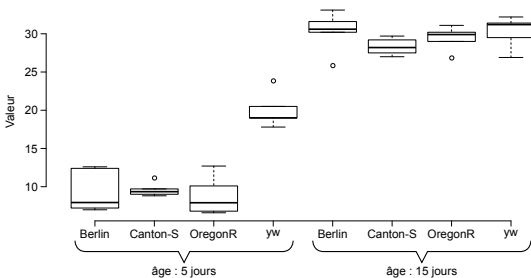
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Plusieurs variables à la fois (ex. : effet de l'âge et de la souche de drosophiles sur une mesure physiologique).



Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

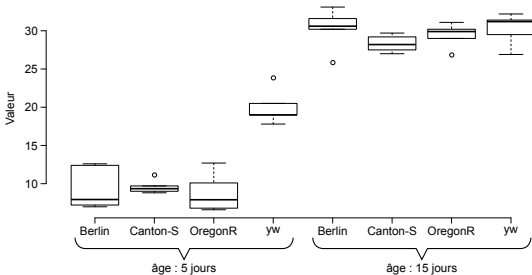
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Plusieurs variables à la fois (ex. : effet de l'âge et de la souche de drosophiles sur une mesure physiologique).



Une ANOVA multidimensionnelle (ici : deux variables → *two-way ANOVA*).

Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t -test

Variantes du t -test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

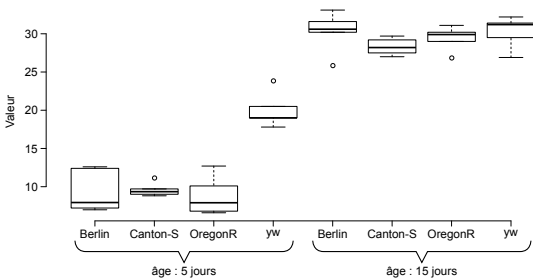
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Plusieurs variables à la fois (ex. : effet de l'âge et de la souche de drosophiles sur une mesure physiologique).



Une ANOVA multidimensionnelle (ici : deux variables → *two-way ANOVA*).

Mêmes exigences que la *one-way ANOVA* : normalité, homoscédasticité, indépendance.

Commandes R ayant généré ce graphe : [\[lien\]](#).

Pourquoi les statistiques ?

La *p*-value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

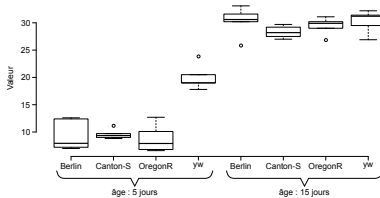
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

ANOVA à deux variables sans interaction : p -values :
 souche : $1,47 \times 10^{-4}$; âge : $< 2 \times 10^{-16}$.



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

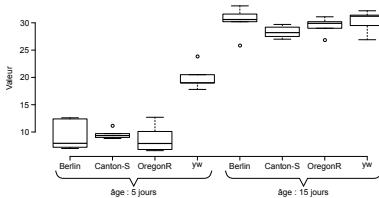
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

ANOVA à deux variables sans interaction : p -values :
 souche : $1,47 \times 10^{-4}$; âge : $< 2 \times 10^{-16}$.



Si chaque variable a un effet, leur interaction pourrait en avoir un.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

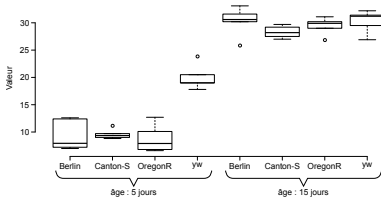
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

ANOVA à deux variables sans interaction : p -values :
 souche : $1,47 \times 10^{-4}$; âge : $< 2 \times 10^{-16}$.



Si chaque variable a un effet, leur interaction pourrait en avoir un.

ANOVA à deux variables avec interaction : p -values :
 souche : $3,17 \times 10^{-7}$; âge : $< 2 \times 10^{-16}$; leur interaction :
 $6,25 \times 10^{-6}$.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

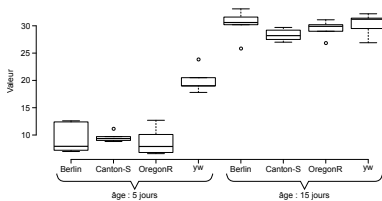
Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

ANOVA à deux variables sans interaction : p -values :
 souche : $1,47 \times 10^{-4}$; âge : $< 2 \times 10^{-16}$.



Si chaque variable a un effet, leur interaction pourrait en avoir un.

ANOVA à deux variables avec interaction : p -values :
 souche : $3,17 \times 10^{-7}$; âge : $< 2 \times 10^{-16}$; leur interaction :
 $6,25 \times 10^{-6}$.

Interprétation : l'âge a un effet, la souche a un effet, et le vieillissement affecte différemment les différentes souches.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Si les conditions d'application de l'ANOVA ne sont pas remplies :

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Si les conditions d'application de l'ANOVA ne sont pas remplies :

- ▶ Une transformation mathématique (ex. : log) pourrait les faire remplir.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Si les conditions d'application de l'ANOVA ne sont pas remplies :

- ▶ Une transformation mathématique (ex. : log) pourrait les faire remplir.
- ▶ Alternatives non-paramétriques (robustes à la non-normalité et hétéroscédasticité) pour la *one-way ANOVA* : test de Kuskal-Wallis (mesures non répétées), test de Friedman (mesures répétées sur chaque sujet).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Designs expérimentaux plus élaborés

Si les conditions d'application de l'ANOVA ne sont pas remplies :

- ▶ Une transformation mathématique (ex. : log) pourrait les faire remplir.
- ▶ Alternatives non-paramétriques (robustes à la non-normalité et hétéroscédasticité) pour la *one-way ANOVA* : test de Kuskal-Wallis (mesures non répétées), test de Friedman (mesures répétées sur chaque sujet).

Si les variables ne sont pas catégorielles (« profession », « souche de la drosophile ») mais numériques avec plus de 2 niveaux : modèles mathématiques (ex. : modèles linéaires) pour extraire l'effet de chaque variable.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conclusion

Statistiques

H. Seitz



Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conclusion

- ▶ Concepts de base, généralisables à de nombreux tests statistiques (p -value, intervalle de confiance, ...).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conclusion

- ▶ Concepts de base, généralisables à de nombreux tests statistiques (p -value, intervalle de confiance, ...).
- ▶ Éléments de vocabulaire (écart-type \neq erreur-type ; normalité, homoscédasticité, ...).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Conclusion

- ▶ Concepts de base, généralisables à de nombreux tests statistiques (p -value, intervalle de confiance, ...).
- ▶ Éléments de vocabulaire (écart-type \neq erreur-type; normalité, homoscédasticité, ...).
- ▶ \longrightarrow être en mesure de trouver par soi-même les informations pour les cas les plus compliqués.

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments

Suppléments

Versions synthétiques de ce cours :

- ▶ Par écrit : [première partie](#) (publiée en juillet 2010 dans *Regard sur la biochimie*), [deuxième partie](#) (publiée en octobre 2010 dans *Regard sur la biochimie*).
- ▶ En vidéo : « [Les statistiques en biologie moléculaire](#) ».

Version anglaise de ce diaporama :

- ▶ [Première partie](#) (généralités, introduction au t-test).
- ▶ [Deuxième partie](#) (distributions classiques, généralités sur les tests statistiques).
- ▶ [Troisième partie](#) (test d'hypothèses multiples, test de corrélation, designs expérimentaux complexes).

Pourquoi les statistiques ?

La p -value

Conditions d'application du t-test

Variantes du t-test

Lois de distribution

Généralisation : les tests statistiques

Tests de comparaison de distributions

Tests d'hypothèses multiples

Tests de corrélation

Designs expérimentaux plus élaborés

Conclusion

Suppléments