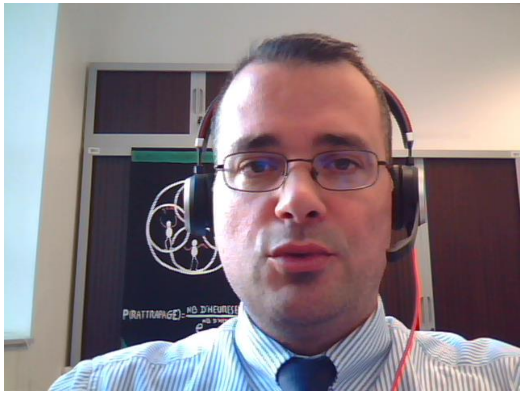
Le Test Binomial

I. Contexte du test
II. Rappels factorielle, arrangement, combinaison
III. Exemple concret pas-à-pas
IV. Généralisation
V. Quelques tests d'adéquation
VI. Généralités sur les tests statistiques



Proportions théoriques et proportions observées

- π = proportion théorique
  p=proportion observée, p_0=une proportion donnée

- Intervalle de fluctuation (théorique) : si π est connue, on peut prédire les valeurs que peuvent prendre des p mesurée dans des échantillons de taille n.
- Intervalle de confiance (pratique) : si dans un échantillon de taille n, on mesure une p, on peut prédire la valeur de π (sous forme de distribution probabilisée ou d'intervalle de confiance)

- Objet de ce chapitre :
  - Si dans *1* échantillon de taille n, on mesure une p, on peut tester si p est « significativement différent » de p_0, autrement dit si π est vraisemblablement différent de p_0
  - Procédé : on formule l'hypothèse que $\pi = p_0$, et on calcule sous cette hypothèse la « p valeur » : probabilité d'observer cette valeur observée de p ou une valeur de p plus éloignée encore de p_0.

Deux conventions pour les études en Santé

- Presque toujours en Santé (sauf tests de non-infériorité) :

- Tests bilatéraux :
  - On ne cherche pas à montrer que p est supérieur ou inférieur à p_0
  - On cherche à montrer que p est différent de p_0
  - Le premier cadre serait avantageux, et pas très honnête car on voit déjà dans quel sens est la différence

- Tests au risque alpha de 5%
  - C'est, de manière consensuelle, le seuil retenu en Santé
  - Une observation qui avait moins d'une chance sur 20 de se produire est considérée comme peu vraisemblable

Rappel sur la Loi Binomiale

- Rappels préalables sur les dénombrements :
- x éléments, combien de suites ordonnées ?

$$x * (x-1) * (x-2) * \cdots * 2 * 1 = x!$$

- n éléments, combien de suites ordonnées de x éléments ?

$$n * (n-1) * \cdots * (n-x+1)$$

$$= \frac{n * (n-1) * \cdots * (n-x+1) * (n-x) * \cdots * 2 * 1}{(n-x) * \cdots * 2 * 1}$$

$$= \frac{n!}{(n-x)!} = A_n^x$$

« arrangement de x éléments parmi n »

- n éléments, combien de mains non-ordonnées de x éléments ?

$$= \frac{A_n^x}{x!} = \frac{n!}{(n-x)!.\,x!} = C_n^x = \binom{n}{x}$$

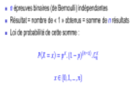
« combinaison de x éléments parmi n »

Loi Binomiale

- *n* épreuves binaires (de Bernoulli) indépendantes
- Résultat = nombre de « 1 » obtenus = somme de *n* résultats
- Loi de probabilité de cette somme :

$$P(X = x) = p^x.(1 - p)^{(n-x)}.C_n^x$$

$$x \in \{0, 1, \dots, n\}$$



Exemple concret avec une loi binomiale

- Exercice inspiré de https://pubmed.ncbi.nlm.nih.gov/22023719/

Multicenter Study > J Sex Med. 2012 Jul;9(7):1860-7. doi: 10.1111/j.1743-6109.2011.02512.x.
Epub 2011 Oct 24.

Sex with animals (SWA): behavioral characteristics and possible association with penile cancer. A multicenter study

Stênio de Cássio Zequi [1], Gustavo Cardoso Guimarães, Francisco Paulo da Fonseca, Ubirajara Ferreira, Wagner Eduardo de Matheus, Leonardo Oliveira Reis, Giuliano Amorim Aita, Sidney Glina, Victor Silvestre Soares Fanni, Marjo Denisson Cardenuto Perez, Luiz Renato Montez Guidoni, Valdemar Ortiz, Lucas Nogueira, Luis Carlos de Almeida Rocha, Gustavo Cuck, Walter Henriques da Costa, Ravendra Ryan Moniz, José Hipólito Dantas Jr, Fernando Augusto Soares, Ademar Lopes

Affiliations + expand

PMID: 22023719 DOI: 10.1111/j.1743-6109.2011.02512.x

Contexte de l'exercice

- Données <u>simplifiées inspirées</u> de l'article :
  - Au Brésil, 30% des hommes avouent avoir eu des relations sexuelles zoophiles
  - Sur 15 patients atteints de cancer du pénis, 9 (60%) avouent avoir eu des relations sexuelles zoophiles
  - Cette proportion observée (9/15) est-elle significativement différente de la proportion attendue (30%) ?
- Résolution avec le Test Binomial :
  - H_0 Hypothèse nulle : ces hommes présentent le même risque individuel que l'ensemble de la population $p_0=0.3$
  - Autrement dit, le hasard suffit à expliquer que p=9/15
  - Autrement dit, la plausibilité de l'observation « 9/15 » est suffisante
  - Soit « p » (petit p, p valeur) la plausibilité de cette observation en supposant que H_0 soit vraie

Résolution

- Pour chaque nombre *x* de patients zoophiles (parmi *n=15*), supposant sous H_0 que *π=0.3*, on peut calculer la probabilité d'observer exactement *x* :

$$P(x) = p^x \times (1-p)^{(n-x)} \times C_n^x$$

$$x \in \{0, 1, \dots, n\}$$

Et donc ici :

$$P(x) = 0.3^x \times 0.7^{(15-x)} \times C_{15}^x$$

$$x \in \{0, 1, \dots, 14, 15\}$$

Résolution

- Voici les probabilités associées à chaque possibilité :

nb zoophiles = x	P(X=x)
0	0.004748
1	0.030520
2	0.091560
3	0.170040
4	0.218623
5	0.206130
6	0.147236
7	0.081130
8	0.034770
9	0.011590
10	0.002980
11	0.000581
12	0.000083
13	0.000008
14	0.000001
15	0.000000
Somme	**1.00**

Valeur qu'on avait le plus de chance d'observer, sous H_0

Valeur qu'on a cependant observée

Attention : elle pourrait être associée à une probabilité faible simplement parce que le nombre de possibilités est élevé. La probabilité associée n'est pas un élément de preuve suffisant.

Résolution

- Voici les probabilités associées à chaque possibilité :

nb zoophiles = x	P(X=x)
0	0.004748
1	0.030520
2	0.091560
3	0.170040
4	0.218623
5	0.206130
6	0.147236
7	0.081130
8	0.034770
9	0.011590
10	0.002980
11	0.000581
12	0.000083
13	0.000008
14	0.000001
15	0.000000
Somme	1.00

Somme = 1,99%

p valeur = probabilité associée à notre observation + probabilités associées aux observations plus écartées encore de la valeur attendue

« Plus écartées » => associées à des probabilités encore plus faibles (test bilatéral)

<u>Interprétation :</u>
Si p < 5% (risque alpha choisi par convention) : observation trop peu probable sous H_0 => on rejette H_0 (au risque alpha : pas certain !)

Si p>5% : observation compatible avec H_0, ce qui ne prouve rien => indétermination

Résolution

- Conclusion :
  - En termes statistiques
    - => section « résultats » du mémoire
    - La proportion de zoophiles observée parmi les patients atteints de cancer du pénis (9/15) est significativement différente de la proportion attendue (30%), au risque alpha de 5%.
  - En termes médicaux
    - => section « discussion » du mémoire
    - Il est possible que la pratique de la zoophilie soit un facteur de risque de cancer du pénis

Le Test Binomial

- Enoncé du problème :
  - On s'appuie sur une probabilité attendue p_0 d'un caractère
  - Dans un échantillon de taille n, on observe x fois ce caractère, soit une proportion observée de x/n qui peut différer de p_0
- Questions (identiques) :
  - La différence observée entre p_0 et x/n dépasse-t-elle le simple effet du hasard (fluctuations aléatoires liées à l'échantillonnage) ?
  - La proportion observée x/n diffère-t-elle significativement de la proportion attendue p_0 ?
  - Notre échantillon est-il réellement issu d'une population caractérisée par $\pi = p_0$?
- Résolution par l'absurde :
  - Imaginons l'hypothèse nulle H_0 : « la probabilité du caractère est bien de p_0. L'échantillon est issu de cette population »
  - Calculons, sous l'hypothèse H_0, la <u>p valeur bilatérale</u> de l'observation, autrement dit la probabilité sous H_0 d'observer ce qu'on a observé OU d'observer encore moins probable. Ces probabilités sont calculées à l'aide d'une loi binomiale.
- Conclusion
  - Si la p valeur est inférieure à 5%, on rejette l'hypothèse H_0
  - Sinon, on ne peut rien conclure.
  - Ce 5% est par convention le seuil α (alpha) ou « risque de première espèce ».

Démarche générale des tests statistiques

Tests exacts (ex : test binomial)

Echantillon

Formulation de 2 hypothèses :

H_0 : hypothèse nulle, simple et permet de calculer la « p valeur »
(ex : µ1=µ2, µ=valeur, etc.)

H_1 : hypothèse alternative (tout le reste => calcul de proba impossible)

Supposant H_0 vraie, calcul de la p valeur (vraisemblance)

Si $p<\alpha$

Rejet H_0
Acceptation H_1

Si $p \geq \alpha$

Impossible de conclure

α=seuil, généralement 5%

Le plus souvent, calcul direct impossible

Echantillon

Formulation de 2 hypothèses :

H_0

H_1

Supposant H_0 vraie, calcul d'une « statistique de test »

A l'aide d'une table de loi, déduction d'une p valeur

Si $p<\alpha$

Rejet H_0
Acceptation H_1

Si $p \geq \alpha$

Impossible de conclure

Cette étape peut également être réalisée avec le raisonnement de « valeur critique »
= valeur de la stat de test correspondant exactement à p=5%…

Quelques tests d'adéquation

- Comparer une proportion observée à une proportion attendue :
  - Test binomial (test exact)
  - Test du Chi² d'adéquation (avec la Loi du Chi²)
- Comparer une moyenne observée à une moyenne attendue :
  - Test de Student (avec la Loi de Student)
- Comparer deux moyennes appariées « avant-après » ou « gauche-droite » :
  - Calculer tout d'abord la différence après-avant ou droite-gauche
  - La valeur moyenne attendue de cette différence est zéro => Test de Student, avec moyenne attendue = 0

Merci de votre attention !

En France, les sévices de nature sexuelle sur les animaux sont interdits par la loi du 9 mars 2004 et passibles de deux ans d'emprisonnement et 30 000 € d’amende.