Tests d'hypothèses

Inférence à plusieurs échantillons

Exemple

Un gérontologue étudiant le processus de vieillissement a étudié l'impact du maintien d'une corpulence "mince et en forme" sur la durée de vie. Elle a réparti aléatoirement des rats nouveau-nés sur trois régimes alimentaires:

1. accès illimité à la nourriture,
2. 90% de la quantité normale,
3. 80% de la quantité normale.

Tout au long de leur vie, elle a enregistré leur durée de vie. Les résultats suggèrent-ils un lien entre l'alimentation et la durée de vie?

L'analyse de la variance

Illimité	Régime 90%	Régime 80%
2.5	2.7	3.1
3.1	3.1	2.9
2.3	2.9	3.8
1.9	3.7	3.9
2.4	3.5	4.0

Exemple

Un paléontologue étudiant des fossiles de l'Ordovicien a collecté quatre échantillons du trilobite Paradoxides (membres d'un sous-phylum éteint des Arthropodes) provenant des mêmes strates rocheuses, mais de différents endroits. Il a mesuré la longueur du céphalon (en mm) de tous les spécimens qu'il a collectés et a enregistré les données ci-dessous. Y a-t-il des différences significatives dans la longueur du céphalon entre ces sites?

L'analyse de la variance

	Site			
	I	II	III	IV
n	8	5	11	15
$\bar{X}$	7.0	9.0	8.0	5.0
s	2.1	2.1	2.2	2.2
$\sum X_i$	56.0	45.0	88.0	75.0

Exemple

Un endocrinologue étudiant les effets génétiques et environnementaux sur la production d'insuline des tissus pancréatiques a élevé cinq portées de souris expérimentales. À l'âge de 2 mois, il a sacrifié les souris, disséqué les tissus pancréatiques et traité les spécimens avec une solution de glucose. La quantité d'insuline libérée par ces spécimens a été enregistrée en pg/ml. Y a-t-il des différences significatives dans la libération d'insuline entre les portées?

L'analyse de la variance

portée				
1	**2**	**3**	**4**	**5**
9	2	3	4	8
7	6	5	10	10
5	7	9	9	12
5	11	10	8	13
3	5	6	10	11

ANOVA I

Hypothèses du modèle

1. Les k échantillons représentent des échantillons aléatoires indépendants tirés de k populations avec des moyennes $\mu_1, \mu_2, \ldots, \mu_k$ où $\mu_1, \mu_2, \ldots, \mu_k$ sont des constantes inconnues.
2. Chacune des k populations est normale.
3. Chacune des k populations a la même variance σ^2.

Pour k moyennes il y a

$$\binom{k}{2} = \frac{k\,(k-1)}{2}$$

comparaisons par paires.

ANOVA I

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

H_1 : au moins une paire de μ n'est pas égale.

ANOVA I

Treatment level					
1	2	$\cdots$	i	$\cdots$	k
X_{11}	X_{21}	$\ldots$	X_{i1}	$\ldots$	X_{k1}
X_{12}	X_{22}	$\ldots$	X_{i2}	$\ldots$	X_{k2}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
X_{1n_1}	X_{2n_2}	$\ldots$	X_{in_1}	$\ldots$	X_{kn_k}

Notation:

$$T_{i.} = \sum_{j=1}^{n_i} X_{ij}$$

est la somme de toutes les observations du traitement i. Ainsi, la moyenne de l'échantillon pour le $i^{\text{ème}}$ traitement devient

$$\bar{X}_{i.} = \frac{T_{i.}}{n_i}$$

Deux points sont utilisés pour indiquer une somme sur les deux indices. Par exemple, la somme de toutes les observations est

$$T_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}$$

puisque le nombre total d'observations est $N = \sum_{i=1}^{k} n_i$, alors la moyenne globale est

$$\bar{X}_{..} = \frac{T_{..}}{N}$$

ANOVA I

Une disposition de données étendue pourrait prendre la forme

Treatment level					
1	2	$\dots$	i	$\dots$	k
X_{11}	X_{21}	$\dots$	X_{i1}	$\dots$	X_{k1}
X_{12}	X_{22}	$\dots$	X_{i2}	$\dots$	X_{k2}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
X_{1n_1}	X_{2n_2}	$\dots$	X_{in_1}	$\dots$	X_{kn_k}
$T_{1.}$	$T_{2.}$	$\dots$	$T_{i.}$	$\dots$	$T_{k.}$
$\bar{X}_{1.}$	$\bar{X}_{2.}$	$\dots$	$\bar{X}_{i.}$	$\dots$	$\bar{X}_{k.}$

Voici les statistiques récapitulatives pour l'exemple

Illimité	Régime 90%	Régime 80%	Sommaire
$T_{1.} = 12.2$	$T_{2.} = 15.9$	$T_{3.} = 17.7$	$T_{..} = 45.8$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$N = 15$
$\bar{X}_{1.} = 2.44$	$\bar{X}_{2.} = 3.18$	$\bar{X}_{3.} = 3.54$	$\bar{X}_{..} = 3.05$
$s_1^2 = 0.19$	$s_2^2 = 0.17$	$s_3^2 = 0.25$	

La variabilité totale des 15 observations peut être résumée comme suit:

$$\sum_{i=1}^{3}\sum_{j=1}^{5}\left(X_{ij}-\bar{X}_{..}\right)^{2}$$

Cette somme est appelée somme totale des carrés ou $\mathrm{SC}_{\mathrm{Total}}$.

Definition

L'écart d'une observation par rapport à la moyenne générale peut être divisé en écart de la moyenne de traitement par rapport à la moyenne générale et en écart d'une observation par rapport à sa moyenne de traitement

$$\left(X_{ij}-\bar{X}_{..}\right)=\left(X_{i.}-\bar{X}_{..}\right)+\left(X_{ij}-\bar{X}_{i.}\right)$$

Partitionner les sommes des carrés

$$
\begin{aligned}
\mathrm{SS}_{\mathrm{Total}} &= \sum_i \sum_j \left(X_{ij} - \bar{X}_{..}\right)^2 \\
&= \sum_i \sum_j \left[\left(X_{i.} - \bar{X}_{..}\right) + \left(X_{ij} - \bar{X}_{i.}\right)\right]^2 \\
&= \sum_i \sum_j \left[\left(X_{i.} - \bar{X}_{..}\right)^2 - 2\left(X_{i.} - \bar{X}_{..}\right)\left(X_{ij} - \bar{X}_{i.}\right) + \left(X_{ij} - \bar{X}_{i.}\right)^2\right] \\
&= \sum_i \sum_j \left(X_{i.} - \bar{X}_{..}\right)^2 - \sum_i \sum_j 2\left(X_{i.} - \bar{X}_{..}\right)\left(X_{ij} - \bar{X}_{i.}\right) \\
&\quad + \sum_i \sum_j \left(X_{ij} - \bar{X}_{i.}\right)^2
\end{aligned}
$$

ANOVA I

Puisque $\bar{X}_{i.}$ et $\bar{X}_{..}$ sont constants par rapport a j

$$\sum_i \sum_j 2\left(X_{i.} - \bar{X}_{..}\right)\left(X_{ij} - \bar{X}_{i.}\right) = 2\sum_i \left[\left(X_{i.} - \bar{X}_{..}\right)\sum_j \left(X_{ij} - \bar{X}_{i.}\right)\right]$$

la somme des écarts des observations par rapport a la moyenne de l'échantillon est toujours égale a 0, c'est-à-dire $\sum_j \left(X_{ij} - \bar{X}_{i.}\right) = 0$. Par conséquent

Definition

$\mathrm{SS}_{\mathrm{Total}}$, la variabilité totale, peut être divisée en $\mathrm{SS}_{\mathrm{Trait}}$, la variabilité due aux différences entre les traitements, et $\mathrm{SS}_{\mathrm{Error}}$, la variabilité inexplicable au sein des traitements.

$$\begin{array}{ccccc} \sum_i \sum_j \left(X_{ij} - \bar{X}_{..}\right)^2 & = & \sum_i \sum_j \left(X_{i.} - \bar{X}_{..}\right)^2 & + & \sum_i \sum_j \left(X_{ij} - \bar{X}_{i.}\right)^2 \\ \mathrm{SC}_{\mathrm{Total}} & = & \mathrm{SC}_{\mathrm{Trait.}} & + & \mathrm{SC}_{\mathrm{Erreur}} \end{array}$$

Exemple

Supposons que nous menions une expérience pour tester les effets de trois régimes alimentaires (A, B et C) sur la prise de poids des petits cochons d'Inde. Tous les petits ont été élevés, après le sevrage, pendant quatre semaines avec une alimentation standard pour cochons d'Inde. Ils ont ensuite été répartis au hasard en trois groupes, chacun recevant un régime différent. La prise de poids de chaque individu après six semaines supplémentaires avec les différents régimes est enregistrée en grammes dans le tableau.

ANOVA I

	A	B	C
	X_{1j}	X_{2j}	X_{3j}
	35	22	23
	39	24	30
	27	29	28
	47	25	31
Moyenne	37	25	28
Moyenne générale: 30.0			

Table 1: Gain de poids de chaque chiot cobaye par régime.

ANOVA I

la variabilité totale est calculée comme suit

$$\begin{aligned}
\mathrm{SC}_{\mathrm{Total}} &= \sum_i \sum_j \left(X_{ij} - \bar{X}_{..}\right)^2 \\
&= (35-30)^2 + (39-30)^2 + \cdots + (28-30)^2 + (31-30)^2 \\
&= 584
\end{aligned}$$

Supposons ensuite que dans une version différente de cette expérience, les moyennes de traitement soient les mêmes, mais qu'au sein de chaque groupe de traitement particulier, les effets soient identiques.

	A	**B**	**C**
	$\bar{X}_{1.}$	$\bar{X}_{2.}$	$\bar{X}_{3.}$
	37	25	28
	37	25	28
	37	25	28
	37	25	28
Moyenne	37	25	28
Moyenne générale: 30.0			

ANOVA I

Seule une partie de la variabilité de l'expérience originale demeure. Cette variabilité (réduite) est mesurée par une somme des carrés corrigée.

$$\sum_i \sum_j (X_{i.} - 30)^2 = 4\,(37-30)^2 + 4\,(25-30)^2 + 4\,(28-30)^2 = 312$$

L'analyse finale examine les données originales "corrigées" pour les moyennes de traitement individuelles. Chaque observation est remplacée par son écart par rapport à sa moyenne de traitement $X_{ij} - \bar{X}_{i.}$.

	A	**B**	**C**
	$X_{1j} - 37$	$X_{2j} - 25$	$X_{3j} - 28$
	-2	-3	-5
	2	-1	2
	-10	4	0
	10	0	3
Moyenne	0	0	0
Moyenne générale: 0			

Le traitement et les moyennes globales sont 0. Ce qui reste, c'est la variabilité au sein de chaque traitement. Pourquoi les cobayes ayant suivi le régime A ont-ils pris du poids différemment? Il existe une myriade de raisons possibles, mais elles n'ont pas été étudiées dans cette expérience. Il s'agit d'une variabilité inexpliquée. Elle est également mesurée par une somme des carrés corrigée.

$$\begin{aligned}\sum_i \sum_j \left[\left(X_{ij} - \bar{X}_{i.}\right) - 0\right]^2 &= (-2-0)^2 + (2-0)^2 + \cdots + (0-0)^2 + (3-0)^2 \\ &= 272\end{aligned}$$

car

$$\sum_i \sum_j \left[\left(X_{ij} - \bar{X}_{i.}\right) - 0\right]^2 = \sum_i \sum_j \left(X_{ij} - \bar{X}_{i.}\right)^2$$

Dans l'analyse ci-dessus, il n'y avait que deux sources de variabilité : l'alimentation (traitement) et le bruit (erreur). Ces sources doivent prendre en compte toute la variabilité d'origine, $\mathrm{SS}_{\mathrm{Total}}$. En utilisant les équations ci-dessus, nous trouvons

$$\mathrm{SC}_{\mathrm{Trait}} + \mathrm{SC}_{\mathrm{Erreur}} = 312 + 272 = 584 = \mathrm{SC}_{\mathrm{Total}}$$

Ainsi, nous avons vérifié l'égalité dans cet exemple particulier.

Nous avons les formules de calcul permettant de gagner du temps pour chacune des sommes des carrés en ANOVA:

$$\mathrm{SC}_{\mathrm{Total}} = \sum_i \sum_j X_{ij}^2 - \frac{\left(\sum_i \sum_j X_{ij}\right)^2}{N} = \text{SC non corrigé} - \text{terme de correction}$$

Definition

La variabilité totale peut être calculée comme

$$\mathrm{SC}_{\mathrm{Total}} = \sum_i \sum_j X_{ij}^2 - \frac{T_{..}^2}{N}$$

Ensuite

$$\mathrm{SC}_{\mathrm{Trait}} = \sum_i \left[\frac{\left(\sum_j X_{ij}\right)^2}{n_i} \right] - \frac{\left(\sum_i \sum_j X_{ij}\right)^2}{N}$$

Definition

La variabilité entre les traitements peut être calculée comme

$$\mathrm{SC}_{\mathrm{Trait}} = \sum_i \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N}$$

ANOVA I

Enfin

$$\mathrm{SC}_{\mathrm{Erreur}} = \sum_i \left[\sum_j X_{ij}^2 - \frac{\left(\sum_j X_{ij} \right)^2}{n_i} \right] = \sum_i \mathrm{CSS}_i$$

Definition

La variabilité inexplicable au sein des traitements peut également être calculée comme suit:

$$\mathrm{SC}_{\mathrm{Erreur}} = \sum_i (n_i - 1) \, s_i^2$$

ANOVA I

Retour au premier exemple, nous avons

$$\sum_i \sum_j X_{ij}^2 = 145.44, \quad \sum_i \sum_j X_{ij} = 45.8, \quad \sum_i n_i = 15,$$

$$\mathrm{SS}_{\mathrm{Total}} = 145.44 - \frac{(45.8)^2}{15} = 145.88 - 139.84 = 5.60$$

La valeur 5.60 représente la variabilité totale des 15 durées de vie.

$$\mathrm{SS}_{\mathrm{Trait}} = \mathrm{SS}_{\mathrm{Reg}} = \left[\frac{(12.2)^2}{5} + \frac{(15.9)^2}{5} + \frac{(17.7)^2}{5}\right] - \frac{(45.8)^2}{15} = 3.15$$

La valeur 3.15 représente la part de la variabilité totale qui s'explique par les différences alimentaires.

$$\mathrm{SS}_{\mathrm{Erreur}} = (5-1)(0.19) + (5-1)(0.17) + (5-1)(0.25) = 2.44$$

Jusqu'à l'erreur d'arrondi: $\mathrm{SS}_{\mathrm{Total}} = \mathrm{SS}_{\mathrm{Reg}} + \mathrm{SS}_{\mathrm{Erreur}}$

Diviser le $\mathrm{SC}_{\mathrm{Total}}$ par $N-1$ nous donnerait une mesure de la variabilité totale existant dans l'ensemble de données combinées.

Avec k traitements, $\mathrm{SS}_{\mathrm{Trait}}$ a $k-1$ degrés de liberté, tandis que $\mathrm{SS}_{\mathrm{Erreur}}$ a $\sum_i (n_i - 1) = N - k$ degrés de liberté; notez que

$$N - 1 = (k - 1) + (N - k)$$

ANOVA I

La somme des carrés corrigée regroupée est définie comme

$$\mathrm{SC}_{\mathrm{Erreur}} = \sum_i \mathrm{CSS}_i$$

Les carrés moyens de l'erreur ($\mathrm{CM}_{\mathrm{Erreur}}$) sont

$$\frac{\mathrm{SC}_{\mathrm{Erreur}}}{\sum_i (n_i - 1)} = \frac{\mathrm{SC}_{\mathrm{Erreur}}}{N - k}$$

et représente notre meilleure estimation de la variance σ^2 commune à toutes les k populations.
La valeur attendue de $\mathrm{CM}_{\mathrm{Erreur}}$ est σ^2, c'est-à-dire,

$$E\left(\mathrm{CM}_{\mathrm{Erreur}}\right) = \sigma^2$$

ANOVA I

$\mathrm{SC}_{\mathrm{Trait}}$ divisé par ses degrés de liberté est appelé les Carrés Moyens de Traitement :

$$\mathrm{CM}_{\mathrm{Trait}} = \frac{\mathrm{SC}_{\mathrm{Trait}}}{k-1}$$

La valeur attendue de $\mathrm{CM}_{\mathrm{Trait}}$ est

$$E\left(\mathrm{MS}_{\mathrm{Trait}}\right) = \sigma^2 + \sum_i \frac{n_i\left(\mu_i - \mu\right)^2}{k-1}$$

Si H_0 est vraie, alors $(\mu_i - \mu) = 0$, par conséquent:

$$\sum_i \frac{n_i\left(\mu_i - \mu\right)^2}{k-1} = 0$$

et $E\left(\mathrm{MS}_{\mathrm{Trait}}\right) = \sigma^2 + 0 = \sigma^2$

Le test F peut maintenant être utilisé pour tester la véracité de H_0 :

$\mu_1 = \mu_2 = \cdots = \mu_k$. Le rapport des deux carrés moyens

$$\frac{\mathrm{MS}_{\mathrm{Trait}}}{\mathrm{MS}_{\mathrm{Erreur}}}$$

suit une loi de F avec $v_1 = k - 1$; $v_2 = N - k$. Si H_0 est vrai, alors

$$E\left(F\right) = \frac{\sigma^2}{\sigma^2} = 1$$

Si H_0 est faux, alors

$$E\left(F\right) = \frac{\sigma^2 + \sum_i \frac{n_i\left(\mu_i - \mu\right)^2}{k-1}}{\sigma^2} > 1$$

Source de variation	Somme des carrés	d.d.l.	CM	F	v.c.
Traitements	$\mathrm{SC}_{\mathrm{Trait}}$	$k-1$	$\frac{\mathrm{SC}_{\mathrm{Trait}}}{k-1}$	$\frac{\mathrm{MC}_{\mathrm{Trait}}}{\mathrm{MC}_{\mathrm{Erreur}}}$	Voir tableau.
Erreur	$\mathrm{SC}_{\mathrm{Erreur}}$	$N-k$	$\frac{\mathrm{SC}_{\mathrm{Erreur}}}{N-k}$		
Total	$\mathrm{SC}_{\mathrm{Total}}$	$N-1$			

ANOVA I

Les hypothèses sont

$$H_0 : \mu_{\mathrm{U}} = \mu_{90\%} = \mu_{80\%}$$

$$H_1 : \text{At least one pair of the } \mu_i\text{'s is not equal}$$

Source de variation	Somme des carrés	d.d.l.	CM	F	v.c.
Régimes	3.15	2	1.575	7.76	3.89
Erreur	2.44	12	0.203		
Total	5.60	14			

Étant donné que $7.76 > 3.89$, les carrés moyens des régimes sont significativement plus grands que les carrés moyens de l'erreur, indiquant qu'au moins une paire de régimes n'est pas égale.

Les hypothèses sont

$$H_0 : \mu_{\text{I}} = \mu_{\text{II}} = \mu_{\text{II}} = \mu_{\text{IV}}$$
$$H_1 : \text{At least one pair of the } \mu_i\text{'s is not equal}$$

	Site			
	I	II	III	IV
n_i	8	5	11	15
$\bar{X}_i$	7.0	9.0	8.0	5.0
s_i	2.1	2.1	2.2	2.2
$T_{i.}$	56.0	45.0	88.0	75.0

De plus, $\sum n_i = 39$ et $\sum T_i = T_{..} = 264$. Ensuite, nous avons besoin des trois sommes de carrés : $\mathrm{SC}_{\mathrm{Total}}$, $\mathrm{SC}_{\mathrm{Trait}}$ et $\mathrm{SC}_{\mathrm{Erreur}}$.

ANOVA I

$$\begin{aligned}
\mathrm{SC}_{\mathrm{Trait}} = \mathrm{SC}_{\mathrm{Sites}} &= \sum_i \left[\frac{T_{i.}^2}{n_i}\right] - \frac{T_{..}^2}{N} \\
&= \left[\frac{(56.0)^2}{8} + \frac{(45.0)^2}{5} + \frac{(88.0)^2}{11} + \frac{(75.0)^2}{15}\right] - \frac{(264)^2}{39} \\
&= 1876 - 1787.08 \\
&= 88.92
\end{aligned}$$

$$\begin{aligned}
\mathrm{SC}_{\mathrm{Erreur}} &= \sum_i (n_i - 1)\, s_i^2 \\
&= 7\,(2.1)^2 + 4\,(2.1)^2 + 10\,(2.2)^2 + 14\,(2.2)^2 \\
&= 164.67
\end{aligned}$$

Étant donné que les sommes de carrés sont additives, $\mathrm{SC}_{\mathrm{Total}} = 88.92 + 164.67 = 253.59$.

Source de variation	Somme des carrés	d.d.l.	CM	F	v.c.
Sites	88.92	3	29.640	6.30	2.92
Erreur	164.67	35	4.705		
Total	253.59	38			

Étant donné que $6.30 > 2.92$, nous rejetons H_0 et concluons que la longueur du céphalon à certains sites est significativement différente.

ANOVA II

Dans l'exemple 3, l'endocrinologue a utilisé cinq portées de souris pour étudier la libération d'insuline, visant à détecter une variation significative qui pourrait indiquer une composante génétique. Les souris individuelles ont été sacrifiées, et les hypothèses de l'expérience sont :

H_0 : Pas de variabilité significative entre les portées, $\sigma^2_{\text{portées}} = 0$

H_1 : Variabilité significative entre les portées, $\sigma^2_{\text{portées}} > 0$

et nous laisserons $\alpha = 0.05$. Les données sont

	Portée				
	1	2	3	4	5
	9	2	3	4	8
	7	6	5	10	10
	5	7	9	9	12
	5	11	10	8	13
	3	5	6	10	11
$T_{i.}$	29	31	33	41	54
$\bar{X}_i$	5.8	6.2	6.6	8.2	10.8

Les statistiques récapitulatives de base sont

$$\sum_i \sum_j X_{ij}^2 = 1634, \quad \sum_i \sum_j X_{ij} = T_{..} = 188, \quad N = 25.$$

ANOVA II

Alors

$$\begin{aligned}\mathrm{SC}_{\text{Total}} &= \sum_i \sum_j X_{ij}^2 - \frac{T_{..}^2}{N} \\ &= 1634 - \frac{(188)^2}{25} \\ &= 220.24\end{aligned}$$

et

$$\begin{aligned}\mathrm{SC}_{\text{Trait}} = \mathrm{SC}_{\text{Portes}} &= \sum_i \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} \\ &= \frac{(29)^2 + (31)^2 + (33)^2 + (41)^2 + (54)^2}{5} - \frac{(188)^2}{25} \\ &= 83.84\end{aligned}$$

par conséquent

$$\mathrm{SC}_{\text{Erreur}} = \mathrm{SC}_{\text{Total}} - \mathrm{SC}_{\text{Trait}} = 220.24 - 83.84 = 136.40$$

La valeur attendue de $\mathrm{CM}_{\mathrm{Trait}}$ pour ce test est

$$E\left(\mathrm{CM}_{\mathrm{Trait}}\right) = \sigma^2 + n_0 \sigma_{\mathrm{Tr}}^2$$

où

$$n_0 = \frac{N - \sum_{i=1}^{k} \frac{n_i^2}{N}}{k-1}$$

et σ_{Tr}^2 est la variabilité due aux traitements (portées dans ce cas). La valeur attendue de $\mathrm{CM}_{\mathrm{Erreur}}$ pour ce test est

$$E\left(\mathrm{CM}_{\mathrm{Erreur}}\right) = \sigma^2$$

Ainsi, si H_0 était vrai, alors $\sigma_{\mathrm{Tr}}^2 = 0$ et la valeur attendue pour F serait $E(F) = \frac{\sigma^2}{\sigma^2} = 1$.

ANOVA II

La table d'ANOVA pour l'exemple 3 est donnée ci-dessous :

Source de variation	Somme des carrés	d.d.l.	CM	F	v.c.
Parmi les portées	83.84	4	20.96	3.07	2.87
Erreur	136.40	20	6.82		
Total	220.24	24			

Étant donné que $3.07 > 2.87$, le test F est significatif. Nous rejetons H_0 et acceptons H_1; il existe une variabilité significative entre les portées.

Exercices

Exercice 1

Rhagoletis pomonella, la mouche de la pomme, est un ravageur sérieux des pommes commerciales. On pense qu'elle se nourrit du houblon. Un entomologiste testant des mesures de contrôle pour R. pomonella a utilisé les appâts suivants dans des pièges standard:

1. A: appât protéique seul
2. B: appât protéique + 10% carbonate d'ammonium
3. C: appât protéique + 10% acétate d'ammonium
4. D: 10 g d'acétate d'ammonium seul

Les chiffres de capture des mouches sont donnés. Supposons une distribution normale et analysons-les de manière appropriée.

Exercices

	Traintement			
	A	B	C	D
	131	152	205	201
	150	165	230	185
	175	149	200	175
	143	181	195	206
$\bar{X}_i$	149.8	161.8	207.5	191.8
CSS_i	1034.75	638.75	725.00	614.75
$\sum\sum X = 2843$			$\sum\sum X^2 = 516,663$	

Exercice 2

La drépanocytose est une maladie génétiquement transmise caractérisée par des globules rouges en forme de faucille. Dans une étude sur ses liens avec d'autres maladies, les chercheurs ont mesuré les niveaux d'hémoglobine chez 30 sujets répartis en trois groupes : normal, drépanocytose (homozygotes), et porteurs de la drépanocytose (hétérozygotes). Existe-t-il des différences significatives dans les niveaux d'hémoglobine entre les groupes?

	Sujets normaux	Drépanocytose	Trait drépanocytaire
Moyenne	13.6	9.8	12.7
Écart-type	1.6	1.6	1.5
$\mathrm{SC}_{\mathsf{Trait}} = 78.87$		$\mathrm{SC}_{\mathsf{Erreur}} = 66.33$	

Exercice 3

Le directeur de ventes d'un laboratoire pharmaceutique veut savoir s'il existe des différences significatives entre les régions en terme de niveau d'accueil d'un nouveau produit. Les résultats suivants ont été obtenus auprès d'un échantillon aléatoire de clients:

Niveau d'accueil	Régions			
	Nord	Est	Sud	Ouest
Faible	22	35	0	5
Modéré	84	55	8	24
Élevé	25	17	22	12

Le niveau d'accueil dépend-t-il de la région ?

Exercice 4

24 têtes d'ovin ont reçu 6 alimentations différentes pour constituer 4 répétitions et on a enregistré les gains moyens quotidiens en poids suivants:

Alim. 1	Alim. 2	Alim. 3	Alim. 4	Alim. 5	Alim. 6
590	460	600	640	690	690
760	430	460	660	600	650
700	540	610	720	550	680
640	470	510	580	480	740

Au seuil de 5%, existe-t-il une différence significative quant à l'effet des différentes alimentations sur le gain moyen quotidien en poids des ovins?

Exercice 5

Le tableau suivant donne le nombre d'étudiants qui ont été brillants et médiocres devant trois examinateurs:

	Examin. 1	Examin. 2	Examin. 3	Total
Brillants	50	47	56	153
Médiocres	5	14	8	27
Total	55	61	64	180

Au seuil de 5%, testez l'hypothèse selon laquelle le nombre d'étudiants médiocres est le même pour chaque examinateur.

Exercice 6

Un fabricant japonais de puces électroniques envisage d'implanter une usine au Maroc pour desservir le marché nord-africain, hésitant entre Tanger, Casablanca et Eljadida. Selon lui, le critère déterminant pour l'emplacement de cette usine est l'assiduité des travailleurs. Le fabricant a visité cinq grandes usines de chaque ville, obtenant les taux d'absentéisme par 3500 journées de travail des administrateurs. Les résultats sont résumés dans le tableau ci-dessous.

Ville	Échantillon	Total
Tanger	141; 127; 111; 124; 144	$T_1 = 647$
Casablanca	157; 131; 105; 132; 163	$T_2 = 688$
Eljadida	183; 161; 145; 157; 189	$T_3 = 835$
J=3	N=15	T=2170

A un seuil de 5%, peut-on conclure que le taux d'absentéisme au travail est le même en moyenne dans ces 3 villes?

Merci. Des questions?