Statistiques Inférentielles

31 janvier 2021

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Introduction - vocabulaire

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Deux méthodes pour étudier une population en stats :

Exhaustive (recensement)

On examine chacun des éléments. Trop long, trop cher.

Sondage

Seule une partie est étudiée.

- Échantillonnage : sélectionner une sous partie (représentative). **Hors programme**
- Estimation : induire des infos. sur la pop totale à partir de l'échantillon. **Notre objectif**

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Principe

Population P d'effectif N.

Caractère étudié de moyenne m et d'écart type σ
Comment les retrouver à partir de l'échantillon ?

Échantillons

k échantillons de P, tous de taille n.
$\overline{X} = (\overline{x_1}, \cdots, \overline{x_k})$: série stat : la *distribution des moyennes.*

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Contexte

Population P

m : inconnu

σ : inconnu

échantillon taille n

$\overline{x}$ connu

σ' connu

Théorème :

$$E(\overline{X}) = m \qquad \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

La v.a. $\dfrac{\overline{X} - m}{\sigma/\sqrt{n}}$ suit approximativement $\mathcal{N}(0,1)$.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Estimation ponctuelle

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Moyenne, écart type, variance

- $\overline{x}$ est un estimateur ponctuel de m
- $\sqrt{\dfrac{n}{n-1}}\sigma'$ est un estimateur ponctuel de σ

Fréquence

Caractère binaire (oui/non) de fréquence p dans P
La fréquence f du caractère dans l'éch. est un estimateur ponctuel de p.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Estimation d'une moyenne par intervalle de confiance

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Contexte

Population P
m : inconnu
σ : **connu**

échantillon taille n
$\overline{X}$ connu

$$P(-t \leq T \leq t) = 2\Pi(t) - 1$$

où $T = \dfrac{\sqrt{n}}{\sigma}(\overline{X} - m)$ suit $\mathcal{N}(0, 1)$

Intervalle de confiance à 95%

Il y a 95% de chance pour que m appartienne à

$$\left[\overline{x} - 1,96\frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{x} + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

Niveau de confiance c et coefficient critique t

90%	95%	99%
1.64	1.96	2.58

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Cas général

On pose "confiance" $= 2\Pi(t) - 1$ qu'on résout en t.
(*fracnormale*, *invnormale*, table $\mathcal{N}(0,1)$ etc.)
L'intervalle de confiance de la moyenne m de P est alors :

$$\left[\overline{x} - t\frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{x} + t\frac{\sigma}{\sqrt{n}}\right]$$

Estimation de la fréquence par intervalle de confiance

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Contexte

Population P p : inconnue échantillon taille n fréquence f connue

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Théorème

L'intervalle de confiance de la fréquence p de P est alors :

$$\left[f - t\sqrt{\frac{f(1-f)}{n-1}} \quad ; \quad f + t\sqrt{\frac{f(1-f)}{n-1}} \right]$$

On obtient t de la même manière que pour la moyenne.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Test de validité d'hypothèse

Réglementer la prise de décision

Introduction des stats inférentielles dans un cadre réglementaire.
On définit des règles pour :

1. prélever des échantillons,
2. réaliser des mesures sur ces échantillons,
3. décider pour toute la population.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Exemple de test : comparaison d'une moyenne à une valeur fixée

Contrat

Les pièces fournies doivent avoir pour masse moyenne 780 grammes.

Réception d'un lot de 500 pièces du fournisseur A

On prélève un échantillon : $n = 36, m = 774,7$ g.

Doit-on accepter la livraison ?

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Intervalle de fluctuation

Même démarche que pour l'intervalle de confiance.

Dans 95% des cas, la moyenne observée $\overline{x}$ fluctue dans

$$\left[m - 1,96\frac{\sigma}{\sqrt{n}} \quad ; \quad m + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

Notre exemple, échantillon de taille $n = 36$

IF95%=$[775,92; 784,08]$

Hypothèse nulle

« on suppose que la moyenne de la population (la livraison) est 780 g »
Notée : $H_0 : m = 780$

Hypothèse alternative

C'est l'hypothèse contre laquelle on va tester H_0.
Notre test sera ici **bilatéral** : la moyenne m peut être supérieure ou inférieure à m.
$H_1 : m \neq 780$

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Règle de décision :

On fixe la règle suivante : on prélève avec remise un échantillon aléatoire non exhaustif de taille $n = 36$ et on calcule $\overline{x}$.

- Si $\overline{x} \in [775,92; 784,08]$ (IF95%), alors on accepte H_0 - et la livraison,
- Si $\overline{x} \notin [775,92; 784,08]$ (IF95%), alors on rejette H_0 - et la livraison.

Décision :

$774.7 \notin [775,92; 784,08]$ donc on rejette H_0 et la livraison. Au seuil de 5% on considère que les pièces n'ont pas une moyenne de 780g et on refuse la livraison du fournisseur A.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Erreur de première espèce

On a pris 5% (noté généralement α) de risque de rejetter une livraison qui pouvait être valable. Notre échantillon avait 5% de chance d'être exceptionnel...

Erreur de seconde espèce

Inversement, **chaque fois qu'on accepte** H_0, on court le risque d'accepter une livraison ne répondant pas au contrat. L'échantillon peut « tomber » dans l'intervalle de fluctuation alors que $m \neq 780$...
On note β ce risque de seconde espèce.

Lien entre α (risque de première espèce) et β

Quand la taille de l'échantillon est fixée, diminuer α revient généralement à augmenter β.
Souvent les erreurs sont d'importance inégale : **on limite la plus grave.**
Ex : Il vaut parfois mieux traiter préventivement une maladie dont on n'est pas certain plutôt que d'attendre qu'il soit trop tard : on limitera β.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Résumé : Test de validité d'hypothèse

1. Construction du test :
   - Choix de H_O et de H_1.
   - Région critique pour le seuil α donné.
   - Énoncé de la règle de décision : si le paramètre de l'échantillon est dans la région critique, on rejette H_0, sinon on l'accepte.
2. Utilisation du test :
   - Calcul du paramètre de l'échantillon.
   - Application de la règle de décision.

Exemple : comparaison de moyennes entre elles

Statistiques Inférentielles

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Réception d'un lot de 800 pièces du fournisseur B

On prélève un échantillon : $n = 50, m = 779.6$ g.

Cette différence entre les échantillons de A et de B est-elle significative ?

Construction d'un test, au seuil de 5% pour décider s'il y a une différence significative entre les lots.

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Contexte

Livraison A m_A : inconnue $\sigma_A = 12,5$	$n_A = 36$ $\overline{x}_A = 774,7$ $\sigma'_A = 12,36$

Livraison B m_B : inconnue $\sigma_B = 12,1$	$n_B = 50$ $\overline{x}_A = 779,6$ $\sigma'_A = 11,99$

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Théorie

On suppose que $\overline{X}_A$ et $\overline{X}_B$ sont indépendantes. Ces v.a. suivent des lois normales donc $D = \overline{X}_A - \overline{X}_B$ aussi.

$$D \sim \mathcal{N}(m_B - m_A; s) \text{ où } s = \sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} \approx 2,7.$$

Introduction - vocabulaire
Principe
Estimation ponctuelle
Estimation d'une moyenne par intervalle de confiance
Estimation de la fréquence par intervalle de confiance
Test de validité d'hypothèse
Exemple de test : comparaison d'une moyenne à une valeur fixée
Exemple : comparaison de moyennes entre elles
Exemple : test unilatéral

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Test de $m_A = m_B$

1. Construction du test :
   - H_O : $m_A = m_B$ et de H_1 : $m_A \neq m_B$.
   - Région critique au seuil $\alpha = 0,05$
     Sous H_0, $D \sim \mathcal{N}(0; 2,7) \Leftrightarrow D/2,7 \sim \mathcal{N}(0; 1)$
     $p(-1,96 \leq D/2.7 \leq 1,96) = 0,95$
     $\Leftrightarrow p(-5,29 \leq D \leq 5,29) = 0,95$
   - Règle de décision : si $d = \overline{x}_B - \overline{x}_A \in [-5,29; 5,29]$ on accepte H_0 sinon on rejette H_0.
2. Utilisation du test :
   - Calcul de d : $d = 779,6 - 774,7 = 4,9$
   - Application : $4,9 \in [-5,29; 5,29]$, on accepte H_0
     Au seuil de 5% il n'y a pas de différence significative entre les moyennes de 2 fournisseurs.

Exemple : test unilatéral

Statistiques Inférentielles

Introduction - vocabulaire

Principe

Estimation ponctuelle

Estimation d'une moyenne par intervalle de confiance

Estimation de la fréquence par intervalle de confiance

Test de validité d'hypothèse

Exemple de test : comparaison d'une moyenne à une valeur fixée

Exemple : comparaison de moyennes entre elles

Exemple : test unilatéral

Peut-on tester si $m_B > m_A$?

1. Construction du test :
   - H_O : $m_A = m_B$ et de H_1 : $m_B > m_A$.
   - Région critique au seuil $\alpha = 0,05$
     Sous H_0, $D \sim \mathcal{N}(0; 2,7) \Leftrightarrow D/2,7 \sim \mathcal{N}(0; 1)$
     $p(D/2.7 \geq t) = 0,95$ si $t = 1,645$
     $\Leftrightarrow p(D \geq 4,44) = 0,95$
   - Règle de décision : si $d = \overline{x}_B - \overline{x}_A \leq 4,44$ on accepte H_0 sinon on accepte H_1.
2. Utilisation du test :
   - Calcul de d : $d = 779,6 - 774,7 = 4,9$
   - Application : $4,9 > 4.44$, on accepte H_0.
     Au seuil de 5% la moyenne de B est significativement supérieure à celle de A.