

SOLUTIONNAIRE : LOGICIELS STATISTIQUES

EXERCICES

- (1) Un "chasseurs de têtes"¹ a comme mission de trouver des candidats pour combler un poste dans une entreprise spécialisée dans la prospection de pétrole. Pour sélectionner les candidats qui n'occupent pas déjà un poste équivalent dans le domaine, un questionnaire est envoyé à tous les candidats potentiels. Or on croit qu'un bon candidat aura une moyenne de 80 à ce questionnaire. Pour valider cette hypothèse une enquête est menée auprès des personnes occupants un poste équivalent. Voici les résultats pour cet échantillon :

91, 80, 74, 71, 67, 88, 74, 71, 77, 72, 80, 72, 84, 71, 88

Peut-on dire au niveau 5% que les candidats ayant les capacités requises ont un score moyen de 80 à ce test ? (Faire le test avec Excel)

On veut confronter les hypothèses suivantes

$$H_0 : \mu = 80$$

$$H_1 : \mu \neq 80$$

au niveau 5%.

On utilise le test de comparaison d'observations appariées et le test est de rejeter H_0 si $\hat{\alpha} \leq \alpha = 5\%$. Comme l'échantillon est de taille relativement petite il faut supposer que les valeurs sont approximativement normales.

La sortie EXCEL pour ce test est la suivante :

Test d'égalité des espérances: observations appariées		
	Score	Norme
Moyenne	77,3333333	80
Variance	55,6666667	0
Observations	15	15
Coefficient de corrélation de Pearson	#DIV/0!	
Différence hypothétique des moyennes	0	
Degré de liberté	14	
Statistique t	-1,38425708	
P(T<t) unilatéral	0,0939745	
Valeur critique de t (unilatéral)	1,76131012	
P(T<t) bilatéral	0,187949	
Valeur critique de t (bilatéral)	2,14478668	

On observe $\hat{\alpha} = 0.188$ pour un test bilatéral donc il faut accepter H_0 au niveau 5%. On peut dire que la moyenne du score à ce test pour ceux qui ont les qualifications pour remplir ce poste est effectivement de 80.

Remark 1 Avec les informations données dans ce problème il n'est pas possible de faire un test

¹ Les chasseurs de têtes ont pour mission de trouver des candidats pour combler un poste au sein d'une entreprise. Habituellement les postes à combler demandent des qualifications spéciales et les candidats sont peu nombreux et surtout très en demande.

bilatéral : rien dans les données du problème permet de dire qu'un score plus élevé permet de sélectionner des candidats plus "adéquats". De plus, la question est de savoir si les candidats parfaitement adaptés ont un score de 80, rien de plus.

- (2) Sur un sac de pétoncles U-10 préemballés, il est mentionné 330 g. Pour vérifier les dires du fabricant on prend un échantillon de 25 sacs et on pèse exactement le contenu. Voici les valeurs observées :

327 335 327 330 324 331 319 332 341 334

338 327 326 325 329 331 324 330 333 333 326 334 344 320 330

Peut-on dire, en moyenne, que la quantité inscrite est la bonne au niveau 5% ? (faire le test avec EXCEL)

On veut confronter les hypothèses

$$H_0 : \mu = 330$$

$$H_1 : \mu \neq 330$$

Le niveau choisi est de $\alpha = 5\%$.

Le test est celui de Student pour un échantillon (commande "échantillons pairés" de EXCEL). On doit alors supposer que les valeurs sont approximativement normales. La sortie EXCEL donne

Test d'égalité des espérances: observations pairées		
	Variable 1	Variable 2
Moyenne	330	330
Variance	35	0
Observations	25	25
Coefficient de corrélation de Pearson	#DIV/0!	
Différence hypothétique des moyennes	0	
Degré de liberté	24	
Statistique t	0	
P(T<=t) unilatéral	0,5	
Valeur critique de t (unilatéral)	1,71088207	
P(T<=t) bilatéral	1	
Valeur critique de t (bilatéral)	2,06389855	

Le test est de rejeter H_0 si $\hat{\alpha} \leq \alpha = 5\%$. On observe $\hat{\alpha} = 1$ donc on accepte H_0 au niveau 5%.

Remark 2 *Le cas ici est très particulier et extrêmement rare puisque la valeur observée est EXACTEMENT égale à la valeur supposée selon H_0 . Comme le principe du test est de considérer H_0 vraie a priori et de la rejeter si les observations la rende "ridicule" il est tout à fait raisonnable de la conserver.*

- (3) Dans un bureau de vérification comptable, on cherche à comparer la productivité des femmes par rapport à celle des hommes dans le traitement des dossiers de vérification. Voici le nombre de dossiers vérifié en une semaine par un échantillon de 11 femmes

8, 16, 17, 8, 14, 16, 10, 16, 6, 9, 9

et le nombre vérifier par un échantillon de 12 hommes

10, 8, 7, 5, 9, 10, 8, 3, 5, 7, 6, 10

Comparer la productivité moyenne au niveau 5% à l'aide de Excel.

On veut confronter les hypothèses $H_0 : \mu_H = \mu_F$ contre $H_1 : \mu_H \neq \mu_F$. En supposant la normalité des observations le test d'égalité des variances donne un niveau empirique (expérimental) $\hat{\alpha}' = 0.067$ donc on peut conclure que les variances sont différentes (rejeter l'égalité si $\hat{\alpha}' < 0.1$ et on observe $\hat{\alpha}'_U = .03374$ d'où $\hat{\alpha}_B = 0.067$)

Test d'égalité des variances (F-Test)		
	Femmes	Hommes
Moyenne	11,7272727	7,33333333
Variance	16,6181818	5,15151515
Observations	11	12
Degré de liberté	10	11
F	3,22588235	
P(F<=f) unilatéral	0,03374166	
Valeur critique pour F (unila	2,85362486	

Le test pour comparer les moyennes est le test de Student pour données avec variances différentes. On observe par le logiciel EXCEL $\hat{\alpha} = 0.0065$ donc on rejette H_0

Test d'égalité des espérances: deux observations de varianc		
	Femmes	Hommes
Moyenne	11,7272727	7,33333333
Variance	16,6181818	5,15151515
Observations	11	12
Différence hypothétique des moy	0	
Degré de liberté	15	
Statistique t	3,15463477	
P(T<=t) unilatéral	0,00327201	
Valeur critique de t (unilatéral)	1,75305033	
P(T<=t) bilatéral	0,00654403	
Valeur critique de t (bilatéral)	2,13144954	

- (4) Un groupe de 100 personnes a été recruté pour valider une nouvelle application pour téléphone intelligent dans le but d'aider les touristes à voyager en métro à Montréal. Aucun des touristes n'a utilisé l'ancienne application. Le groupe est séparé en deux, 50 d'entre eux utiliseront l'ancienne application tandis que les 50 autres utiliseront la nouvelle application.

Au terme d'un essai de 3 jours chaque testeur note sur 100 l'intérêt de l'application, 100 étant une très grande satisfaction... Faire le test pour établir au niveau 5% si la nouvelle application est meilleure.

Les hypothèses à confronter sont $H_0 : \mu_n = \mu_a$ contre $H_1 : \mu_n > \mu_a$. Puisque les tailles d'échantillon sont grandes on utilise le test de Student pour variances différentes. Selon les sorties EXCEL, les observations vont dans le sens de H_1 puisque $\bar{x}_n = 59,78$ et $\bar{x}_a = 58,72$. Le niveau expérimental ou empirique est de $\hat{\alpha} = 0.3650$ pour un test unilatéral donc on accepte H_0 au niveau

- (5) Un viticulteur participe à un concours de vin pour sa cuvée de rouge Château bonami. Son voisin et principal concurrent présente aussi un vin rouge, le domaine de l'amitié. Il y a 10 juges qui donne des notes de 1 à 100 pour chaque vin. Voici les résultats des 10 juges dans le même ordre pour chaque vin :

Bonami : 75, 75, 72, 78, 78, 75, 72, 80, 81, 73

Amitié : 76, 82, 83, 85, 80, 79, 83, 86, 84, 84

Le viticulteur peut-il se targuer d'avoir un meilleur vin que son voisin ? (niveau 10%)

Les hypothèses à confronter sont $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$, où $D = \text{Bonami} - \text{Amitié}$. Le logiciel EXCEL donne la sortie suivante

Test d'égalité des espérances: observations paires		
	Bonami	Amitié
Moyenne	75,9	82,2
Variance	10,3222222	9,28888889
Observations	10	10
Coefficient de corrélation de Pearson	0,26325522	
Différence hypothétique des moyennes	0	
Degré de liberté	9	
Statistique t	-5,23990189	
P(T<=t) unilatéral	0,00026742	
Valeur critique de t (unilatéral)	1,83311292	
P(T<=t) bilatéral	0,00053483	
Valeur critique de t (bilatéral)	2,26215716	

Puisque les données ne vont pas dans le sens de H_1 ($x_D = 75.9 - 82.2 = -6.3 \neq 0$) on accepte H_0 directement.

- (6) Un sondage auprès des consommateurs français montre que ces derniers perçoivent une caisse d'épargne comme étant une institution plus appropriée à des petits prêts aux particuliers tandis que les banques sont plus axées sur les prêts importants aux entreprises.

Pour déboulonner ce mythe, une association des caisses d'épargne demande à un chercheur universitaire de faire une enquête pour comparer les prêts aux entreprises effectués par les banques et les caisses d'épargne. La méthodologie consiste à prendre 25 prêts aux entreprises au hasard dans des banques et à prendre le même nombre de prêts aux entreprises dans des caisses d'épargne. Voici les résultats en milliers d'euros :

Caisses : 46, 155, 42, 6, 77, 109, 79, 72, 123, 145, 78, 90, 101, 124, 111, 28, 126, 37, 32, 106, 12, 4, 83, 132, 72,

Banques : 12, 112, 199, 169, 15, 27, 50, 52, 189, 401, 6, 19, 213, 292, 80, 4, 40, 182, 88, 161, 51, 68, 4, 49, 42,

Vérifier si le mythe est vraie au niveau 5%.

Les hypothèses statistiques sont $H_0 : \mu_c = \mu_b$ contre $H_1 : \mu_c < \mu_b$. Puisque les tailles sont petites on doit supposer que les valeurs sont normales puis faire le test d'égalité des variances.

Cela conduit à utiliser le test de Student pour variances différentes :

Test d'égalité des espérances: deux observations de variances différen		
	<i>Caisse</i>	<i>Banque</i>
Moyenne	79,6	101
Variance	1956,41667	10239,4167
Observations	25	25
Différence hypothétique des moyennes	0	
Degré de liberté	33	
Statistique t	-0,96889795	
P(T<=t) unilatéral	0,16982208	
Valeur critique de t (unilatéral)	1,69236026	
P(T<=t) bilatéral	0,33964417	
Valeur critique de t (bilatéral)	2,03451529	

Le test est de rejeter H_0 si $\hat{\alpha} \leq \alpha = 5\%$. Selon la sortie EXCEL version 2007 les données vont dans le sens de H_1 et on obtient $\hat{\alpha} = 0.1698$ donc on accepte H_0 et on peut dire qu'il n'y a pas de différences entre les deux types de banques.

- (7) Un créateur de site web vous affirme, avec chiffres à l'appui, que l'année dernière il y a eu une augmentation de 20% du chiffre d'affaire pour ses clients dans le secteur des services de soins du corps (Saunas, spa, massage, soins de la peau, etc.)

Vous pensez cependant que cela peut avoir une cause autre que la création d'un site internet. Pour valider cette hypothèse vous prenez un échantillon de 20 entreprises dans le secteur qui n'ont pas de site internet puis vous obtenez l'augmentation, en % du chiffre d'affaire pour l'année dernière. En combinant cette information avec les données du créateur de site web (25 observations)

Vérifier si le site WEB est générateur de plus de profit en moyenne que le marché au niveau 5%.

Les hypothèses statistiques sont $H_0 : \mu_w = \mu_s$ contre $H_1 : \mu_w > \mu_s$.

En supposant la normalité des observations en considérant que les variances sont semblables ($\hat{\alpha} = 0.93$) le test de Student pour variances égales est de rejeter H_0 si $\hat{\alpha} \leq \alpha = 5\%$.

On observe $\hat{\alpha} = 0.1343$ donc on accepte H_0