Psy1004 – Section 4:

Statistiques inductives

Plan du cours:
- Varia
- 4.0: Idée générale
- 4.1: Structure d'un test
  - a) les hypothèses
  - b) le seuil
  - c) le test statistique
  - d) appliquer le test et conclure
- 4.2: Tests bicaudal vs. unicaudal
- 4.3: Intervalle de confiance
- 4.4: Le travail du statisticien
- 4.5: Test binomial sur une proportion
- 4.6: Test binomial sur une proportion utilisant l'approximation normale
- 4.7: Test des signes
- 4.8: Test de la médiane
- 4.9: Exemples

Disponible sur: http://mapageweb.umontreal.ca/cousined/home/course/PSY1004

Varia

- Aujourd'hui:
  - Remise du TP1,
  - Énoncé du TP2 disponible sur le site web
- Autre chose?
- Questions sur le cours 3: Probabilités

4.0: Idée générale (1/3)

Vos intuitions

Le vrai monde

Votre hypothèse formalisée

H$_0$: $\mu_{stress} = 100$
H$_1$: $\mu_{stress} > 100$

où μ est la consommation d'alcool, et 100 (disons), la consommation de base d'alcool.

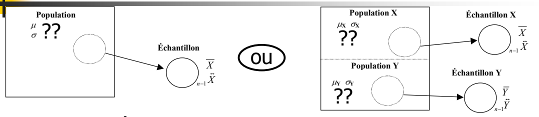
Décider des conditions d'observation
Collecter un échantillon représentatif

Le monde contrôlé

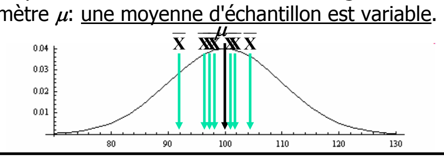
Décision à savoir si l'hypothèse formelle est rejetée ou pas

Test statistique

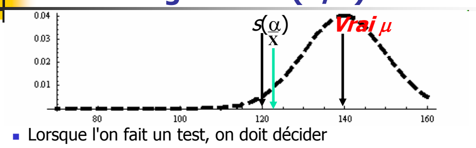
4.0: Idée générale (2/3)



- Les paramètres d'une population sont rarement connus…
- Les statistiques d'un échantillon sont connues.
- Jusqu'à quel point peut-on inférer l'un à partir de l'autre? une moyenne est rarement **exactement** égale au paramètre μ: <u>une moyenne d'échantillon est variable</u>.



4.0: Idée générale (3/3)



- Lorsque l'on fait un test, on doit décider
  si $\bar{X}$ confirme que μ est de 100, et jusqu'à quel point.
- La façon de procéder est de se donner un critère de décision, une valeur critique $s(\alpha)$:
  si $\bar{X}$ excède $s(\alpha)$, on dira que $\bar{X}$ est incompatible avec l'idée μ = 100

<u>Décision</u>	Dans la réalité, l'hypothèse est:	
<u>On peut conclure:</u>	vraie (i.e. μ = 100)	fausse (i.e. $\mu \neq$ 100)
que l'hypothèse n'est pas fausse quand:	$\bar{X}$ est en deçà de $s(\alpha)$	(ERREUR β!)
que l'hypothèse est fausse quand:	(ERREUR α!)	$\bar{X}$ est au delà de $s(\alpha)$

4.1: Structure d'un test
a) les hypothèses (1/1)

- Poser les hypothèses statistiques H_0 (hypothèse nulle) et H_1 (hypothèse alternative).
  - L'hypothèse nulle est toujours très précise, alors que
  - L'hypothèse alternative inclus une infinité de cas possibles.
  - Les hypothèses portent toujours sur la population, sur des paramètres de la population qui sont inconnus.

  - Ex: H_0: μ = 100
    H_1: μ > 100 ← une infinité de cas

  - Ex: H_0: σ = 10 (l'écart type d'une population)
    H_1: σ > 10 ← une infinité de cas

  - Ex: H_0: asymétrie = 0 (L'asymétrie d'une population)
    H_1: asymétrie > 0 ← une infinité de cas

Structure d'un test
et
Test Z

Disponible sur: http://mapageweb.umontreal.ca/cousined/home/course/PSY1004

4.1: Structure d'un test
b) le seuil (1/2)

- Comment choisir s(α) de telle façon que l'on ne commettra *pas* l'erreur α *trop souvent*?
- Qu'entend-t-on par: "*pas … trop souvent*"?
- Exemples de seuils (présume que H_0 est vraie ET normal):

$\overline{\mathbf{X}} - \mu > 1SE_{\overline{\mathbf{X}}}$ peut se produire 31% du temps par hasard: trop fréquent!

$\overline{\mathbf{X}} - \mu > 2SE_{\overline{\mathbf{X}}}$ peut se produire 4.6% du temps par hasard: rarement (mieux).

(vérifiez ces pourcentages)

- La valeur "1" ou "2" ci-haut est la *valeur critique*, exprimée en unité d'erreur type.

4.1: Structure d'un test
b) le seuil (2/2)

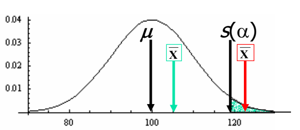
- 4.6% n'est pas un chiffre rond, on préfère 5% (1%, 10%) Trouver la *valeur critique* $s(\alpha)$ telle que:

$$\overline{\mathbf{X}} - \mu > s(\alpha) SE_{\overline{\mathbf{X}}}$$

se produise par hasard seulement 5% du temps (si H_0 est vraie ET normale) (doit être presque 2).

- Réponse: $s(\alpha)$ doit être placé à 1.64 erreur type de μ.

(ici, erreur type = 10)



4.1: Structure d'un test
c) le test statistique

- La décision se fait suivant la condition:

rejet de H_0 si $\overline{\mathbf{X}} - \mu > s(\alpha) SE_{\overline{\mathbf{X}}}$

- ou encore, si on reformule la condition:

rejet de H_0 si $\dfrac{\overline{\mathbf{X}} - \mu}{SE_{\overline{\mathbf{X}}}} > s(\alpha)$

- Un grand nombre de tests sont de cette forme:
  La distance entre l'hypothèse et l'observé, mesuré en terme d'erreur type.

4.1: Structure d'un test
d) appliquer le test et conclure

- Exemple: Soit un échantillon de 25 sujets qui a une moyenne de 106, et un écart type de 10, on peut conclure
  - si les hypothèses H_0 : μ = 100, H1 : μ > 100, et le seuil de 5%
  - (et si la population est normale), que…
  - l'hypothèse est … {rejetée/non rejetée} ?

{l'erreur type est de 2 (i.e. 10/√25), et donc la distance à μ, de 3}

- Interprétation:
  "La moyenne observée dans notre échantillon diffère <u>significativement</u> de 100 (<u>z</u> = 3, <u>p</u> < .05)"

Le résultat trouvé

La table utilisée par le test

Le seuil α; l'alphabet grec n'existait pas sur les vieilles dactylo…

On n'utilise pas les symboles H_0 et H_1 dans un rapport scientifique.

Un test a été effectué

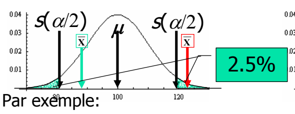
4.2: Tests bicaudal vs. unicaudal

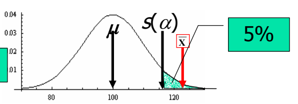
Dans certains cas, nos hypothèses de recherches limitent les hypothèses statistiques, par ex:

H_0: μ = 100	H_0: μ = 100
H_1: $\mu \neq$ 100	H_1: μ > 100

Dans ces cas, on peut utiliser un test plus précis, ayant

2 valeurs critiques	1 valeur critique
bicaudal/bidirectionnel, | |	unicaudal/unidirectionnel





Par exemple:

si α = 10%:	s(α/2) = ?	s(α) = ?
si α = 5%:	s(α/2) = 1.96	s(α) = ?
si α = 1%:	s(α/2) = ?	s(α) = ?

4.3: Intervalle de confiance

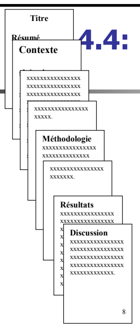
- Un intervalle dans lequel μ, si pas exactement 100, doit se trouver, <u>selon toutes probabilités</u>. Qu'entend-t-on par "selon toutes probabilités"?
  - Très souvent; En général, on prend $1 - \alpha$, par exemple, 95%.

  <u>Il s'agit de la même chose que le test statistique!</u>
- Par exemple, si α = 5% (et normal), $s(\alpha/2)$ = 1.96.
  - non rejet de H_0 si $| \overline{\mathbf{X}} - \mu | > s(\alpha/2)\, SE_{\overline{\mathbf{X}}}$

  → μ se trouve à ± 1.96 $SE_{\overline{\mathbf{X}}}$ de $\overline{\mathbf{X}}$, si μ n'est pas exactement 100

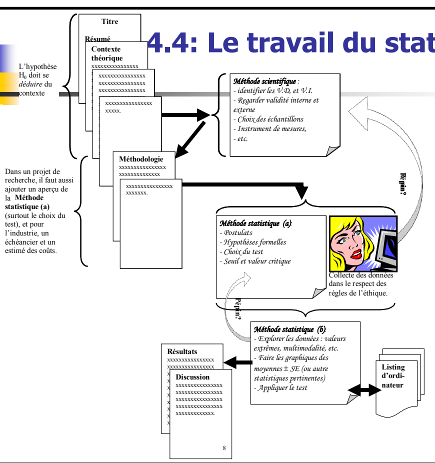
L'intervalle de confiance à 95%	L'intervalle de confiance à 1 - α
$\left[\overline{\mathbf{X}} - 1.96\, SE_{\overline{\mathbf{X}}} , \overline{\mathbf{X}} + 1.96\, SE_{\overline{\mathbf{X}}}\right]$ contient μ dans 95% des cas (i.e. 19 fois sur 20).	$\left[\overline{\mathbf{X}} - s(\alpha/2)\, SE_{\overline{\mathbf{X}}} , \overline{\mathbf{X}} + s(\alpha/2)\, SE_{\overline{\mathbf{X}}}\right]$ contient μ dans 1-α des cas.

4.4: Le travail du statisticien (1/2)



La lecture d'un rapport de recherche laisse croire qu'écrire une recherche (thèse, mémoire) est un travail linéaire, sans détour…

4.4: Le travail du statisticien (2/2)



Trois types de tests binomiaux.

Disponible sur: http://mapageweb.umontreal.ca/cousined/home/course/PSY1004

4.5: Test binomial sur une proportion a) méthode mélangeante

Postulats pour que le test soit valable:

- chaque observation est binaire: succès ou échec
- notons le nombre de succès sur n observations: $\boldsymbol{m}$

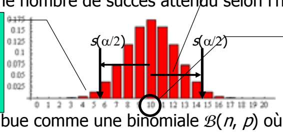
Test du genre:

rejet de H_0 si $|\boldsymbol{m} - np| \geq d(\alpha/2)$

distance $\boldsymbol{d}(\alpha/2)$ entre la valeur probable et les frontières

- où $n\,p$ est le nombre de succès attendu selon l'hypothèse

La distance critique est 5 car probabilité d'obtenir 0 ou 1 ou 2 ou 3 ou 4 ou 5 = 0.00+0.00+0.00+0.00+0.00+0.01 = 2.1% < $\alpha/2$



Par exemple, si $p = ½$ et $n = 20$, $n\,p = 10$.

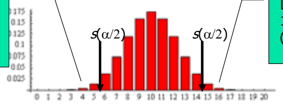
- $\boldsymbol{m}$ se distribue comme une binomiale $\mathcal{B}(n, p)$ où p est le contenu de l'hypothèse H_0.
- On choisi donc s(α) dans une table $\mathcal{B}(n, p)$ tel que l'aire à l'extérieur du critère ne dépasse pas α.

4.5: Test binomial sur une proportion b) méthode avec zone de rejet

Test du genre:

rejet de H_0 si $\boldsymbol{m} \in zone\ rejet$

La zone de rejet contient les histogrammes tel que leur total ne dépasse pas $\alpha/2$ à gauche et $\alpha/2$ à droite (si bicaudal)



Les histogrammes pour 15 à 20 totalisent 2.1% (rajouter l'histogramme 14 excède 2.5%)

- $\boldsymbol{m}$ se distribue comme une binomiale $\mathcal{B}(n, p)$ où p est le contenu de l'hypothèse H_0.
- Par exemple, pour $n = 20$ et $p = ½$, la zone de rejet est {0, 1, 2, 3, 4, 5 et 15, 16, 17, 18, 19, 20}

4.5: Test binomial sur une proportion c) exemple

Soit une pièce de monnaie que l'on soupçonne truquée

a) H_0: $p = ½$, H_1: $p \neq ½$ {donc bicaudal}

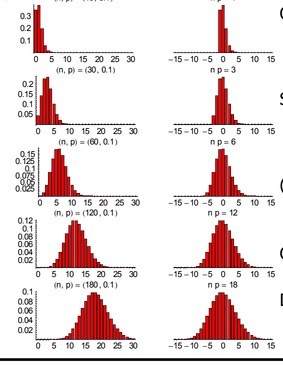
b) Seuil à 5% {a priori, avant de voir les données}

Collecte des données: $n = 20$, $\boldsymbol{m} = 6$

c) Chaque lancée est binaire ➔ $s(\alpha/2)$ est lue sur une table binomiale, et utilise le test: rejet de H_0 si $\boldsymbol{m} \in zone\ rejet$

d) Calcul (de la zone de rejet) et conclusion: "Sur 20 lancés, nous avons obtenu 6 piles. La pièce ne diffère pas significativement d'une pièce non truquée ($\mathcal{B}(20,½) = 6$, $\underline{p} > .05$)"

4.6: Test binomial sur une proportion utilisant l'approximation normale (1/3)



On remarque que plus $n\,p$ est grand, plus la distribution binomiale tend à être symétrique (colonne de gauche, où $p = 1/10$, très petit)

Si on centre à zéro la moyenne de ces graphes, on a des histogrammes qui ressemblent de plus en plus à une courbe normale standardisée

(en fait, l'asymétrie tend vers zéro et la kurtose vers 3, peu importe p);

Ceci a lieu quand:

$n > 20$ et $n\,p > 10$;

Dans ces cas, on dit que la normale approxime la distribution binomiale

4.6: Test binomial sur une proportion utilisant l'approximation normale (2/3)

Postulats pour que le test soit valable:

- chaque observation est binaire: succès ou échec;
- $n > 20$ et $n\,p > 10$;
- notons la proportion de succès sur n observations: $\bar{\mathbf{X}} = {}^{m}\!/_{n}$

Test du genre (si bicaudal):

rejet de H_0 si $\dfrac{|\bar{\mathbf{X}} - p|}{\sqrt{p(1-p)/n}} > s(\alpha/2)$

- où $n\,p$ est le nombre de succès attendu selon l'hypothèse
- La partie de gauche se distribue ≈ comme une normal $\mathcal{N}(0, 1)$.
- On choisi donc $s(\alpha/2)$ dans une table $\mathcal{N}(0, 1)$ tel que l'aire à l'extérieur du critère égale α.

4.6: Test binomial sur une proportion utilisant l'approximation normale (3/3)

Exemple 2:

Soit la même pièce de monnaie que l'on soupçonne truquée

a) H_0: $p = ½$, H_1: $p \neq ½$ {donc bicaudal}

b) Seuil à 5% {a priori, avant de voir les données}

Collecte des données: $n = 200$, $\boldsymbol{m} = 90$

c) Chaque lancée est binaire, mais $n > 20$ et $n\,p > 10$ ➔ $s(\alpha/2)$ est lue sur une table normale $\mathcal{N}(0,1)$, et utilise le test: rejet de H_0 si

$$\frac{|\bar{\mathbf{X}} - p|}{\sqrt{p(1-p)/n}} > s(\alpha/2)$$

d) Calcul et conclusion: "Sur 200 lancés, nous avons obtenu 90 piles. La pièce ne diffère pas significativement d'une pièce non truquée (z = 1.41, p > .05)"

4.7: Test des signes a) idée

Lorsque l'on veut tester des données du type avant-après.

Postulats pour que le test soit valable:

- La population n'est pas du tout normale, soit très asymétrique ou multimodale (sinon, voir Section 5)
- notons par un + les scores qui sont meilleurs après qu'avant, et par un – les autres scores, et notons le nombre de + par $\boldsymbol{m}$.
- Le signe est binaire, et sa probabilité doit être de ½.

Continuez avec un test

- binomial utilisant l'approximation normale si $n > 20$ et $n\,p > 10$
- binomial standard sinon.

4.7: Test des signes b) exemple

Sommes-nous plus intelligents les mercredi que les lundi?

On mesure le QI de 20 personnes, une fois un mercredi, une fois un lundi (ordre aléatoire).

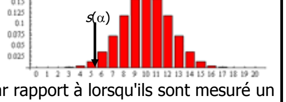
a) H_0: $QI_{lundi} = QI_{mercredi}$; H_1: $QI_{lundi} < QI_{mercredi}$;

b) seuil $\alpha = 5\%$

Collecte des données: $\boldsymbol{m} = 7$ sur $n = 20$ personnes

c) Choisir le test: test des signes (test binomial avec $p = ½$)

a) Calcul de la zone de rejet: {0, 1, 2, 3, 4, 5}; conclure: "Sur 20 personnes, 7 ont montré une baisse du QI lorsque mesuré un mercredi par rapport à lorsqu'ils sont mesuré un lundi, ce qui n'est pas significatif (B(20, ½) = 7, p > .05)."



4.8: Test de la médiane
a) la même idée

Lorsque l'on veut tester une tendance centrale d'un échantillon de score.

Postulats pour que le test soit valable:

- La population n'est pas du tout normale, soit très asymétrique ou multimodale (sinon, un test sur une moyenne est préférable; voir Section 5)
- notons par un + les scores qui sont supérieurs à la médiane, et par un – les autres scores, et notons le nombre de + par ***m***.
- Le signe est binaire, et sa probabilité doit être de ½.

Continuez avec un test

- binomial utilisant l'approximation normale $n > 20$ et $n\,p > 10$
- binomial standard sinon.

4.9: Exemples

1. Soit ces données
   X = { 1,1,1,1,1,1,1,1,1,2,2,2,2,2,3,3,3,4,5,7,9,15},
   est-ce que la médiane peut-être de 2?
   Quel signe met-on pour une donnée exactement de 2?

2. Soit ces données avant-après
   X = {1,1,1,1,1,2,2,8,9,9,9,9,11,12} et
   Y = {0,1,0,2,1,2,4,9,10,11,8,7,13,11},
   est-ce que les scores ont changés après comparés à avant?
   Quel signe met-on si le score est inchangé?