

Travaux Dirigés n°3

Dispersion · Forme · Régression linéaire · V de Cramer

Modules | Durée : 1h30 | Documents autorisés : cours

Dr. Anicet EBOU, ediman.ebou@inphb.ci

Compétences évaluées

- Calculer et interpréter les mesures de dispersion (variance, écart-type, EIQ, CV)
- Identifier la forme d'une distribution et son asymétrie à partir de données ou d'un graphique
- Ajuster et interpréter une droite de régression linéaire simple
- Calculer et interpréter le V de Cramer pour deux variables qualitatives

Exercice 1 Dispersion des rendements de maïs (5 pts)

Contexte. Un technicien agricole mesure le rendement en grain (t/ha) sur **10 parcelles** de maïs de la région de Bouaké :

Parcelle	1	2	3	4	5	6	7	8	9	10
Rendement (t/ha)	2,8	3,1	3,4	3,0	4,2	3,6	2,9	3,8	3,5	3,7

1. Calculez la **moyenne** $\bar{x}$ et la **médiane** de ces rendements.
2. Complétez le tableau ci-dessous, puis calculez la **variance** s^2 et l'**écart-type** s .

Parcelle i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	2,8		
2	3,1		
3	3,4		
4	3,0		
5	4,2		
6	3,6		
7	2,9		
8	3,8		
9	3,5		
10	3,7		
Total		0 (vérif.)	

3. Triez les rendements par ordre croissant, puis calculez Q_1 , Q_3 et l'écart interquartile (EIQ).

Rappel. Avec $n = 10$ valeurs triées : Q_1 = moyenne des valeurs aux rangs 2,5 (interpolation) et Q_3 = moyenne des valeurs aux rangs 7,5.

4. Calculez le **coefficient de variation** (CV). Ces rendements sont-ils homogènes ? Justifiez.
5. Vérifiez vos calculs avec R et reproduisez le code ci-dessous. Que fournit la fonction `summary()` ?

```

1 rendement <- c(2.8, 3.1, 3.4, 3.0, 4.2, 3.6, 2.9, 3.8, 3.5, 3.7)
2 summary(rendement)
3 var(rendement)
4 sd(rendement)
5 IQR(rendement)

```

Exercice 2 Forme et asymétrie - Revenus de producteurs de cacao (4 pts)

Contexte. Une enquête porte sur les **revenus annuels nets** (en milliers de FCFA) de 12 producteurs de cacao de la région du Haut-Sassandra :

320	350	370	380	390	400	420	440	450	480	510	1800
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

1. Calculez la **moyenne** et la **médiane** de ces revenus.
2. Comparez les deux valeurs. Quelle est la **forme probable** de la distribution ? Justifiez sans dessin.
3. La valeur 1 800 est-elle une valeur aberrante ? Calculez les bornes de la règle des moustaches ($Q_1 - 1,5 \times \text{EIQ}$ et $Q_3 + 1,5 \times \text{EIQ}$) pour répondre.
4. En R, tracez un **histogramme** et une **boîte à moustaches**. Que révèlent-ils sur la forme ?

```
1 revenus <- c(320, 350, 370, 380, 390, 400, 420, 440, 450, 480,
2           510, 1800)
3 par(mfrow = c(1, 2))
4 hist(revenus, main = "Revenus cacao", xlab = "Milliers FCFA", col
5     = "#1D6F42", border = "white")
6 boxplot(revenus, main = "Boite a moustaches", col = "#1A4F8A",
7     horizontal = TRUE)
```

Exercice 3 Régression linéaire - Dose d'azote et rendement du riz (6 pts)

Contexte. Un essai agronomique teste l'effet de la dose d'engrais azoté (kg N/ha) sur le rendement en riz (t/ha) sur 8 parcelles expérimentales à la station de recherche de Bouaké :

Parcelle	1	2	3	4	5	6	7	8
Dose azote x (kg N/ha)	0	20	40	60	80	100	120	140
Rendement y (t/ha)	2,1	2,8	3,4	4,0	4,6	5,1	5,5	5,8

1. Calculez $\bar{x}$, $\bar{y}$, puis complétez le tableau de calcul ci-dessous.

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	0	2,1				
2	20	2,8				
3	40	3,4				
4	60	4,0				
5	80	4,6				
6	100	5,1				
7	120	5,5				
8	140	5,8				
Σ			0	0		

2. Calculez le **coefficient de corrélation de Pearson** r . Interprétez.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

3. Calculez la pente $\hat{b}$ et l'ordonnée à l'origine $\hat{a}$ de la droite des moindres carrés.

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad \hat{a} = \bar{y} - \hat{b} \bar{x}$$

4. Écrivez l'équation de la droite $\hat{y} = \hat{a} + \hat{b}x$. Quelle augmentation de rendement prédit-elle pour une dose supplémentaire de 10 kg N/ha ?

5. Estimez le rendement prédit pour une dose de 90 kg N/ha. Ce point est-il dans l'intervalle des données observées (interpolation) ? Serait-il prudent de prédire pour 200 kg N/ha ?

6. En R, ajustez le modèle et tracez la droite de régression.

```

1 dose <- c(0, 20, 40, 60, 80, 100, 120, 140)
2 rdt <- c(2.1, 2.8, 3.4, 4.0, 4.6, 5.1, 5.5, 5.8)
3
4 # Nuage et droite
5 plot(
6   dose, rdt, pch=19, col="#1A4F8A",
7   xlab="Dose azote (kg N/ha)", ylab="Rendement (t/ha)",
8   main="Riz - Essai azote Bouake"
9 )
10 modele <- lm(rdt ~ dose)
11 abline(modele, col="#1D6F42", lwd=2)
12
13 # Résumé du modèle
14 summary(modele)

```

Relevez les valeurs de $\hat{a}$, $\hat{b}$ et R^2 dans la sortie. Interprétez R^2 .

Exercice 4 V de Cramer - Variété d'igname et qualité sanitaire (5 pts)

Contexte. Dans le cadre d'une étude post-récolte à Yamoussoukro, 90 tubercules d'igname sont classés selon leur **variété** et leur **qualité sanitaire** :

	Saine	Altération légère	Altération grave	Total
Florido	24	10	6	40
Gnidrou	15	12	8	35
Krenglè	6	5	4	15
Total	45	27	18	90

1. Calculez le **profil ligne** (fréquences conditionnelles en %) pour chaque variété.

	Saine	Alt. légère	Alt. grave	Total
Florido				100 %
Gnidrou				100 %
Krenglè				100 %

2. Calculez le **profil marginal** des colonnes (répartition globale de la qualité sanitaire, sans distinguer la variété). Ces profils lignes diffèrent-ils du profil marginal ? Qu'est-ce que cela suggère ?

3. Calculez les **effectifs théoriques** $\hat{n}_{ij} = n_{i\bullet} \times n_{\bullet j} / n$ sous hypothèse d'indépendance.

	Saine	Alt. légère	Alt. grave
Florido			
Gnidrou			
Krenglè			

4. Calculez le χ^2 puis le **V de Cramer**. Concluez sur la force de l'association.

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} ; \quad V = \sqrt{\frac{\chi^2}{n \times \min(r-1, c-1)}} \quad (r = 3 \text{ lignes}, c = 3 \text{ colonnes})$$

5. Reproduisez l'analyse sous R.

```
1 variete <- c(rep("Florido",40), rep("Gnidrou",35), rep("Krengle",15))
2 qualite <- c(
3   rep("Saine",24), rep("Leg",10), rep("Grave",6),      # Florido
4   rep("Saine",15), rep("Leg",12), rep("Grave",8),      # Gnidrou
5   rep("Saine",6), rep("Leg",5), rep("Grave",4)        # Krengle
6 )
7 tab <- table(variete, qualite)
8 addmargins(tab)
9 round(prop.table(tab, margin=1) * 100, 1) # profils lignes
10
11 # V de Cramer
12 chi2 <- chisq.test(tab, correct=FALSE)$statistic
13 V <- sqrt(chi2 / (sum(tab) * min(nrow(tab)-1, ncol(tab)-1)))
14 cat("V de Cramer =", round(V, 3), "\n")
```

Relevez la valeur de V affichée. Correspond-elle à votre calcul manuel ?

Question de synthèse (bonus - 2 pts)

Pour aller plus loin. Un chercheur compare deux variétés de sorgho (A et B) sur 20 parcelles chacune.

- Variété A : rendement moyen 3,2 t/ha, écart-type 0,2 t/ha
- Variété B : rendement moyen 3,5 t/ha, écart-type 1,1 t/ha

Il conclut : “Variété B est meilleure car sa moyenne est plus élevée.”

Question. Cette conclusion est-elle complète ? Calculez le CV pour chaque variété et rédigez une recommandation argumentée d'au moins 5 lignes intégrant les concepts de

dispersion vus en cours.