

# TP2\_SEM

2025-10-06

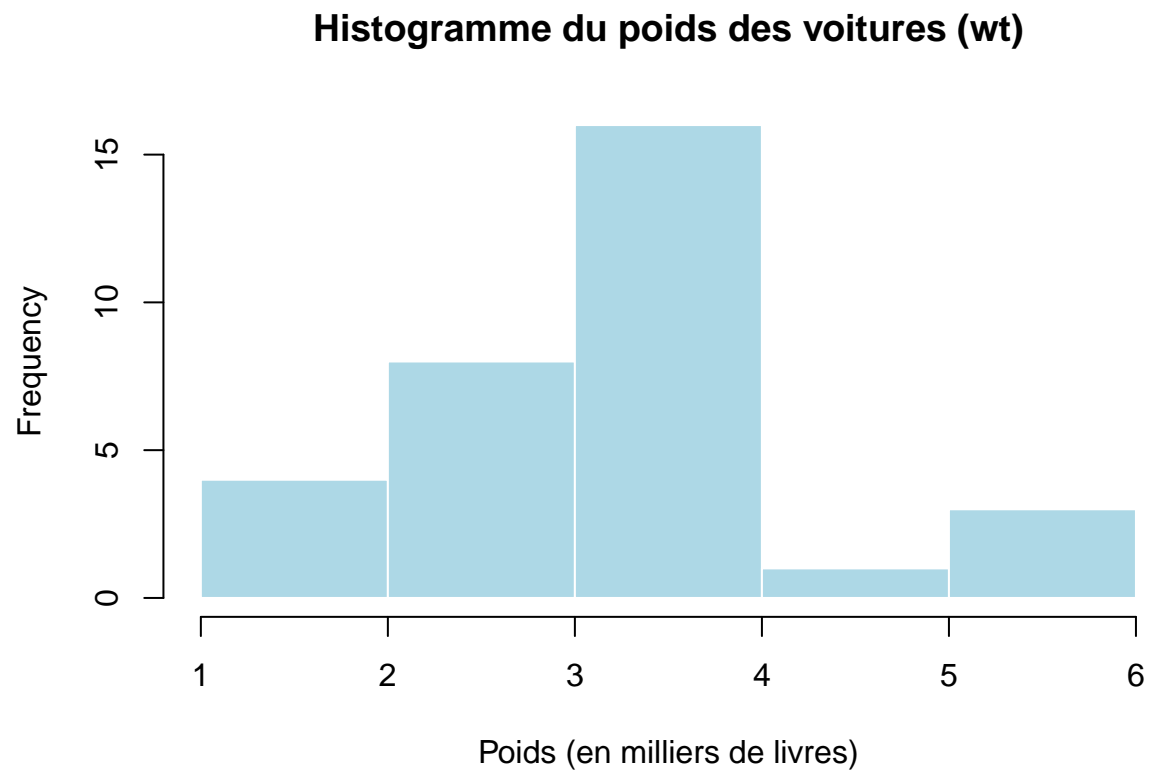
## 1. Représentations tabulaires et graphiques

```
# Charger le jeu de données intégré  
data(mtcars)
```

```
# 1.1 Tableau de contingence entre 'cyl' et 'am'  
tab_cont <- table(mtcars$cyl, mtcars$am)  
tab_cont
```

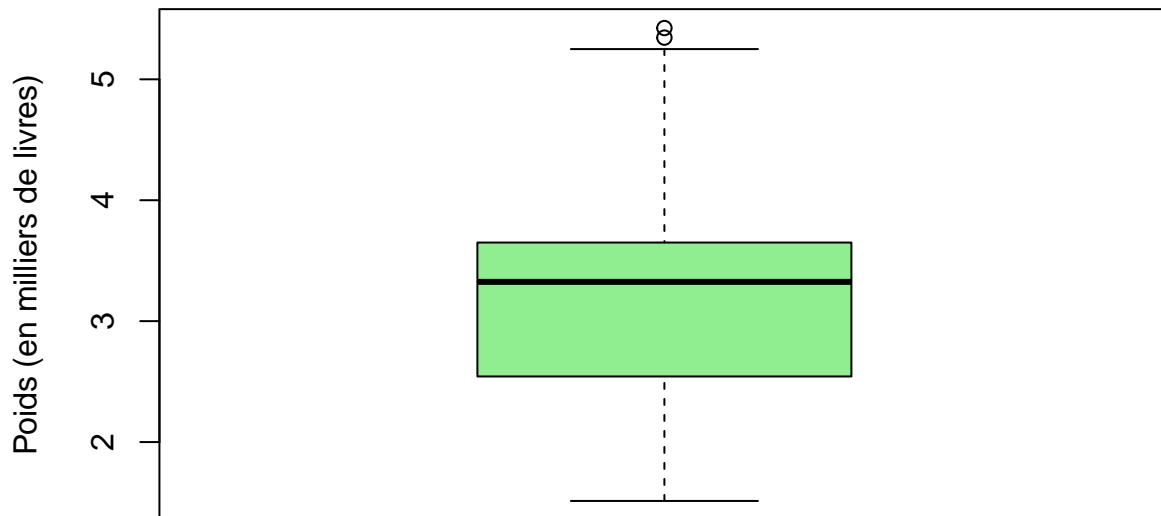
```
##  
##      0  1  
##   4  3  8  
##   6  4  3  
##   8 12  2
```

```
# 1.2 Histogramme des poids ('wt') avec 5 classes (bins)
hist(mtcars$wt, breaks = 5, main = "Histogramme du poids des voitures (wt)",
     xlab = "Poids (en milliers de livres)", col = "lightblue", border = "white")
```



```
# 1.3 Diagramme en boîte du poids ('wt')
boxplot(mtcars$wt, main = "Diagramme en boîte du poids des voitures",
        ylab = "Poids (en milliers de livres)", col = "lightgreen")
```

## Diagramme en boîte du poids des voitures



## 2. Corrélation

```
# Données
a <- c(11, 15, 23, 46, 52, 75)
w <- c(34, 49, 58, 62, 69, 64)
```

```
# 2.1 Formule de corrélation de Pearson
```

```
# Calcul "à la main"
cov_aw <- cov(a, w)
sd_a <- sd(a)
sd_w <- sd(w)
r_aw <- cov_aw / (sd_a * sd_w)
r_aw
```

```
## [1] 0.7906587
```

```
# Vérification avec la fonction cor()
cor(a, w)
```

```
## [1] 0.7906587
```

## 2.2 Formule de corrélation de Spearman :

**Calcul à la main :** On a  $d = \sum_{i=1}^6 (r(a_i) - r(w_i))^2 = 2$  avec  $r(a_i)$  le rang de la valeur  $a_i$  lorsque l'on ordonne la liste  $(a_1, a_2, \dots, a_6)$  par ordre croissant.  $r(w_i)$  est défini de façon analogue.

On a donc

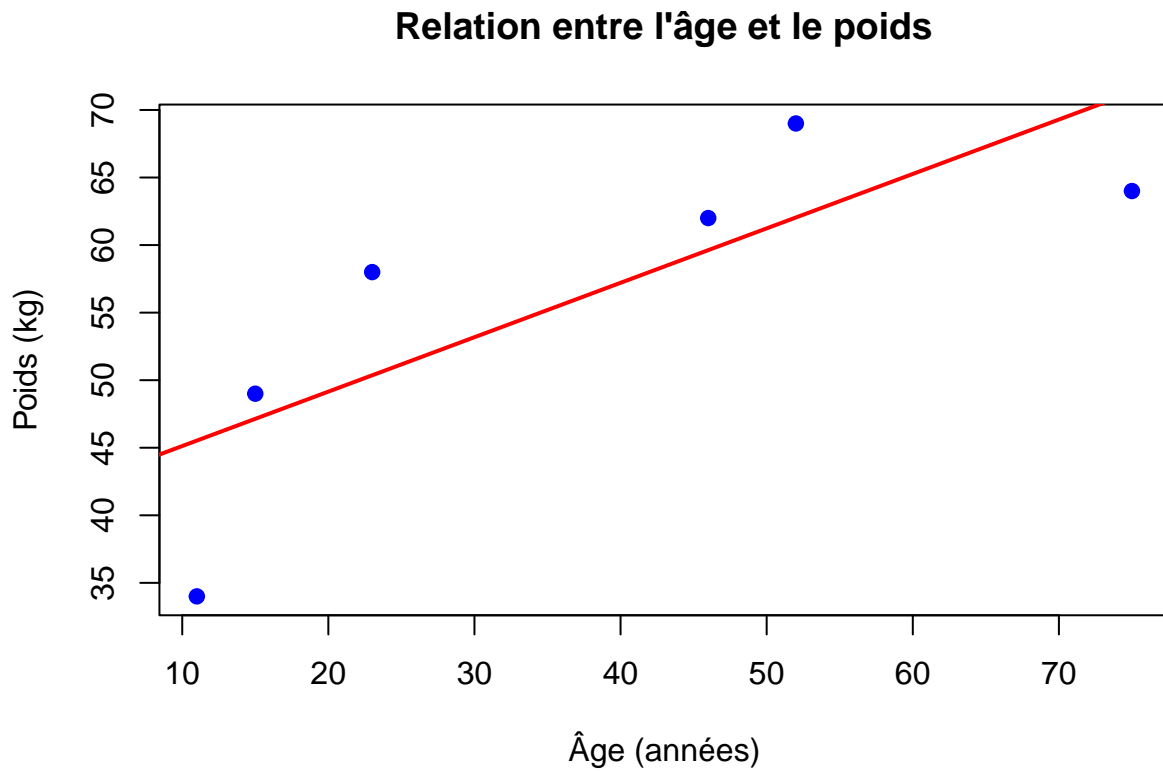
$$\rho = 1 - 6 \frac{d}{6(6^2 - 1)} = 1 - \frac{2}{35} = \frac{33}{35} \approx 0,943$$

```
# Vérification avec la fonction cor()
cor(a, w, method = "spearman")
```

```
## [1] 0.9428571
```

```
# Relation entre âge et poids
plot(a, w, main = "Relation entre l'âge et le poids",
     xlab = "Âge (années)", ylab = "Poids (kg)", pch = 19, col = "blue")

# Optionnel : ajouter une droite de tendance
abline(lm(w ~ a), col = "red", lwd = 2)
```



Peu de données pour un résultat correct mais si on devait examiner quelque chose, on peut voir qu'avec

l'âge, le poids augmente et semble commencer à se stabiliser vers la cinquantaine. **ATTENTION** : ceci est juste un échantillon de seulement 6 personnes, il faudrait un échantillon beaucoup plus grande pour conclure quoique ce soit.

### 3. Data Frames

```
# Chager le jeu données puis le convertir
state.x77
```

##	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost
## Alabama	3615	3624	2.1	69.05	15.1	41.3	20
## Alaska	365	6315	1.5	69.31	11.3	66.7	152
## Arizona	2212	4530	1.8	70.55	7.8	58.1	15
## Arkansas	2110	3378	1.9	70.66	10.1	39.9	65
## California	21198	5114	1.1	71.71	10.3	62.6	20
## Colorado	2541	4884	0.7	72.06	6.8	63.9	166
## Connecticut	3100	5348	1.1	72.48	3.1	56.0	139
## Delaware	579	4809	0.9	70.06	6.2	54.6	103
## Florida	8277	4815	1.3	70.66	10.7	52.6	11
## Georgia	4931	4091	2.0	68.54	13.9	40.6	60
## Hawaii	868	4963	1.9	73.60	6.2	61.9	0
## Idaho	813	4119	0.6	71.87	5.3	59.5	126
## Illinois	11197	5107	0.9	70.14	10.3	52.6	127
## Indiana	5313	4458	0.7	70.88	7.1	52.9	122
## Iowa	2861	4628	0.5	72.56	2.3	59.0	140
## Kansas	2280	4669	0.6	72.58	4.5	59.9	114
## Kentucky	3387	3712	1.6	70.10	10.6	38.5	95
## Louisiana	3806	3545	2.8	68.76	13.2	42.2	12
## Maine	1058	3694	0.7	70.39	2.7	54.7	161
## Maryland	4122	5299	0.9	70.22	8.5	52.3	101
## Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103
## Michigan	9111	4751	0.9	70.63	11.1	52.8	125
## Minnesota	3921	4675	0.6	72.96	2.3	57.6	160
## Mississippi	2341	3098	2.4	68.09	12.5	41.0	50
## Missouri	4767	4254	0.8	70.69	9.3	48.8	108
## Montana	746	4347	0.6	70.56	5.0	59.2	155
## Nebraska	1544	4508	0.6	72.60	2.9	59.3	139
## Nevada	590	5149	0.5	69.03	11.5	65.2	188
## New Hampshire	812	4281	0.7	71.23	3.3	57.6	174
## New Jersey	7333	5237	1.1	70.93	5.2	52.5	115
## New Mexico	1144	3601	2.2	70.32	9.7	55.2	120
## New York	18076	4903	1.4	70.55	10.9	52.7	82
## North Carolina	5441	3875	1.8	69.21	11.1	38.5	80
## North Dakota	637	5087	0.8	72.78	1.4	50.3	186
## Ohio	10735	4561	0.8	70.82	7.4	53.2	124
## Oklahoma	2715	3983	1.1	71.42	6.4	51.6	82
## Oregon	2284	4660	0.6	72.13	4.2	60.0	44
## Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126
## Rhode Island	931	4558	1.3	71.90	2.4	46.4	127
## South Carolina	2816	3635	2.3	67.96	11.6	37.8	65
## South Dakota	681	4167	0.5	72.08	1.7	53.3	172
## Tennessee	4173	3821	1.7	70.11	11.0	41.8	70

## Texas	12237	4188	2.2	70.90	12.2	47.4	35
## Utah	1203	4022	0.6	72.90	4.5	67.3	137
## Vermont	472	3907	0.6	71.64	5.5	57.1	168
## Virginia	4981	4701	1.4	70.08	9.5	47.8	85
## Washington	3559	4864	0.6	71.72	4.3	63.5	32
## West Virginia	1799	3617	1.4	69.48	6.7	41.6	100
## Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149
## Wyoming	376	4566	0.6	70.29	6.9	62.9	173
##	Area						
## Alabama	50708						
## Alaska	566432						
## Arizona	113417						
## Arkansas	51945						
## California	156361						
## Colorado	103766						
## Connecticut	4862						
## Delaware	1982						
## Florida	54090						
## Georgia	58073						
## Hawaii	6425						
## Idaho	82677						
## Illinois	55748						
## Indiana	36097						
## Iowa	55941						
## Kansas	81787						
## Kentucky	39650						
## Louisiana	44930						
## Maine	30920						
## Maryland	9891						
## Massachusetts	7826						
## Michigan	56817						
## Minnesota	79289						
## Mississippi	47296						
## Missouri	68995						
## Montana	145587						
## Nebraska	76483						
## Nevada	109889						
## New Hampshire	9027						
## New Jersey	7521						
## New Mexico	121412						
## New York	47831						
## North Carolina	48798						
## North Dakota	69273						
## Ohio	40975						
## Oklahoma	68782						
## Oregon	96184						
## Pennsylvania	44966						
## Rhode Island	1049						
## South Carolina	30225						
## South Dakota	75955						
## Tennessee	41328						
## Texas	262134						
## Utah	82096						
## Vermont	9267						

```
## Virginia      39780
## Washington    66570
## West Virginia 24070
## Wisconsin     54464
## Wyoming       97203
```

```
# Vérifier le type
class(state.x77)
```

```
## [1] "matrix" "array"
```

```
# Conversion si nécessaire
state_df <- as.data.frame(state.x77)
```

```
# Vérifier à nouveau
class(state_df)
```

```
## [1] "data.frame"
```

```
# Ajouter un attribut "states" avec les noms de lignes
state_df$states <- rownames(state_df)
```

```
# Supprimer les noms de lignes
rownames(state_df) <- NULL
```

```
# Afficher les 6 premières lignes
head(state_df)
```

```
##   Population Income Illiteracy Life Exp Murder HS Grad Frost   Area   states
## 1      3615   3624         2.1   69.05   15.1   41.3    20  50708  Alabama
## 2       365   6315         1.5   69.31   11.3   66.7   152 566432   Alaska
## 3      2212   4530         1.8   70.55    7.8   58.1    15 113417   Arizona
## 4      2110   3378         1.9   70.66   10.1   39.9    65  51945   Arkansas
## 5     21198   5114         1.1   71.71   10.3   62.6    20 156361 California
## 6      2541   4884         0.7   72.06    6.8   63.9   166 103766   Colorado
```

```
# Filtrer les États selon les deux conditions
etat_filtre <- subset(state_df, Income > 4300 & Population > 1000)
```

```
# Nombre d'États
nrow(etat_filtre)
```

```
## [1] 22
```

```
# Liste des États concernés
etat_filtre$states
```

```
## [1] "Arizona"      "California"    "Colorado"      "Connecticut"
## [5] "Florida"      "Illinois"      "Indiana"        "Iowa"
## [9] "Kansas"       "Maryland"      "Massachusetts"  "Michigan"
## [13] "Minnesota"    "Nebraska"      "New Jersey"     "New York"
## [17] "Ohio"         "Oregon"        "Pennsylvania"   "Virginia"
## [21] "Washington"   "Wisconsin"
```

```
# Trier les États par revenu décroissant
state_sorted <- state_df[order(-state_df$Income), ]
```

```
# Afficher les 5 premiers
head(state_sorted[, c("states", "Income")], 5)
```

```
##           states Income
## 2      Alaska   6315
## 7 Connecticut   5348
## 20    Maryland   5299
## 30 New Jersey   5237
## 28     Nevada   5149
```

```
# Créer la variable catégorielle selon les intervalles
state_df$frost_cat <- cut(state_df$Frost,breaks = c(-1, 30, 90, 190),
  labels = c("faible", "intermédiaire", "élevé"))
```

```
# Afficher les États de la catégorie "faible"
subset(state_df, frost_cat == "faible", select = c(states, Frost, frost_cat))
```

```
##           states Frost frost_cat
## 1    Alabama    20    faible
## 3    Arizona    15    faible
## 5 California    20    faible
## 9    Florida    11    faible
## 11   Hawaii      0    faible
## 18 Louisiana    12    faible
```