

Master1 de  
Sciences et Pratiques d'Éducation et de Formation  
4PAKQNT1 Méthodes quantitatives

Jean-Claude Régnier

**Document autocorrectif et auto-évaluatif**

# Situation Problème

## Document n° 1 : situation de l'étude

Afin de mieux cerner la population des usagers de son site Web, le service Réseau d'une administration française a conduit une étude à partir de données issues des « traces laissées par ses visiteurs » complétées par une enquête par questionnaire. Pour cela il a procédé au tirage aléatoire d'une semaine de l'année 2003 : il a obtenu celle du lundi 3 novembre 2003 au dimanche 9 novembre 2003. Le groupe des 1024 visiteurs de cette période est considéré comme un échantillon aléatoire avec remise représentatif de la population des usagers du site. Les diverses méthodes mises en œuvre ont permis de recueillir des résultats relatifs aux variables suivantes.

<i>code de la variable</i>	<i>descriptif succinct</i>
V01	Genre : <b>1</b> = homme <b>2</b> = femme
V02	Âge
V03	Statut : <b>ST1</b> = Célibataire <b>ST2</b> = Marié(e), Vie maritale, PACS <b>ST3</b> = Divorcé(e) <b>ST4</b> = Veuf (ve)
V03A	Nombre d'enfant(s)
V04	Secteur d'activité professionnelle
V05	Type de Navigateur utilisé : <b>N1</b> = Internet Explorer 5 <b>N2</b> = Internet Explorer 4 <b>N3</b> = Netscape Navigator 4.X <b>N4</b> = Netscape 3.X <b>N5</b> = Autre
V06	Nombre de pages ouvertes lors de la consultation
V07	Durée de la consultation mesurée en 1/1000 <sup>ème</sup> de seconde
V08	Qualité de l'apport informationnel : <b>0</b> = Sans avis <b>1</b> = Très forte <b>2</b> = Plutôt forte <b>3</b> = Plutôt faible <b>4</b> = Très faible
V9	Lieu de consultation ( lieu du poste client): <b>L1</b> = à la maison <b>L2</b> = lieu public <b>L3</b> = chez un ami <b>L4</b> = au travail <b>L5</b> = autre, lequel ?

Nous rapportons ci-dessous les tableaux statistiques des variables

**(Tableau n° T1) Tableau statistique de la variable V01**

Modalités de V01	homme	femme
effectifs	463	561

**(Tableau n° T2) Tableau statistique de la variable V02**

Valeurs de V02	[15 ; 20[	[20 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[	[40 ; 45[	[45 ; 50[	[50 ; 55[	[55 ; 60[
effectifs	14	84	512	140	124	96	32	16	6

**(Tableau n° T3) Tableau statistique de la variable V03**

Modalités de V03	ST1	ST2	ST3	ST4
effectifs	768	102	128	26

**(Tableau n° T5) Tableau statistique de la variable V05**

Modalités de V05	N1	N2	N3	N4	N5
effectifs	256	230	150	180	208

**(Tableau n° T6) Tableau statistique de la variable V06**

Valeurs de V06	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
effectifs	51	90	102	205	77	90	51	51	51	38	90	38	13	26	13	0	38

**(Tableau n° T8) Tableau statistique de la variable V08**

Modalités de V08	0	1	2	3	4
effectifs	210	176	305	141	192

**(Tableau n° T9) Tableau statistique de la variable V09**

Modalités de V09	L1	L2	L3	L4	L5
effectifs	610	204	102	60	48

**(Tableau n° T10) Tableau statistique « nombre de visiteurs/visites quotidiennes »**

Jour de la semaine	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
effectifs	146	156	182	136	135	138	131

## Document n° 2 : traitement demandé

### TQ1 De quoi s'agit-il ? A propos de l'enquête par questionnaire et l'analyse des traces.

Q101- Après avoir pris connaissance des informations fournies dans le **document n°1**, identifier et caractériser "succinctement" et clairement, les éléments de cette étude statistique : **population, échantillon, individus, variables**.

Considérons l'enquête par questionnaire visant la construction de données relatives aux usages du site Internet, mise en oeuvre dans cette étude.

Q102 Par une présentation synthétique, exposer les principales étapes de cette **enquête par questionnaire** et les précautions majeures à prendre.

Q103 Construire 5 **questions** relevant chacune d'un format différent que vous préciserez.

Q104- Quelles informations sont fournies par les **tableaux statistiques** T2, T8 et T10.

Q105- Sur l'une des **variables qualitatives** de votre choix, réaliser une courte synthèse s'appuyant sur un graphique de votre choix afin de rendre compte des informations apportées.

### TQ2-Etude de la variable V06 « nombre de pages consultées lors de la visite »

Q201 Quel est le **nombre moyen** de pages consultées par visite sur la période avec quel **écart-type** ?

Q202 Construire un **diagramme** de la distribution des fréquences

Q203 Quelle est la valeur du **mode** ?

Q204 Quelles sont les valeurs respectives des trois **quartiles** ?

Q205- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

### TQ3 – Fréquentation du site.

Q301- Sur l'échantillon, quelle est **proportion** des visites effectuées durant le week-end ?

Q302- Donner une **estimation ponctuelle** de la proportion des visites effectuées durant le week-end, dans la population.

Q303- Donner une **estimation par intervalle de confiance** à 95%, de la proportion des visites effectuées durant le week-end, dans la population.

Q304- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

**TQ4- Étude de l'homogénéité de l'évaluation de la qualité informationnelle selon le genre**

Q401- Peut-on considérer qu'il y a une différence d'appréciation entre les hommes et les femmes ?

Qualité informationnelle	0	1	2	3	4	
homme	78	70	165	60	90	
femme	132	106	140	81	102	

Tester les hypothèses correspondantes ( $H_0$  et  $H_1$ ) au seuil de 5% à l'aide du **test du Khi-deux d'homogénéité**.

Q402- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

**TQ5- Étude de la répartition des visiteurs/visites quotidiennes.**

Q501- A partir des informations obtenues sur l'échantillon et consignées dans le tableau n° T10, peut-on considérer que les visites sont également réparties sur la semaine ? (il s'agit de tester l'hypothèse  $H_0$  contre  $H_1$  à l'aide du **test du Khi-deux d'adéquation** avec un niveau de risque maximum de 5%)

Q502- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

*TQ6-Etude de la variable V02*

Q601 Quel est l'**âge moyen** des visiteurs sur la période avec quel **écart-type** ?

Q602 Construire l'**histogramme** de la distribution des fréquences, puis le polygone des fréquences.

Q603 Déterminer l'**âge modal**.

Q604 Déterminer l'**âge médian**.

Q605 Donner une **estimation par intervalle de confiance** à 95 % de l'âge moyen des visiteurs sur la population des internautes.

Q606 A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

*TQ7-Etude de la corrélation entre les variables V02 et V06*

Sur un échantillon aléatoire de 10 individus, nous avons recueilli l'information suivante :

V02=Age	17	27	30	31	35	43	48	52	54	65
V06=Nombre de pages	14	15	12	10	9	6	7	5	4	4

Q701 Construire le **nuage statistique** ?

Q702 Quelle interprétation donnez-vous de la forme de ce nuage ?

Q703 Calculer le **coefficient de corrélation** linéaire de Bravais Pearson.

Q704 A partir de cette étude, énoncer sous une forme synthétique **vosre conclusion** et **vosre interprétation**.

**GRILLE d'auto-évaluation**

code	items repérés et sur lesquels le jugement est porté	validité du résultat				validité du raisonnement ou pertinence			
		R+	R-	E	N	R+	R-	E	N
Q101	- analyse "succincte" de l'étude statistique	R+	R-	E	N	R+	R-	E	N
Q101.1	- Étapes d'une enquête par questionnaire : construction	R+	R-	E	N	R+	R-	E	N
Q101.2	- Étapes d'une enquête par questionnaire : passation	R+	R-	E	N	R+	R-	E	N
Q101.3	- Étapes d'une enquête par questionnaire : traitement	R+	R-	E	N	R+	R-	E	N
Q102.1	- Question 1 (format1 = fermée nominale ; pertinence)	R+	R-	E	N	R+	R-	E	N
Q102.2	- Question 2 (format2 = fermée ordinale ; pertinence)	R+	R-	E	N	R+	R-	E	N
Q102.3	- Question 3 (format3 = ; pertinence)	R+	R-	E	N	R+	R-	E	N
Q102.4	- Question 4 (format1 ; pertinence)	R+	R-	E	N	R+	R-	E	N
Q102.5	- Question 5 (format1 ; pertinence)	R+	R-	E	N	R+	R-	E	N
Q104	- informations T2, T8 T10	R+	R-	E	N	R+	R-	E	N
Q105	- synthèse relative à une variable qualitative	R+	R-	E	N	R+	R-	E	N
Q201.1	- moyenne de V06	R+	R-	E	N	R+	R-	E	N
Q201.2	- variance, écart-type de V06	R+	R-	E	N	R+	R-	E	N
Q202	- diagramme	R+	R-	E	N	R+	R-	E	N
Q203	- mode	R+	R-	E	N	R+	R-	E	N
Q204	- quartiles	R+	R-	E	N	R+	R-	E	N
Q205	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q301	- proportion sur l'échantillon	R+	R-	E	N	R+	R-	E	N
Q302	- estimation ponctuelle	R+	R-	E	N	R+	R-	E	N
Q303	- estimation par intervalle 95%	R+	R-	E	N	R+	R-	E	N
Q304	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q401.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q401.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q401.3	- mise en œuvre du test	R+	R-	E	N	R+	R-	E	N
Q402	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q501.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q501.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q501.3	- mise en œuvre du test	R+	R-	E	N	R+	R-	E	N
Q502	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q601.1	- moyenne de V02	R+	R-	E	N	R+	R-	E	N
Q601.2	- écart-type de V05	R+	R-	E	N	R+	R-	E	N
Q602.1	- histogramme	R+	R-	E	N	R+	R-	E	N
Q602.2	- polygone des fréquences	R+	R-	E	N	R+	R-	E	N
Q603	- mode	R+	R-	E	N	R+	R-	E	N
Q604	- médiane (interpolation linéaire)	R+	R-	E	N	R+	R-	E	N
Q605.1	- calculs statistiques sur un échantillon	R+	R-	E	N	R+	R-	E	N
Q605.2	- détermination d'un intervalle de confiance	R+	R-	E	N	R+	R-	E	N
Q605.3	- interprétation	R+	R-	E	N	R+	R-	E	N
Q606	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q701	- construction du nuage	R+	R-	E	N	R+	R-	E	N
Q702	- interprétation du nuage	R+	R-	E	N	R+	R-	E	N
Q703.1	- Covariance	R+	R-	E	N	R+	R-	E	N
Q703.2	- Coefficient de corrélation BP	R+	R-	E	N	R+	R-	E	N
Q704	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N

	Tous les items repérés et sur lesquels le jugement est porté	validité du résultat				Validité du raisonnement ou pertinence			
		R+	R-	E	N	R+	R-	E	N
S01	modalités du jugement nombre de...								
S02	résultat & raisonnement nombre de...	R+	1	R-	0,5	E	0	N	0
S03	scores...								

Score total obtenu  $S = (0,5) \times n_{R-} + (1) \times n_{R+} =$

Score maximum total  $S_{\max} = 44 + 44 = 88$

Taux de réussite  $T = S/S_{\max} = \boxed{\phantom{000}}$  niveau de réussite note = ... /20

### Explicitation du codage utilisé

code des modalités du jugement	modalités du jugement	Caractérisation	pondération
R+	réussite forte	- lorsque le résultat est celui attendu par l'évaluateur - lorsque le raisonnement est logiquement vrai	+1
R-	réussite faible	Il s'agit de repérer le déjà-réussi sans laisser la place à l'ancrage de sous-compréhension. - lorsque le résultat n'est pas celui attendu par l'évaluateur sans erreur manifeste de raisonnement - lorsque le raisonnement ne comporte que des erreurs "d'esprit fatigué"	+0,5
E	échec par erreur	- lorsque le résultat est faux mathématiquement - lorsque le raisonnement est logiquement faux	0
N	échec par non-réponse	- lorsque le résultat n'est pas fourni - lorsque le raisonnement n'est pas explicité - lorsque le résultat ou le raisonnement ne sont pas interprétables par l'évaluateur	0

### Évaluation des items "CONCLUSION-INTERPRETATION"

Conformément aux objectifs visés en cours, l'item "conclusion" correspond à deux niveaux :

- *niveau 1* (représenté par le point de vue "**résultat**") correspond à une **description** lisible et "bien" rédigée des résultats obtenus dans les traitements abordés. Ces résultats sont rapportés dans le contexte de la situation problème. (Par exemple : les nombres sont donnés avec un ordre de grandeur compatible avec le contexte, avec leurs unités, ou les modalités sont explicitées en clair...)

- *niveau 2* (représenté par le point de vue "**raisonnement**") correspond à l'énoncé d'une **interprétation** personnelle fondée sur les résultats de l'étude et exprimée de façon pertinente.

## Document n° 3 : Corrigé guide du traitement demandé

### TQ1- De quoi s'agit-il ?

Q101- Après avoir pris connaissance des informations fournies dans le **document n°1**, identifier et caractériser "succinctement" et clairement, les éléments de cette étude statistique : population, échantillon, individus, variables.

#### Population :

Ensemble des usagers d'un site Web (**individus**) d'une administration française

#### Échantillon :

Sous-ensemble des 1204 individus ayant accédé durant une semaine aléatoirement extraite

#### Variables.

Variables quantitatives		Variables qualitatives		
discrètes	continues	ordinales	nominales	Textuelles
[V03A] [V06]	[V02]→âge [V07]→durée de consultation	[V08] partiellement ordonnée	[V01]→genre [V03] [V04] [V05] [V08] [V09]	

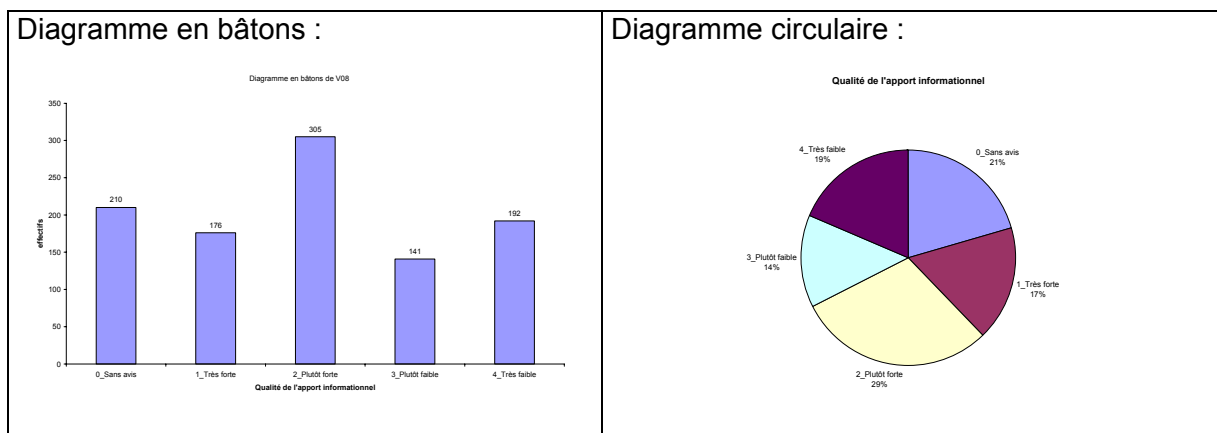
Q102- Sur l'une des variables qualitatives de votre choix, réaliser une courte synthèse s'appuyant sur un graphique de votre choix afin de rendre compte des informations apportées.

Nous choisissons la variable V08 = *Qualité de l'apport informationnel*

Cette question obtient 1024 réponses réparties comme l'indique le tableau statistique

Modalités de V08	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Effectifs	210	176	305	141	192	1024
Fréquences	0,20507813	0,171875	0,29785156	0,13769531	0,1875	1
Fréquences (%)	20,51%	17,19%	29,79%	13,77%	18,75%	100%

que nous traduisons graphiquement :



Il ressort que la majorité relative 29,79% (mode = plutôt forte) est constituée par le groupe des usagers qui évaluent la qualité de l'apport informationnel à niveau plutôt fort.

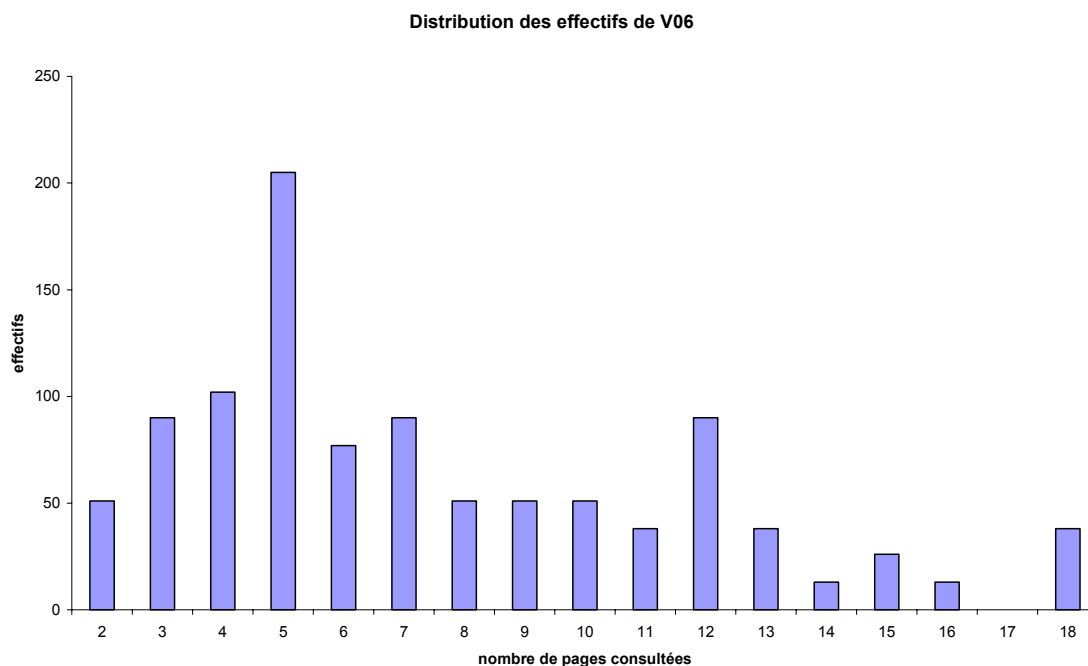
Q103- Quelles informations sont fournies par les tableaux T2, T8 et T10.

**TQ2-Etude de la variable V06 « nombre de pages consultées lors de la visite »**

Q201 Quelle est la valeur du nombre moyen de pages consultées par visite sur la période et celle de l'écart-type ?

	valeurs	effectifs	$n_k x_k$	$n_k(x_k-m)^2$
	2	51	102	1570,264172
	3	90	270	1862,265358
	4	102	408	1284,606468
<b>mode</b>	<b>5</b>	<b>205</b>	<b>1025</b>	<b>1331,787586</b>
	6	77	462	184,7128792
	7	90	630	27,10910797
	8	51	408	10,3813591
	9	51	459	107,4008904
	10	51	510	306,4204216
	11	38	418	452,6023178
	12	90	1080	1783,163795
	13	38	494	1129,180443
	14	13	182	541,0290413
	15	26	390	1443,51902
	16	13	208	928,4899788
	17	0	0	0
	18	38	684	4150,625755
<b>Somme</b>		<b>1024</b>	<b>7730</b>	<b>Somme 17113,55859</b>
		<b>moyenne</b>	<b>7,548828125</b>	<b>variance 16,71245956</b>
				<b>écart-type 4,088087519</b>

Q202 Construire un diagramme de la distribution des fréquences



Q203 Quelle est la valeur du **mode** ? La valeur la plus fréquente est 5 pages consultées

Q204 Quelles sont les valeurs respectives des trois **quartiles** ?

Si nous appliquons les règles explicitées ci-dessous :

<b>N =</b>	<b>Q1</b>	<b>Q2</b>	<b>Q3</b>
N = 4q	entre la valeur de rang q et celle de rang q+1	entre la valeur de rang 2q et celle de rang 2q+1	entre la valeur de rang 3q et celle de rang 3q+1
N = 4q + 1	entre la valeur de rang q et celle de rang q+1	la valeur de rang 2q+1	entre la valeur de rang 3q+1 et celle de rang 3q+2
N = 4q + 2	la valeur de rang q+1	entre la valeur de rang 2q+1 et celle de rang 2q+2	la valeur de rang 3q+2
N = 4q + 3	la valeur de rang q+1	la valeur de rang 2q+2	la valeur de rang 3q+3

nous obtenons :  $\frac{1024}{4} = 256$  et nous sommes dans le cas n°1 avec q=256. Ainsi le premier

quartile est la valeur  $Q1 = \frac{\text{valeur\_de\_rang.256} + \text{valeur\_de\_rang.257}}{2} = \frac{5 + 5}{2} = 5$

Le second quartile Q2 ou médiane vaut

$Q2 = \frac{\text{valeur\_de\_rang.512} + \text{valeur\_de\_rang.513}}{2} = \frac{6 + 6}{2} = 6$

Le troisième quartile Q3 vaut  $Q3 = \frac{\text{valeur\_de\_rang.768} + \text{valeur\_de\_rang.769}}{2} = \frac{10 + 11}{2} = 10,5$

Q205- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

### TQ3 – Fréquentation du site.

Q301- Sur l'échantillon, quelle est proportion des visites effectuées durant le week-end ?

Q302- Donner une estimation ponctuelle de la proportion des visites effectuées durant le week-end, dans la population.

Q303- Donner une estimation par intervalle de confiance à 95%, de la proportion des visites effectuées durant le week-end, dans la population.

Pour résoudre ce problème, nous avons recours à l'approche théorique suivante :

#### Estimation d'une proportion $\pi$

##### Conditions d'utilisation:

- L'échantillon est obtenu aléatoirement par n tirages avec remise

**dans notre situation n = 1024**

-Pour des conditions optimales, la taille n de l'échantillon devrait être supérieure à 100 et l'estimation ponctuelle devrait être comprise entre 0,1 et 0,9 , sinon il conviendrait de consulter des documents de statistique précisant d'autres conditions

**Estimateur:**

La formule donnant "proportion des cas favorables calculée à partir des valeurs observées sur l'échantillon":  $F_n = R_n/n$  où  $R_n$  est la variable qui associe à chaque échantillon le nombre d'observations ayant le caractère étudié parmi les  $n$  observations de l'échantillon.  $X = nF_n = R_n = B(n, \pi)$  variable binomiale telle que l'on approche

$$Y = \frac{R_n - n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{n(\frac{R_n}{n} - \pi)}{\sqrt{n\pi(1-\pi)}} = \frac{(\frac{R_n}{n} - \pi)}{\frac{1}{\sqrt{n}}\sqrt{\pi(1-\pi)}} \text{ par une variable de Laplace-Gauss LG}(0;1)$$

**Estimation ponctuelle:**

$$p = f_n$$

où  $f_n$  est la valeur de  $F_n$  calculée à partir de l'échantillon obtenu à l'issue de  $n$  tirages

La variance est estimée ponctuellement

**Estimation par intervalle de confiance bilatéral symétrique:**

On choisit un niveau de confiance  $1 - \alpha$ . On détermine la valeur  $k$  à partir d'une table de la variable de Laplace-Gauss LG(0;1). telle que  $\text{Prob}(LG(0;1) < k) = 1 - \alpha/2$

Les deux situations usuelles sont:

- Intervalle de confiance à 95% :  $k=1,9600$
- Intervalle de confiance à 99% :  $k=2,5758$

$$f_n - k\sqrt{\frac{f_n(1-f_n)}{n-1}} \leq \pi \leq f_n + k\sqrt{\frac{f_n(1-f_n)}{n-1}}$$

Dans notre situation, l'échantillon est constitué de 1024 individus (visiteur/visite) dont  $138+131 = 269$  correspondent aux deux jours du week-end.  $R_n=269$ , soit une proportion de  $\frac{269}{1024} = 0.262653... \approx 26,27\%$ . Nous utilisons cette valeur comme estimation ponctuelle de la

proportion théorique des visites réalisées le week-end sur l'ensemble de toutes les visites possibles. Si maintenant nous souhaitons attacher un niveau de confiance à cette estimation, nous utilisons la formule ci-dessus qui nous conduit à :

$$0,2627 - 1,96\sqrt{\frac{0,2627(1-0,2627)}{1024-1}} \leq \pi \leq 0,2627 + 1,96\sqrt{\frac{0,2627(1-0,2627)}{1024-1}}$$

Ainsi à un niveau de confiance de 95%, nous pouvons affirmer que la proportion théorique  $p$  est contenue dans l'intervalle  $[0,2357 ; 0,2897]$ , c'est à dire entre 23,57% et 28,97%.

Q304- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

## TQ4- Étude de l'homogénéité de l'évaluation de la qualité informationnelle selon le genre

Q401- Peut-on considérer qu'il y a une différence d'intérêt entre les hommes et les femmes ?

Qualité informationnelle	0	1	2	3	4	
homme	78	70	165	60	90	463
femme	132	106	140	81	102	561
	210	176	305	141	192	1024

Tester les hypothèses correspondantes (Ho et H1) au seuil de 5% à l'aide du test d'homogénéité du Khi-deux.

Nous utilisons la procédure de calcul que nous explicitons par le biais des divers tableaux :

Tableau de **contingence (effectifs observés)**

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	78	70	165	60	90	463
Femme	132	106	140	81	102	561
	210	176	305	141	192	1024

Tableau des **fréquences observées**

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	0,08	0,07	0,16	0,06	0,09	0,45
Femme	0,13	0,10	0,14	0,08	0,10	0,55
	0,21	0,17	0,30	0,14	0,19	1,00

Tableau des **fréquences conditionnelles**

*Profil ligne*

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	0,17	0,15	0,36	0,13	0,19	1,00
Femme	0,24	0,19	0,25	0,14	0,18	1,00
	0,21	0,17	0,30	0,14	0,19	1,00

Tableau des **fréquences conditionnelles**

*Profil colonne*

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	0,37	0,40	0,54	0,43	0,47	0,45
Femme	0,63	0,60	0,46	0,57	0,53	0,55
	1,00	1,00	1,00	1,00	1,00	1,00

Tableau des **effectifs théoriques** sous Ho

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	94,95	79,58	137,91	63,75	86,81	463,00
Femme	115,05	96,42	167,09	77,25	105,19	561,00
	210,00	176,00	305,00	141,00	192,00	1024,00

Tableau des **contributions absolues**

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	3,0262	1,1528	5,3234	0,2209	0,1170	9,84
Femme	2,4976	0,9514	4,3935	0,1823	0,0966	8,12
	5,52	2,10	9,72	0,40	0,21	<b>d<sup>2</sup> = 17,96</b>

Tableau des **contributions relatives**

	0_Sans avis	1_Très forte	2_Plutôt forte	3_Plutôt faible	4_Très faible	
Homme	16,85%	6,42%	29,64%	1,23%	0,65%	54,79%
Femme	13,90%	5,30%	24,46%	1,02%	0,54%	45,21%
	30,75%	11,72%	54,10%	2,25%	1,19%	100,00%

**Phi<sup>2</sup> = 0,017540818**

**Prob( Khi-Deux > 17,96 ) = 0,00125549**

Q402- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

Sous l'hypothèse  $H_0$  d'homogénéité, avec un niveau de risque  $\alpha = 5\%$ , la valeur critique obtenu par lecture de la table du Khi-Deux ( $ddl=4$ ) est de 9,49. Au vu des résultats, il advient que nous pouvons rejeter l'hypothèse d'homogénéité au profit de l'hypothèse d'hétérogénéité avec un risque d'erreur de première espèce de 5%, puisque la valeur empirique  $d^2 = 17,96$

Nous observons par ailleurs que la probabilité d'avoir un résultat au moins aussi grand est plutôt faible puisqu'elle de l'ordre de 0,0012.

### TQ5- Étude de la répartition des visiteurs/visites quotidiennes.

Q501- A partir des informations obtenues sur l'échantillon et consignées dans le tableau n° T10, peut-on considérer que les visites sont également réparties sur la semaine ? (il s'agit de tester l'hypothèse  $H_0$  contre  $H_1$  à l'aide du test du Khi-deux d'adéquation avec un niveau de risque maximum de 5%)

### Adéquation à la distribution de fréquences uniforme

**Situation-problème paradigmatique** : étude de l'équilibrage d'un dé à 6 faces

Il s'agit de se donner une méthode pour décider entre deux hypothèses alternatives concernant l'état d'un dé cubique

$H_0$ le dé est bien équilibré	$H_0$ la distribution de fréquences des résultats est uniforme, les résultats sont équiprobables
$H_1$ le dé est truqué	$H_1$ au moins un des six résultats a une probabilité différente d'apparition des cinq autres

Pour pouvoir décider, on recourt à une expérience aléatoire.

On jette au hasard et de manière identique 336 fois le dé	Extraction d'un échantillon aléatoire avec remise de taille $n=336$ , de la population constituée par les six faces $P = \{1 ; 2 ; 3 ; 4 ; 5 ; 6\}$
---	--

Description des résultats obtenus sur l'échantillon :

Faces	Face1	Face2	Face3	Face4	Face5	Face6
Effectifs observés	47	49	78	56	60	46

Description des résultats théoriques que l'on peut ESPÉRER si le dé est parfaitement équilibré, sur cet échantillon

Faces	Face1	Face2	Face3	Face4	Face5	Face6
Effectifs théoriques	56	56	56	56	56	56

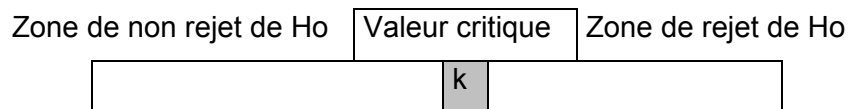
Mesure synthétique  $D^2$  de l'écart entre ces deux tableaux

Faces	Face1	Face2	Face3	Face4	Face5	Face6
Écart exact	$\frac{(47-56)^2}{56}$	$\frac{(49-56)^2}{56}$	$\frac{(78-56)^2}{56}$	$\frac{(56-56)^2}{56}$	$\frac{(60-56)^2}{56}$	$\frac{(46-56)^2}{56}$
Écart approché	1,446	0,875	8,643	0	0,286	1,786

Valeur de  $D^2$  dans cette expérience  $d^2 \approx 13,0357143$

Règle de décision : on utilise la table du Khi-deux ( $\chi^2$ ) de ddl = 6-1= 5

Il faut fixer le niveau de risque $\alpha$ puis lire dans la table. En choisissant un de deux niveaux habituels 5% ou 1%, nous obtenons	0,05	11,0704826
	0,01	15,0863174



Appliquons cette procédure avec les données fournies dans le tableau T10 :

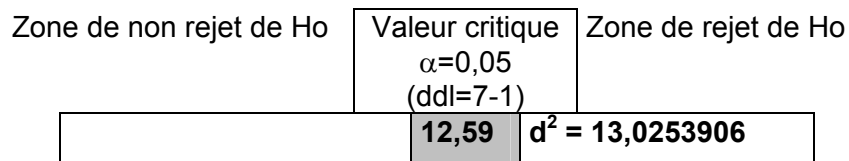
**test d'adéquation**

modalités	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche	effectif total
effectifs théoriques	146,29	146,29	146,29	146,29	146,29	146,29	146,29	1024
sous Ho observés	146	156	182	136	135	138	131	1024

$$d^2 = 0,00055804 + 0,64508929 + 8,71930804 + 0,72321429 + 0,87067522 + 0,46930804 + 1,59723772$$

$$d^2 = 13,0253906$$

$$\text{Prob}(D^2 > 13,0253906) = 0,01115239$$



Q502- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

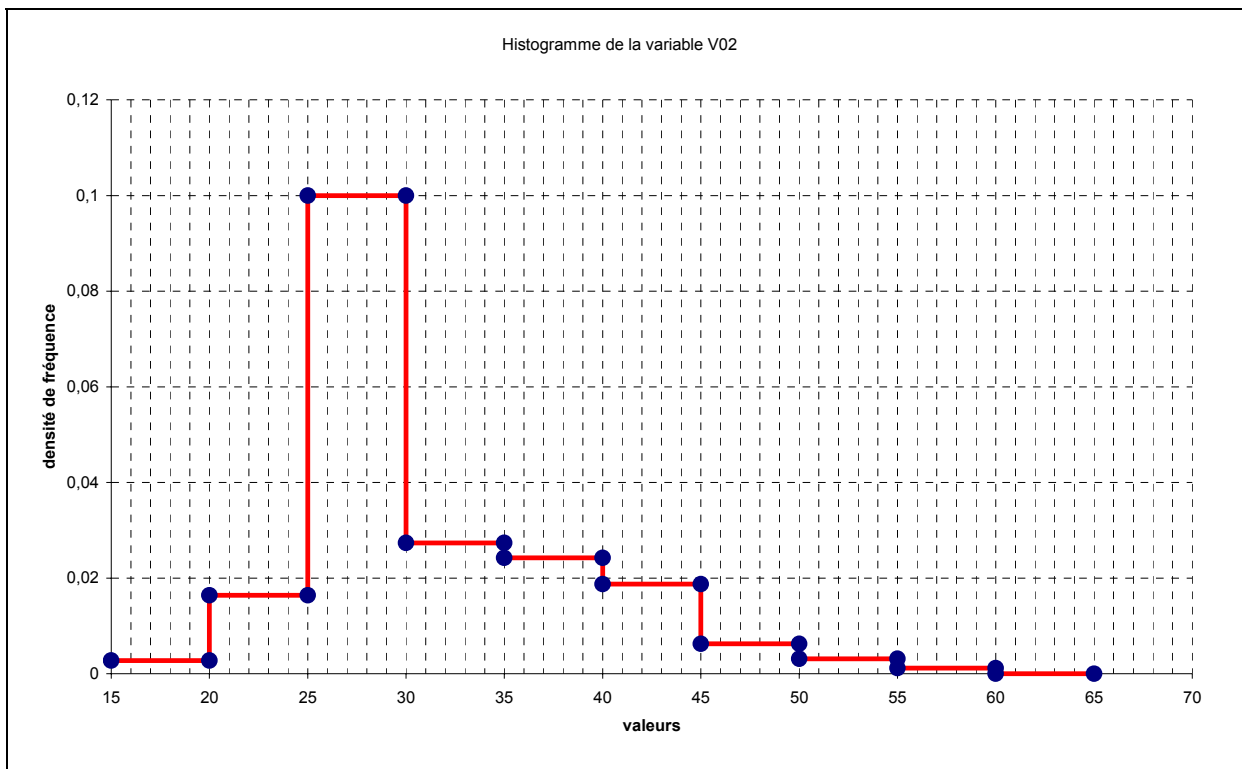
Au seuil de risque de 5%, nous rejetons l'hypothèse d'équi-répartition des visites sur la semaine. Il apparaît une journée de pointe significative, celle du mercredi.

TQ6-Etude de la variable V02

Q601 Quel est l'âge moyen des visiteurs sur la période avec quel écart-type ?

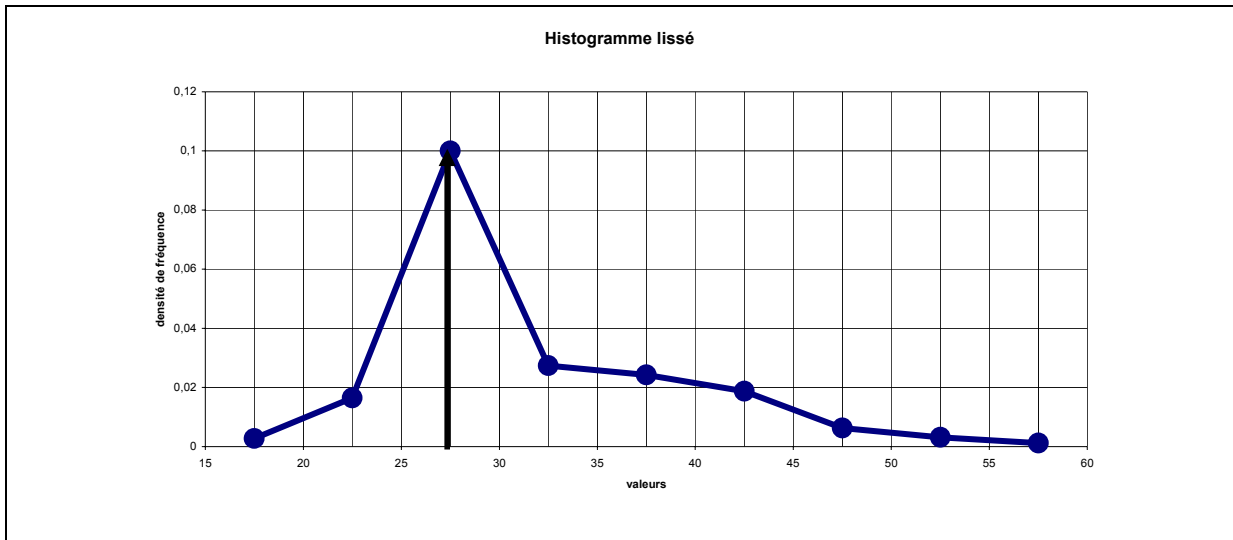
[a ; b[	Valeurs centrale s	effectifs	$n_k c_k$	$n_k(c_k-m)^2$
[15 ; 20[	17,5	14	245	2722,60437
[20 ; 25[	22,5	84	1890	6721,563721
[25 ; 30[	27,5	512	14080	7969,53125
[30 ; 35[	32,5	140	4550	155,7312012
[35 ; 40[	37,5	124	4650	4545,74585
[40 ; 45 [	42,5	96	4080	11731,78711
[45 ; 50[	47,5	32	1520	8248,095703
[50 ; 55[	52,5	16	840	7092,797852
[55 ; 60[	57,5	6	345	4073,080444
<b>Somme</b>		<b>1024</b>	<b>32200</b>	<b>Somme 53260,9375</b>
		<b>moyenne 31,4453</b>		<b>variance 52,01263428</b>
				<b>écart-type 7,211978527</b>

Q602 Construire l'histogramme de la distribution des fréquences, puis le polygone des fréquences.

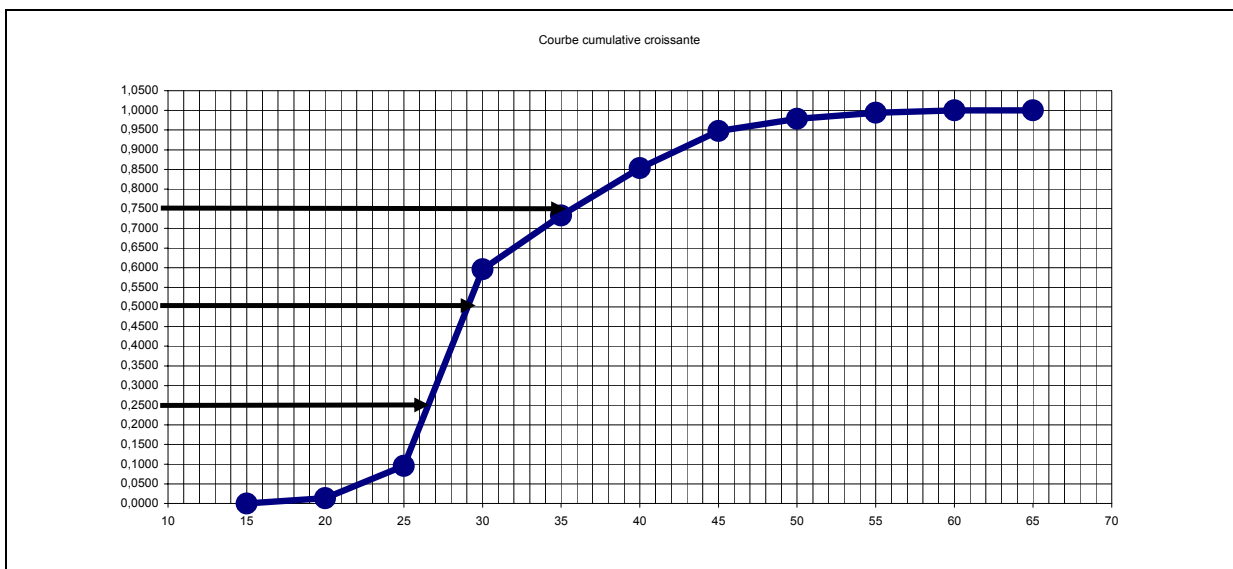


**Q603 Déterminer l'âge modal.**

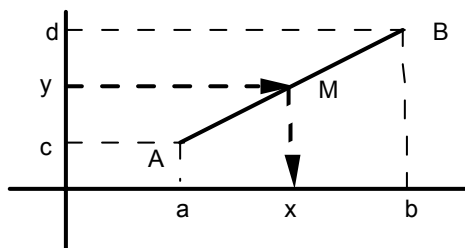
Comme le montre le graphique ci-dessous, une estimation ponctuelle de la valeur modale donne 27,5 ans.



**Q604 Déterminer l'âge médian.**



L'estimation de l'âge médian par interpolation linéaire donne de manière évidente que cette valeur est dans l'intervalle [25 ; 30[ Par une démarche fondée sur le raisonnement schématisé ci-dessous.



$$\frac{d-c}{b-a} = \frac{y-c}{x-a} = \frac{d-y}{b-x}$$

$$\frac{610-98}{30-25} = \frac{512-98}{x-25}$$

- a = 25**
- b = 30**
- x = Q2**
- c = 98**
- d = 610**
- y = 512**

nous obtenons

$$\frac{2070}{512} = x - 25$$

**x=29,0429...**

Comme le montre le graphique et la procédure algébrique ci-dessus, une estimation ponctuelle de la valeur médiane donne environ 29 ans.

Q605 Donner une **estimation par intervalle de confiance** à 95 % de l'âge moyen des visiteurs sur la population des internautes.

### Estimation d'une moyenne $\mu$ et d'une variance $\sigma^2$

#### Conditions d'utilisation:

- La variance  $\sigma^2$  de la population est inconnue
- L'échantillon est obtenu aléatoirement par n tirages avec ou sans remise dans une population de taille inconnue ou avec remise dans une population de taille N
- La taille n de l'échantillon est supérieure à 30 ou quelconque si la variable est distribuée selon une loi gaussienne sur la population

#### Estimateur:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (X_i - \bar{X})^2$$

La variable  $Y = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$  est une variable de Student  $T_{(n-1)}$  à n-1 ddl

#### Estimation ponctuelle de la moyenne:

Elle est obtenue à partir de la "moyenne calculée avec les valeurs observées sur

l'échantillon", c'est à dire le résultat de  $\bar{X}$  sur l'échantillon:  $m = \frac{1}{n} \sum_{i=1}^{i=n} x_i$

#### Estimation ponctuelle de la variance de la population:

Elle est obtenue à partir de la "variance calculée avec les valeurs observées sur l'échantillon", c'est à dire le résultat de  $S^2$  sur l'échantillon

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - m)^2 = \frac{n}{n-1} \sigma_{\text{échantillon}}^2$$

#### Estimation par intervalle de confiance bilatéral symétrique:

On choisit un niveau de confiance  $1 - \alpha$ . On détermine la valeur k à partir d'une table de la variable de Student  $T_{n-1}$  à n-1 degrés de liberté telle que  $\text{Prob}\{T_{n-1} < k\} = 1 - \frac{\alpha}{2}$

$$m - k \frac{s}{\sqrt{n}} < \mu < m + k \frac{s}{\sqrt{n}}$$

Il suffit alors de remplacer dans la formule ci-dessus les valeurs par les calculs effectués tout au long de ce traitement, soit  $m \approx 31,44$  ;  $n=1024$  ;  $s \approx 7,22$  et  $k=1,96$  (approximation par Laplace Gauss)

$$\left[ 31,44 - 1,96 \frac{7,22}{\sqrt{1024}} ; 31,44 + 1,96 \frac{7,22}{\sqrt{1024}} \right]$$

Ainsi avec un niveau de confiance de 95%, nous estimons que l'âge moyen est situé dans l'intervalle : [ 30,99 ; 31,99] soit entre 31 ans et 32 ans.

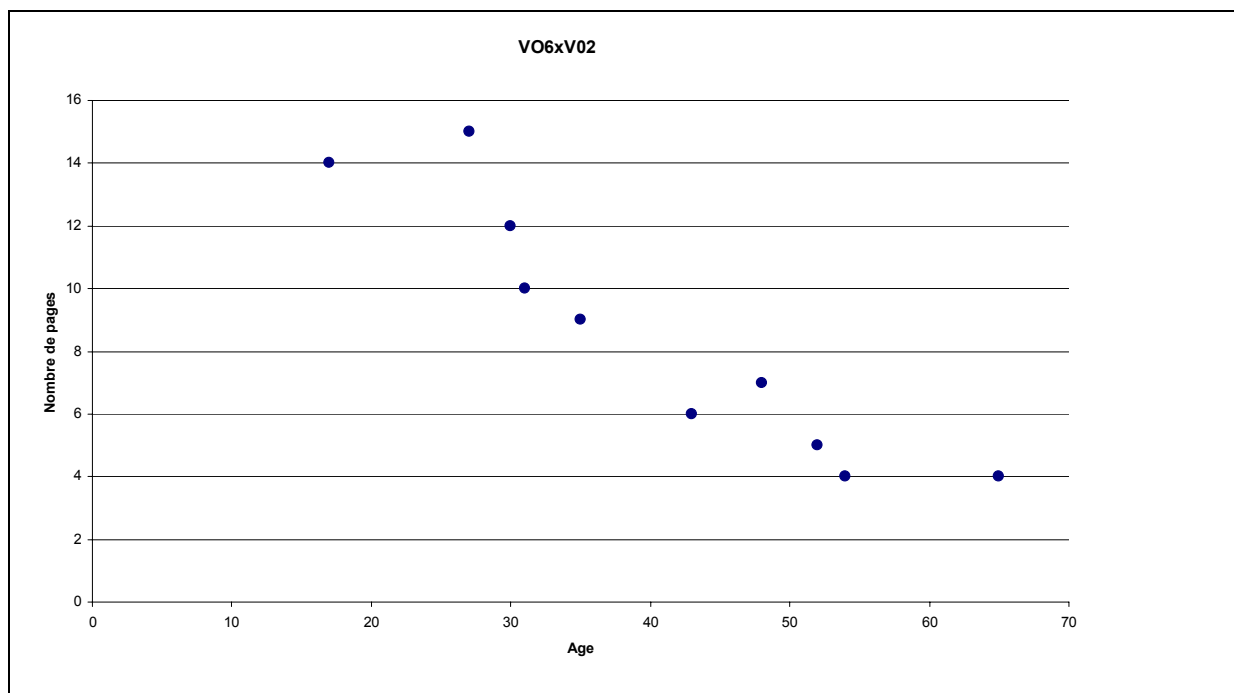
Q606 A partir de cette étude, énoncer sous une forme synthétique **vosre conclusion** et **vosre interprétation**.

### TQ7-Etude de la corrélation entre les variables V02 et V06

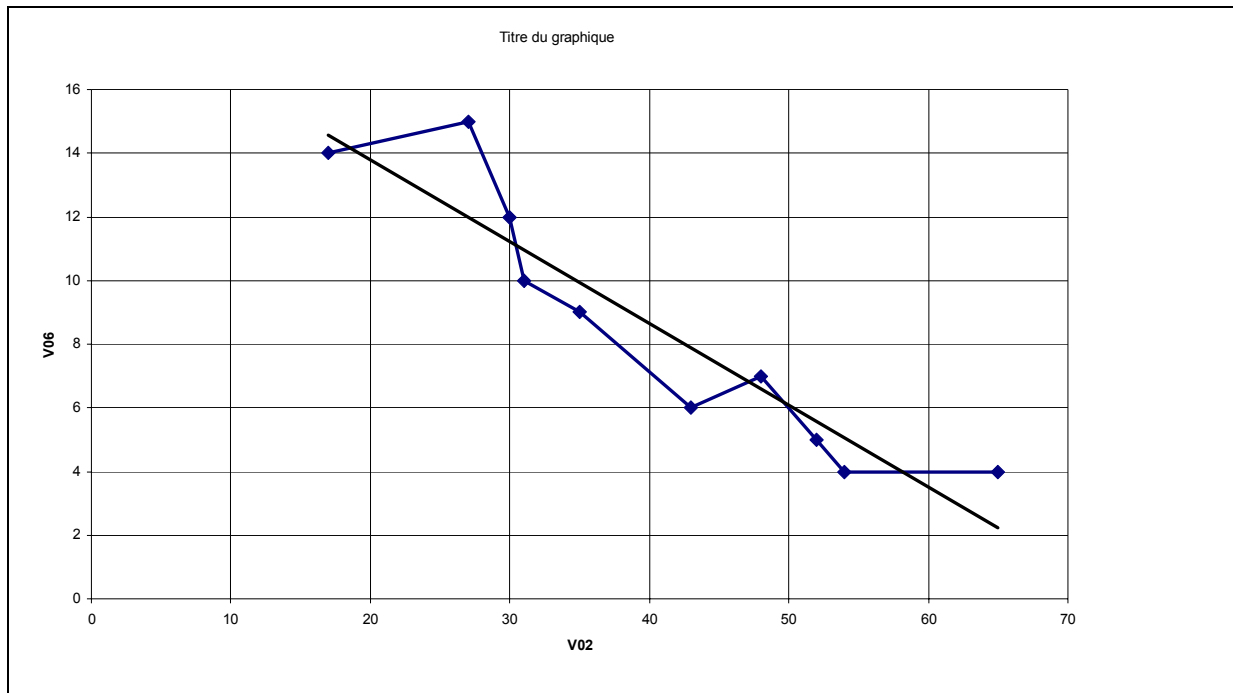
Sur un échantillon aléatoire de 10 individus, nous avons recueilli l'information suivante :

V02=Age	17	27	30	31	35	43	48	52	54	65
V06=Nombre de pages	14	15	12	10	9	6	7	5	4	4

Q701 Construire le **nuage statistique** ?



Q702 Quelle interprétation donnez-vous de la forme de ce nuage ?



Q703 Calculer le **coefficient de corrélation** linéaire de Bravais Pearson.

$$\text{Covariance (V02, V06)} = -49,92$$

$$\text{moyenne V02} = 40,2$$

$$\text{variance V02} = 194,16$$

$$\text{moyenne V06} = 8,6$$

$$\text{variance V06} = 14,84$$

$$\text{Coefficient de corrélation de Bravais Pearson} = -0,92998886$$

$$\text{Équation de la droite de régression : } Y = (-0,25710754) X + (18,9357231)$$

Q704 A partir de cette étude, énoncer sous une forme synthétique **votre conclusion** et **votre interprétation**.