



Corrigé de l'examen CT de statistique

À lire !!

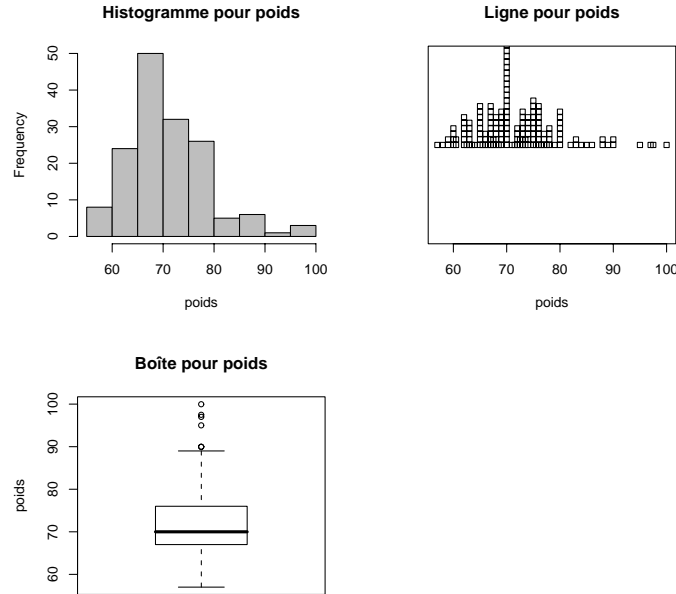
Un mouvement de rétentions de notes s'organise à Lyon I, pour exprimer un certain nombre de mécontentements. Ce mouvement ne se limite ni à la rétention de notes, ni à Lyon et concerne un très grand nombre d'universités et d'UFR en France. Pour en comprendre les fondements (et notamment pourquoi ce mouvement concerne aussi les étudiants), voir : <http://douaalter.lautre.net/mobilisation/index.html>.

Correction de l'exercice 1.

- (1) • On étudie la variable quantitative (ou numérique) 'poids'. On procèdera donc comme dans le chapitre 3 du document de cours.
- Les différents résultats déterminés par \mathcal{R} sont donnés dans le tableau suivant

noms	valeurs
moyenne	71.812903
sd	7.980119
Q_1 (quartile à 25 %)	67
médiane	70
Q_3 (quartile à 75 %)	76
minimum	57
maximum	100
nombre	155

•



Voir les trois graphiques ci-dessus pour la variable 'poids'. Le nombre de données étant important (ici 155), la ligne de point n'est pas très pertinente.

(2) On procède au *test de Student pour un échantillon* (à la moyenne).

On fait l'hypothèse nulle $H_0 : \mu = \mu_0$, avec $\mu_0 = 65$. On cherche à montrer que la moyenne de la loi normale, dont proviendraient les données de l'échantillon étudié, est plus petite que μ_0 . On fait donc l'hypothèse alternative suivante : $H_1 : \mu < \mu_0$.

Grâce à \mathcal{R} , on trouve la valeur suivante de la statistique

$$t = \frac{\bar{m} - \mu_0}{sd/\sqrt{n}} = 10.628909$$

La probabilité critique $P(T \leq t)$ (pour la loi de Student à $ddl = 154$ degrés de libertés) est égale à

$$p_c = 1$$

Puisque p_c est strictement supérieure au niveau de signification 0.05, on accepte l'hypothèse nulle H_0 . Ainsi, H_0 est vraie et donc *la moyenne est égale à $\mu_0 = 65$* , au risque 0.05.

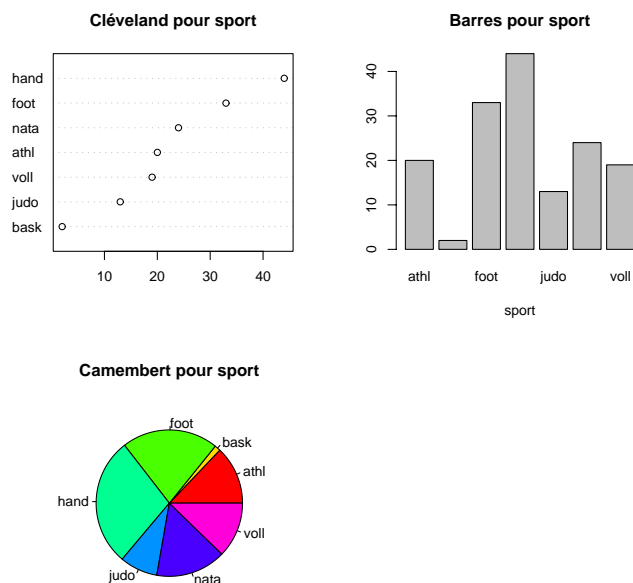
Correction de l'exercice 2.

- (1)
- On étudie la variable qualitative (ou catégorielle) 'sport'. On procèdera donc comme dans le chapitre 5 du document de cours.
 - Les effectifs et les pourcentages déterminés par \mathcal{R} sont donnés dans le tableau suivant
 - On peut aussi afficher les effectifs et les pourcentages ordonnés : Les effectifs et les pourcentages déterminés par \mathcal{R} sont donnés dans le tableau suivant

	effectifs	pourcentages
athl	20	12.903
bask	2	1.290
foot	33	21.290
hand	44	28.387
judo	13	8.387
nata	24	15.484
voll	19	12.258

	effectifs	pourcentages
bask	2	1.290
judo	13	8.387
voll	19	12.258
athl	20	12.903
nata	24	15.484
foot	33	21.290
hand	44	28.387

•



Voir les trois graphiques (ordonnés) ci-dessus pour la variable 'sport'.

- (2) On pourra aussi consulter la correction de la question 2 de l'exercice de l'examen CCF de statistiques du 8 décembre 2008 (disponible sur <http://utbmjb.chez-alice.fr/UFRSTAPS/index.html>).

Cela n'était pas traité explicitement dans le cours, mais en appliquant les idées des chapitres 5 et 6 du document de cours avec un peu d'astuce, on pouvait s'en sortir! On renvoie aussi aux pages 203 et 304 (définition 12.5) de l'ouvrage de S. Chamepely [Cha04].

Si 'MORPHOSPORT' est le nom de la variable dans laquelle on a importé le fichier 'MORPHOSPORT.xls', on obtient le nombre et la proportion de judokas en tapant sous Rgui

```
sum(MORPHOSPORT$sport=="judo")
sum(MORPHOSPORT$sport)
pi<-sum(MORPHOSPORT$sport=="judo")/length(MORPHOSPORT$sport)
```

La proportion observée de judokas est donc $\pi = 0.083871$.

On choisit un risque $\alpha = 0.05$. Nous allons adapter la définition 6.28 page 78 du document de cours à une variable aléatoire qui n'est pas normale mais qui suit une loi binomiale de paramètres $\pi = 0.083871$ et $n = 155$. Comme dans la section 5.7 page 40, on approchera la loi suivie par la proportion par une loi normale (puisque n est "grand").

On fait donc l'hypothèse H_0 que la proportion de succès (observée) π suit une loi normale de moyenne $\mu = \pi_0$ et d'écart-type $\sqrt{\pi_0(1-\pi_0)/n}$. Ainsi la statistique

$$z = \frac{\pi - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

suit une loi normale centrée réduite. Puisque l'on veut montrer que $\pi > \pi_0 = 0.07$, l'hypothèse alternative choisie est $H_1 : \pi > \pi_0$. Comme dans la définition 6.28 page 78 du document de cours (ou définition 12.5 page 204 de [Cha04]), on calcule la probabilité que $P(Z > z)$, où Z suit une loi normale centrée réduite.

Sous \mathbb{R} , on obtient successivement

$$z = \frac{\pi - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.083871 - 0.07}{\sqrt{\frac{0.07 \times (1-0.07)}{155}}} = 0.676834,$$

puis

$$p_c = 0.24926.$$

Enfin, comme dans la définition 6.28 page 78 du document de cours, on constate que $p_c \geq \alpha = 0.05$; ainsi, on acceptera H_0 donc la proportion observée est, au risque 0.05, égale à 0.07.

Correction de l'exercice 3.

On renvoie à la section 6.5 page 75 du chapitre 6 page 61. On souhaite tester la moyenne du groupe des ampoules par rapport à la norme $\mu_0 = 750$. On veut vérifier que le groupe des ampoules est "sous" la moyenne (pour le retirer le cas échéant). On teste donc l'hypothèse nulle $H_0 : \mu = \mu_0$ contre $H_1 : \mu < \mu_0$. On choisit le seuil $\alpha = 0.1$.

On procède au *test de Student pour un échantillon* (à la moyenne).

On fait l'hypothèse nulle $H_0 : \mu = \mu_0$ avec $\mu_0 = 750$. On cherche à montrer que la moyenne de la loi normale, dont proviendraient les données de l'échantillon étudié, est plus petite que μ_0 . On fait donc l'hypothèse alternative suivante : $H_1 : \mu < \mu_0$.

Grâce à \mathbb{R} , on trouve la valeur suivante de la statistique

$$t = \frac{m - \mu_0}{sd/\sqrt{n}} = -3.162278$$

La probabilité critique $P(T \leq t)$ (pour la loi de Student à $ddl = 9$ degrés de libertés) est égale à

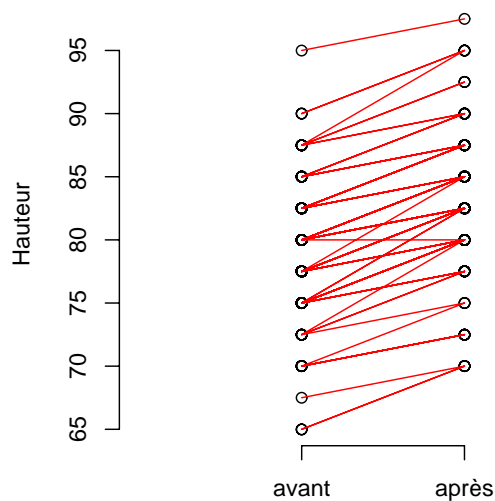
$$p_c = 0.00575399$$

Puisque p_c est inférieure au égal au niveau de signification 0.05, on rejette l'hypothèse nulle H_0 . Ainsi, H_1 est vraie et *la moyenne est plus petite que $\mu_0 = 750$* , au risque 0.05. La durée de vie des ampoules entreposées est donc significativement réduite.

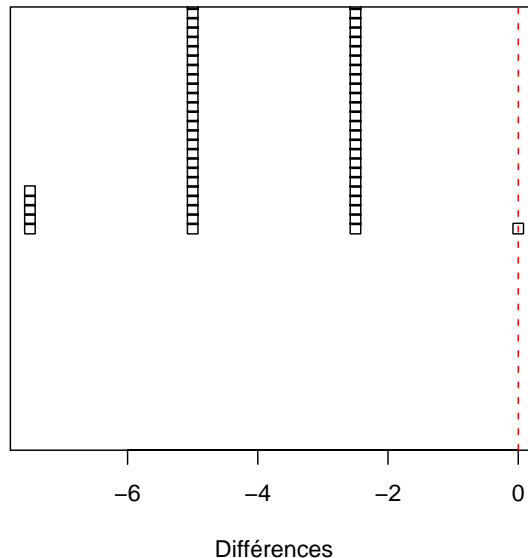
Correction de l'exercice 4.

(1) On renvoie au chapitre 7 page 89 du polycopié de cours.

- Grâce à la fonction `parallelplot.R`, on peut tracer le graphe des différences ci-dessus.



On pouvait aussi afficher une ligne de point



Sur ces graphes, il semblerait que les différences soient négatives.

- Confirmons cela par un test d'hypothèse. On veut montrer que la hauteur de saut a augmenté significativement. On choisit le seuil conventionnel $\alpha = 0.05$.

On procède au *test de Student apparié* (à la différence des moyennes).

On fait l'hypothèse nulle $H_0 : \mu_x - \mu_y = 0$. On cherche à montrer que la moyenne de la loi normale, dont proviendraient les différences entre le premier et le second échantillon (qui sont appariés) est strictement négative. On fait donc l'hypothèse alternative suivante :

$$H_{a(2)} : \mu_x - \mu_y < 0$$

Grâce à \mathbb{R} , on trouve la valeur suivante de la statistique

$$t = \frac{m_d}{sd_d/\sqrt{n}} = -27.566358$$

La probabilité critique $P(T \leq t)$ (pour la loi de Student à $ddl = 99$ degrés de libertés) est égale à

$$p_c = 1.51954e - 48$$

Puisque p_c est inférieure au égal au niveau de signification 0.05, on rejette l'hypothèse nulle H_0 . Ainsi, H_1 est vraie et la moyenne μ_x du premier échantillon est strictement inférieure à celle du second échantillon μ_y , au risque 0.05. Ainsi, le stage semble être positif!

- (2) En un an, si la taille augmente de 4 cm en un an, elle augmentera en moyenne de $4 * 3/4 = 3$ cm entre septembre 2005 et juin 2006, augmentation qui semble être inférieure à la moyenne des différences de hauteur de saut (égale à 4.1)! Autrement dit, la progression enregistrée à la suite du stage n'est peut-être due qu'à la croissance et non aux vertus de l'enseignement du sport au collège!

Pour être sûr de la progression, il faudrait faire par exemple un test sur la hauteur de saut obtenue par normalisation en la divisant par la taille de la jeune fille.

Références

[Cha04] Stéphane Champely. *Statistique vraiment appliquée au sport*. de Boeck, 2004. disponible à la BU de Lyon I sous la cote 519.5 CHA.