

MAT 2779, Introduction à la biostatistique

Devoir 4 - solutionnaire

Il y a 7 questions pour un *total* de 100 points.

Partie I)

Question 1) [5 points] Soit p le vrai taux d'efficacité de ce nouveau médicament. Nous voulons tester

$$H_0 : p = 0,66 \quad \text{contre} \quad H_1 : p > 0,66.$$

La valeur observée de la statistique du test est

$$z_0 = \frac{\hat{p} - 0,66}{\sqrt{0,66(1 - 0,66)/n}} = 1,10,$$

où $n = 170$ et le taux d'efficacité de l'échantillon est $\hat{p} = 119/170 = 0,7$. C'est une alternative à la droite, alors la valeur P est $P(Z > 1,10) = 1 - \Phi(1,10) = 1 - 0,8643 = 0,1357$. À $\alpha = 0,1$, on a $P > \alpha$, alors les preuves contre H_0 ne sont pas significatives. Nous ne pouvons pas conclure que le nouveau médicament est plus efficace.

Question 2) [5 points] Soit μ la poids moyen post-éruption (en lb). On veut tester

$$H_0 : \mu = 1,9 \quad \text{contre} \quad H_1 : \mu > 1,9.$$

Puisque la population est normale (avec σ inconnu), on utilise une statistique t . Sa valeur observée est

$$t_0 = \frac{\bar{x} - 1,9}{s/\sqrt{n}} = 0,547,$$

où $\bar{x} = 2,3$, $s = 4$, et $n = 30$. Puisque c'est une alternative unilatérale à la droite, alors la valeur P est $P(T > 0,547)$, où T suit une loi $T(29)$. Du Tableau 17.4, on a $P(T > 0,256) = 0,40$ et $P(T > 0,683) = 0,25$. Alors, $0,25 < P(T > 0,547) < 0,40$ et $0,25 < P < 0,40$.

À $\alpha = 0,05$, on a $P > \alpha$. Alors, les preuves contre H_0 ne sont pas significatives. On ne peut pas conclure que le poids moyen post-éruption est supérieure à 1,9 lb.

Question 3) [20 points]

(a) On veut

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 1,3583$$

(b) Pour tester $H_0 : \mu_1 - \mu_2 = 0$ contre $H_0 : \mu_1 - \mu_2 \neq 0$, où μ_i la quantité moyenne d'acide ascorbique pour la marque i , on utilise un statistique t . Sa valeur observée est

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} = -6,05.$$

C'est une alternative bilatérale, alors la valeur P est $2P(T > |-6,05|) = 2P(T > 6,05)$, où T suit une loi $T(n_1 + n_2 - 2) = T(28)$. Du Tableau 17.4, on a $P(T > 2,763) = 0,005$. Alors, $P(T > 6,05) < 0,005$ et ceci implique que $P < 2(0,005) = 0,01$.

À $\alpha = 0,05$, on a $P < \alpha$. Alors, les preuves contre H_0 sont significatives. On peut conclure que les deux marques ont des quantités moyennes différentes d'acide ascorbique par tablette.

(c) Un IC à 95% pour $\mu_1 - \mu_2$ est

$$\bar{x}_1 - \bar{x}_2 \pm t_{0,025;n_1+n_2-2} s_p \sqrt{1/n_1 + 1/n_2} = [-4,01, -1,98],$$

où $t_{0,025;n_1+n_2-2} = t_{0,025;28} = 2,048$.

Question 4) [20 points] On veut tester $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. Puisqu'on suppose que D est normalement distribuée, alors on utilise une statistique t . Sa valeur observée est

$$t_0 = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = 2,71,$$

où $\bar{d} = 0,06$, $s_d = 0,07$ et $n = 10$. Puisque l'alternative est à la droite, alors la valeur P est $P(T > 2,71)$, où T suit une loi $T(9)$. Du Tableau 17.4, on a $P(T > 2,262) = 0,025$ et $P(T > 2,821) = 0,01$. Alors, $0,01 < P(T > 2,71) < 0,025$ et $0,01 < P < 0,025$.

À $\alpha = 0,01$, on a $P > \alpha$. Les preuves contre H_0 ne sont pas significatives. Alors, on ne peut pas conclure que l'utilisation des engrais traditionnels donne un rendement plus élevé en moyenne.

Question 5) [20 points]

- (a) Soit μ_1 le débit salivaire moyen de la population de patients avec une paralysie cérébrale et μ_2 le débit moyen de la population des sujets du groupe témoin (sans une paralysie cérébrale). On veut tester $H_0 : \mu_1 - \mu_2 = 0$ contre $H_1 : \mu_1 - \mu_2 \neq 0$.
- (b) Il y a des tendances linéaires dans les diagrammes de probabilité à l'échelle normale. Alors, c'est raisonnable de supposer que les populations sont normales. Le débit salivaire est beaucoup plus dispersé pour le groupe témoin en comparaison au débit salivaire chez les patients avec une paralysie cérébrale. On observe ceci à la fois dans les diagrammes à boîte et moustaches comparatifs (c'-à-d les tailles des boîtes sont très différentes) et dans les diagrammes de probabilité à l'échelle normale superimposés (c'-à-d les pentes sont très différentes). Il n'est donc pas raisonnable de supposer que les variances des population sont égaux.
- (c) De la partie (b), on suppose que les populations sont normales avec des variances égaux. Alors, nous allons utiliser l'approximation de Welch. La valeur observée de la statistique du test est

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = -3,04,$$

où $\bar{x}_1 = 0,53$, $\bar{x}_2 = 0,63$, $s_1 = 0,04$, $s_2 = 0,1$, $n_1 = 20$ et $n_2 = 10$. C'est une alternative bilatérale, alors la valeur P est $2P(T > |-3,04|) = 2P(T > 3,04)$, où T suit approximativement une loi $T(\nu)$, où

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} = 10,47.$$

En arrondissant vers le bas, on utilise $\nu = 10$ degrés de liberté. Du Tableau 17.4, on a $P(T > 2,624) = 0,01$ et $P(T > 3,169) = 0,005$. Alors, $0,005 < P(T > 3,04) < 0,01$. Alors, $0,01 < P < 0,02$.

À $\alpha = 5\%$, on a $P < \alpha$. Alors, les preuves contre H_0 sont significatives. On peut conclure que ces deux groupes ont un débit salivaire moyen de la population différent.

(d) Un IC à 95% pour $\mu_1 - \mu_2$ est

$$\bar{x}_1 - \bar{x}_2 \pm t_{0,025;10} \sqrt{s_1^2/n_1 + s_2^2/n_2} = [-0,173; -0,027],$$

où $t_{0,025;10} = 2,228$.

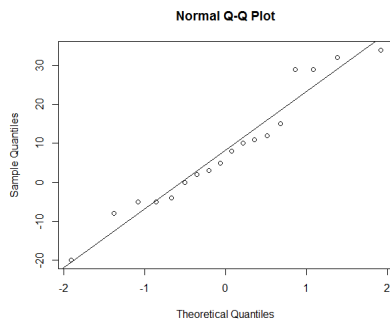
Partie II)

Question 6) [15 points]

(a) Nous importons les données du fichier `triglyceride-FR.txt` et nous attribuons ces données au jeu de données `donnees`. On affiche les noms des colonnes et on attribue les colonnes aux vecteur numériques `x` et `y`, respectivement. En outre, on calcul la différence `d=x-y` et on produit un diagramme quantile-quantile pour la différence des mesures appariées.

```
> donnees=read.table(file.choose(),header=TRUE,sep="\t")
> names(donnees)
[1] "Pre" "Post"
> x=donnees$Pre
> y=donnees$Post
> d=x-y
> qqnorm(d)
> abline(mean(d),sd(d))
```

Le diagramme quantile-quantile pour la différence entre les mesures pré-intervention et post-intervention est ci-bas. Il y a une tendance linéaire dans le diagramme. Alors, c'est raisonnable de suppose que la différence des mesures appariées est normalement distribuée.



- (b) On veut vérifier l'hypothèse de recherche que les niveaux de triglycérides vont diminuer en moyenne. Alors, nous testons $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$. Puisque la différence D suit une loi normale, alors on peut utiliser un test- T pour des mesures appariées. La valeur P du test est 0,01676.

```
> t.test(x,y,paired=TRUE,alternative="greater")
```

Paired t-test

```
data: x and y
t = 2.3129, df = 17, p-value = 0.01676
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.03802      Inf
sample estimates:
mean of the differences
      8.222222
```

À $\alpha = 0,05$, on a $P < \alpha$. Alors, nous avons des preuves significatives contre H_0 . On peut conclure que les niveaux de triglycérides diminuent en moyenne après l'intervention.

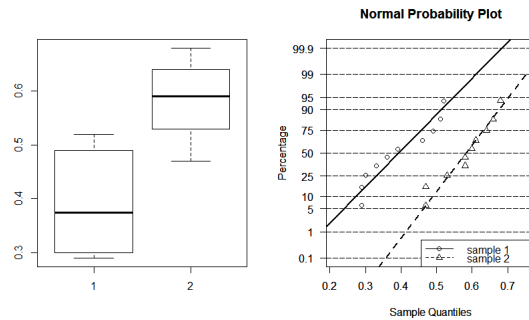
- (c) Nous sommes 95% que μ_D est entre 0,72 and 15,7.

```
> t.test(x,y,paired=TRUE)$conf.int
[1] 0.7219486 15.7224958
attr(,"conf.level")
[1] 0.95
```

Question 7) [15 points]

- (a) Ci-bas, nous trouvons des diagrammes à boîte et moustaches comparatifs et des diagrammes de probabilités à l'échelle normale superimposés. Il y a des tendances linéaires dans les diagrammes de probabilité avec des petites déviations dans les extrémités pour le groupe témoin. C'est raisonnable de supposer que les deux populations sont normales. Le groupe

témoin est légèrement plus dispersé par rapport au groupe de traitement, mais la différence de variabilité n'est pas frappante. De plus, les droites dans les diagrammes de probabilités sont approximativement parallèles. Nous n'avons pas de fortes preuves solides contre la condition d'égalité des variances. Il est raisonnable de supposer que les variances de ces deux populations sont égaux.



- (b) Soit μ_1 le vrai poids moyen d'une tige pour le groupe témoin et soit μ_2 le vrai poids moyen de la tige chez les tiges qui reçoivent l'azote. Nous voulons tester

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{et} \quad H_1 : \mu_1 - \mu_2 < 0.$$

De la partie (a), on suppose que les populations sont normales de variances égaux. Alors, on utilise le test t de Student à deux échantillons. La valeur P est $4,5 \times 10^{-5}$. Puisque $P < \alpha$, à $\alpha = 5\%$, les preuves contre H_0 sont significatives. Nous pouvons conclure que le traitement d'azote produit des tiges de poids supérieures en moyenne en comparaison au traitement sans azote.

```
> t.test(x1,x2,var.equal=TRUE,alternative="less")
```

```
Two Sample t-test
```

```
data: x1 and x2
t = -5.0126, df = 18, p-value = 4.517e-05
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.122963
sample estimates:
mean of x mean of y
 0.394      0.582
```

Remark: Si votre conclusion de la partie (a) est que les populations ont des variance inégaux, alors vous devriez utiliser l'approximation de Welch. Ceci veut dire ne pas utiliser l'argument `var.equal=TRUE` dans la fonction `t.test`.

```
> t.test(x1,x2,alternative="less")
```

```
Welch Two Sample t-test
```

```
data: x1 and x2
t = -5.0126, df = 17.002, p-value = 5.335e-05
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
    -Inf -0.1227554
sample estimates:
mean of x mean of y
    0.394    0.582
```

- (c) Si on n'utilise pas l'argument `alternative` dans la fonction `t.test`, on obtient un intervalle de confiance à 95% pour $\mu_1 - \mu_2$. Nous sommes confiant à 95% que $\mu_1 - \mu_2$ est entre -0,267 and -0,109.

```
> t.test(x1,x2,var.equal=TRUE)$conf.int
[1] -0.2667962 -0.1092038
attr(,"conf.level")
[1] 0.95
```

Remarque: Si votre conclusion de la partie (a) est que les populations ont des variance inégaux, l'intervalle est la même que ci-haut. Mais, vous ne devrez pas avoir l'argument `var.equal=TRUE` dans la fonction `t.test`.

```
> t.test(x1,x2)$conf.int
[1] -0.2671293 -0.1088707
attr(,"conf.level")
[1] 0.95
```