

Valeurs manquantes

Tests de Welch et de Little

Véronique Tremblay

Tester le mécanisme MCAR

- Plusieurs tests de Welch
- Test de Little (1988)

(Test de t pour variance inégales)

Séparer les cas observés et manquants pour une variable
et tester les différences de moyennes pour les autres
variables.

Test de Welch

- Peut nous aider à identifier les variables pour lesquelles ajuster dans les procédures de traitement de données manquantes. 😊
- Implique beaucoup de tests. 😞
- Ne tient pas compte des corrélations entre les variables. 😞
- Ne garantie pas l'hypothèse MCAR. 😞

Test de Little

Compare en un seul test les moyennes de chaque patron de non-réponse.

Lire les sections 3.1, 3.2 et 3.3 de Little (1988)

Patron de non-réponse

Identifiant	Véhicule	Vitesse	Distraktion	Ceinture
1	Auto	35	Aucune	Oui
2	Auto	80		
3	Vélo	22	Écouteurs	Oui
4	Auto	45	Maquillage	
5	Auto	50	Nourriture	Oui
6	Auto	35	Aucun	
7	Auto	50	Aucun	

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Test de Little: Hypothèses

H_0 : Les données sont MCAR ($\bar{y}_{obs_j} = \hat{\mu}_{obs}$)

H_1 : Les données ne sont pas MCAR ($\bar{y}_{obs_j} \neq \hat{\mu}_{obs}$)

Test de Little: Statistique de test

$$d^2 = \sum_{j=1}^J m_j (\bar{y}_{obs,j} - \hat{\mu}_{obs,j}) \hat{\Sigma}_j^{-1} (\bar{y}_{obs,j} - \hat{\mu}_{obs,j})^\top$$

où $j = 1, \dots, J$ dénote le patron de non-réponse, $\bar{y}_{obs,j}$ est l'EMV de la moyenne pour les variables observées du patron j , $\hat{\mu}_{obs,j}$ est l'EMV global de la moyenne $\hat{\Sigma}_j$ l'EMV global de la matrice de covariance du patron J .

Test de Little: Règle de décision

Sous H_0 ,

$$d^2 \sim \chi^2_{\sum_{j=1}^J p_j - p}$$

où p_j est le nombre de variables complètement observées dans le patron j et p est le nombre de variables du jeu de données.

Test de Little: Interprétation di résultat

- Si la p-valeur $<$ seuil choisi, on rejette H_0 et on conclut que les données ne sont pas MCAR.
- Si la p-valeur $>$ seuil choisi...

Test de Little

- Le test ne nous indique pas quelles variables sont problématiques. 😞
- Le test suppose la même matrice de covariance pour tous les patrons de non-réponse. 😞
- Des études de simulations ont mis en évidence un manque de puissance de ce test. 😞
- Ne permet pas de garantir l'hypothèse MCAR. 😞

- Welch
- Little (1988)

Little, Roderick JA. 1988. “A Test of Missing Completely at Random for Multivariate Data with Missing Values.” *Journal of the American Statistical Association* 83 (404): 1198–1202.