

Initiation à l'analyse de la variance à un facteur (ANOVA) avec des données d'automesure tensionnelles (AMT)

Pr Jean R. LOBRY

Un jeu de données minimaliste est utilisé pour illustrer l'analyse de la variance : 6 auto-mesures (répétées 3 fois) de la pression artérielle (systolique et diastolique). Les deux facteurs considérés sont le jour de la mesure (trois modalités) et la tranche horaire (matin et soir).

1 Les données

EN cas de suspicion d'hypertension artérielle, votre médecin traitant vous demandera peut-être de procéder à la mesure de votre tension à domicile (voir la figure 1 page 2). L'objectif est de lutter contre l'effet « blouse blanche » qui est à la pratique médicale ce que la relation d'incertitude d'HEISENBERG est à la physique. Pour une introduction aux automesures tensionnelles (AMT) je recommande la thèse de doctorat ès sciences de Jean-Marc BOIVIN [1] dont la lecture est accessible même pour un non spécialiste.

LES unités utilisées sont particulières : des millimètres de mercure¹ avec 100 mmHg comme ordre de grandeur. Elles se raccrochent au système international par la relation :

$$1 \text{ mmHg} = \frac{101325}{760} \text{ Pa}$$

CE sont donc des unités de pression (N.m^{-2}). Parler de *tension* artérielle n'est pas complètement un abus de langage : dans la mesure où la surface interne du système de circulation sanguin est approximativement constante, la pression est, à une constante multiplicative près, égale à la tension (N) qui s'exerce sur les artères.

AVEC 10^2 mmHg comme ordre de grandeur nous sommes dans le domaine des basses pressions. Une pression atmosphérique typique de 1000 hPa est de

¹Dans le langage courant on utilise plutôt des cmHg : « Il a 5 de tens' » pour désigner quelqu'un de particulièrement apathique. L'avantage des mmHg est de ne pas avoir à gérer le séparateur décimal et les arrondis.

RELEVÉ D'AUTOMESURE TENSIONNELLE

Pour affirmer le diagnostic d'une hypertension artérielle et pour mettre en place un traitement adapté et efficace, le médecin propose la réalisation d'un relevé d'automesure par le patient lui-même. En effet, la tension artérielle d'une personne en bonne santé n'est pas stable. Elle fluctue en fonction de l'environnement. Émotion, stress, activité physique... influent directement sur la pression artérielle.

QUAND EFFECTUER LA MESURE DE VOTRE TENSION ?

La règle des 3 mesures :

- La tension artérielle est mesurée 3 fois le matin au début du petit-déjeuner avant toute prise de médicaments et 3 fois le soir avant le coucher (avec un intervalle de 2 minutes entre chaque mesure) ;
- Pendant 3 jours consécutifs.

Modalités de mesure de la tension : en position assise, après 5 minutes de repos.

COMMENT REMPLIR LE RELEVÉ D'AUTOMESURE TENSIONNELLE ?

Inscrivez tous les chiffres qui apparaissent sur l'écran du tensiomètre pour la pression systolique et diastolique.

- Systolique = pression systolique = pression maximum
- Diastolique = pression diastolique = pression minimum

Calculez votre moyenne :

- moyenne systolique > additionnez l'ensemble des mesures systoliques et divisez par 18
- moyenne diastolique > additionnez l'ensemble des mesures diastoliques et divisez par 18

En cas de doute, votre médecin effectuera cette moyenne. **BRAUN ExactFit™ 5 connect BU46350**

	JOUR 1		JOUR 2		JOUR 3	
	systolique	diastolique	systolique	diastolique	systolique	diastolique
Matin	09h00		08h30		09h00	
Mesure 1	127	84	111	70	114	76
Mesure 2	128	79	112	68	115	78
Mesure 3	115	77	109	68	116	70
Soir	18h00		17h00		17h00	
Mesure 1	144	93	133	84	147	91
Mesure 2	140	90	124	83	144	90
Mesure 3	139	87	122	88	139	88

MOYENNE SYSTOLIQUE	MOYENNE DIASTOLIQUE
127	81
[113-141]	[72-90]
(d = 0,05)	

Nom : [REDACTED] Prénom : [REDACTED]

Période du relevé : du 2026-07-10 au 2026-07-12

Traitement : État méant

Montrez ce document à votre médecin lors de la prochaine consultation

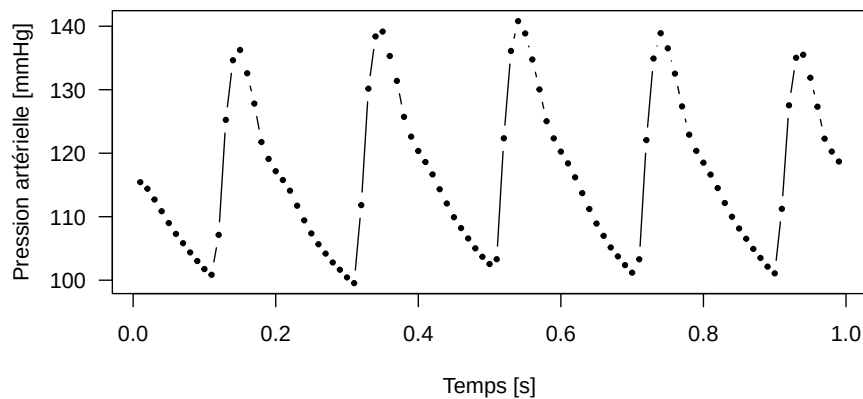
FIGURE 1 : Exemple de formulaire à remplir en FRANCE pour une automesure de la pression artérielle.

l'ordre de 10^7 mmHg. Pour les arbres des régions tempérées le potentiel hydrique foliaire est de l'ordre de -2.5 MPa en milieu de journée [4], soit environ $-0.3 \cdot 10^9$ mmHg.

La pression artérielle évolue périodiquement avec les pulsations cardiaques. Le paquet `bp` [12] nous permet d'illustrer ce phénomène avec des données [2] d'une haute résolution temporelle (100 Hz) chez le rat.

```
library(bp) ; data("bp_rats")
with(subset(bp_rats, rat_ID == "rat_01" & rat_type == "ss_ls" & time_sec < 1), {
  plot(time_sec, ABP, type = "b", cex = 0.5, pch = 19, las = 1,
       xlab = "Temps [s]", ylab = "Pression artérielle [mmHg]",
       main = "Évolution de la pression artérielle chez un rat")
})
```

Évolution de la pression artérielle chez un rat



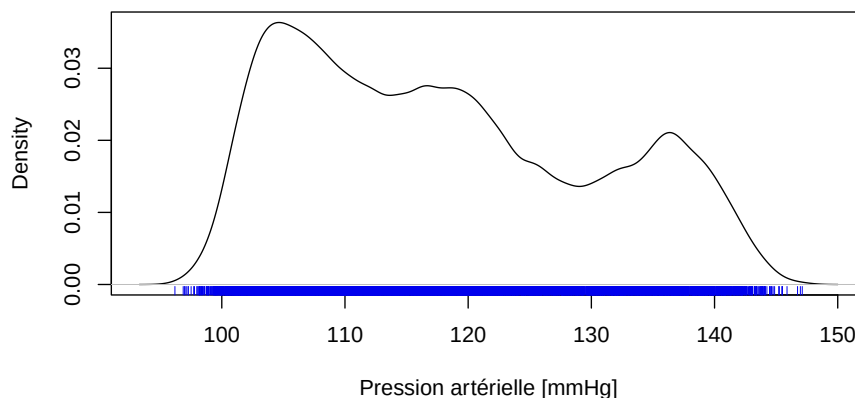
CE signal continu va être résumé par ses extrema : le maximum, la pression artérielle *systolique* (ici PAS = 140 mmHg), et le minimum, la pression artérielle *diastolique* (ici PAD = 100 mmHg). On note que au passage que les valeurs observées chez le rat sont voisines de celles de l'homme mais que le rythme cardiaque avec 300 pulsations par minutes est bien plus élevé. La courbe est asymétrique avec une vitesse de croissance plus forte (en valeur absolue) que la vitesse de décroissance. Cette asymétrie explique que pour calculer, approximativement, la pression artérielle moyenne (PAM) à partir de PAS et PAD on donne plus de poids à PAD :

$$\text{PAM} = \frac{1}{3} \text{PAS} + \frac{2}{3} \text{PAD} \quad (1)$$

EN effet, la distribution des valeurs de pression artérielle n'est pas uniforme : il y a un biais en faveur des valeurs les plus faibles, d'où le poids plus important donné à PAD.

```
with(subset(bp_rats, rat_ID == "rat_01" & rat_type == "ss_ls" & time_sec < 60), {
  plot(density(ABP, adjust = 0.5),
       main = "Distribution de la pression artérielle chez un rat pendant une minute",
       xlab = "Pression artérielle [mmHg]")
  rug(ABP, col = "blue2")
})
```

Distribution de la pression artérielle chez un rat pendant une minute



IMPORTER les données de pression artérielle de la figure 1 page 2 (reprises dans la table 1 page 4) dans l'objet `tension` avec le code suivant :

```
chmin <- "https://esb.univ-lyon1.fr/donnees/tension/"
load(url(paste0(chmin, "tension.Rda")))
```

	jour	horaire	systolique_mmHg	diastolique_mmHg	fac
1	J1	AM	127	84	J1AM
2	J1	AM	128	79	J1AM
3	J1	AM	115	77	J1AM
4	J1	PM	144	93	J1PM
5	J1	PM	140	90	J1PM
6	J1	PM	139	87	J1PM
7	J2	AM	111	70	J2AM
8	J2	AM	112	68	J2AM
9	J2	AM	109	68	J2AM
10	J2	PM	133	84	J2PM
11	J2	PM	124	83	J2PM
12	J2	PM	122	88	J2PM
13	J3	AM	114	76	J3AM
14	J3	AM	115	78	J3AM
15	J3	AM	116	70	J3AM
16	J3	PM	147	91	J3PM
17	J3	PM	144	90	J3PM
18	J3	PM	139	88	J3PM

TABLE 1 : Les données d'automesure tensionnelle.

2 Analyse de la répétabilité

Le fait de faire trois mesures consécutives constitue une expérience de *répétabilité* visant à se faire une idée de la fiabilité des mesures. On apprécie généralement la variabilité des résultats avec le coefficient de variation de

PEARSON [11], $^{\text{P}}\text{CV}$, qui est le rapport de l'écart-type de l'échantillon, s_x , sur la moyenne de l'échantillon, $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad ^{\text{P}}\text{CV}(x) = \frac{s_x}{\bar{x}}$$

C'EST donc une statistique de la dispersion *relative* des mesures : en divisant par la moyenne on s'est affranchi des effets d'échelle. On préférera ici utiliser la variante plus moderne du coefficient de variation de KVÅLSETH [6], $^{\text{K}}\text{CV}$, dont la relation avec le coefficient de variation de PEARSON est donnée par :

$$^{\text{K}}\text{CV} = \sqrt{\frac{^{\text{P}}\text{CV}^2}{1 + ^{\text{P}}\text{CV}^2}}$$

LES avantages du coefficient de variation de KVÅLSETH par rapport à celui de PEARSON sont qu'il est borné entre 0 et 1, qu'il a de meilleures propriétés de convergence [7] et qu'il a une échelle d'interprétation verbale particulièrement simple à retenir :

$^{\text{K}}\text{CV}$	Variabilité relative
0.0 - 0.2	Très faible
0.2 - 0.4	Faible
0.4 - 0.6	Modérée
0.6 - 0.8	Forte
0.8 - 1.0	Très forte

LE code  suivant permet de calculer le coefficient de variation de KVÅLSETH avec la fonction `KCV()`.

```
sdn <- function(x, i, ...){
  n <- sum(!is.na(x)) # number of non missing values
  return(sqrt((n - 1)/n)*sd(x[i], ...))
}
PCV <- function(x, i, ...){
  barx <- mean(x[i], ...)
  if(isTRUE(all.equal(barx, 0))) return(0)
  return(sdn(x[i], ...)/barx)
}
KCV <- function(x, i, ...){
  PCV2 <- PCV(x, i, ...)^2
  return(sqrt(PCV2/(1 + PCV2)))
}
```

SI on calcule le $^{\text{K}}\text{CV}$ sur nos données de répétabilité on constate que toutes les valeurs sont inférieures à 5 %, nous sommes dans le domaine des *très faibles* variations relatives dans l'échelle verbale de KVÅLSETH :

```
tension$fac <- with(tension, as.factor(paste0(jour, horaire)))
with(tension, tapply(systolique_mmHg, fac, KCV))
      J1AM      J1PM      J2AM      J2PM      J3AM      J3PM
0.047837185 0.015319102 0.011269337 0.037842795 0.007099791 0.023015983
with(tension, tapply(diastolique_mmHg, fac, KCV))
      J1AM      J1PM      J2AM      J2PM      J3AM      J3PM
0.03677411 0.02720648 0.01372893 0.02540647 0.04547985 0.01390816
```

LES résultats de l'expérience de répétabilité sont donc tout à fait satisfaisants, nous pouvons agréger les données en prenant la moyenne des mesures répétées. La table `dft` contient les données que nous allons analyser par la suite.

```

sys.mean <- with(tension, tapply(systolique_mmHg, fac, mean))
dia.mean <- with(tension, tapply(diastolique_mmHg, fac, mean))
dft <- data.frame(list(systolique_mmHg = sys.mean, diastolique_mmHg = dia.mean,
  fac = names(sys.mean)))
dft$jour <- substr(dft$fac, 1, 2)
dft$horaire <- substr(dft$fac, 3, 4)
dft

```

	systolique_mmHg	diastolique_mmHg	fac	jour	horaire
J1AM	123.3333	80.00000	J1AM	J1	AM
J1PM	141.0000	90.00000	J1PM	J1	PM
J2AM	110.6667	68.66667	J2AM	J2	AM
J2PM	126.3333	85.00000	J2PM	J2	PM
J3AM	115.0000	74.66667	J3AM	J3	AM
J3PM	143.3333	89.66667	J3PM	J3	PM

NOTONS que pour calculer les moyennes demandées (figure 1 page 2) ce regroupement ne change rien par rapport à procéder avec les données brutes puisqu'il y a toujours le même nombre de répétitions. Il n'en va pas de même pour ce qui est du calcul des intervalles de confiance : ne pas regrouper les données de répétabilité conduirait à une sous-estimation de l'amplitude d'iceux.

3 Analyse de la variance

L'ANALYSE de la variance, ANOVA pour *ANalysis Of VAriance*, porte un nom assez peu intuitif puisqu'il s'agit somme toute d'une analyse pour comparer des moyennes. C'est une généralisation du test de comparaison de moyennes de STUDENT [5] où la variable qualitative définissant les groupes à comparer n'est pas une simple variable binaire mais peut comporter un nombre plus élevé de modalités.

À quoi peut bien servir l'analyse de variance en pratique ? Notre objectif ici est de synthétiser les données par le calcul d'une moyenne. Mais est-ce bien raisonnable s'il y a une hétérogénéité dans nos données ? Par exemple, certains protocoles² d'AMT recommandent de censurer les données du premier jour [1, p. 24]. Une des finalités de l'ANOVA c'est de rechercher des facteurs de nuisance qui pourraient perturber l'interprétation des résultats.

3.1 Impact du jour de mesure

AVANT de se lancer dans un test, il faut toujours commencer par regarder les données. Comme il y a peu de points, on utilise un diagramme du type graphe de CLEVELAND [3] produit par la fonction de base `stripchart()`.

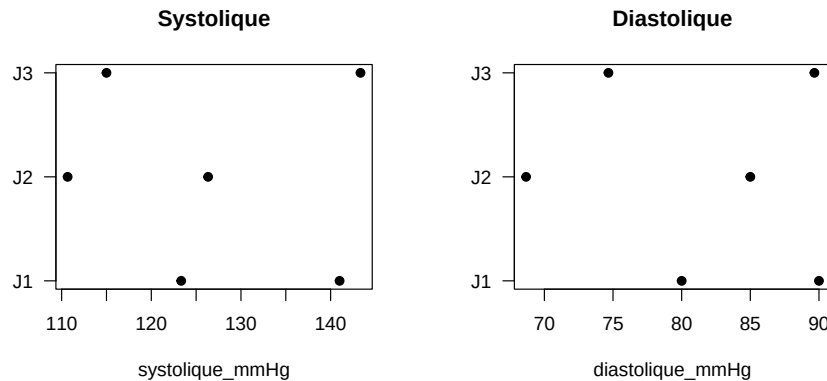
```

par(mfrow = c(1, 2), oma = c(0, 0, 1, 0), las = 1)
stripchart(systolique_mmHg~jour, dft, pch = 19, main = "Systolique")
stripchart(diastolique_mmHg~jour, dft, pch = 19, main = "Diastolique")
title(main = "Impact du jour de mesure sur la pression artérielle", outer = TRUE)

```

²En l'occurrence celui de l'*European Society of Hypertension* ([8, 9] cités par [1])

Impact du jour de mesure sur la pression artérielle



Il est clair qu'il n'y a pas d'effet marqué du jour sur les valeurs de la pression artérielle. On pourrait s'arrêter là, mais par acquit de conscience on fait quand même rouler les tests.

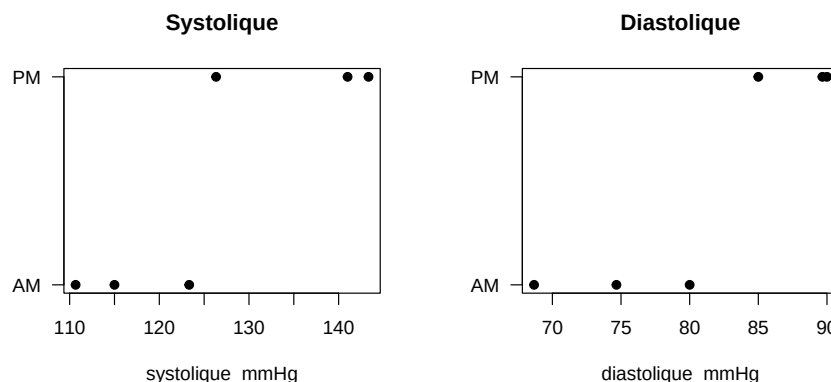
```
anova(lm(systolique_mmHg~jour, dft))
Analysis of Variance Table
Response: systolique_mmHg
      Df Sum Sq Mean Sq F value Pr(>F)
jour    2  206.37   103.19   0.4551  0.672
Residuals 3  680.17   226.72
anova(lm(diastolique_mmHg~jour, dft))
Analysis of Variance Table
Response: diastolique_mmHg
      Df Sum Sq Mean Sq F value Pr(>F)
jour    2   68.778   34.389   0.3487  0.7309
Residuals 3  295.889   98.630
```

Avec un risque de première espèce de 5 %, les données expérimentales ne permettent pas de rejeter l'hypothèse nulle de l'absence d'effet du jour de la mesure sur les valeurs de la pression artérielle. On l'accepte donc, faute de mieux, avec un risque de seconde espèce inconnu.

3.2 Impact de la tranche horaire

```
par(mfrow = c(1, 2), oma = c(0, 0, 1, 0))
stripchart(systolique_mmHg-horaire, dft, pch = 19, las = 1, main = "Systolique")
stripchart(diastolique_mmHg-horaire, dft, pch = 19, las = 1, main = "Diastolique")
title(main = "Impact de la tranche horaire sur la pression artérielle", outer = TRUE)
```

Impact de la tranche horaire sur la pression artérielle



CETTE fois il semblerait bien qu'il y ait quelque chose, avec des pressions artérielles systématiquement plus élevées le soir que le matin. Cela n'a rien d'étonnant puisqu'il est bien connu que la pression artérielle fluctue selon un rythme circadien (*e.g.* [10, fig. 4]). Voyons ce que nous dit l'analyse de la variance.

```
anova(lm(systolique_mmHg~horaire, dft))
Analysis of Variance Table
Response: systolique_mmHg
      Df Sum Sq Mean Sq F value    Pr(>F)
horaire 1  633.80   633.80  10.031 0.03395 *
Residuals 4  252.74    63.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(lm(diastolique_mmHg~horaire, dft))
Analysis of Variance Table
Response: diastolique_mmHg
      Df Sum Sq Mean Sq F value    Pr(>F)
horaire 1  284.741  284.741   14.25 0.01952 *
Residuals 4   79.926   19.981
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

QUE ce soit pour la pression artérielle systolique ou diastolique, avec un risque de première espèce de 5 %, les données expérimentales sont en contradiction avec l'hypothèse nulle de l'absence d'effet de la tranche horaire, nous la rejetons donc avec un risque contrôlé. Autrement dit nous avons mis en évidence un effet statistiquement significatif de la tranche horaire sur la valeur de la pression artérielle.

MAIS statistiquement significatif ne veut pas dire que ce soit forcément biologiquement pertinent, il faut absolument examiner la *taille d'effet*. Pour la pression systolique on a une valeur moyenne de 116.3 mmHg le matin, contre 136.9 mmHg le soir, soit une augmentation de 17.7 %. Pour la pression diastolique on a une valeur moyenne de 74.4 mmHg le matin, contre 88.2 mmHg le soir, soit une augmentation de 18.5 %. En résumé, on a une augmentation d'environ un cinquième des pressions artérielles entre le matin et le soir. Est-ce que ce $\frac{1}{5}$ est important ? Si on examine les échelles verbales utilisées par les sociétés savantes on constate qu'elle utilisent généralement une discrétisation des valeurs de pression artérielle avec un pas de 20 mmHg. Un cinquième d'augmentation sur une valeur typique de 100 mmHg signifie un changement d'une tranche, donc oui, c'est loin d'être négligeable.

L'ANALYSE de la variance nous a donc permis, soulignons-le à partir d'un jeu de données de taille extrêmement réduite, de mettre en évidence un potentiel paramètre de nuisance pour la détermination de la pression artérielle : l'heure à laquelle est effectuée la mesure. Ce n'est pas un hasard si tous les protocoles d'AMT [1, p. 24] demandent faire plusieurs mesures par jour pour pallier cet inconvénient.

4 Synthèse des données

DANS le formulaire (figure 1 page 2) il est simplement demandé de calculer la moyenne des pressions systoliques et diastolique. On peut compléter ici en donnant l'intervalle de confiance ($\alpha = 0.05$) des ces valeurs.

```
with(dft, mean(systolique_mmHg))
[1] 126.6111
with(dft, c(t.test(systolique_mmHg)$conf))
[1] 112.6372 140.5851
with(dft, mean(diastolique_mmHg))
[1] 81.33333
with(dft, c(t.test(diastolique_mmHg)$conf))
[1] 72.37105 90.29562
```

ON peut compléter en calculant la pression artérielle moyenne calculée selon l'équation 1 page 3.

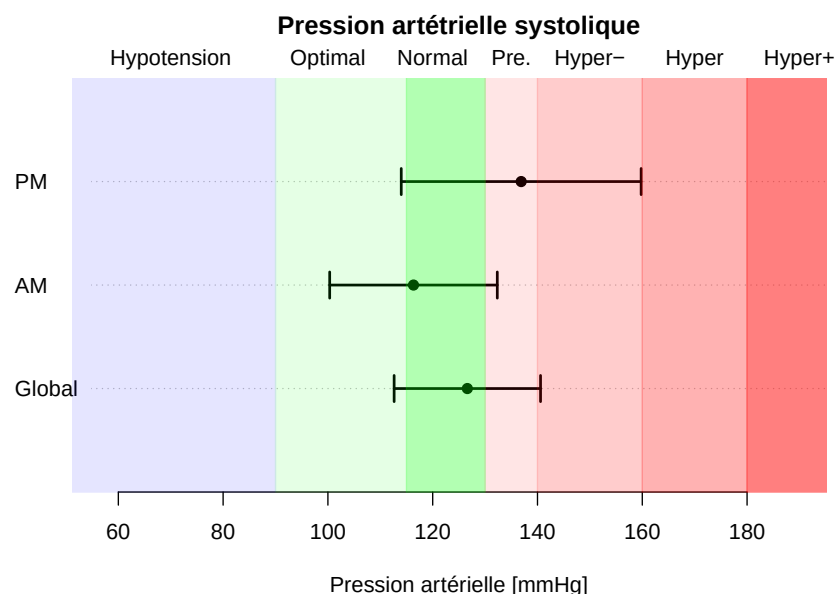
```
dft$PAM <- (dft$systolique_mmHg + 2*dft$diastolique_mmHg)/3
with(dft, mean(PAM))
[1] 96.42593
with(dft, c(t.test(PAM)$conf))
[1] 85.89396 106.95790
```

COMME l'analyse de la variance à un facteur nous a montré qu'il y avait un effet significatif de la tranche horaire, on pourrait décider de faire les calculs séparément. Mais les intervalles de confiances sont là pour nous rappeler qu'il n'y a alors plus beaucoup de points et que l'incertitude devient élevée.

```
with(subset(dft, horaire == "AM"), mean(systolique_mmHg))
[1] 116.3333
with(subset(dft, horaire == "AM"), c(t.test(systolique_mmHg)$conf))
[1] 100.3411 132.3256
with(subset(dft, horaire == "PM"), mean(systolique_mmHg))
[1] 136.8889
with(subset(dft, horaire == "PM"), c(t.test(systolique_mmHg)$conf))
[1] 113.9963 159.7815
with(subset(dft, horaire == "AM"), mean(diastolique_mmHg))
[1] 74.44444
with(subset(dft, horaire == "AM"), c(t.test(diastolique_mmHg)$conf))
[1] 60.35955 88.52934
with(subset(dft, horaire == "PM"), mean(diastolique_mmHg))
[1] 88.22222
with(subset(dft, horaire == "PM"), c(t.test(diastolique_mmHg)$conf))
[1] 81.27782 95.16663
with(subset(dft, horaire == "AM"), mean(PAM))
```

```
[1] 88.40741
with(subset(dft, horaire == "AM"), c(t.test(PAM)$conf))
[1] 73.76472 103.05010
with(subset(dft, horaire == "PM"), mean(PAM))
[1] 104.4444
with(subset(dft, horaire == "PM"), c(t.test(PAM)$conf))
[1] 92.23408 116.65481
```

CETTE synthèse numérique est un peu indigeste, il conviendrait de rechercher une représentation graphique pour visualiser les résultats. Par exemple, on pourrait proposer pour la pression systolique le graphique suivant en mettant en arrière-plan l'échelle verbale donnée sur la page française de wikipédia :



Références

- [1] J.M. Boivin. *Automesure tensionnelle - Implémentation en Médecine Générale - Intérêts et limites*. PhD thesis, Université de Lorraine, France, 2013.
- [2] S.M. Bugenhagen, A.W. Cowley Jr, and D.A. Beard. Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the dahl ss rat. *Physiological genomics*, 42(1) :23–41, 2010.
- [3] W.S. Cleveland. *The elements of graphing data*. Wadsworth, Monterey, CA, USA, 1985.
- [4] H. Cochard and A. Granier. Fonctionnement hydraulique des arbres forestiers. *Revue forestière française*, 51(2) :121–134, 1999.
- [5] Student [Gosset, W.S.]. The probable error of a mean. *Biometrika*, 6 :1–25, 1908.

- [6] T.O. Kvalseth. Coefficient of variation : the second-order alternative. *Journal of Applied Statistics*, 44 :402–415, 2017.
- [7] J.R. Lobry, M.-C. Bel-Venner, M. Bogdziewicz, A. Hacket-Pain, and S. Venner. The CV is dead, long live the CV! *Methods in Ecology and Evolution*, 14 :2780–2786, 2023.
- [8] G. Parati, G.S. Stergiou, R. Asmar, G. Bilo, P. De Leeuw, Y. Imai, K. Kario, E. Lurbe, A. Manolis, T. Mengden, et al. European society of hypertension guidelines for blood pressure monitoring at home : a summary report of the second international consensus conference on home blood pressure monitoring. *Journal of hypertension*, 26(8) :1505–1526, 2008.
- [9] G. Parati, G.S. Stergiou, R. Asmar, G. Bilo, P. De Leeuw, Y. Imai, K. Kario, E. Lurbe, A. Manolis, T. Mengden, et al. European society of hypertension practice guidelines for home blood pressure monitoring. *Journal of human hypertension*, 24(12) :779–785, 2010.
- [10] G.K. Paschos and G.A. FitzGerald. Biological clocks : a cardiovascular perspective. *Dialogues in Cardiovascular Medicine*, 15 :7–24, 2010.
- [11] K. Pearson. VII. mathematical contributions to the theory of evolution.—III. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, (187) :253–318, 1896.
- [12] J. Schwenck and I. Gaynanova. *bp : Blood Pressure Analysis in R*, 2025. R package version 2.1.1.