Université de Picardie Jules Verne

UFR d'économie et de gestion

Probabilités

Licence 2 - Semestre 4

Exercices d'entrainement **Analyse de la variance** **Corrigés**

Correction 1

Population 1 : les hommes dans l'industrie agro-alimentaire
Population 2 : les femmes dans l'industrie agro-alimentaire

Variable X_1 : salaire, variable aléatoire de moyenne μ_1 et de variance σ_1^2 (inconnues).
Variable X_2 : salaire, variable aléatoire de moyenne μ_2 et de variance σ_2^2 (inconnues).

L'échantillon (e_1) de X_1 est de taille $n_1 = 25 + 15 + 15 + 5 + 5 = 65$.
L'échantillon (e_2) de X_2 est de taille $n_2 = 5 + 10 + 15 + 5 + 0 = 35$.

L'estimation ponctuelle de μ_1 est faite par $\overline{x_1}$:
$\overline{x_1} = \dfrac{25 \times 12 + 15 \times 16 + 15 \times 20 + 5 \times 27 + 5 \times 41}{65} = \dfrac{1180}{65} \approx 18,154.$
De plus, $\overline{x_1^2} = \dfrac{25 \times 12^2 + 15 \times 16^2 + 15 \times 20^2 + 5 \times 27^2 + 5 \times 41^2}{65} = \dfrac{25490}{65} \approx 392,15.$
Et donc, $s_{c,1}^2 = \dfrac{65}{64}(392,15 - 18,154^2) \approx 63,57$ qui est un estimateur de σ_1^2.

De même, l'estimation ponctuelle de μ_2 est faite par $\overline{x_2}$:
$\overline{x_2} = \dfrac{5 \times 12 + 10 \times 16 + 15 \times 20 + 5 \times 27 + 0 \times 41}{65} = \dfrac{655}{35} \approx 18,714.$
De plus, $\overline{x_2^2} = \dfrac{5 \times 12^2 + 10 \times 16^2 + 15 \times 20^2 + 5 \times 27^2 + 0 \times 41^2}{65} = \dfrac{12935}{35} \approx 369,2857.$
Et donc, $s_{c,2}^2 = \dfrac{35}{34}(369,2857 - 18,714^2) \approx 19,6218$ qui est un estimateur de σ_2^2.

Test par ANOVA de $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$.

On compare $k = 2$ moyennes avec deux échantillons avec $n = n_1 + n_2 = 65 + 35 = 100$.
De plus, $\overline{x} = \dfrac{1180 + 655}{100} = 18,35.$

On a $s_r^2 = \dfrac{1}{100 - 2}\,(64 \times 63,57 + 34 \times 19,62) \approx 48,32$
Et $s_F^2 = \dfrac{1}{2 - 1}\,(65 \times (18,154 - 18,35)^2 + 35 \times (18,714 - 18,35)^2) \approx 7,1456$
Et donc $f = \dfrac{7,1456}{48,32} \approx 0,14787.$
Pour $\alpha = 0,05$ et $(k - 1, n - k) = (2 - 1, 100 - 2) = (1, 98)$ degrés de liberté, grâce à la table de Fisher-Snédécor, on trouve que $f_{\max}$ est compris entre $3,94$ et $3,96$.
Comme $f < f_\alpha$, on ne peut rejeter H_0 au risque $\alpha = 0,05$: il n'y a pas de différence significative.

Correction 2

Population 1 : La surface traitée avec une concentatrion de 200 mg/l de désherbant.
Population 2 : La surface traitée avec une concentatrion de 300 mg/l de désherbant.
Population 3 : La surface traitée avec une concentatrion de 400 mg/l de désherbant.

Variables X_1, X_2 et X_3 : le temps de désagrégation pour chacune de ces concentrations, variables aléatoires de moyennes respectives μ_1, μ_2 et μ_3 et de variance respectives σ_1^2, σ_2^2 et σ_3^2 (toutes inconnues).

L'échantillon (e_1) de X_1 est de taille $n_1 = 10$, de moyenne $\overline{x_1} = 19$ et de variance corrigée $s_{c,1}^2 = \dfrac{10}{9} \times 18 = 20$.

L'échantillon (e_2) de X_2 est de taille $n_2 = 15$, de moyenne $\overline{x_2} = 21$ et de variance corrigée $s_{c,2}^2 = \dfrac{15}{14} \times 21 = 15$.

L'échantillon (e_3) de X_3 est de taille $n_3 = 5$, de moyenne $\overline{x_3} = 27$ et de variance corrigée $s_{c,2}^2 = \dfrac{5}{4} \times 24 = 30$.

Test de Bartlett (nécessaire pour mettre en œuvre l'ANOVA puisque l'on n'a aucune information sur les variances) : on trouve $\lambda \approx 1,066$ et $b \approx 0,81$. Grâce à la table du χ^2 avec $d.l.l = 2$ et $\alpha = 0,005$, on trouve $b_{\max} = 5,991$. Puisque $b < b_{\max}$, on peut supposer que les variances σ_1^2, σ_2^2 et σ_3^2 sont égales.

On peut donc tester, par ANOVA, $H_0 : \mu_1 = \mu_2 = \mu_3$ contre H_1 : au moins de deux de ces moyennes sont différentes.

On compare $k = 3$ moyennes avec trois échantillons avec $n = n_1 + n_2 + n_3 = 10 + 15 + 5 = 30$.

De plus, $\overline{x} = \dfrac{10 \times 19 + 15 \times 21 + 5 \times 27}{30} = \dfrac{640}{30} \approx 21,33$.

On a $s_r^2 = \dfrac{1}{30-3}\left(9 \times 20 + 14 \times 15 + 4 \times 30\right) \approx 18,89$

Et $s_F^2 = \dfrac{1}{3-1}\left(10 \times (19 - 21,33)^2 + 15 \times (21 - 21,33)^2 + 5 \times (27 - 21,33)^2\right) \approx 108,33$

Et donc $f = \dfrac{108,33}{18,89} \approx 5,73$.

Pour $\alpha = 0,05$ et $(k-1, n-k) = (3-1, 30-3) = (2, 27)$ degrés de liberté, grâce à la table de Fisher-Snédécor, on trouve $f_{\max} \approx 3,35$.

Comme $f > f_\alpha$, on rejette H_0 au risque $\alpha = 0,05$: la différence de moyenne est significative.

Correction 3

Population 1 : Les caddies en petite surface.
Population 2 : Les caddies en moyenne surface.
Population 3 : Les caddies en grande surface.

Variables X_1, X_2 et X_3 : la valeur du caddie, variables aléatoires de moyennes respectives μ_1, μ_2 et μ_3 et de variance respectives σ_1^2, σ_2^2 et σ_3^2 (toutes inconnues).

L'échantillon (e_1) de X_1 est de taille $n_1 = 10$, de moyenne $\overline{x_1} = \dfrac{5 \times 30 + 3 \times 70 + 2 \times 120}{5+3+2} = 60$ (euros) et de variance corrigée $s_{c,1}^2 = \dfrac{10}{9}\left(\dfrac{5 \times 30^2 + 3 \times 70^2 + 2 \times 120^2}{5+3+2} - 60^2\right) \approx 1333,33$.

L'échantillon (e_2) de X_2 est de taille $n_2 = 40$, de moyenne $\overline{x_2} = 78,75$ (euros) et de variance corrigée $s_{c,2}^2 \approx 1293$.

L'échantillon (e_3) de X_3 est de taille $n_3 = 100$, de moyenne $\overline{x_3} = 72$ (euros) et de variance corrigée $s^2_{c,2} \approx 816$.

Puisque les variances sont supposées égales, on ne fait pas le test de Bartlett.

On teste, par ANOVA, H_0 : $\mu_1 = \mu_2 = \mu_3$ contre H_1 : au moins de deux de ces moyennes sont différentes.

On compare $k = 3$ moyennes avec trois échantillons avec $n = 150$.

De plus, le montant moyen du caddy est $\overline{x} = 73$.

On obtient $s^2_r \approx 979,847$ et $s^2_F \approx 1556,25$. Et donc $f \approx 1,59$.

Pour $\alpha = 0,05$ et $(k-1, n-k) = (3-1, 150-3) = (2, 147)$ degrés de liberté, grâce à la table de Fisher-Snédécor, on trouve que $f_{\max}$ est compris entre $3,09$ et $3,06$.
Comme $f < f_\alpha$, on ne peut pas rejeter H_0 au risque $\alpha = 0,05$: la différence de moyenne n'est pas significative.

Correction 4

Il s'agit d'un test d'Anova à deux facteurs.

	Adjuvant	Age					$\overline{x_{i,j}}$	$\overline{(x_{i,j})^2}$	$s^2_{i,j}$	$s^2_{c,i,j}$
$x_{(1,1)}$	A	< 50	5	6	12	9	8	71,5	7,5	10
$x_{(1,2)}$	A	≥ 50	1	2	5	4	3	11,5	2,5	3,33
$x_{(2,1)}$	B	< 50	7	8	10	11	9	83,5	2,5	3,33
$x_{(1,2)}$	B	≥ 50	4	4	3	9	5	30,5	5,5	7,33

Si l'on considère l'ensemble des données, on obtient :
- la moyenne : $\overline{x} = 6,25$
- la moyenne des carrés : $\overline{x^2} = 49,25$
- la variance corrigée : $s^2_t \approx 10,87$.

Le nombre de modalités de chacun des facteurs est $r = 2$ et $q = 2$. Dans chacune des classes, on a un échantillon de $n = 4$.

De plus, $s^2_r = 6$.
Moyenne "Adjuvant A" : $\overline{x_{1,\star}} = 5,5$.
Moyenne "Adjuvant B" : $\overline{x_{2,\star}} = 7$.
Moyenne "moins de 50 ans" : $\overline{x_{\star,1}} = 8,5$.
Moyenne "50 ans ou plus" : $\overline{x_{\star,2}} = 4$.
On obtient $s^2_A = 9$, $s^2_B = 81$ et $s^2_{AB} = 1$.

1. On a $f_A = 1,5$, $v_1 = r - 1 = 1$, $v_2 = (n-1)rq = 12$ et $f_{\max} = 4,75$.

   Puisque $f_A < f_{\max}$, on ne peut pas rejeter $H_{0,A}$: l'adjuvant A n'a pas d'influence sur le taux d'anticorps.

2. On a $f_B = 13,5$, $v_1 = q - 1 = 1$, $v_2 = (n-1)rq = 12$ et $f_{\max} = 4,75$.

   Puisque $f_B > f_{\max}$, on rejette $H_{0,B}$: l'âge a une influence sur le taux d'anticorps.

3. On a $f_{AB} \approx 0,16$, $v_1 = (r-1)(q-1) = 1$, $v_2 = (n-1)rq = 12$ et $f_{\max} = 4,75$.

   Puisque $f_{AB} < f_{\max}$, on ne peut pas rejeter $H_{0,A}$: il n'y a pas d'interaction significative.

Correction 5

Il s'agit d'un test d'anova à deux facteurs avec $n = 1$. En particulier, pour tout couple (i, j), on a $\overline{x_{i,j}} = x_{i,j}$.
Le type de vaccins possède $r = 4$ modalités et l'âge possède $q = 3$ modalités.

Age \ Vaccin	a	b	c	d	Moyenne $\overline{x_{i,*}}$
20 ans ou moins	80	90	95	95	90
compris entre 21 et 59 ans	75	90	90	85	85
60 ans ou plus	70	75	70	60	68,75
Moyenne $\overline{x_{*,j}}$	75	85	85	80	

On obtient $\overline{x} = 81,25$, $s_A^2 = 68,75$ et $s_B^2 = 493,5$.

$\overline{x_{i,j}} - \overline{x_{i,\star}} - \overline{x_{\star,j}} + \overline{x}$	a	b	c	d
20 ans ou moins	$-3,75$	$-3,75$	$1,25$	$6,25$
compris entre 21 et 59 ans	$-3,75$	$1,25$	$1,25$	$1,25$
60 ans ou plus	$7,5$	$2,5$	$-2,5$	$-7,5$

$s_{AB}^2 \approx 35,42$

1. On a $f_A \approx 1,94$ $v_1 = r - 1 = 3$, $v_2 = (r-1)(q-1) = 6$ et $f_{\max} = 4,76$.

   Puisque $f_A < f_{\max}$, on ne peut pas rejeter $H_{0,A}$: le type de vaccin n'a pas d'influence sur le taux de protection.

2. On a $f_B \approx 13,94$, $v_1 = q - 1 = 2$, $v_2 = (r-1)(q-1) = 6$ et $f_{\max} = 5,14$

   Puisque $f_B \geq f_{\max}$, on rejette $H_{0,B}$: l'âge a une influence sur le taux de protection.