Université de Batna

Faculté de médecine

Corrigé type de TD 11 de Bio-statistique

(Anova 1 facteur)

Remarque : Dans les solutions des exercices, pour la comparaison de plusieurs variances on utilise la 2ème méthode pour <u>le calcul numérique</u> par SPSS et la 3ème méthode ou la 4ème méthode pour <u>le calcul à la main.</u>

<u>Règle :</u>
Attention ! Il faut toujours mettre la variance la plus grande au numérateur.

<u>Exercice 1 :</u>

a) Les conditions de validité de test de l'Anova :
   - L'indépendance des échantillons
   - Normalité de la distribution des mesures
   - L'homogénéité des variances (Variances égales)

b) Test d'égalité des variances : trois variances, donc on teste les variances deux à deux :

	N	Moyennes	Variances S
Sans substance	5	1.6	1.8
Avec l'alumine	6	4	2
Avec phosphates	5	2.6	1.3

- 1er échantillon et 2ème échantillon :
  1. **Choix des hypothèses :**

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

  2. **Calcul de la statistique de test observée :**$T_0 = \frac{S_2^2}{S_1^2} = \frac{2}{1.8} = 1.11$
  3. **Le seuil critique de Fisher-Snedecor :** $\boldsymbol{F_{5;4;0.95}} = \mathbf{6.26}$
  4. **Décision :** $\boldsymbol{F_{5;4;0.95}} > \mathrm{T}_0$

  On accepteH_0, variances égales.

- 2ème échantillon et 3ème échantillon :

1) **Choix des hypothèses :**

$$H_0 : \sigma_2^2 = \sigma_3^2$$
$$H_1 : \sigma_2^2 \neq \sigma_3^2$$

2) **Calcul de la statistique de test observée :** $T_0 = \frac{S_2^2}{S_3^2} = \frac{2}{1.3} = 1.54$
3) **Le seuil critique de Fisher-Snedecor :** $\boldsymbol{F_{5;4;0.95}} = \mathbf{6.26}$
4) **Décision :** $\boldsymbol{F_{5;4;0.95}} > T_0$

On accepte H_0, variances égales.

On confirme enfin par un test l'égalité des variances du 1[ier] et du 3[ème] échantillon.

c) **<u>Test de comparaison des moyennes (Anova 1 facteur) :</u>**

- Substance : Facteur (variable indépendante qualitative) avec 3 niveaux ou échantillons.
- Quantité d'anticorps : Variable dépendante expliquée quantitative

1) **Choix des hypothèses :**

H_0 : Les moyennes égales

H_1 : Au moins l'une des trois est différente.

2) **Calcul de la statistique de test observée :**

Tableau de calcul à la main :

	Sans substance	Avec l'alumine	Avec des phosphates	
x_{ij}	1 ; 3 ; 3 ; 0 ; 1	2 ; 4 ; 5 ; 4 ; 3 ; 6	1 ; 4 ; 2 ; 3 ; 3	
N_i	5	6	5	N=16
$x_{i.} = \sum_{j=1}^{N_i} x_{ij}$	8	24	13	$x_{..} = \sum_{i=1}^{3}\sum_{j=1}^{N_i} x_{ij}$=45
$x_{i.}^2$	64	576	169	
$\frac{x_{i.}^2}{N_i}$	12.8	96	33.8	$\sum_{i=1}^{3}\frac{x_{i.}^2}{N_1}$ =142.6
$\sum_{j=1}^{N_i} x_{ij}^2$	20	106	39	$\sum_{i=1}^{3}\sum_{j=1}^{N_i} x_{ij}^2$= 165

$$\mathrm{SCE}_t = \sum_{i=1}^{3}\sum_{j=1}^{N_i} x_{ij}^2 - \frac{x_{..}^2}{N} = 165 - \frac{45^2}{16} = 38.4375.$$

$$SCE_{fa} = \sum_{i=1}^{3} \frac{x_{i.}^2}{N_i} - \frac{x_{..}^2}{N} = 142.6 - \frac{45^2}{16} = 16.0375$$
$$SCE_r = SCE_T - SCE_{fa} = 22.4$$

Alors :

$$T_0 = \frac{CM_{fa}}{CM_r} = \frac{SCE_{fa}/_{k-1}}{SCE_r/_{N-k}} = \frac{16.0375/_2}{22.4/_{13}} = 4.65$$

3) Identification de seuil critique **:**

$\boldsymbol{F_{(k-1);(N-k);0.95} = F_{2;13;0.95} = \frac{3.89+3.68}{2}}$ **=3.785**

4) Décision : $\boldsymbol{T_0 > F_{2;13;0.95}}$

On rejette H_0, alors au moins une moyenne est différente, c'est-à-dire la présence ou non de la substance influe sur l'efficacité du vaccin.

- Maintenant, ce qui concerne la 2ème partie de la question c, on va faire un test de l'anova 1 facteur :

<u>Test de comparaison de deux moyennes (Anova 1 facteur) :</u>

- Substance ajoutée : Facteur (variable indépendante qualitative) avec 2 niveaux ou échantillons {avec de l'alumine ; avec des phosphates}
- Quantité d'anticorps : Variable dépendante expliquée quantitative

1) **Choix des hypothèses :**

H_0: Les moyennes égales

H_1: Les moyennes différentes

2) Calcul de la statistique de test observée :

Tableau de calcul à la main :

	Avec l'alumine	Avec des phosphates	
x_{ij}	2 ; 4 ; 5 ; 4 ; 3 ; 6	1 ; 4 ; 2 ; 3 ; 3	
N_i	6	5	N=11
$x_{i.} = \sum_{j=1}^{N_i} x_{ij}$	24	13	$x_{..} = \sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}$=37
$x_{i.}^2$	576	169	
$\frac{x_{i.}^2}{N_i}$	96	33.8	$\sum_{i=1}^{2} \frac{x_{i.}^2}{N_1}$ =129.8

$\sum_{j=1}^{N_i} x_{ij}^2$	106	39	$\sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}^2$ = 145

$$\text{SCE}_t = \sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}^2 - \frac{x_{..}^2}{N} = 145 - \frac{37^2}{11} = 20.545$$

$$\text{SCE}_{\text{fa}} = \sum_{i=1}^{2}\frac{\text{x}_{\text{i.}}^2}{\text{N}_\text{i}} - \frac{\text{x}_{..}^2}{\text{N}} = 129.8 - \frac{37^2}{11} = 5.345$$

$$\text{SCE}_r = SCE_T - SCE_{fa} = 15.2$$

Alors :

$$\text{T}_0 = \frac{\text{CM}_{fa}}{\text{CM}_\text{r}} = \frac{\text{SCE}_{\text{fa}}/_{\text{k}-1}}{\text{SCE}_\text{r}/_{\text{N}-\text{k}}} = \frac{5.345/_1}{15.2/_9} = 3.165$$

3) Identification de seuil critique **:**

$$\boldsymbol{F_{(k-1);(N-k);0.95} = F_{1;9;0.95} = 5.12}$$

4) Décision : $\boldsymbol{T_0 < F_{1;9;0.95}}$

On accepteH_0, alors les moyennes sont égales, c'est-à-dire la nature de la substance n'influe pas sur l'efficacité du vaccin.

Tableau de variation (Anova par SPSS)

	Somme des carrés	ddl	Carré moyen	F	Signification
Intergroupes	5.345	1	5.345	3.165	0.109
Intragroupes	15.2	9	1.689		
Total	20.545	10			

Exercice 2 :

- **Test de comparaison des moyennes (Anova 1 facteur) :**
  - Traitement : Facteur (variable indépendante qualitative) avec 3 niveaux ou échantillons.
  - Temps séparant la prochaine crise d'asthme : Variable dépendante expliquée quantitative

Les conditions de validité de test de l'Anova : **Satisfaites par énoncé.**

- L'indépendance des échantillons
- Normalité de la distribution des mesures
- L'homogénéité des variances (Variances égales)

1) **Choix des hypothèses :**

H_0: Les moyennes égales

H_1: Au moins l'une des trois est différente.

2) Calcul de la statistique de test observée :

Tableau de calcul par SPSS :(Anova 1 facteur)

	Somme des carrés	ddl	Carré moyen	F	Signification
Intergroupes	1398.572	2	699.286	**5.32**	.009
Intragroupes	5258.033	40	131.451		
Total	6656.605	43			

$$\mathrm{SCE}_{\mathrm{fa}} = 1398.572$$
$$\mathrm{SCE}_r = 5258.033$$

Alors :

$$\mathrm{T}_0 = \frac{\mathrm{CM}_{fa}}{\mathrm{CM}_\mathrm{r}} = \frac{\mathrm{SCE}_{\mathrm{fa}}/_{\mathrm{k}-1}}{\mathrm{SCE}_\mathrm{r}/_{\mathrm{N}-\mathrm{k}}} = \frac{699.286}{131.451} = 5.32$$

3) Identification de seuil critique **:**
$\boldsymbol{F}_{(\boldsymbol{k}-\mathbf{1});(\boldsymbol{N}-\boldsymbol{k});\mathbf{0.95}} = \boldsymbol{F}_{\mathbf{2;40;0.95}}$ =**3.23**

4) Décision : $\boldsymbol{T}_{\mathbf{0}} > \boldsymbol{F}_{\mathbf{2;40;0.95}}$
On rejetteH_0, alors au moins une moyenne est différente.
On conclut que les traitements ont une efficacité différente.

<u>Exercice 3 :</u>

Enceinte	Non enceinte
$\overline{X}_A$=4,29 $\mathrm{S}_\mathrm{A}^2 = 0.65$	$\overline{\mathrm{X}}_\mathrm{B}$=2.23 $\mathrm{S}_\mathrm{B}^2 = 0.44$

a) Test de comparaison de deux moyennes : Anova 1 facteur

ON doit réaliser un test d'homogénéité des variances :

1. **Choix des hypothèses :**

$$H_0: \sigma_A^2 = \sigma_B^2$$
$$H_1: \sigma_A^2 \neq \sigma_B^2$$

2. **Calcul de la statistique de test observée :**$T_0 = \frac{S_A^2}{S_B^2} = \frac{0.65}{0.44}$ =1.48
3. **Le seuil critique de Fisher-Snedecor :** $\boldsymbol{F}_{\mathbf{12;14;0.95}} = \frac{\mathbf{2.69+2.48}}{\mathbf{2}}$**= 2.585**

Décision : $\boldsymbol{F}_{\mathbf{12;14;0.95}} > \mathrm{T}_0$

On accepteH_0, variances égales *

En revenant au test de comparaison :

- Etat de femme : Facteur (variable indépendante qualitative) avec 2 niveaux ou échantillons {enceintes non enceintes}
- Activité de l'enzyme : Variable dépendante expliquée quantitative

- Les conditions de validité de test de l'Anova :
  - L'indépendance des échantillons **(Satisfaite par énoncé)**
  - Normalité de la distribution des mesures **(Satisfaite par énoncé)**
  - L'homogénéité des variances (Variances égales) **(Satisfaite par*)**

Tableau de calcul à la main :

	Enceinte	Non enceinte	
x_{ij}	4.2 ; 5.5 ; 4.6 ; 5.4 ; 3.9 ; 5.4 ; 2.7 ; 3.9 ; 4.1 ; 4.1 ; 4.6 ; 3.9 ; 3.5	1.5 ; 1.6 ; 1.4 ; 2.9, 2.2, 1.8, 2.7 ; 1.9, 2.2 ; 2.8 ; 2.1 ; 1.8 ; 3.7 ; 1.8 ; 3.1	
N_i	13	15	N = 28
$x_{i.} = \sum_{j=1}^{N_i} x_{ij}$	55.8	33.5	$x_{..} = \sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}$=89.3
$x_{i.}^2$	3113.64	1122.25	
$\frac{x_{i.}^2}{N_i}$	239.51	74.82	$\sum_{i=1}^{2}\frac{x_{i.}^2}{N_1}$ = 314.33
$\sum_{j=1}^{N_i} x_{ij}^2$	247.32	81.03	$\sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}^2$=328.35

$$\mathrm{SCE}_t = \sum_{i=1}^{2}\sum_{j=1}^{N_i} x_{ij}^2 - \frac{x_{..}^2}{N} = 328.35 - \frac{89.3^2}{28} = 43.55$$

$$\mathrm{SCE}_{\mathrm{fa}} = \sum_{i=1}^{2}\frac{\mathrm{x}_{\mathrm{i.}}^2}{\mathrm{N_i}} - \frac{\mathrm{x}_{..}^2}{\mathrm{N}} = 314.33 - \frac{89.3^2}{28} = 29.53$$

$$\mathrm{SCE}_r = SCE_T - SCE_{fa} = 14.02$$

$$T_0 = \frac{CM_{fa}}{CM_r} = \frac{SCE_{fa}/_{k-1}}{SCE_r/_{N-k}} = \frac{29.53/_1}{14.02/_{26}} = 54.76$$

4) Identification de seuil critique **:**

- Significative : $\alpha = 0.05$

  $\boldsymbol{F_{(k-1);(N-k);0.95} = F_{1;26;0.95} = 4.23}$

  4) Décision : $\boldsymbol{T_0 > F_{1;26;0.95}}$
  On rejetteH_0, les moyennes sont différentes.

- Hautement Significative : $\alpha = 0.01$
  $\boldsymbol{F_{(k-1);(N-k);0.95} = F_{1;26;0.99} = 7.72}$

  Décision : $\boldsymbol{T_0 > F_{1;26;0.95}}$

<u>Décision statistique</u> : La statistique de test se trouve dans la zone de rejet de H_0 pour $\alpha = 0.05\ et\ \alpha = 0.01.$

<u>Décision pratique</u> : La différence est hautement significative entre les deux populations.

(La grossesse a une influence hautement significative sur l'activité de la PDE.

b) **<u>Test de comparaison de 5 moyennes (Anova 1 facteur) :</u>**
- Age de grossesse : Facteur (variable indépendante) avec 5 niveaux ou échantillons.
- Dosage de l'enzyme : Variable dépendante expliquée quantitative

- Les conditions de validité de test de l'Anova :(Satisfaites par énoncé)
  - L'indépendance des échantillons
  - Normalité de la distribution des mesures**.**
  - L'homogénéité des variances (Variances égales)

1) **Choix des hypothèses :**
   H_0: Les moyennes égales
   H_1: Au moins l'une des trois est différente.

4) Calcul de la statistique de test observée :

- ***Tableau de calcul par SPSS :(Anova 1 facteur)***

	Somme des carrés	ddl	Carré moyen	F	Signification
Intergroupes	54.094	4	13.524	**5.189**	.009
Intragroupes	143.334	55	2.606		
Total	197.428	59			

$$SCE_{fa} = 54.094$$
$$SCE_r = 143.334$$

Alors :

$$T_0 = \frac{CM_{fa}}{CM_r} = \frac{SCE_{fa}/_{k-1}}{SCE_r/_{N-k}} = \frac{13.524}{2.606} = 5.189$$

5) Identification de seuil critique :

$$\boldsymbol{F_{(k-1);(N-k);0.95} = F_{4;55;0.95} = \frac{2.56+2.53}{2} = 2.545}$$

4) Décision : $\boldsymbol{T_0 > F_{4;55;0.95}}$

On rejette H_0, alors au moins une moyenne est différente.

On conclut que l'âge de grossesse a une influence sur l'activité de l'enzyme.

<u>Exercice 4 :</u>

c) **<u>Test de comparaison de 4 moyennes (Anova 1 facteur) :</u>**

- Milieu nutritif : Facteur (variable indépendante qualitative) avec 4 niveaux ou échantillons.
- Taille des colonies : Variable dépendante expliquée quantitative

- Les conditions de validité de test de l'Anova :
  - L'indépendance des échantillons **(Satisfaite par énoncé)**
  - Normalité de la distribution des mesures **(Satisfaite par énoncé)**
  - L'homogénéité des variances (Variances égales) **(Satisfaite par**)**

** **Test de comparaison de variances :**

$$S^2_{max} = S^2_B = 3.13$$
$$S^2_{max} = S^2_A = 1.31$$

On applique la **méthode approximative :**

1. **Choix des hypothèses :**

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C = \sigma^2_D$$

H_1: Au moins une variance est différente

2. **Calcul de la statistique de test observée :**

$$T_0 = \frac{S_{max}^2}{S_{min}^2} = \frac{3.13}{1.31} = 2.39$$

3. **Le seuil critique de Fisher-Snedecor : $\boldsymbol{F_{7;6;0.95}} = \mathbf{4.21}$**
4. **Décision :**

$$\boldsymbol{F_{7;6;0.95}} > \mathrm{T}_0$$

On accepteH_0, variances égales (homogénéité des variances)

En revenant aux étapes du test de l'Anova :

1) **Choix des hypothèses :**
   H_0: Les moyennes égales
   H_1: Au moins l'une des cinq est différente.
2) **Calcul de la statistique de test observée :**

Tableau de calcul à la main :

	A	B	C	D	
x_{ij}	9.5 11.5 9 12 11.5 10 11	8 6.5 10 7 11.5 9.5 10.5 10	7.5 5 6 5.5 8.5 6.5 9	14 11.5 12 11 13 15	
N_i	7	8	7	6	N=28
$x_{i.} = \sum_{j=1}^{N_i} x_{ij}$	74.5	73	48	76.5	$x_{..} = \sum_{i=1}^{4}\sum_{j=1}^{N_i} x_{ij}$=272
$x_{i.}^2$	5550.25	5329.00	2304.00	5852.25	
$\frac{x_{i.}^2}{N_i}$	792.89	666.125	329.14	975.375	$\sum_{i=1}^{4}\frac{x_{i.}^2}{N_1}$ =2763.53
$\sum_{j=1}^{N_i} x_{ij}^2$	800.75	688	343	987.25	$\sum_{i=1}^{4}\sum_{j=1}^{N_i} x_{ij}^2$= 2819

$$\mathrm{SCE}_t = \sum_{i=1}^{4}\sum_{j=1}^{N_i} x_{ij}^2 - \frac{x_{..}^2}{N} = 2819 - \frac{272^2}{28} = 176.71$$

$$\mathrm{SCE_{fa}} = \sum_{i=1}^{4}\frac{x_{i.}^2}{N_i} - \frac{x_{..}^2}{N} = 2763.53 - \frac{272^2}{28} = 121.24$$

$$\mathrm{SCE}_r = SCE_T - SCE_{fa} = 55.47$$

Alors :

$$\mathrm{T}_0 = \frac{\mathrm{CM}_{fa}}{\mathrm{CM_r}} = \frac{\mathrm{SCE_{fa}}/_{\mathrm{k-1}}}{\mathrm{SCE_r}/_{\mathrm{N-k}}} = \frac{121.24/_3}{55.47/_{24}} = 17.49$$

Tableau de calcul par SPSS :(Anova 1 facteur)

	Somme des carrés	ddl	Carré moyen	F	Signification
Intergroupes	121.25	3	40.417	**17.489**	0.000
Intragroupes	55.464	24	2.311		
Total	176.714	27			

1) Identification de seuil critique **:**

$$\boldsymbol{F_{(k-1);(N-k);0.95} = F_{3;24;0.95} = 3.01}$$

4) Décision : $\boldsymbol{T_0 > F_{3;24;0.95}}$

On rejetteH_0, alors au moins une moyenne est différente.

On conclut que le milieu nutritif à un effet sur le diamètre des colonies.