

Nom :
Prénom :

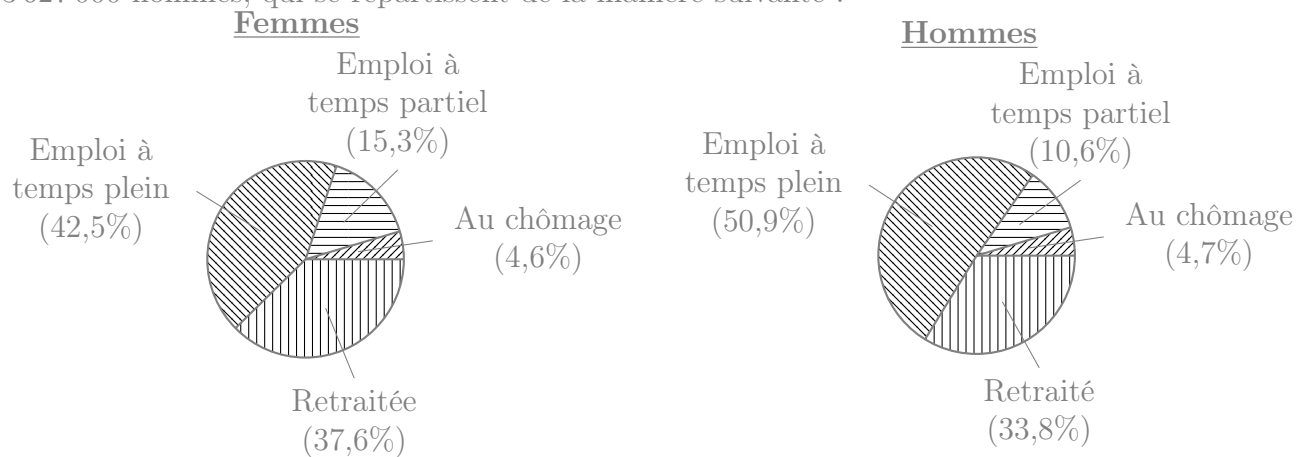
Le formulaire et la calculatrice sont autorisés. Vous rendrez l'énoncé auquel vous pouvez joindre une copie si vous le souhaitez. Vous pouvez soit répondre directement sur l'énoncé, soit détailler certaines questions sur la copie si vous avez besoin de plus de place. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.

Les inégalités de genre se manifestent dans de nombreux aspects de nos vies (santé, violences sexistes et sexuelles, monde du travail, etc). Par exemple, le site web du ministère <https://www.egalite-femmes-hommes.gouv.fr/> regroupe des statistiques très diverses à ce sujet. En particulier le salaire moyen des hommes est (en France) supérieur de 32,3% à celui des femmes, ce qui correspond pour partie au fait de ne pas occuper les mêmes postes et d'être moins souvent à temps partiel.

On se concentrera ici sur quelques indicateurs adaptés au programme de ce contrôle : Le premier exercice retranscrit des données réelles (données de 2021/2022, pour la population française), tandis que les exercices 2 et 3 produisent des données fictives qui se veulent proches de la réalité française.

Exercice 1 : Situation d'emploi

On choisit de regrouper les personnes actives (en situation d'emploi ou au chômage) et les retraité-e-s (au sein de la population française), obtenant ainsi un total de 23 946 000 femmes et 23 627 000 hommes, qui se répartissent de la manière suivante :



1. Pourquoi choisir de représenter ces données sous forme de “camemberts”, et pas sous la forme d’un “diagramme en bâton” ou d’un “histogramme” ?

La variable est qualitative, et les autres types de diagrammes (diagramme en bâton, histogramme, polygone des fréquences cumulées) ne sont possibles que pour des variables quantitatives.

2. Pour les femmes Françaises, calculer les effectifs de chacune des quatre modalités (chômage / emploi à temps partiel / emploi à plein temps / retraitée). *On pourra les indiquer dans le tableau ci-dessous*

On calcule les effectifs en multipliant chaque pourcentage par le nombre total de femmes considérées (c'est à dire 23 946 000). Par exemple le nombre de femmes au chômage est environ $23\,946\,000 \times \frac{4,6}{100} = 1\,101\,516$. On obtient de même :

Modalités	Emploi à temps plein	Emploi à temps partiel	Retraité	Au chômage
Fréquence	42,5%	15,3%	37,6%	4,6%
Effectif	$\simeq 10\,177\,000$	$\simeq 3\,664\,000$	$\simeq 9\,004\,000$	$\simeq 1\,102\,000$

3. Au sein des femmes qui exercent un emploi, quelle proportion exerce à temps partiel ?

On a obtenu qu'environ $10\,177\,000 + 3\,664\,000 = 13\,841\,000$ femmes exercent un emploi, parmi lesquelles environ 3 664 000 l'exercent à temps partiel. La proportion demandée est donc environ $\frac{3\,664\,000}{13\,841\,000} \simeq 0,265$.

4. Pour les hommes français, calculer les effectifs de chacune des quatre modalités (chômage / emploi à temps partiel / emploi à plein temps / retraitée). Indiquer ensuite quelle est, parmi les hommes qui exercent un emploi, la proportion qui exercent à temps partiel.

On calcule les effectifs en multipliant chaque pourcentage par le nombre total d'hommes considérés (c'est à dire 23 627 000) :

Modalités	Emploi à temps plein	Emploi à temps partiel	Retraitée	Au chômage
Fréquence	50,9%	10,6%	33,8%	4,7%
Effectif	$\simeq 12\,026\,000$	$\simeq 2\,504\,000$	$\simeq 7\,986\,000$	$\simeq 1\,110\,000$

On a obtenu qu'environ $12\,026\,000 + 2\,504\,000 = 14\,530\,000$ hommes exercent un emploi, parmi lesquelles environ 2 504 000 l'exercent à temps partiel. Donc

Parmi les hommes qui exercent un emploi, la proportion de temps partiels est environ $\frac{2\,504\,000}{14\,530\,000} \simeq 0,172$.

Exercice 2 : Revenu On cherche désormais à savoir si les écarts de revenus entre femmes et hommes s'expliquent entièrement par leur temps de travail (temps partiel / temps plein). Pour exclure cet effet du temps de travail, on décide dans cet exercice d'exprimer les salaires en équivalent temps plein (ÉQTP) ; par exemple, en France, le salaire ÉQTP moyen des hommes est de 2795€ mensuels net.

Les données ci-dessous indiquent le salaire net mensuel de 250 femmes françaises, en équivalent temps plein :

Revenu net (en €/mois)	[1400 ; 1600[[1600 ; 2000[[2000 ; 2800[[2800 ; 4400[[4400 ; 6000[
Effectif	44	75	78	37	16
Fréquence	0,176	0,300	0,312	0,148	0,064
Fréquence cumulée	0,176	0,476	0,788	0,936	1,000

1. Calculer la moyenne et l'écart-type du salaire mensuel (en ÉQTP) des femmes de cet échantillon. *Indiquez les calculs effectués, y compris les calculs intermédiaires).*

$$\begin{aligned}
 \text{moyenne : } m(X) &= \frac{\sum c_i n_i}{n} = \frac{1500 \times 44 + 1800 \times 75 + 2400 \times 78 + \dots + 5200 \times 16}{250} = \frac{604\,600}{250} = 2\,418,4 \text{ €} \\
 m(X^2) &= \frac{\sum c_i^2 n_i}{n} = \frac{1500^2 \times 44 + 1800^2 \times 75 + 2400^2 \times 78 + \dots + 5200^2 \times 16}{250} = \frac{1\,703\,440\,000}{250} \\
 \text{Var}(X) &= m(X^2) - m(X)^2 = \frac{1\,703\,440\,000}{250} - \left(\frac{604\,600}{250}\right)^2 = 965\,101,44 \\
 \text{Écart-type : } s(X) &= \sqrt{\text{Var}(X)} \simeq 982,4 \text{ €}
 \end{aligned}$$

2. Calculer les fréquences et les fréquences cumulées. *On pourra simplement les ajouter au tableau des données au début de l'exercice.*

3. Donner la médiane du salaire mensuel (en ÉQTP) des femmes de cet échantillon.

$$\begin{aligned}
 &\text{Classe de la médiane : } [2000 ; 2800[\\
 \text{Méd} &\simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 2000 + \frac{2800 - 2000}{0,788 - 0,476} (0,5 - 0,476) \simeq 2061,538 \text{ €}
 \end{aligned}$$

Exercice 3 : Échantillon de 8 femmes et 8 hommes

On considère désormais un échantillon de femmes et d'hommes occupant le même type de poste dans le même secteur d'activité. On leur a demandé leur genre (femme/homme), leur durée de travail annuelle, notée X (impactée notamment par le recours au temps partiel et/ou par la durée du contrat s'il dure moins d'un an), leur salaire annuel net noté Y , si ils/elles sont à la tête d'une famille monoparentale, et enfin si ils/elles ont déjà été victime d'au moins une agression dans les transports (le cas échéant on leur demande combien) :

Femmes								
	Mme A.	Mme B.	Mme C.	Mme D.	Mme E.	Mme F.	Mme G.	Mme H.
Mère de famille monoparentale	non	oui	non	non	non	non	non	non
Nombre d'agressions subies dans les transports	0	0	0	0	0	0	plus de 3	0
Durée de travail (X)	1 607 h	1 607 h	1 339 h	1 286 h	134 h	1 473 h	804 h	671 h
Salaire annuel (Y)	25,3 k€	24,4 k€	20,9 k€	19,7 k€	3,2 k€	21,6 k€	12,3 k€	10,2 k€
Rangs x'_i	7,5	7,5	5	4	1	6	3	2
Rangs y'_i	8	7	5	4	1	6	3	2

Hommes								
	Mr S.	Mr T.	Mr U.	Mr V.	Mr W.	Mr X.	Mr Y.	Mr Z.
Père de famille monoparentale	non	non	non	non	non	non	non	non
Nombre d'agressions subies dans les transports	0	0	0	0	0	0	0	0
Durée de travail (X)	603 h	804 h	1 607 h	857 h	536 h	1 607 h	1 722 h	402 h
Salaire annuel (Y)	9,5 k€	15 k€	25,7 k€	13,7 k€	9,1 k€	24,1 k€	28,4 k€	7,2 k€

1. Quelles sont les variables statistiques étudiées ? Indiquer la nature de chaque variable.

Les variables sont :

- Genre : qualitative (nominale)
- Monoparentalité : qualitative (nominale)
- Nombre d'agressions subies dans les transports : quantitative discrète
- Durée de travail : quantitative (continue)
- Salaire Annuel : quantitative (discrète, car on a toujours un nombre entier de centimes)

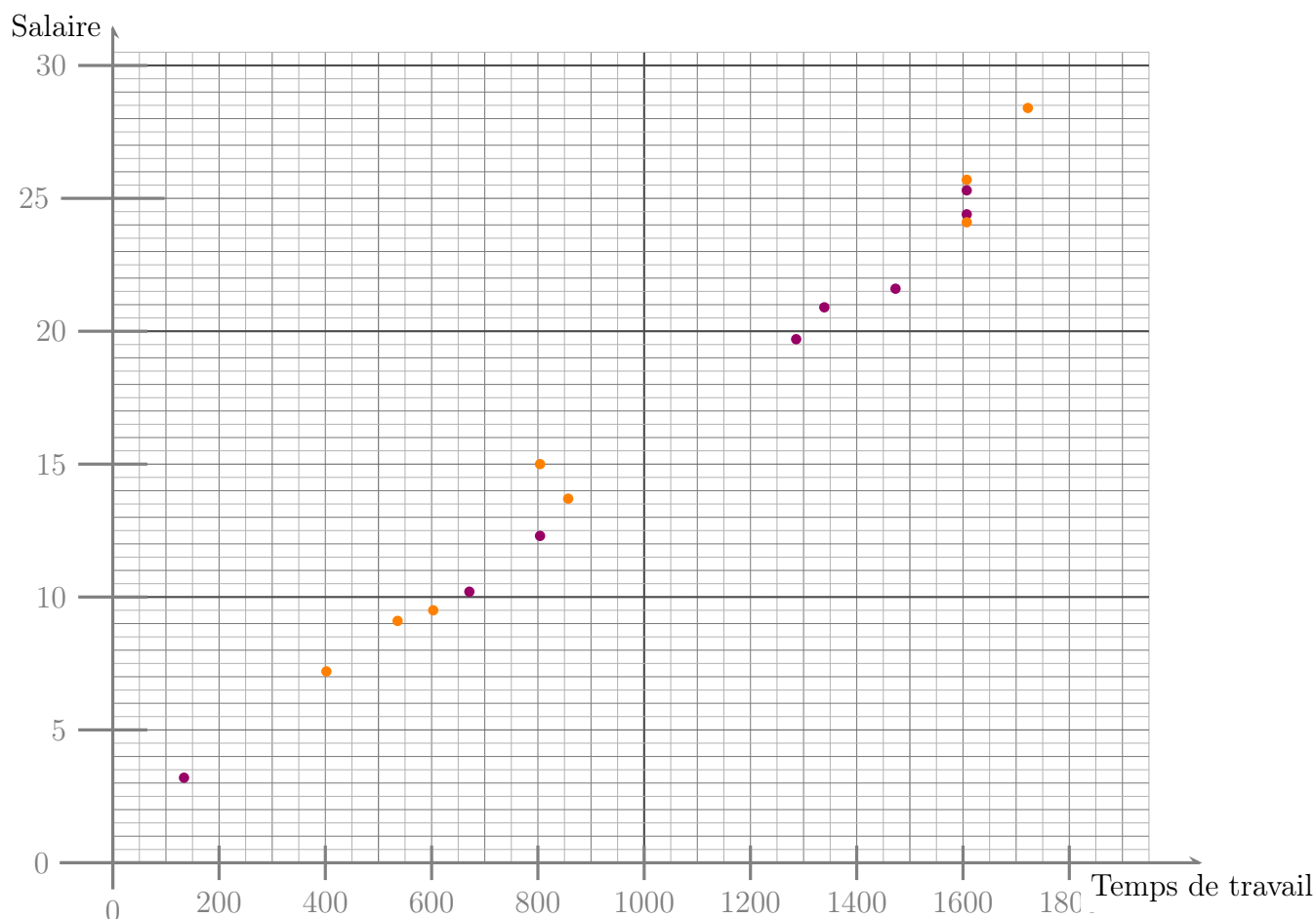
2. Quel est salaire médian des femmes de cet échantillon ?

La médiane est la valeur numéro $\frac{8+1}{2} = 4,5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 4 et 5. C'est donc le milieu entre 19,7 et 20,9 c'est-à-dire $\frac{19,7+20,9}{2} = 20,3$ k€.

3. Dans la suite de cet exercice on s'intéresse uniquement au lien entre la durée de travail et le salaire perçu.

- a) Tracer le nuage de points des données correspondant à cet échantillon de 16 personnes. Vous utiliserez une couleur différente pour les hommes et les femmes (ou marquez les points avec un type de croix différentes, par exemple “×” et “+”).

Voir les points ci-dessous, où on a représenté les femmes en mauve et les hommes en orange.



b) On se concentre désormais uniquement sur les femmes de l'échantillon :

- i) Calculer le coefficient de corrélation de rangs (“de Spearman”) entre la durée de travail et le salaire perçu, pour les femmes de cet échantillon. Que pouvez-vous conclure au vu de la valeur de ce coefficient ?

On calcule tout d'abord les rangs x'_i et y'_i , entrés dans la table au début de l'exercice.

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left(6 \times \frac{(7,5-8)^2 + (7,5-7)^2 + \dots + (2-2)^2}{8(8^2-1)} \right) \simeq 0,994$$

On conclut qu'il y a un très fort lien entre la durée du travail et le salaire perçu par les femmes de cet échantillon, et que le salaire a tendance à augmenter quand le temps de travail augmente.

- ii) Calculer le coefficient de corrélation linéaire entre la durée de travail et le salaire perçu, pour les femmes de cet échantillon. Que pouvez-vous conclure au vu de la valeur de ce coefficient ? On demande dans cette question de présenter les calculs intermédiaires effectués.

$$\begin{aligned}\text{moyenne : } m(X) &= \frac{\sum x_i}{n} = \frac{1\,607+1\,607+1\,339+\dots+671}{8} = \frac{8\,921}{8} \simeq 1\,115,13 \\ m(X^2) &= \frac{\sum x_i^2}{n} = \frac{1\,607^2+1\,607^2+1\,339^2+\dots+671^2}{8} = \frac{11\,895\,957}{8} \\ \text{Var}(X) &= m(X^2) - m(X)^2 = \frac{11\,895\,957}{8} - \left(\frac{8\,921}{8}\right)^2 \simeq 243\,490,86 \\ \text{Écart-type : } s(X) &= \sqrt{\text{Var}(X)} \simeq 493,45\end{aligned}$$

$$\begin{aligned}\text{moyenne : } m(Y) &= \frac{\sum x_i}{n} = \frac{25,3+24,4+20,9+\dots+10,2}{8} = \frac{137,6}{8} = 17,2 \\ m(Y^2) &= \frac{\sum x_i^2}{n} = \frac{25,3^2+24,4^2+20,9^2+\dots+10,2^2}{8} = \frac{2\,792,48}{8} \\ \text{Var}(Y) &= m(Y^2) - m(Y)^2 = \frac{2\,792,48}{8} - \left(\frac{137,6}{8}\right)^2 = 53,22 \\ \text{Écart-type : } s(Y) &= \sqrt{\text{Var}(Y)} \simeq 7,3\end{aligned}$$

$$\begin{aligned}m(XY) &= \frac{\sum x_i y_i}{n} = \frac{1\,607 \times 25,3 + 1\,607 \times 24,4 + \dots + 671 \times 10,2}{8} = \frac{182\,166,2}{8} = 22\,770,775 \\ \text{Cov}(X,Y) &= m(XY) - m(X)m(Y) = \frac{182\,166,2}{8} - \frac{8\,921}{8} \frac{137,6}{8} = 3\,590,625 \\ r(X,Y) &= \frac{\text{Cov}(X,Y)}{s(X)s(Y)} = \frac{3\,590,625}{493,45 \times 7,3} \simeq 0,996\,8.\end{aligned}$$

On conclue qu'il y a un très fort lien (et cette fois-ci on peut affirmer qu'il est linéaire) entre la durée du travail et le salaire perçu par les femmes de cet échantillon, et que le salaire a tendance à augmenter quand le temps de travail augmente.

- iii) Si une femme travaille 1527 heures par an sur ce type de poste dans ce secteur d'activité, quel salaire peut-elle s'attendre à percevoir ?

Pour estimer Y lorsque $X = 1527$, on doit utiliser la droite $D_{Y|X}$ (on peut le faire car il y a une forte corrélation linéaire).

$$\text{on pose } a = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \simeq \frac{3\,590,625}{243\,490,86} \simeq 0,01475 \text{ et } b = m(Y) - a m(X) \simeq 17,2 - 0,01475 \times 1\,115,125 \simeq 0,7519$$

D'où l'équation de la droite $D_{Y|X} : Y = 0,01475 X + 0,7519$

Donc pour $x = 1527$ h, on s'attend à $y = 0,01475 \times 1\,527 + 0,7519 \simeq 23,3\text{k€}$.

Remarque : Les données de cet exercice sont peu concluantes (échantillon de seulement 8 femmes et 8 hommes), mais d'après les statistiques nationales, à type de poste égal et temps de travail égal, les hommes gagnent environ 4,5% de plus que les femmes.

Mais comme les femmes sont en fait plus souvent à temps partiel, et n'accèdent pas aux même type de postes, le salaire moyen des hommes (dans le secteur privé) est en fait supérieur de 32,3% à celui des femmes. (Ce qui revient au même que de dire que le salaire moyen des femmes est inférieur de 24,4% à celui des hommes, dans le secteur privé.)