Statistique pour Business et Machine learning
Partie II: Statistiques déductives

Tien-Nam Le

tien-nam.le@ens-lyon.fr

slides: http://perso.ens-lyon.fr/tien-nam.le/su

Sciences U University

18 Octobre, 2018

Plan

Partie II. Statistiques déductives

Échantillonnage

1. Théorème central limite
2. Proportion d'échantillonnage

Estimateurs ponctuels

1. Estimateurs sans biais
2. Estimation de la moyenne et de la variance

Intervalles de confiance

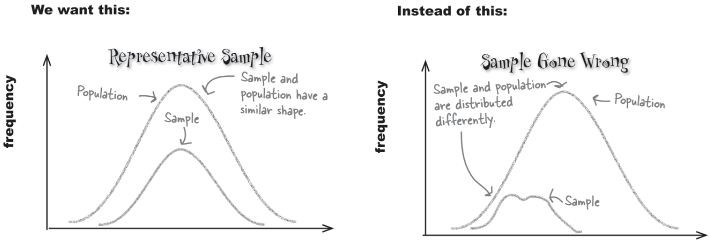
1. Estimation de la moyenne lorsque la variance est connue
2. Estimation de la moyenne lorsque la variance est inconnue

Tests d'hypothèses

1. niveau de signification, p-valeur
2. Proportion de la population

Échantillonnage

Échantillonnage



A partir de maintenant, un échantillon $X = \{X_1, ..., X_n\}$signifie toujours:

- Chaque X_i est indépendant des autres
- Chaque X_i suit au hasard la distribution de la population

Cette méthode d'échantillonnage s'est révélée bonne.

Rappel

Paramètres importants:

- Population
  - moyenne: μ
  - variance: σ^2
  - écarts types: σ

- Échantillon
  - moyenne: $\overline{X} = \frac{\sum_i X_i}{n}$
  - variance: $s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$
  - écarts types: $s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}}$

Moyenne de l'echantillon

Moyenne de l'echantillon $\overline{X}$ est une variable aléatoire avec

- Éspérence:

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{\sum_i X_i}{n}\right] = \frac{1}{n}\sum_i \mathbb{E}[X_i] = \frac{1}{n} n\mu = \mu$$

- Variance: $Var(\overline{X}) = \frac{\sigma^2}{n} \to 0$ when $n \to \infty$.
- écarts types: $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

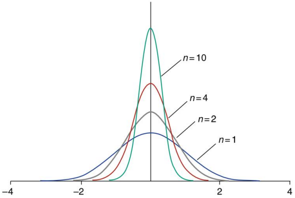


FIGURE 7.1
Densities of sample means from a standard normal population.

Note: La variance ci-dessus
est la variance de la moyenne de l'echantillon
($Var(\overline{X})$, un numéro), pas
variance de l'échantillon (s^2, une variable aléatoire).

Exercise: Le montant d'argent retiré lors de chaque transaction chez un guichet automatique d'une succursale de la Bank of America équivaut à 80 USD et à l'écart type à 40 USD. Quels sont la moyenne et l'écart type du montant moyen retiré lors des 20 prochaines transactions?

Théorème central limite

- **Théorème central limite:** quand n est grand ($n \geq 30$), $\overline{X}$ respecte la loi normale avec la moyenne μ et l'écart type $\sigma/\sqrt{n}$.

- Cela implique

$$\mathbb{P}(\overline{X} \leq a) = \mathbb{P}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{a - \mu}{\sigma/\sqrt{n}}\right) \approx \mathbb{P}\left(Z \leq \frac{a - \mu}{\sigma/\sqrt{n}}\right),$$

  où $Z \sim \mathcal{N}(0, 1)$.

- Similaire pour $\mathbb{P}(\overline{X} \geq a)$ et $\mathbb{P}(a \leq \overline{X} \geq b)$.

- Calculer la probabilité de Z à https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Exercise:
1. *Une compagnie d'assurance compte 10 000 assurés automobiles. Si la La réclamation annuelle prévue par titulaire de police est de 260 \$ avec un écart-type de 800 \$, approximativement la probabilité que la réclamation annuelle totale dépasse 2,8 millions de dollars.*

2. *Les taux de cholestérol sanguin d'une population de travailleurs ont une moyenne de 202 et un écart-type de 14. Si un échantillon de 36 travailleurs est sélectionné, approximez la probabilité que la moyenne de leur taux de cholestérol dans le sang soit comprise entre 198 et 206.*

Cas particulier: proportion d'échantillonnage

C'est un cas particulier très courant: lorsque chaque personne de la population choisit soit A ou B.

- la proportion de la population (i.e. pourcentage) choisissant A est p et choisissant B est $1 - p$ (p peut être inconnu).
- proportion d'échantillonnage $\overline{X}$: proportion d'un échantillon choisissant A.
- $\overline{X}$ respecte la loi binomiale avec l'éspérance p et la variance $\dfrac{p(1-p)}{n}$
- Et ox peut être approché par la loi normale comme dans le slide précédente.

Exercice: *supposons que 46% de la population envisage de voter pour le candidat A lors d'une prochaine élection. Si un échantillon aléatoire de taille 200 est choisi, alors*
a) quelle est la valeur attendue et l'écart type de la proportion de ceux qui, dans l'échantillon, sont en faveur du candidat A?
(b) quelle est la probabilité qu'au moins 100 favorisent le candidat A?

Exercises

1. Une agence de publicité a lancé une campagne pour introduire un produit. À la fin de sa campagne, il a affirmé qu'au moins 25% de tous les consommateurs connaissaient maintenant le produit. Pour vérifier cette affirmation, le producteur a échantillonné au hasard 1 000 consommateurs et a découvert que 232 d'entre eux connaissaient le produit. Si 25% de tous les consommateurs connaissaient réellement le produit, quelle est la probabilité que seulement 232 (c'est-à-dire 232 ou moins) sur un échantillon aléatoire de 1 000 consommateurs soient familiers?

2. Supposons que 12% des membres d'une population sont gauchers. Dans un échantillon aléatoire de 100 individus de cette population,
(a) Trouvez la moyenne et l'écart type du nombre de gauchers.
(b) Trouvez la probabilité que ce nombre soit compris entre 10 et 14 inclus.

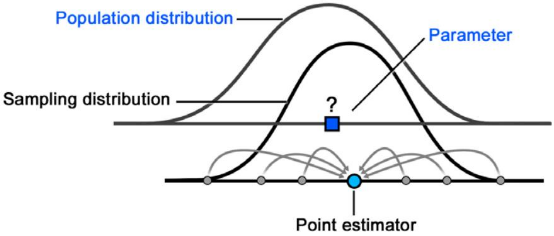
3. Si X est binomial avec les paramètres $n = 80$ et $p = 0.4$, approximez les probabilités suivantes.
(a) $\mathbb{P}(X > 34)$
(b) $\mathbb{P}(X \leq 42)$
(c) $\mathbb{P}(25 \leq X \leq 39)$

Estimateurs ponctuels

Estimateurs ponctuels

Objectif: approximativement une paramètre de la population, ex. la moyenne de la population μ, la variance de la population σ^2.

- *Estimateur* (*point estimator*):approximer un paramètre de la population θ par un paramètre de l'échantillon.
- Exemple: Pour estimer la moyenne de la population μ: Nous avons échantillon $X = \{X_1, ..., X_n\}$, nous pouvons choisir la moyenne de l'échantillon $\overline{X}$ comme estimateur ponctuels de μ.



Estimateurs ponctuels

Remarque: Il existe de nombreux estimateurs de points pour un paramètre de population.

Exemple: Pour estimer la moyenne de la population μ, on peut choisir:

- $\hat{\mu} = \overline{X}$, moyenne de la donnée
- $\hat{\mu} = \mathit{median}(X)$, médianne de la donnée
- $\hat{\mu} = \dfrac{\max(X) + \min(X)}{2}$

Note : Some estimators are good, some are bad.

Question: Comment choisir un bon paramètre?

Biais

- Un estimateur $\hat{\theta}$est une variable aléatoire. Sa valeur varie de échantillon à échantillonner.
- Le *biais* (*bias*) de l'estimateur $\hat{\theta}$ est

$$\textit{biais}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

- Un estimateur $\hat{\theta}$ est *sans biais* (*unbiased*) if $\mathbb{E}[\hat{\theta}] = \theta$, i.e. $\textit{biais}(\hat{\theta}) = 0$.
- L'écart type de $\hat{\theta}$, $\sigma(\hat{\theta})$ est généralement appelé l'*Erreur type* (*standard error*) of $\hat{\theta}$.
- Un estimateur sans biais $\hat{\theta}$ est *consistent* (*consistent*) si $\sigma(\hat{\theta}) \to 0$ (également $\textit{Var}(\hat{\theta}) \to 0$) quand la taille de l'échantillon $\to \infty$.

Exercises

$\overline{X}$ un estimateur sans biaisé de μ? Quelle est l'erreur type de $\overline{X}$? Est-ce que $\overline{X}$ est un estimateur consistent?

Une étude proposée pour estimer le taux moyen de cholestérol chez les adultes qui travaillent exige un échantillon de 1 000. Si nous voulons réduire l'erreur type résultante par un facteur de 4, quelle taille d'échantillon est nécessaire?

Une enquête est en cours de planification pour découvrir la proportion de la population favorable à un nouveau lien scolaire. Quelle est la taille d'un échantillon requis pour être certain que l'erreur type de l'estimateur résultant est inférieure ou égale à 0,1?

Estimation de la variance de la population σ^2

Objectif: Trouver un estimateur pour la variance de population σ^2. Cela dépend si vous connaissez la moyenne de la population μ ou pas.

- Si la moyenne de la population μ est *connue*, utilisez $\frac{\sum (X_i - \mu)^2}{n}$, qui est un estimateur sans biais.
- Si la moyenne de la population μ est *inconnue*, utilisez variance de l'échantillon $s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$, qui est un estimateur sans biais.
- Ne jamais utiliser $\frac{\sum (X_i - \overline{X})^2}{n}$, parce que c'est un estimateur biaisé.

Exercises

1. Les données suivantes se rapportent aux quantités (en tonnes) de produits chimiques produits quotidiennement dans une usine de produits chimiques.

$$776, 810, 790, 788, 822, 806, 795, 807, 812, 791$$

Utilisez-les pour estimer la moyenne et la variance de la production quotidienne.

2. La cohérence est d'une grande importance dans la fabrication des balles de baseball, car on ne veut pas que les balles soient trop vives ou trop mortes. Les balles sont testées en les laissant tomber à une hauteur standard, puis en mesurant la hauteur à laquelle elles rebondissent. Un échantillon de 30 boules a généré les statistiques récapitulatives suivantes:

$$\sum_{i}^{30} X_i = 52,1 \quad \text{et} \quad \sum_{i}^{30} X_i^2 = 136.2.$$

Estimez l'écart-type de la taille du rebond.

Intervalle de Confiance

Intervalle de confiance

- Au lieu d'estimer un paramètre par un point, nous utilisons plutôt un intervalle.
- L'estimateur d'intervalle typique de θ est $(\hat{\theta} - c, \hat{\theta} + c)$ pour certains c.
- L'intervalle a *confiance* (*confidence*) 95% si l'intervalle $(\hat{\theta} - c, \hat{\theta} + c)$ contient θ avec probabilité 95%.
  Remarque: Ça ne veut pas dire $(\theta - c, \theta + c)$ contains $\hat{\theta}$ avec probabilité 95%.
- La longueur de l'intervalle dans ce cas est $2c$
- c est choisi par le niveau de confiance (95% par exemple)
- Si on veut augmenter 95% (à 99% par exemple) alors c doit être augmenté, c'est-à-dire que l'intervalle est plus long.

Case 1. Estimez la moyenne μ lorsque la variance σ^2 est connue

Théorème: Supposons que la variance de population σ^2 soit connue (et $n \geq 30$ idéalement), puis

$$\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

est un 95% intervalle confiance de μ. En d'autres termes,

$$\mathbb{P}\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

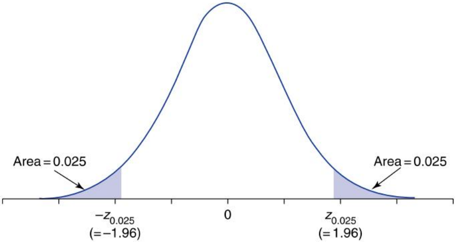


FIGURE 8.1

$P\{|Z| \leq 1.96\} = P\{-1.96 \leq Z \leq 1.96\} = 0.95.$

- 90% : 1.645 95% : 1.96 99% : 2.576

Case 1. Estimez la moyenne μ lorsque la variance σ^2 est connue

Exercice: Pour estimer μ, la teneur moyenne en nicotine d'une cigarette nouvellement commercialisée, 44 de ces cigarettes sont choisies au hasard et leur teneur en nicotine est déterminée. Supposons que l'expérience passée montre que l'écart type de la teneur en nicotine d'une cigarette est égal à 0,7 milligramme.
Si le résultat moyen en nicotine est de 1,74 milligrammes, quel est l'estimateur d'intervalle de confiance à 95% de μ?

- La longueur de l'intervalle de confiance de 95 % est $2 \times 1.96 \frac{\sigma}{\sqrt{n}}$
- Si nous voulons une longueur inférieure à b, alors n doit être au moins $\left(\frac{2 \times 1.96}{b}\right)^2$

Exercice (suite): Quelle est la taille d'un échantillon nécessaire pour que la longueur de l'intervalle de confiance de 95% soit inférieure ou égale à 0,3 milligramme?

Exercises

1. L'expérience a montré que le poids des saumons élevés dans une écloserie commerciale est normal, la moyenne variant d'une saison à l'autre mais l'écart-type restant fixé à 0,3 livre. Si nous voulons avoir la certitude à 90% que notre estimation du poids moyen d'un saumon est correcte à moins de ± 0.1 livres, quelle est la taille d'un échantillon? Et si nous voulons être sûrs à 99%?

2. On sait que l'écart type de la durée de vie d'un certain type d'ampoule est égal à 100 heures. Un échantillon de 169 ampoules de ce type avait une durée de vie moyenne de 1350 heures. Trouvez une estimation d'intervalle de confiance à 90% de la durée de vie moyenne de ce type de bulbe.

3. Une étude pilote a révélé que l'écart type des gains mensuels des travailleurs dans l'industrie chimique était de 180 dollars. Quelle taille d'échantillon faut-il choisir pour obtenir un estimateur du salaire moyen qui, avec 90% de confiance, sera correct à ± 20 \$?

Cas particulier: estimation de la proportion de la population

Rappelez la proportion d'échantillonnage: lorsque les gens choisissent entre deux options A ou B.

- La proportion de la population est p
- $\overline{X}$ a l'éspérance p et la variance $\dfrac{p(1-p)}{n}$

Un estimateur d'intervalle de confiance de 95 % de p est

$$\left(\overline{X} - 1.96\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \overline{X} + 1.96\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right)$$

Advantage: Dans ce cas, nous n'avons pas besoin de connaître la variance de la population σ.

Exercise: Sur un échantillon aléatoire de 100 étudiants dans une université, 82 ont déclaré qu'ils étaient non-fumeurs. Sur cette base, construisez une estimation d'intervalle de confiance à 99% de p, la proportion de tous les étudiants de l'université qui ne sont pas fumeurs.

Exercises

Exercise 1. *Un importateur de vin a la possibilité d'acquérir un important lot de vin Château Lafite Rothschild de 1947. En raison de l'âge du vin, certaines bouteilles se sont peut-être transformées en vinaigre. Cependant, le seul moyen de déterminer si une bouteille est toujours bonne consiste à l'ouvrir et à en boire. En conséquence, l'importateur a pris des dispositions avec le vendeur pour sélectionner et ouvrir au hasard 20 bouteilles. Supposons que 3 de ces bouteilles sont gâtées. Construisez une estimation d'intervalle de confiance à 95% de la proportion de l'envoi gâché.*

The number of necessary samples to ensure the length of the 95% interval at most b is

$$n > \left(\frac{1.96}{b}\right)^2$$

Exercise 2. *Quel est le plus petit nombre de certificats de décès que nous devons échantillonner au hasard pour estimer la proportion de la population américaine qui meurt du cancer, si nous voulons que l'estimation soit correcte à 0,01 avec 95% de confiance.*

Case 2. Estimez la moyenne μ lorsque la variance σ^2 est inconnue

- si la variance σ^2 de la population est inconnue, calculez la variance de l'échantillon $s^2 = \dfrac{\sum(X_i - \overline{X})^2}{n-1}$ à la place.
- Choisir $\alpha = 1 - \frac{95}{100} = 0.05$ pour confiance 95% (et $\alpha = 0.1$ pour 90%, etc)
- Rechercher de la valeur $t_{n-1,\alpha}$ sur la tableau de Student's t: http://www.sthda.com/french/wiki/table-de-student-ou-table-t
- L'intervalle de confiance est

$$\left(\overline{X} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}\right)$$

Exercice: Le Centre national des statistiques de l'éducation a récemment choisi un échantillon aléatoire de 2 000 étudiants nouvellement diplômés et a interrogé chacun d'entre eux sur le temps qu'il lui fallait pour obtenir son diplôme. Si la moyenne de l'échantillon était de 5,2 ans avec un écart type de 1,2 an, construisez une estimation d'intervalle de confiance à 95% de la durée moyenne d'achèvement des études de tous les nouveaux diplômés..

Exercises

1. Le responsable du service des expéditions d'une société de vente par correspondance située à New York a reçu des plaintes concernant le temps requis par les clients californiens pour recevoir leurs commandes. Pour en savoir plus sur ce problème potentiel, le responsable a choisi un échantillon aléatoire de 12 commandes, puis a vérifié le nombre de jours nécessaires pour recevoir chacune de ces commandes. Les données résultantes ont été

$$15, 20, 10, 11, 7, 12, 9, 14, 12, 8, 13, 16$$

Trouvez une estimation d'intervalle de confiance de 90% pour la durée moyenne.

2. Une grande entreprise assure elle-même son grand parc de voitures contre les collisions. Pour déterminer son coût moyen de réparation par collision, il a choisi au hasard un échantillon de 16 accidents. Si le coût moyen des réparations dans ces accidents est de 2200 USD avec un écart-type d'échantillon de 800 USD, trouvez une estimation de l'intervalle de confiance de 90% du coût moyen par collision.

3. pour déterminer la durée moyenne d'un appel téléphonique passé en milieu de journée, la compagnie de téléphone a sélectionné au hasard un échantillon de 1 200 appels de ce type. La moyenne d'échantillon de ces appels est de 4,7 minutes et l'écart type d'échantillon est de 2,2 minutes. Trouvez une estimation d'intervalle de confiance à 95% de la longueur moyenne de tous ces appels.

Test d'hypothèse

Test d'hypothèse

- Une *Hypothèse* (*hypothesis*) est un énoncé concernant un paramètre de population (tel que μ ou variable σ^2).
- L'*hypothèse nulle* (*null-hypothesis*) H_0. Ex. $H_0 : \mu \leq 1.5$
- L'*hypothèse alternative* (*alternative hypothesis*) H_1 est l'hypothèse complémentaire à l'hypothèse nulle. Ex. $H_1 : \mu > 1.5$
- Objectif: Utiliser les données pour tester pour accepter l'hypothèse nulle H_0 ou rejeter l'hypothèse nulle (c.-à-d. Accepter l'hypothèse alternative H_1)

Exercice: Une société pharmaceutique britannique, Glaxo Holdings, a récemment mis au point un nouveau médicament contre les migraines. Parmi les affirmations de Glaxo concernant son médicament appelé somatriptan, il a été indiqué que le temps moyen nécessaire pour qu'il pénètre dans le sang est inférieur à 10 minutes. Pour convaincre la Food and Drug Administration de la validité de cette affirmation, Glaxo a mené une expérience sur un groupe de personnes souffrant de migraine choisi au hasard. Pour prouver l'affirmation de la société, qu'aurait dû prendre Glaxo comme hypothèse nulle et comme hypothèse alternative?

Critère: Accepter H_0 95 % des fois

Niveau de signification (*significance-level*) α: La probabilité H_0 est rejetée. Par exemple: $\alpha = 0.05$ signifie 95% accepté H_0 (similaire à un intervalle de confiance de 95%).

- Hypothèses: $H_0 : \mu = \mu_0$ et $H_1 : \mu \neq \mu_0$ pour une valeur μ_0.
- Population variance σ^2 est connue
- niveau de signification: $\alpha = 0.05$
- Rejeter H_0 si
$$\frac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| \geq z_{\alpha/2} = 1.96$$
- Accepter H_0 autrement.

Exercise:Supposons que l'intensité du signal reçu soit normalement répartie avec la moyenne μ et l'écart type 4. On suppose que μ est égal à 10. Vérifier si cette hypothèse est plausible si le même signal est indépendamment reçu 20 fois et la moyenne des 20 valeurs reçues est de 11,6. Utilisez le 5 % niveau de signification.

Exercises

1. Une société de crédit-bail part de l'hypothèse que le nombre annuel de miles parcourus dans ses voitures louées a une moyenne de 13 500 et un écart type de 4 000 miles. Pour voir si cette hypothèse est valide, un échantillon aléatoire de 36 voitures d'un an a été vérifié. Quelle conclusion pouvez-vous tirer si le kilométrage moyen de ces 36 voitures est de 15 233?

2. Les autorités routières affirment que les feux de signalisation sont rouges pendant une durée moyenne de 30 secondes et que l'écart type est de 1,4 seconde. Pour tester cette affirmation, un échantillon de 40 feux de signalisation a été vérifié. Si le temps moyen des 40 feux rouges observés était de 32,2 secondes, pouvons-nous conclure, au seuil de signification de 5%, que les autorités sont inexactes? Qu'en est-il du niveau de signification de 1%?

Trouver le bon niveau de signification

Rappel:

- Rejeter H_0 si

$$\frac{\sqrt{n}}{\sigma}\left|\overline{X}-\mu_0\right| \geq z_{\alpha/2}$$

- Accepter H_0 autrement.

En pratique, le niveau de signification n'est souvent pas défini à l'avance.

p-valeur (*p-value*) p = niveau de significance α où obtenir la ligne de démarcation accepter-rejeter:

$$\frac{\sqrt{n}}{\sigma}\left|\overline{X}-\mu_0\right| = z_{p/2}$$

Calculer la probabilité de Z à https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Exercice: supposons que la moyenne des 20 valeurs de l'exemple précédent soit égale à 10,8. Quelle est la valeur de p pour l'hypothèse $\mu = 10$? (Rappelons que $\sigma = 4$)

Exercises

1. Les données historiques indiquent que l'utilisation de l'eau par les ménages a tendance à être normalement répartie avec une moyenne de 360 gallons et un écart-type de 40 gallons par jour. Pour voir si cela est toujours le cas, un échantillon aléatoire de 200 ménages a été choisi. On a alors constaté que la consommation moyenne d'eau par jour dans ces ménages était égale à 374 gallons par jour.
(a) Ces données sont-elles cohérentes avec la distribution historique? Utilisez le niveau de signification de 5%.
(b) Quelle est la p-valeur?

2. Pour tester l'hypothèse $H_0 : \mu = 105$ contre $H_1 : \mu \neq 105$ un échantillon de taille 9 est choisi. Si la moyenne de l'échantillon est $\overline{X} = 100$, recherchez la valeur p si l'écart-type de la population est connu comme étant.
(a) $\sigma = 5$
(b) $\sigma = 10$
Dans quels cas l'hypothèse nulle serait-elle rejetée au seuil de signification de 5%?

Proportion de la population

- Rappel: p est la proportion de la population favorable à A sur B.
- Supposons que l'hypothèse nulle $H_0 : p \leq p_0$ et $H_1 : p > p_0$ pour une valeur p_0.
- Donc p-value $= \mathbb{P}\Big(Binom(n, p_0) \geq \overline{X}\Big)$
- Calculer $Binom(n, p_0)$ à https://stattrek.com/online-calculator/binomial.aspx

Exercise: Un éducateur réputé affirme que plus de la moitié de la population adulte américaine est préoccupée par le manque de programmes éducatifs diffusés à la télévision. Pour recueillir des données sur cette question, un service de vote national a choisi et interrogé au hasard 920 personnes. Si 478 (52%) des personnes interrogées se disent préoccupées par le manque d'émissions éducatives à la télévision, cela prouve-t-il la prétention de l'éducateur?

Exercises

1. Un fabricant de puces informatiques affirme qu'au moins 2% des puces produites sont défectueuses. Une entreprise d'électronique, impressionnée par cette affirmation, a acheté une grande quantité de puces. Pour déterminer si la réclamation du fabricant est plausible, la société a décidé de tester un échantillon de 400 de ces puces. S'il y a 13 puces défectueuses (3,25%) sur ces 400, est-ce que cela réfute (au niveau d'importance de 5%) la revendication du fabricant?

2. Un médicament standard est connu pour être efficace dans 72% des cas dans lesquels il est utilisé pour traiter une infection donnée. Un nouveau médicament a été mis au point et les tests ont montré qu'il était efficace dans 42 cas sur 50. Cette preuve est-elle suffisamment solide pour prouver que le nouveau médicament est plus efficace que l'ancien? Trouvez la p-valeur pertinente.

3. Un économiste pense qu'au moins 60% des immigrants récemment arrivés qui travaillent dans le secteur de la santé aux États-Unis depuis plus d'un an ont le sentiment d'être sous-employés en ce qui concerne leur formation. Supposons qu'un échantillon aléatoire de taille 450 indique que 294 personnes (65,3%) ont le sentiment d'être sous-employées. Ces preuves sont-elles suffisamment solides, au seuil de signification de 5%, pour prouver que l'économiste a raison? Qu'en est-il du seuil de signification de 1%.