GM – Probabilités et Statistique

http://moodle.epfl.ch/course/view.php?id=18431

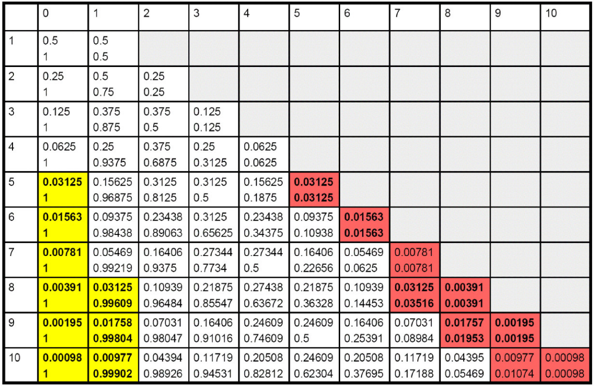
Cours 8

- Tests d'hypothèses – Introduction
- Test d'hypothèses pour une moyenne
- Test d'hypothèses pour une proportion
- (Test d'hypothèses pour 2 moyennes / proportions *indépendantes* – vidéo seulemet)

Tests d'hypothèses – Introduction

- Prenons l'expérience de lancer une pièce
- En particulier, examinons la question de savoir si la pièce est *équilibrée* (même probabilité de montrer ‘pile’ ou ‘face’)
- Formellement : on veut faire l'*inférence* au sujet de $P(\text{pile})$
- **Prenez une pièce de monnaie**, lancer plusieurs fois (soit 10 fois)
- En supposant qu'elle est équilibrée ($P(\text{pile}){=}0.5$), on essaie de déterminer si cette hypothèse est *compatible* avec les observations

10 jets

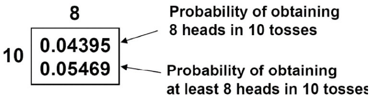


Number of heads

Number of tosses

	0	1	2	3	4	5	6	7	8	9	10
1	0.5 1	0.5 0.5									
2	0.25 1	0.5 0.75	0.25 0.25								
3	0.125 1	0.375 0.875	0.375 0.5	0.125 0.125							
4	0.0625 1	0.25 0.9375	0.375 0.6875	0.25 0.3125	0.0625 0.0625						
5	**0.03125** **1**	0.15625 0.96875	0.3125 0.8125	0.3125 0.5	0.15625 0.1875	**0.03125** **0.03125**					
6	**0.01563** **1**	0.09375 0.98438	0.23438 0.89063	0.3125 0.65625	0.23438 0.34375	0.09375 0.10938	**0.01563** **0.01563**				
7	**0.00781** **1**	0.05469 0.99219	0.16406 0.9375	0.27344 0.7734	0.27344 0.5	0.16406 0.22656	0.05469 0.0625	0.00781 0.00781			
8	**0.00391** **1**	**0.03125** **0.99609**	0.10939 0.96484	0.21875 0.85547	0.27438 0.63672	0.21875 0.36328	0.10939 0.14453	**0.03125** **0.03516**	**0.00391** **0.00391**		
9	**0.00195** **1**	**0.01758** **0.99804**	0.07031 0.98047	0.16406 0.91016	0.24609 0.74609	0.24609 0.5	0.16406 0.25391	0.07031 0.08984	**0.01757** **0.01953**	**0.00195** **0.00195**	
10	**0.00098** **1**	**0.00977** **0.99902**	0.04394 0.98926	0.11719 0.94531	0.20508 0.82812	0.24609 0.62304	0.20508 0.37695	0.11719 0.17188	0.04395 0.05469	0.00977 0.01074	0.00098 0.00098

🟥🟨 **Significant evidence ($p<0.05$) that the coin is biased towards head or tail.**



8

10

0.04395
0.05469

Probability of obtaining 8 heads in 10 tosses

Probability of obtaining at least 8 heads in 10 tosses

Test d'hypothèses

Définition : Une *hypothèse* est *une déclaration au sujet d'un* **paramètre** de la population

- Deux **hypothèses** concurrentes
  - H (ou H_0) : l'*hypothèse NULLE*, généralement plus conservatrice
  - A (ou H_A) : l'*hypothèse ALTERNATIVE*, généralement plus intéressante
- Exemples de l'hypothèse NULLE :
  - La pièce est équilibrée
  - Une nouvelle machine n'est pas mieux (ou pire) que la machine actuelle
- Exemples de l'hypothèse ALTERNATIVE :
  - La pièce est biaisée (soit vers pile ou face)
  - La pièce est biaisée (pile est favorisé)
  - La nouvelle machine est plus efficace que l'ancienne

Statistique de test

- Afin de prendre une décision, il faut évaluer dans quelle mesure la valeur observée est de ce qu'on attend si l'hypothèse NULLE H est vraie – c.-à-d., on a besoin d'une **statistique de test** (ST) T
- La statistique de test T est choisie de manière que les valeurs 'aberrantes' (trop grande/petite) suggèrent que l'hypothèse NULLE H est fausse
- On calcul T à partir de l'échantillon, la valeur observée est notée t_{obs}

Exemple

Exemple 8.1 Pour un EAS de 25 fermes dans un comté particulier, l'effet de la pulvérisation de pesticides a été évaluée en mesurant les rendements des cultures (boisseaux à l'acre) sur des bandes traitées et non traitées dans un champ dans chaque ferme.

Les données :

moyenne des diffs de l'échantillon $\overline{x}= 4.7$ boisseaux à l'acre
l'écart-type des diffs de l'échantillon : $s = 6.5$ boisseaux à l'acre

Supposons qu'un gain de 2 boisseaux à l'acre devrait payer les coûts de la pulvérisation. Est-ce qu l'échantillon fournit des preuves solides que la pulvérisation est rentable *? ?*

Alors, on va concrétiser le raisonnement pour aborder cette question.

Étapes d'un test d'hypothèses (I)

1. **Identifier** le paramètre de la population
   - Ici, le paramètre testé est $\mu =$ la moyenne différence de rendement dans la population
2. **Formuler** les hypothèses NULLE et ALT
   - $H : \mu = 2$ (ou $\mu \leq 2$)
     $A : \mu > 2$
3. Calculer la **statistique de test (ST)**
   - $t_{obs} = (4.7 - 2)/(6.5/\sqrt{25}) = 2.08$

Hypothèse vérité vs décision

Decision Truth	not rejected	rejected
true H	☺ specificity	✗ Type I error (False +) α
false H	✗ Type II error (False -) β	☺ Power 1 - β; sensitivity

Un peu de terminologie

- La probabilité de rejeter l'hypothèse NULLE qui est réellement *vraie* est égale à α ; ce type d'erreur est appelé *erreur de type I* ou *faux positif* (un résultat à un test déclaré positif *à tort*)
- La probabilité de *NE PAS* rejeter l'hypothèse NULLE qui est réellement *fausse* est égale à β ; ce type d'erreur est appelé *erreur de type II* ou *faux négatif* (un résultat à un test déclaré negatif *à tort*)
- La probabilité de rejeter l'hypothese NULLE correctement (c.-à-d., quand elle est vraiment fausse) est appelée la *puissance*

p-valeur

- On décide de *rejetter* (ou pas) l'hypothèse NULLE H, selon la probabilité d'obtenir une valeur de T *au moins aussi extrême* (aussi loin de ce qu'il serait au hasard ou encore plus loin, dans le sens de l'ALT) que la valeur réellement observée, *en supposant que la nulle est vraie*
- Cette probabilité est appelée la *valeur (ou niveau) de signification observée*, ou *p-valeur* p_{obs}
- Plus la valeur de p_{obs} est petite, plus de doute que l'hypothèse NULLE H est vraie
- Un résultat dont la p-valeur est inférieure à un *seuil* (ou *niveau*) de signification α est dit 'statistiquement significatif' à ce niveau
- **À noter** : signification statistique $\neq$ signification pratique $\neq$ signification scientifique

p-value – interprétation

- Tout comme dans l'interprétation d'un IC, l'interprétation d'une p-valeur est un peu délicat
- En particulier, la p-valeur ne nous dit **pas** la probabilité que l'hypothèse NULLE est vraie
- La p-valeur représente la probabilité qu'on obtiendrait une différence (entre l'observée et l'espérance sous l'hypothèse nulle) aussi grande que nous avons obtenue (ou plus) **S'IL N'Y AVAIT VRAIMENT PAS** d'autre effet que la variabilité aléatoire
- (Interprétation frequenciste de probabilité)

Étapes d'un test d'hypothèses (II)

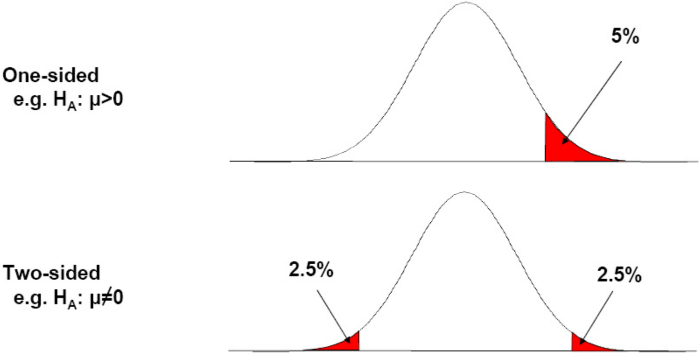
4. Calculer la ***p*-valeur**
   Ici, $p_{obs} = P(Z > 2.08) = 0.02$
5. (En option) *Règle de décision* : on REJETTE l'hypothèse NULLE H si $p_{obs} \leq \alpha$
   (Il s'agit d'un type d'argument par l'absurde)

- Une valeur typique de α est de 0.05, en raison principalement des raisons historiques
- Dans la pratique, vous devez choisir une valeur de α qui est appropriée à la situation
- Ici, si nous utilisons $\alpha = 0.05$, la décision sera **REJETTE *H***
- Au contraire, si nous utilisons $\alpha = 0.01$, la décision est **NE REJETTE PAS *H***

Test unilatéral vs. bilatéral

- La choix de l'hypothèse alternative influence la conclusion
- Si l'ALT est "la pièce est biaisée", on ne précise pas le sens du biais
- Il s'agit d'un *test bilatéral*
- Si α est p.ex. 0.05, cela signifie que nous devons permettre à $\alpha/2$ (0.025) pour le biais vers 'pile' et $\alpha/2$ (0.025) pour le biais vers 'face'
- Si l'ALT est "la pièce est orientée vers pile", nous précisons le sens de la partialité et le test est *unilatéral*

Test unilatéral vs. bilatéral



Exemple

Exemple 8.2 De longue expérience avec un processus de fabrication d'un boisson alcoolisée, il est connu que la distribution des rendements a une moyenne de 500 unités et un écart-type de 100 unités.

Une modification du processus est proposée pour laquelle c'est revendiqué que le rendement moyen va augmenter (l'écart-type ne changera pas). Avec un échantillon aléatoire de 49 observations, un rendement moyen de 535 unités est observée.

Tester cette assertion en utilisant un test d'hypothèses. Que concluez-vous (utilisez $\alpha = 0.05$) ?

Test

1.
2.
3.
4.
5.

Autre exemple

Exemple 8.3 Le rendement moyen du maïs aux États-Unis est d'environ 120 boisseaux à l'acre. Un échantillon aléatoire de 50 agriculteurs de l'Illinois (un état des États-Unis) a produit une moyenne de $\bar{x} = 123{,}6$ boisseaux à l'acre. Supposons que l'ES du rendement de cette population est $\sigma = 10$ boisseaux à l'acre.

Tester si la moyenne du rendement μ de l'Illinois est égale à la moyenne nationale. (Poser vos suppositions.)

Test

1.
2.
3.
4.
5.

PAUSE

Test d'hypothèses pour une proportion p

- On a considéré le *z-test* pour la moyenne μ d'une population
- Ce test est basé sur le TCL
- Statistique de test pour moyenne μ : $\boxed{T = \dfrac{\overline{X} - \mu_H}{S/\sqrt{n}}}$
- On peut également tester *la proportion p d'une population* qui possède un caractère particulier
- Le TCL s'applique pour la proportion observée $\hat{p} = X/n$
- On a vu comment trouver l'écart-type d'une proportion
- Donc, la statistique de test pour le test d'une proportion p est donnée par : $\boxed{T = \dfrac{\hat{p} - p_H}{\sqrt{p_H \cdot (1 - p_H)/n}}}$

Exemple : Test d'hypothèses pour une proportion

Exemple 8.4 Une nouvelle intervention chirurgicale a été développée pour corriger des malformations cardiaques chez les nourrissons avant l'âge d'un mois. Auparavant, la procédure a été utilisée sur les nourrissons qui ont été de plus d'un mois et le taux de réussite était de 91%.

Une étude est menée afin de déterminer si le taux de réussite de la nouvelle procédure est plus de 91% : $n = 200$, dont 187 succès ($=$ 93,5%).

Est la procédure plus efficace pour les nourrissons avant l'âge d'un mois (que pour les plus agés) **? ?**

Test

1
2
3
4
5

Autre exemple

Exemple 8.5 La rédactrice en chef pense que 50% des jeunes mariées sont plus jeunes que leurs mariés.

Elle prend un échantillon aléatoire de 100 mariées dont 53 sont plus jeunes que leurs mariés.

Effectuer un test d'hypothèses afin de déterminer si cette pourcentage est compatible avec 50% ou non. Utiliser $\alpha = 1\%$.

Test

1.
2.
3.
4.
5.

Erreur Standard d'une différence

- Variance de la différence entre deux moyennes (indépendantes) :

- Variance de la différence entre deux proportions (indépendantes) :

Test de comparaison sur 2 échantillons indépendants

- Ci-dessus, nous avons été intéressés par *une seule population*. Souvent, cependant, nous nous intéressons à la **comparaison de deux populations**. Dans ce cas, nous réalisons un test sur *deux échantillons indépendants*.
- Quand on compare deux *moyennes* (ou *proportions*) la notion de base est la même que ci-dessus : pour T, on utilise la *différence standardisée* entre les moyennes des échantillons (ou proportions).
- ST pour la différence des *moyennes* de deux populations indépendantes : $\frac{\overline{X_1} - \overline{X_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$
  (utiliser s au lieu de σ si σ est inconnu)
- ST pour la différence des *proportions* de deux populations indépendantes : $\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}}$

Exemple

Exemple 8.6 Le montant d'un certain élément trace dans le sang varie avec un écart-type de 16 ppm (parties par million) pour les hommes et 10 ppm pour les femmes. Des échantillons aléatoires de 64 hommes et 20 femmes donnent des concentrations de 28 et 33 ppm, respectivement.

Tester si les moyennes de concentrations de l'élément sont les mêmes pour les hommes et les femmes (utiliser $\alpha = 0{,}05$).

Test

1.
2.
3.
4.
5.

Exemple

Exemple 8.7 Lors d'une étude d'une certaine grande université, un chercheur s'intéresse à la théorie que les femmes font leurs études dans le domaine d'ingénierie à un taux plus bas que celui des hommes. Un échantillon aléatoire de 100 hommes a trouvé 28 (28%) faisaient des études d'ingénieur, alors qu'un échantillon aléatoire de 64 femmes a trouvé 12 (18,75%) qui les faisaient.

Tester sa théorie que les femmes étudient l'ingénierie moins frèquemment que les hommes (utiliser $\alpha = 0.05$).

Test

1.
2.
3.
4.
5.

Test de comparaison de 2 proportions, bis

- L'erreur standard (ES) pour la différence de deux proportions (de deux échantillons indépendants) égale :

$$ES = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- Toutefois, l'hypothèse nulle spécifie *l'égalité des proportions*, de sorte que nous pourrions utiliser une estimation basée sur la *proportion globale* (en combinant les échantillons ainsi) :

$$p_{\text{agrégé}} = \frac{x_1 + x_2}{n_1 + n_2},$$

  où $x_i =$ nombre de 'succès' de population i ($i = 1, 2$) et n_i sont les tailles des échantillons

Test de comparaison de 2 proportions, bis (cont)

- Donc, **sous l'hypothèse nulle H :**

$$ES_{ag} = \sqrt{\frac{\hat{p}_{ag}(1-\hat{p}_{ag})}{n_1} + \frac{\hat{p}_{ag}(1-\hat{p}_{ag})}{n_2}}$$

$$= \sqrt{\hat{p}_{ag}(1-\hat{p}_{ag})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- (Pour moi, chacune des méthodes est autorisée : sous l'hypothèse NULLE, la distribution de T sera correcte de toute façon.)
- ST pour la différence des *proportions* de deux populations indépendantes :

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{ag}(1-\hat{p}_{ag})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Relation entre test et IC

- Dans un essai, l'intervalle de confiance visualise la précision avec laquelle l'effet du traitement est connu
- Lorsque l'IC contient la valeur caractéristique de l'hypothèse nulle, il n'est pas possible d'exclure la possibilité que *la vraie valeur soit cette valeur nulle*
- Ainsi : la différence observée *ne peut pas être considérée comme statistiquement significative*
- À l'inverse, si un test est significatif au seuil de α, c'est-à-dire que l'IC à $100(1-\alpha)\%$ *ne contient pas la valeur de l'hypoth'ese nulle*
- Résumé :
    - test *significatif* (seuil α) $\Leftrightarrow$ $100(1-\alpha)\%$ IC *ne contient pas* la valeur nulle
    - test *non significatif* (seuil α) $\Leftrightarrow$ $100(1-\alpha)\%$ IC *contient* la valeur nulle

À propos des échantillons petits...

- Le z-test que nous avons étudié suppose que la distribution d'échantillonnage de la statistique de test T est *normale*
    - soit exactement
    - soit approximativement, par le TCL
- Toutefois, si l'écart-type de la population σ est *inconnu* et la taille de l'échantillon est *faible* (moins de 30, p. ex.) puis la vraie distribution d'échantillonnage de T possède des *queues qui s'étendent plus loin* que la distribution normale
- Dans ce cas, il faut utiliser le *t-test* (la semaine prochaine)

Les pièges des tests d'hypothèses

Faire attention :

- Difficultés d'interprétation des tests sur des *échantillons arbitraires* (pas aléatoires) et les données observationnelles (pas celles d'une expérience)
    - dans la pratique, la plupart des échantillons ne sont pas exactement aléatoires
    - la p-valeur d'un tel échantillon devrait être considéré comme *indicatrice approximative* de l'importance
- Signification statistique vs. signification pratique
    - une petite p-valeur peut provenir d'un écart très faible de l'hypothèse nulle si la taille de l'échantillon est très grande
- Risques de *recherche* de signification (plusieurs tests)
- Ignorer *l'absence* de signification