Statistiques Appliquées

Aide-mémoire
(document distribué pendant le contrôle)

1 Théorie d'échantillonnage

1.1 Une population – un échantillon

Le tableau 1 contient les statistiques d'échantillonnage liées aux paramètres de la population dans le cas d'un échantillonnage aléatoire *avec* remplacement.

Dans le cas d'un échantillonnage *sans* remplacement dans une population finie de taille N, l'écart-type de la population utilisé dans Z – égal à σ ou $\sqrt{\pi(1-\pi)}$ – est multiplié par le facteur $\sqrt{\frac{N-n}{N-1}}$.

1.2 Deux populations – deux échantillons indépendants

Le tableau 2 contient les statistiques d'échantillonnage liées aux paramètres des populations dans le cas d'un échantillonnage aléatoire et indépendant *avec* remplacement.

1.3 Deux populations – deux échantillons appariés

Si les échantillons sont appariés (« avant » / « après »), créer un nouvel échantillon $d_i = x_{1i} - x_{2i}$ et travailler avec $\bar{D}$ (tableau 1).

2 Intervalles de confiance / tests

Le tableau 3 contient les intervalles de confiance des statistiques normalisées ainsi que les décisions de réjet des tests d'hypothèse.

- Intervalle de confiance : niveau de confiance $1-\alpha$
- Tests d'hypothèse : seuil de signification α
- Symétries :
  - Normale centrée réduite : $z_{1-\alpha} = -z_\alpha$
  - Student (ν) : $t_{1-\alpha}(\nu) = -t_\alpha(\nu)$
  - Fischer (ν_1, ν_2) : $f_{1-\alpha}(\nu_1,\nu_2) = 1/f_\alpha(\nu_2,\nu_1)$
- Proportions : remplacer $\pi(1-\pi)$ par $\hat{p}(1-\hat{p})$ comme variance de la population
  - cas spécial : tests avec deux populations et $H_0 : \pi_1 - \pi_2 = 0$
    remplacer $\pi_j(1-\pi_j)$ par $\hat{p}(1-\hat{p})$ où $\hat{p} = (n_1\hat{p}_1 + n_2\hat{p}_2)/(n_1+n_2)$
- Test unilatéral sur une moyenne :
  - $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$, H_1 précise : $\mu = \mu_0 + \delta$
  - Si conditions pour Z : $z_\alpha + z_\beta = \frac{\delta}{\sigma/\sqrt{n}}$
  - Si conditions pour T : $t_\alpha + t_\beta = \frac{\delta}{s/\sqrt{n}}$

Paramètre θ	μ				π	σ^2
Population Écart-type σ	$\approx$ normale connu	— connu	$\approx$ normale inconnu		— —	$\approx$ normale —
Échantillon	—	$n > 30$	$n > 30$	$n < 30$	$n > 30$ [a]	—
Statistique $\hat{\Theta}$	$\bar{X}$				$\hat{P}$	S^2
St. normalisée	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$		$Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	$Z = \frac{\hat{P}-\pi}{\sqrt{\pi(1-\pi)/n}}$	$X^2 = \frac{(n-1)S^2}{\sigma^2}$
Distribution	$N(0,1)$			Student (ν)	$N(0,1)$	khi-deux (ν)
Degrés de liberté	—			$n-1$	—	$n-1$
Mesure $\hat{\theta}$	$\bar{x}$				$\hat{p}$	s^2

Tab. 1: Théorie d'échantillonnage : une population, un échantillon aléatoire, avec remplacement.

[a] En plus : $n\hat{p} \geq 5$, $n(1-\hat{p}) \geq 5$, ni $\hat{p} \approx 0$, ni $\hat{p} \approx 1$.

Paramètre θ	$\mu_2 - \mu_1$				
Populations	$\approx$ normales	—	$\approx$ normales		
Écart-types σ_1, σ_2	connus	connus	inconnus	inc., $\sigma_1 = \sigma_2$ ou $n_1 = n_2$	inc., $\sigma_1 \neq \sigma_2$ et $n_1 \neq n_2$
Échantillons	—	$n_1 > 30$ et $n_2 > 30$	$n_1 > 30$ et $n_2 > 30$	$n_1 < 30$ ou $n_2 < 30$	
Statistique $\hat{\Theta}$	$\bar{X}_2 - \bar{X}_1$				
St. normalisée	$Z = \frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$		$Z = \frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	$T = \frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{S_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$T = \frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
Distribution	$N(0,1)$			Student (ν)	
Degrés de liberté	—			$n_1 + n_2 - 2$	ν^*
Mesure $\hat{\theta}$	$\bar{x}_2 - \bar{x}_1$				

$$S_c^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$ la variance *commune*, calculée à partir des deux échantillons

$$\nu^* = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$ arrondir au nombre entier *inférieur*

Paramètre θ	$\pi_2 - \pi_1$	σ_1^2/σ_2^2
Populations	—	$\approx$ normales
Écart-types σ_1, σ_2	—	—
Échantillons	$n_1 > 30$ et $n_2 > 30$ [a]	—
Statistique $\hat{\Theta}$	$\hat{P}_2 - \hat{P}_1$	S_1^2/S_2^2
St. normalisée	$Z = \frac{(\hat{P}_2-\hat{P}_1)-(\pi_2-\pi_1)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}}$	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$
Distribution	$N(0,1)$	Fischer (ν_1, ν_2)
Degrés de liberté	—	$n_1 - 1, n_2 - 1$
Mesure $\hat{\theta}$	$\hat{p}_2 - \hat{p}_1$	s_1^2/s_2^2

Tab. 2: Théorie d'échantillonnage : deux populations, deux échantillons aléatoires et indépendants, avec remplacement.

[a] En plus : $n_i\hat{p}_i \geq 5$, $n_i(1-\hat{p}_i) \geq 5$, ni $\hat{p}_i \approx 0$, ni $\hat{p}_i \approx 1$ $(i = 1, 2)$.

Statistique normalisée	Intervalle de confiance	Test d'hypothèse $H_0 : \theta = \theta_0$		
		$H_1 : \theta \neq \theta_0$	$H_1 : \theta < \theta_0$	$H_1 : \theta > \theta_0$
Z	$-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$	$z < -z_{\frac{\alpha}{2}}$ ou $z > z_{\frac{\alpha}{2}}$	$z < -z_\alpha$	$z > z_\alpha$
T	$-t_{\frac{\alpha}{2}}(\nu) < t < t_{\frac{\alpha}{2}}(\nu)$	$t < -t_{\frac{\alpha}{2}}(\nu)$ ou $t > t_{\frac{\alpha}{2}}(\nu)$	$t < -t_\alpha(\nu)$	$t > t_\alpha(\nu)$
X^2	$\chi^2_{1-\frac{\alpha}{2}}(\nu) < \chi^2 < \chi^2_{\frac{\alpha}{2}}(\nu)$	$\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(\nu)$ ou $\chi^2 > \chi^2_{\frac{\alpha}{2}}(\nu)$	$\chi^2 < \chi^2_{1-\alpha}(\nu)$	$\chi^2 > \chi^2_\alpha(\nu)$
F	$f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) < f < f_{\frac{\alpha}{2}}(\nu_1,\nu_2)$	$f < f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)$ ou $f > f_{\frac{\alpha}{2}}(\nu_1,\nu_2)$	$f < f_{1-\alpha}(\nu_1,\nu_2)$	$f > f_\alpha(\nu_1,\nu_2)$
Procédure à suivre	calculer les valeurs critiques à partir de la valeur de α choisie			
	remplacer z, t, χ^2 ou f en fonction de θ ; mettre sous la forme : $\theta_L < \theta < \theta_H$	« entrer dans le monde de H_0 » : prendre $\theta = \theta_0$ et calculer z, t, χ^2 ou f à partir des mesures ; le tableau contient les décisions de *rejet* de H_0		
p-value	—	trouver α_p qui rend le seuil de décision égal à z, t, χ^2 ou f		

Tab. 3: Intervalles de confiance et tests d'hypothèse.