- Licence Sciences et technologie
- Mention science de la vie – L2

SCIENCES SORBONNE UNIVERSITÉ

INTRODUCTION A LA MODELISATION EN BIOLOGIE (BM1) - LU2SV382

Modelisation statistique (Cours 3)

Principe de test de hypothèse – comparer le(s) moyenne(s)

Martin LARSEN

www.immulab.fr

Population versus echantillon(s)
Intervalle de pari versus intervalle de confience

Mention science de la vie – L2

Population versus echantillon(s)

Population

Loi connue dans la population
- Loi Binomiale
- Loi Normale
- …

Échantillon

Prédiction de ce qui sera observé dans l'échantillon
⇨ Intervalle de pari

Population versus echantillon(s)

Population

Généralisation / Inférence

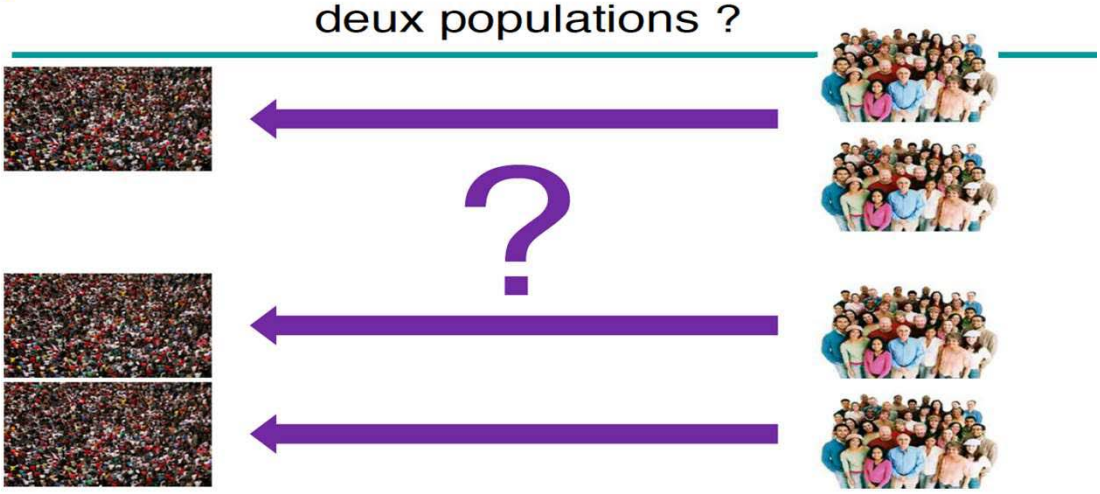
Échantillon

Observation

Estimation

Intervalle de confience and intervalle de variation (pari)

Deux échantillons viennent-ils d'une ou de deux populations ?



Hypothesis test avec l'Intervalle de confience

Population

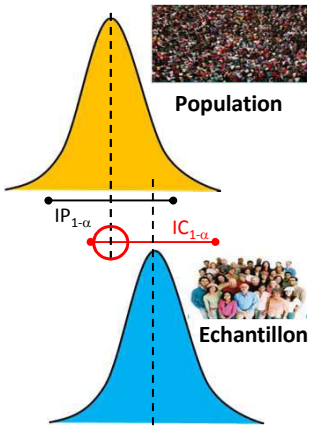
Démarche

Echantillon

Construire un intervalle de confiance a partir des données observées et voir si cela correspond a ce que l'on attend en population.

Exemple: Le traitement A est il plus efficace que le traitement B ?

- Estimer efficacité avec les deux traitements (deux échantillon)
- Calculer la différence ($\bar{X}_A - \bar{X}_B$) et l'intervalle de confiance de cette différence.
- Hypothèse : H_0 : $\mu_A = \mu_B$ alors sous H0 la différence en population ($\mu_A - \mu_B = 0$)
- Voir si l'intervalle de confiance avec un certain risque alpha est compatible avec cette hypothèse ($0 \in IC_{(1-\alpha)}$).



Hypothesis test avec l'Intervalle de confience

Population

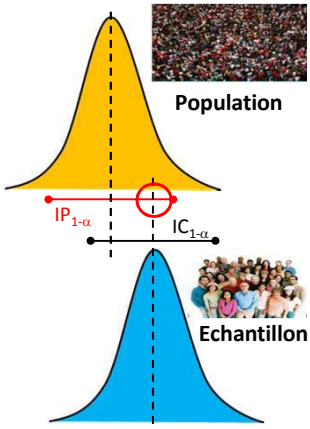
Démarche

Echantillon

Construire un intervalle de pari a partir des hypothèses et voir si les données de l'échantillon correspondent a ce qui est attendu.

Exemple: Le traitement A est il plus efficace que le traitement B ?

- Hypothèse : H_0 : $\mu_A = \mu_B$
  H1: $\mu_A \neq \mu_B$
- Trouver une statistique de test correspondant a cette hypothèse.
- Etablir un intervalle de pari de cette statistique (sous H_0)
- Estimer l'éfficacite avec les deux traitements (deux échantillon)
- Voir si l'intervalle de pari est compatible avec l'estimation (Variable de décision ($\bar{X}_A - \bar{X}_B$) $\in IP_{(1-\alpha)}$).



Moyenne et pourcentage

Intervalle de confiance

Intervalle de Pari

Intervalle de Pari

- $IP_{1-\alpha} = \left[\mu - u_\alpha \frac{\sigma}{\sqrt{n}}; \mu + u_\alpha \frac{\sigma}{\sqrt{n}}\right]$
- $IP_{1-\alpha} = \left[\pi - u_\alpha \sqrt{\frac{\pi(1-\pi)}{n}}; \pi + u_\alpha \sqrt{\frac{\pi(1-\pi)}{n}}\right]$
- A partir de la connaissance de la population on imagine ce que l'on obtiendra dans un échantillon

Intervalle de confiance

- $IC_{1-\alpha} = \left[m - u_\alpha \frac{s}{\sqrt{n}}; m + u_\alpha \frac{s}{\sqrt{n}}\right]$
- $IC_{1-\alpha} = \left[p - u_\alpha \sqrt{\frac{p(1-p)}{n}}; p + u_\alpha \sqrt{\frac{p(1-p)}{n}}\right]$
- A partir de l'observation d'un échantillon on en déduit une information sur la population

Principaux tests univariés sous R

Mention science de la vie – L2

Types of variables

Qualitatif / catégorique:

Ne prend généralement qu'un petit nombre de valeurs (souvent Nominal <Ordinal)

Nominal: Aucun ordre donné (par exemple, «France», «Allemagne», «Italie»)

Ordinal / Rang: valeurs ordonnées mais pas de différence absolue significative entre les valeurs (menu petit, moyen et grand)

Quantitatif / numérique:

Peut prendre un nombre infini de valeurs, qui sont ordonnées et a une différence absolue significative entre les valeurs.

Discret: généralement uniquement des entiers (par exemple, le nombre de quelque chose)

Continu: généralement une mesure (par exemple la distance et le poids)

Choix de test dépend de la caractéristique de l'étude

Level of Measurement	Sample characteristics					Correlation
	1 Sample	2 Sample		K Samples (K>2)		
		Independent	Dependent	Independent	Dependent	
Categorical or Nominal	χ^2 conformity or independence	χ^2 homogeneity	McNemar χ^2	χ^2 homogeneity	Cochran's Q (CMH)	
Rank or ordinal	χ^2 or Mann Whitney U*	Mann Whitney U	Wilcoxon Matched Pairs Signed Ranks	Kruskal Wallis H	Friedman's ANOVA	Spearman's rho
Parametric (interval & Ratio)	Z test or t test	T test between groups	T test within groups	1 way ANOVA between groups	1 way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

* H_0 : μ = valeure theorique, Correspond à un Wilcoxon signed ranks, test, lorsque le data est centré sur la valeur théorique

Principe d'un test statistique – e.g. comparer une moyennes avec valeur theorique

Condition:

Normalité

Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Hypothèse:

H_0: $\mu = \varphi_0$

H_1: $\mu \neq \varphi_0$

Variable de décision

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

$$E(Z) = E\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = \frac{E(\overline{X} - \mu)}{\sigma/\sqrt{n}} = \frac{E(\overline{X}) - \mu}{\sigma/\sqrt{n}} = 0$$

$$\mathrm{Var}(Z) = Var\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = \frac{Var(\overline{X} - \mu)}{\sigma^2/n} = \frac{Var(\overline{X}) + Var(\mu)}{\sigma^2/n} = \frac{Var(\overline{X}) + 0}{Var(\overline{X})} = 1$$

Principe d'un test statistique – e.g. comparer une moyennes avec valeur theorique

Condition:

Normalité

Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Hypothèse:

H_0: $\mu = \varphi_0$

H_1: $\mu \neq \varphi_0$

Variable de décision sous H_0

$$Z_{obs} = \frac{\overline{X} - \varphi_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

Principe d'un test statistique – e.g. comparer une moyennes avec valeur theorique

Condition:
Normalité

Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

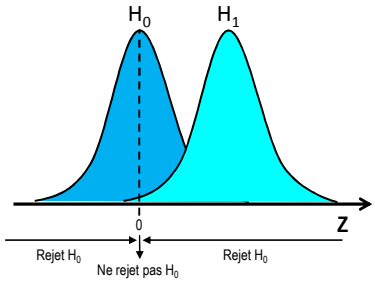
Hypothèse:
H_0: $\mu = \varphi_0$
H_1: $\mu \neq \varphi_0$

Variable de décision sous H_0

$$Z_{obs} = \frac{\overline{X} - \varphi_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)
β=P(ne rejet pas H_0 | H_1 vrai)

Basé sur un seuil fixe (inapproprié)



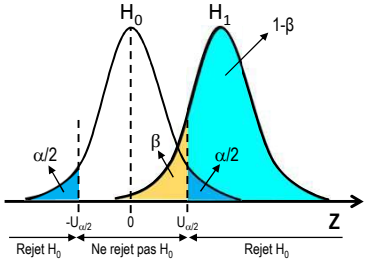
Règle de décision (α et β):
Ne rejet pas H_0: Z_{obs}= 0 α=100%
Rejet H_0: $Z_{obs} \neq 0$ β=0%
Z VAC => P(Z_{obs}=0) = 0%

Conclusion:
Statistique: Rejet ou non-rejet (objectif)
Biologique: Interprétation (subjectif)

Principe d'un test statistique – e.g. comparer une moyennes avec valeur theorique

Condition:
Normalité

Basé sur un intervalle de pari (approprié)



Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

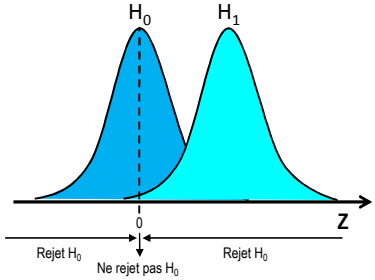
Hypothèse:
H_0: $\mu = \varphi_0$
H_1: $\mu \neq \varphi_0$

Variable de décision sous H_0

$$Z_{obs} = \frac{\overline{X} - \varphi_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)
β=P(ne rejet pas H_0 | H_1 vrai)

Basé sur un seuil fixe (inapproprié)



Règle de décision (α et β):
Ne rejet pas H_0: Z_{obs}= 0 α=100%
Rejet H_0: $Z_{obs} \neq 0$ β=0%
Z VAC => P(Z_{obs}=0) = 0%

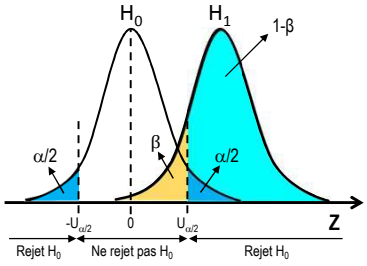
Règle de décision (α et β):
Ne rejet pas H_0: $Z_{obs} \in IP_{1-\alpha}$=[-1.96,1.96] α=5%
Rejet H_0: $Z_{obs} \notin IP_{1-\alpha}$=[-1.96,1.96] β= ?
$IP_{1-\alpha}$ créé sous H_0

Conclusion:
Statistique: Rejet ou non-rejet (objectif)
Biologique: Interprétation (subjectif)

Principe d'un test statistique – e.g. comparer deux moyennes

Condition:
Independence, normalité, homoscedasticité

Basé sur un intervalle de pari (approprié)



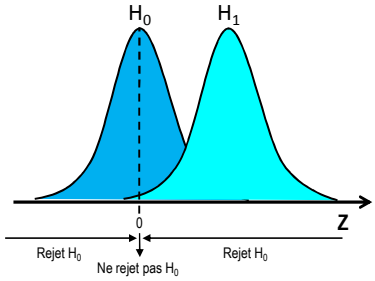
Hypothèse:
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Variable de décision

$$Z = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)
β=P(ne rejet pas H_0 | H_1 vrai)

Basé sur un seuil fixe (inapproprié)



Règle de décision (α et β):
Ne rejet pas H_0: $Z_{obs} \in IP_{1-\alpha}$=[-1.96,1.96] α=5%
Rejet H_0: $Z_{obs} \notin IP_{1-\alpha}$=[-1.96,1.96] β= ?
$IP_{1-\alpha}$ créé sous H_0

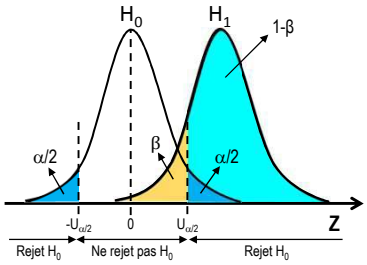
Règle de décision (α et β):
Ne rejet pas H_0: Z_{obs}= 0 α=100%
Rejet H_0: $Z_{obs} \neq 0$ β=0%
Z VAC => P(Z_{obs}=0) = 0%

Conclusion:
Statistique: Rejet ou non-rejet (objectif)
Biologique: Interprétation (subjectif)

Principe d'un test statistique – e.g. comparer deux moyennes

Condition:
Independence, normalité, homoscedasticité

Basé sur un intervalle de pari (approprié)



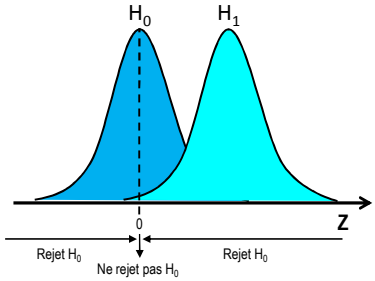
Hypothèse:
H_0: $\mu_1 = \mu_2$ => μ_1-μ_2=0
H_1: $\mu_1 \neq \mu_2$

Variable de décision sous H_0

$$Z_{obs} = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)
β=P(ne rejet pas H_0 | H_1 vrai)

Basé sur un seuil fixe (inapproprié)



Règle de décision (α et β):
Ne rejet pas H_0: $Z_{obs} \in IP_{1-\alpha}$=[-1.96,1.96] α=5%
Rejet H_0: $Z_{obs} \notin IP_{1-\alpha}$=[-1.96,1.96] β= ?
$IP_{1-\alpha}$ créé sous H_0

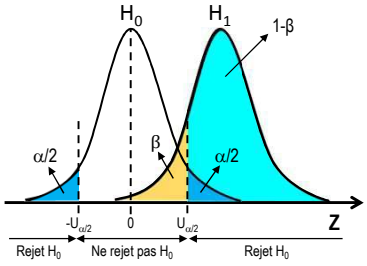
Règle de décision (α et β):
Ne rejet pas H_0: Z_{obs}= 0 α=100%
Rejet H_0: $Z_{obs} \neq 0$ β=0%
Z VAC => P(Z_{obs}=0) = 0%

Conclusion:
Statistique: Rejet ou non-rejet (objectif)
Biologique: Interprétation (subjectif)

Principe d'un test statistique – e.g. comparer deux moyennes

Condition:
Independence, normalité, homoscedasticité

Basé sur un intervalle de pari (approprié)



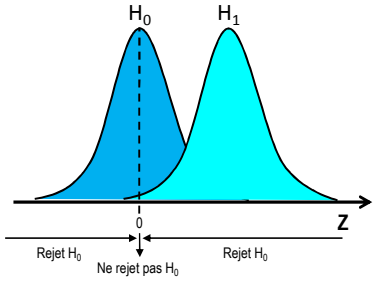
Hypothèse:
H_0: $\mu_1 = \mu_2$ => $\mu_1-\mu_2=0$
H_1: $\mu_1 \neq \mu_2$

Variable de décision sous H_0

$$Z_{obs} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\left(\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)
β=P(ne rejet pas H_0 | H_1 vrai)

Basé sur un seuil fixe (inapproprié)



Règle de décision (α et β):
Ne rejet pas H_0: $Z_{obs} \in IP_{1-\alpha}$=[-1.96,1.96] α=5%
Rejet H_0: $Z_{obs} \notin IP_{1-\alpha}$=[-1.96,1.96] β= ?
$IP_{1-\alpha}$ créé sous H_0

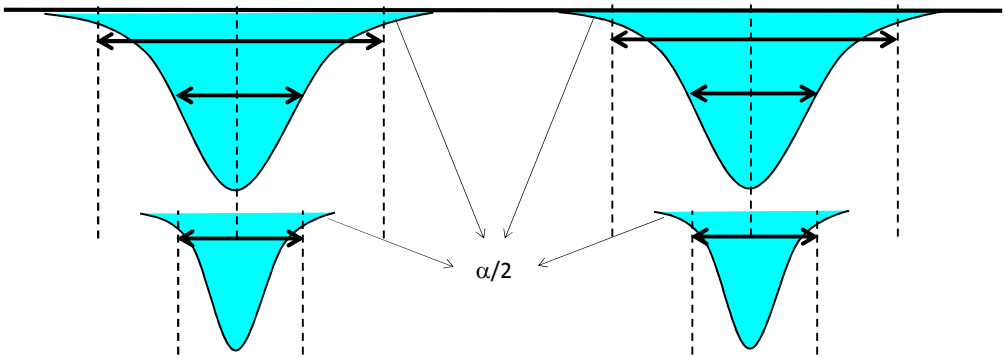
Règle de décision (α et β):
Ne rejet pas H_0: Z_{obs}= 0 α=100%
Rejet H_0: $Z_{obs} \neq 0$ β=0%
Z VAC => P(Z_{obs}=0) = 0%

Conclusion:
Statistique: Rejet ou non-rejet (objectif)
Biologique: Interprétation (subjectif)

Intervalle de pari vs. Intervalle de confiance

Deux façons de change les intervalles:

Paramètre	Intervalle de pari $IP_{1-\alpha}$: $\bar{x} \in [\mu - u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$	Intervalle de confiance $IC_{1-\alpha}$: $\mu \in [\bar{x} - u_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + u_{\alpha/2}\frac{s}{\sqrt{n}}]$
n, α		
α ↑		
n ↑		



Saisie des données pour R

- Deux échantillons où X :

VA représentant la taille d'individus d'une espèce de poissons

- 1. Création de la table sous Excel
- (2. On s 'assure des points au lieu des virgules)
- 3. On enregistre « sous » au format « .txt » avec tabulation

adu1	adu2
46.01668	45.31913
49.47434	47.04900
50.60362	35.20589
42.30919	48.79993
51.14163	40.33424
49.3686	42.51932
42.94062	49.57548
49.82174	52.00621
48.73526	40.59082
64.07578	43.75296
57.55783	51.04380
59.44963	35.72993
52.63457	44.08565
54.87269	42.33216
42.50207	43.10860
39.98742	46.59471
47.96577	43.49734
45.51302	46.27165
57.83556	49.00498
46.34471	41.97896

Saisie des données pour R

- Pour récupérer le fichier créé à partir d'excel …dans un script sous R

nouveau script : test de comparaisons de moyennes

>donnees<-read.table(file.choose(),header=TRUE)

>attach(donnees)

adu1	adu2
46.01668	45.31913
49.47434	47.04900
50.60362	35.20589
42.30919	48.79993
51.14163	40.33424
49.3686	42.51932
42.94062	49.57548
49.82174	52.00621
48.73526	40.59082
64.07578	43.75296
57.55783	51.04380
59.44963	35.72993
52.63457	44.08565
54.87269	42.33216
42.50207	43.10860
39.98742	46.59471
47.96577	43.49734
45.51302	46.27165
57.83556	49.00498
46.34471	41.97896

Comparer deux moyennes (n>30)

Hypothese:

H0 : $\mu_{adu1} = \mu_{adu2}$

H1 : $\mu_{adu1} \neq \mu_{adu2}$

α = 5%

Conditions:

- Comparer deux échantillons indépendant de grande taille (n>30)

Application de TCL pour s'assurer que

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Remplacer σ avec $\hat{\sigma}$ (bonne estimateur si n>30)

Variables:
X VAC « longeur depoisson »

Condition:
Independence

Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Hypothèse:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Variable de décision sous H_0

$$Z_c = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\left(\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)}} \sim N(0,1)$$

α=P(rejet H_0 | H_0 vrai)

β=P(ne rejet pas H_0 | H_1 vrai)

Comparer deux moyennes (n<30)

Hypothese:

H0 : $\mu_{adu1} = \mu_{adu2}$

H1 : $\mu_{adu1} \neq \mu_{adu2}$

α = 5%

Conditions:

- Comparer deux échantillons indépendant de petite taille (n<30)
- Normalité
- Homoscedasticity

Distribution d'echantillonage de la moyenne (DEM) pour s'assurer que $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

Remplacer σ avec $\hat{\sigma}$ (mauvaise estimateur car n<30) demande une changement de loi (Student)

Variables:
X VAC « longeur depoisson »

Condition:
Independence, normalité, homoscedasticité

Ditribution d'echantillonage de la moyenne (toute n)

$$X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Theoreme centrale limite (n>30)

$$X \sim ? \Rightarrow \bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Hypothèse:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Variable de décision sous H_0

$$T_c = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim T(0,1),$$

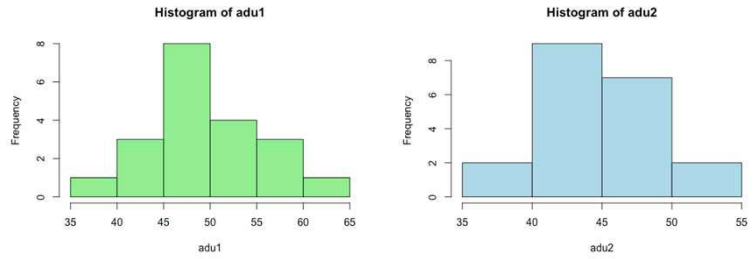
$$ddl = n_1 + n_2 - 2$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1){\hat{\sigma}_1}^2 + (n_2 - 1){\hat{\sigma}_2}^2}{n_1 + n_2 - 2}$$

Histogramme - Normalité

- Comme nous avons 2 échantillons on va réaliser des graphiques avec les deux échantillons pour évaluer la normalité de X dans les deux échantillons
- Script (histogramme des distributions)

```
hist(adu1,col='lightgreen')
hist(adu2,col='lightblue')
```

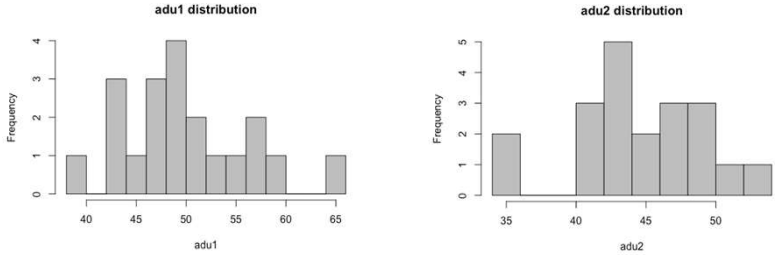


Histogramme - Normalité

- Comme nous avons 2 échantillons on va réaliser des graphiques avec les deux échantillons pour évaluer la normalité de X dans les deux échantillons
- Script (histogramme des distributions)

```
q <- hist(adu1,col='grey', nclass=10, main="adu1 distribution")
q    # qu'est-ce qui est sauvegardé dans q?

q <- hist(adu2,col='grey', nclass=10, main="adu2 distribution")
```



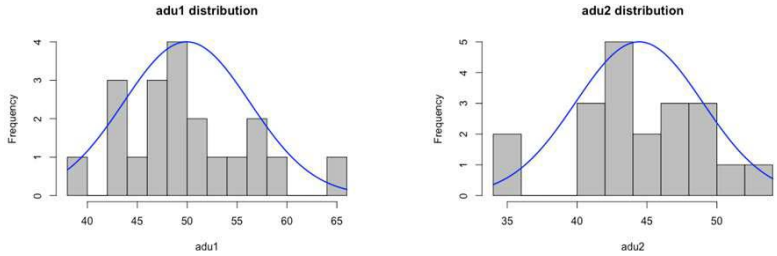
Histogramme - Normalité

- Comme nous avons 2 échantillons on va réaliser des graphiques avec les deux échantillons pour évaluer la normalité de X dans les deux échantillons
- Script (histogramme des distributions)

Line width

```
q <- hist(adu1,col='grey', nclass=10, main="adu1 distribution")
curve(dnorm(x, mean=mean(adu1), sd=sd(adu1))*max(q$counts)/dnorm(0,0,sd(adu1)), col='blue', lwd=2, add=TRUE)

q <- hist(adu2,col='grey', nclass=10, main="adu2 distribution")
curve(dnorm(x, mean=mean(adu1), sd=sd(adu1))*max(q$counts)/dnorm(0,0,sd(adu1)), col='blue', lwd=2, add=TRUE)
```



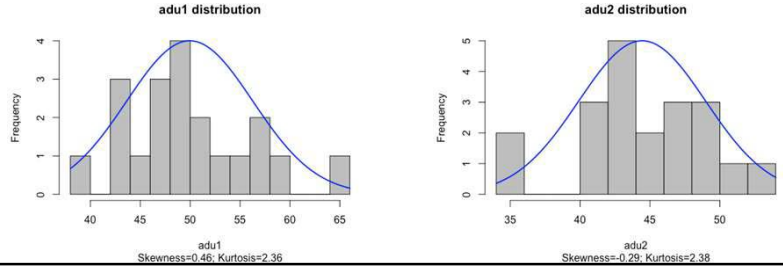
Histogramme - Normalité

- Comme nous avons 2 échantillons on va réaliser des graphiques avec les deux échantillons pour évaluer la normalité de X dans les deux échantillons
- Script (histogramme des distributions)

$$skewness = \frac{\sum(x-\overline{x})^3/N}{\hat{\sigma}^3}$$ = 0 pour X ~ N(μ,σ)

$$kurtosis = \frac{\sum(x-\overline{x})^4/N}{\hat{\sigma}^4}$$ = 3 pour X ~ N(μ,σ)

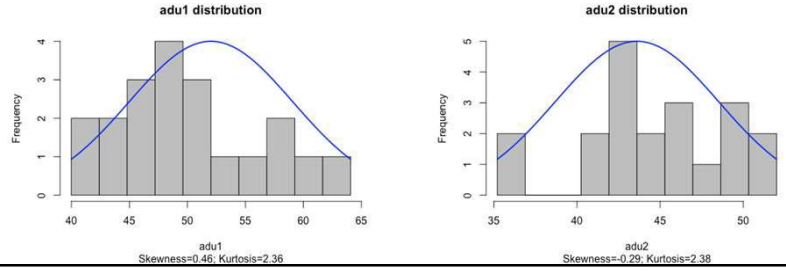
```
skewness <- function(x){sum((x-mean(x))^3)/length(x)/sd(x)^3}
kurtosis <- function(x){sum((x-mean(x))^4)/length(x)/sd(x)^4}
subtitle <- function(x){paste("Skewness=",round(skewness(x),digits=2),"; Kurtosis=", round(kurtosis(x), digits=2), sep="")}
q <- hist(adu1,col='grey', nclass=10, main="adu1 distribution", sub=subtitle(adu1))
curve(dnorm(x, mean=mean(adu1), sd=sd(adu1))*max(q$counts)/dnorm(0,0,sd(adu1)), col='blue', lwd=2, add=TRUE)
q <- hist(adu2,col='grey', nclass=10, main="adu2 distribution", sub=subtitle(adu2))
curve(dnorm(x, mean=mean(adu1), sd=sd(adu1))*max(q$counts)/dnorm(0,0,sd(adu1)), col='blue', lwd=2, add=TRUE)
```



Histogramme - Normalité

- Comme nous avons 2 échantillons on va réaliser des graphiques avec les deux échantillons pour évaluer la normalité de X dans les deux échantillons
- Script (histogramme des distributions)

```
histbin <- function(x,nbins,coltmp='grey',maintmp=paste(deparse(substitute(x))," distribution", sep="")){
  q <- hist(x,col=coltmp, breaks=seq(min(x),max(x),l=nbins+1), main=maintmp, sub=subtitle(x),xlab=deparse(substitute(x)))
  curve(dnorm(x, mean=mean(x), sd=sd(x))*max(q$counts)/dnorm(0,0,sd(x)), col='blue', lwd=2, add=TRUE) # to normalize curve ot
        hist: *max(q$counts)/dnorm(0,0,sd(adu2)
  }
histbin(adu1, 10)
histbin(adu2, 10)
```



Normalité

Il existe plusieurs méthodes pour évaluer si les données sont normalement distribuées ou non. Ils se divisent en deux grandes catégories: graphique et statistique. Les techniques les plus courantes sont:

Graphique

- Histograms
- Q-Q probability plots
- Cumulative frequency (P-P) plots

Statistique

- W/S test
- Jarque-Bera test
- Shapiro-Wilks test (shapiro.test)
- Kolmogorov-Smirnov test (ks.test)
- D'Agostino test
- Anderson-Darling (ad.test – package "nortest")
- Liliefors (lillie.test – package "nortest")

Pour installer un package:
install.packages('nortest') #Note! Une fois
Pour activer le package:
library(nortest) #Note! Chaque session R

Quantile-Quantile (QQ)-plot et normalité

- Pour effectuer des statistiques paramétriques, nous devons tester la normalité des deux séries – test graphique et/ou statistique.

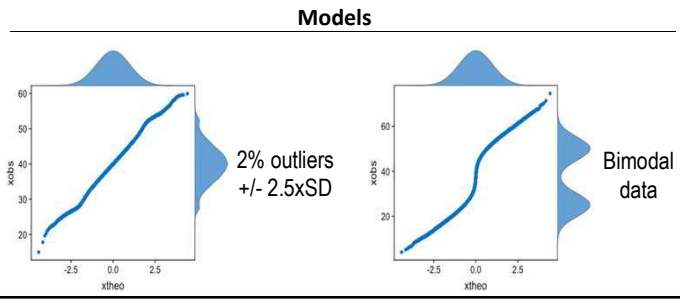
Traçons la distribution des quartiles sous forme linéaire….

qqnorm(adu1,col='red')

qqline(adu1,col='black')

qqnorm(adu2,col='blue')

qqline(adu2,col='black')



Quantile-Quantile (QQ)-plot et normalité

- Pour effectuer des statistiques paramétriques, nous devons tester la normalité des deux séries – test graphique et/ou statistique.

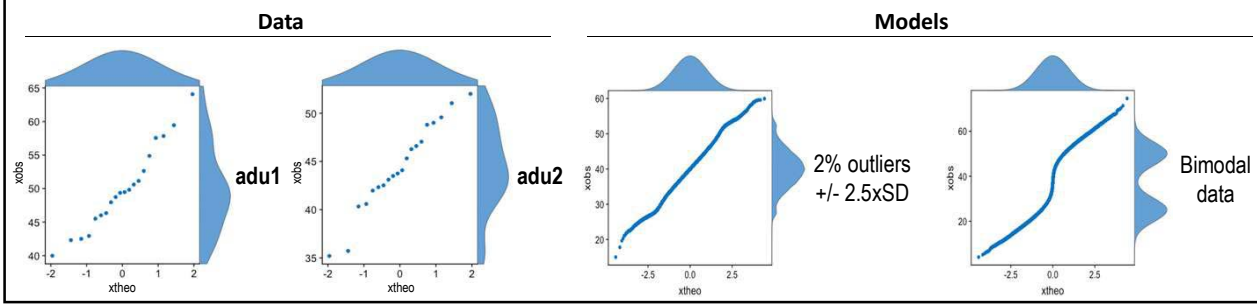
Traçons la distribution des quartiles sous forme linéaire….

qqnorm(adu1,col='red')

qqline(adu1,col='black')

qqnorm(adu2,col='blue')

qqline(adu2,col='black')



Normalité

- Les méthodes graphiques ne sont généralement pas très puissante lorsque la taille de l'échantillon est petite.
- Les histogrammes de adu1 et adu2 du dernier exemple ne semblent pas «normaux», mais ils ne sont pas statistiquement différents de la normale. QQ-plot est plus puissante.
- Test statistique
- Hypothesis
  - H_0 : X suit une loi normale
  - H_1 : X ne suit pas une loi normale

  α = 0.05

Normalité Jarque-Bera test

Jarque-Bera test (skewness and kurtosis)

$$skewness = \frac{\sum(x-\overline{x})^3/N}{\hat{\sigma}^3}$$ = 0 pour X ~ N(μ,σ)

$$kurtosis = \frac{\sum(x-\overline{x})^4/N}{\hat{\sigma}^4}$$ = 3 pour X ~ N(μ,σ)

- Hypothese:
  - H0 : X suit une loi normale
  - H1 : X ne suit pas une loi normale

- Basé sur la statistique JB, qui sous H_0 suit une loi JB. Le test évalue **Skewness** et **Kurtosis** (valeurs extrêmes).

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

- où JB est la variable de décision (VD), S est le skewness et K est le kurtosis.

- Le test de Jarque-Bera ne teste pas à proprement parler si les données suivent une loi normale, mais plutôt si le kurtosis et le coefficient d'asymétrie des données sont les mêmes que ceux d'une loi normale de même espérance et variance.
  - H0 : S = 0 et K = 3
  - H1 : S ≠ 0 ou K ≠ 3

Normalité W/S test

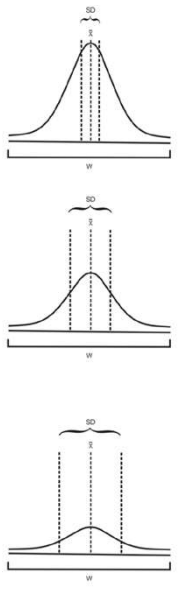
W/S test (Width/standard deviation)

- Un test assez simple qui ne nécessite que l'écart type de l'échantillon et la intervalle absolue de données.
- Ne doit pas être confondu avec le test de Shapiro-Wilk.
- Hypothese:
  - H0 : X suit une loi normale
  - H1 : X ne suit pas une loi normale
- Basé sur la statistique q, qui sous H_0 suit une loi d'etendue de Student ('studentized') (similaire au distribution t). Le test évalue le **Kurtosis** (valeurs extrêmes).

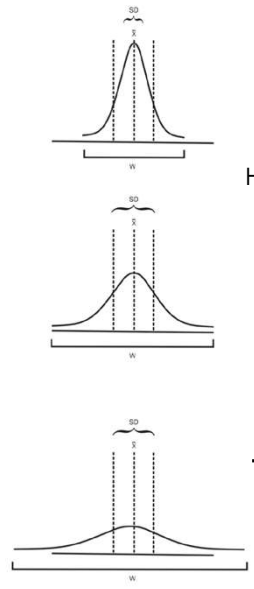
$$q = \frac{w}{s}$$

- où q est la variable de décision (VD), w est la intervalle des données et s est l'écart type.

Normalité W/S test



w = fixe
s = variable



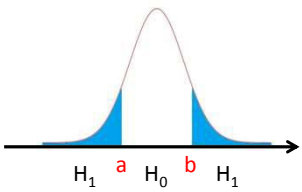
w = variable
s = fixe

Hypothesis:

H_0: X suit une loi normale

H_1: X ne suit pas une loi normale

$$q = \frac{w}{s}$$



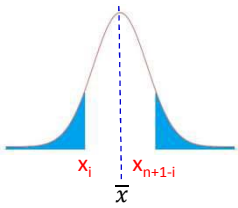
Columns *a* denote the lower boundaries or the left-sided critical values.
Columns *b* denote the upper boundaries or the right-sided critical values.

n	Level of significance α											
	0.000		0.005		0.01		0.025		0.05		0.10	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
3	1.732	2.000	1.735	2.000	1.737	2.000	1.745	2.000	1.758	1.999	1.782	1.997
4	1.732	2.449	1.82	2.447	1.87	2.445	1.93	2.439	1.98	2.429	2.04	2.409
5	1.826	2.828	1.98	2.813	2.02	2.803	2.09	2.782	2.15	2.753	2.22	2.712
6	1.826	3.162	2.11	3.115	2.15	3.095	2.22	3.056	2.28	3.012	2.37	2.949
7	1.871	3.464	2.22	3.369	2.26	3.338	2.33	3.282	2.40	3.222	2.49	3.143
8	1.871	3.742	2.31	3.585	2.35	3.543	2.43	3.471	2.50	3.399	2.59	3.308
9	1.897	4.000	2.39	3.772	2.44	3.720	2.51	3.634	2.59	3.552	2.68	3.449
10	1.897	4.243	2.46	3.935	2.51	3.875	2.59	3.777	2.67	3.685	2.76	3.57
11	1.915	4.472	2.53	4.079	2.58	4.012	2.66	3.903	2.74	3.80	2.84	3.68
12	1.915	4.690	2.59	4.208	2.64	4.134	2.72	4.02	2.80	3.91	2.90	3.78
13	1.927	4.899	2.64	4.325	2.70	4.244	2.78	4.12	2.86	4.00	2.96	3.87
14	1.927	5.099	2.70	4.431	2.75	4.34	2.83	4.21	2.92	4.09	3.02	3.95
15	1.936	5.292	2.74	4.53	2.80	4.44	2.88	4.29	2.97	4.17	3.07	4.02
16	1.936	5.477	2.79	4.62	2.84	4.52	2.93	4.37	3.01	4.24	3.12	4.09
17	1.944	5.657	2.83	4.70	2.88	4.60	2.97	4.44	3.06	4.31	3.17	4.15
18	1.944	5.831	2.87	4.78	2.92	4.67	3.01	4.51	3.10	4.37	3.21	4.21
19	1.949	6.000	2.90	4.85	2.96	4.74	3.05	4.56	3.14	4.43	3.25	4.27
20	1.949	6.164	2.94	4.91	2.99	4.80	3.09	4.63	3.18	4.49	3.29	4.32
25	1.961	6.93	3.09	5.19	3.15	5.06	3.24	4.87	3.34	4.71	3.45	4.53
30	1.966	7.62	3.21	5.40	3.27	5.26	3.37	5.06	3.47	4.89	3.59	4.70
35	1.972	8.25	3.32	5.57	3.38	5.42	3.48	5.21	3.58	5.04	3.70	4.84
40	1.975	8.83	3.41	5.71	3.47	5.56	3.57	5.34	3.67	5.16	3.79	4.96
45	1.978	9.38	3.49	5.83	3.55	5.67	3.66	5.45	3.75	5.26	3.88	5.06
50	1.980	9.90	3.56	5.93	3.62	5.77	3.73	5.54	3.83	5.35	3.95	5.14
55	1.982	10.39	3.62	6.02	3.69	5.86	3.80	5.63	3.90	5.43	4.02	5.22
60	1.983	10.86	3.68	6.10	3.75	5.94	3.86	5.70	3.96	5.51	4.08	5.29
65	1.985	11.31	3.74	6.17	3.80	6.01	3.91	5.77	4.01	5.57	4.14	5.35
70	1.986	11.75	3.79	6.24	3.85	6.07	3.96	5.83	4.06	5.63	4.19	5.41
75	1.987	12.17	3.83	6.30	3.90	6.13	4.01	5.88	4.11	5.68	4.24	5.46
80	1.987	12.57	3.88	6.35	3.94	6.18	4.05	5.93	4.16	5.73	4.28	5.51
85	1.988	12.96	3.92	6.40	3.99	6.23	4.09	5.98	4.20	5.78	4.33	5.56
90	1.989	13.34	3.96	6.45	4.02	6.27	4.13	6.03	4.24	5.82	4.36	5.60
95	1.990	13.71	3.99	6.49	4.06	6.32	4.17	6.07	4.27	5.86	4.40	5.64
100	1.990	14.07	4.03	6.53	4.10	6.36	4.21	6.11	4.31	5.90	4.44	5.68
150	1.993	17.26	4.32	6.82	4.38	6.64	4.48	6.39	4.59	6.18	4.72	5.96
200	1.995	19.95	4.53	7.01	4.59	6.84	4.68	6.60	4.78	6.39	4.90	6.15
500	1.998	31.59	5.06	7.60	5.13	7.42	5.25	7.15	5.47	6.94	5.49	6.72
1000	1.999	44.70	5.50	7.99	5.57	7.80	5.68	7.54	5.79	7.33	5.92	7.11

Source: Sachs, 1972

Normalité – Shapiro-Wilk test

- adu1
  - [1] 46 49 51 42 51 49 43 50 49 64 58 59 53 55 43 40 48 46 58 46
- adu1[order(adu1)]
  - i = [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
  - Adu1 = [1] 40 42 43 43 46 46 46 48 49 49 49 50 51 51 53 55 58 58 59 64
- $SST = \sum_{i=1}^{n}(x_i - \overline{x})^2$
- Si n est pair: m = n/2, alors que si n est impair: m = (n–1)/2
- $b = \sum_{i=1}^{m} a_i(x_{n+1-i} - x_i)$
  - Table for a_i: http://www.real-statistics.com/statistics-tables/shapiro-wilk-table/
  - Note: $a_i > a_{i+1}$ (les extrêmes sont plus importantes - kurtosis)
- VD: $W = b^2/SST$; sous H_0 W suit une loi de Shapiro-Wilk
  - Table for W: http://www.real-statistics.com/statistics-tables/shapiro-wilk-table/



Le test évalue le **Kurtosis** (valeurs extrêmes).

Fonctionne intégrée:

Shapiro.test (adu1)

Shapiro.test (adu2)

Hypothesis:

H_0: X suit une loi normale

H_1: X ne suit pas une loi normale

- Valeur a 5% pour n=20 : W_α=0,905
- $W_{obs} > W_\alpha$ => pas de rejet de la normalité
- Note: non-normal kurtosis => $W_{obs}\downarrow < W_\alpha$

n =	15	16	17	18	19	20	21	22	23	24	25	26
a 1	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407
a 2	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043
a 3	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533
a 4	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151
a 5	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836
a 6	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563
a 7	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316
a 8		0.0196	0.0359	0.0496	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089
a 9				0.0163	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876
a 10						0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672
a 11								0.0122	0.0228	0.0321	0.0403	0.0476
a 12									0.0000	0.0107	0.0200	0.0284
a 13											0.0000	0.0094

Homoscédasticité

- Pour l'égalité des variances, on va appliquer le test de Fisher-Snedecor
- Hypothesis:
  - H_0: $\sigma_1^2 = \sigma_2^2$
  - H_1: $\sigma_1^2 \neq \sigma_2^2$
- Variable de decision: $F = \frac{\sigma_1^2}{\sigma_2^2}$ suit loi de F(v_1, v_2) sous H_0

var.test (adu1,adu2)

p.value ?????

Test de hypothese - comparer le(s) moyenne(s)

Mention science de la vie – L2

Comparer une moyenne avec une valeur théorique – Student T Test statistique

- Comparaison moyenne à une norme ou une référence :

>Vec<-c(2.4,2.5,3.1,1.9,3.5) # $\overline{X}$ = 2.68
Création du vecteur de données

>shapiro.test(vec) # p-val = 0.87
>t.test(vec,mu=2)

Les sorties possibles et extractions avec $
tes<- t.test(vec,mu)
tes$p.value # p = 0.07209211
tes$parameter # df = 4 (n-1)
tes$conf.int # [1.902595; 3.457405]
tes$statistic # t = 2.428571

Condition:
Normalité

Hypothèse:
H_0: $\mu_1 = 2$
H_1: $\mu_1 \neq 2$

Sous H_0

$$T_c = \frac{\overline{X_1} - 2}{\frac{\hat{\sigma}_1}{\sqrt{n_1}}} \sim \mathrm{T}(0,1),$$

$$\mathrm{ddl} = \mathrm{n}_1 - 1$$

Comparer deux moyennes – Student T Test statistique

- On peut faire un test t classique de comparaison de moyennes sous la condition de normalité et d'égalité des variances (homoscedasticity) que l'on indique comme argument
  - test <- t.test (adu1,adu2,var.equal=TRUE)

- Et on récupère le ddl (20+20-2) avec Two Sample t-test

data: adu1 and adu2

t = 3.1683, **df = 38**, p-value = 0.003023

alternative hypothesis: true difference in means

is not equal to 0

95 percent confidence interval:

1.99 9.04

sample estimates:

mean of x mean of y

49.96 44.44

Condition:

Independence, normalité, homoscedasticité

Hypothèse:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Sous H_0

$$T_c = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathrm{T}(0,1),$$

$$\mathrm{ddl} = \mathrm{n}_1 + \mathrm{n}_2 - 2$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)\hat{\sigma}_1{}^2 + (n_2 - 1)\hat{\sigma}_2{}^2}{n_1 + n_2 - 2}$$

Alternative Student T Test statistique

- Hypothèses unilatérales

>?t.test

Default S3 method:

- test <- t.test(x,y , alternative = ***c("two.sided", "less", "greater")***, mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)
- Explore test$....

- Les autres arguments possibles

mu =0

Paired =FALSE ou TRUE

Var.equal=FALSE (Welch's t-test) ou TRUE (t-test)

Conf.level=0.95

Test apparié

- Imaginons maintenant que les deux échantillons sont appariés …

t.test(adu1,adu2,paired=T)

CONCLUSION ?

Conditions:
Dépendance, normalité

Hypothèse:
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Test non paramétrique – Mann Whitney Wilcoxon

- Supposons que la normalité et/ou que l'égalité des variances ne sont pas respectées pour les 2 échantillons indépendant
- On doit alors faire un test non paramétrique :
  - Mann-Whitney Wilcoxon

Souvent appeler ≪ Test des médianes ≫, mais qui va en fait comparer comment les valeurs sont rangées entre les deux échantillons (tests des distributions des rangs)

>wilcox.test(adu1,adu2)

CONCLUSION ?

Test appariée non-paramétrique:

Wilcoxon signed rank test (appariée)

>wilcox.test(adu1,adu2, paired=T)

X_1 VAC échantillon 1 (E1)
X_2 VAC échantillon 2 (E2)

Conditions:
Independence

Hypothèse:
H_0: $\mu_{Rang1} = \mu_{Rang2}$
ou $P(x_1>x_2)=0.5$
H_1: $\mu_{Rang1} \neq \mu_{Rang2}$
ou $P(x_1>x_2)\neq 0.5$

Variable de décision:
W_1 = somme des rangs E1
W_2 = somme des rangs E2

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2 = n_1 n_2 - U_1$$

$U_1+U_2 = n_1 n_2$
Petits effectifs: $U_{obs} = \min(U_1,U_2)$
Grand effectifs (n>20):

$$Z_{obs} = \frac{W_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} \sim N(0,1) \text{ sous } H_0$$

Test non paramétrique – Mann Whitney Wilcoxon

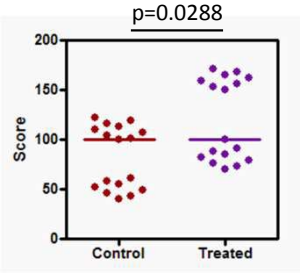
Souvent appeler ≪ Test des médianes ≫, mais qui va en fait comparer comment les valeurs sont rangées entre les deux échantillons (tests des distributions des rangs)

Des problemes avec les tests de rang.

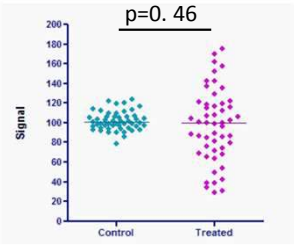
1. Compare les médianes des rangs
2. Ne compare pas les médianes de non-rang
3. Ne compare pas les distributions

Si les échantillon suit le même distribution (pas forcement normale) le test permettre de conclure sur le différence de médiane ou moyenne non-rang.

p=0.0288



p=0. 46



X_1 VAC échantillon 1 (E1)
X_2 VAC échantillon 2 (E2)

Conditions:
Independence

Hypothèse:
H_0: $\mu_{Rang1} = \mu_{Rang2}$
ou $P(x_1>x_2)=0.5$
H_1: $\mu_{Rang1} \neq \mu_{Rang2}$
ou $P(x_1>x_2)\neq 0.5$

Variable de décision:
W_1 = somme des rangs E1
W_2 = somme des rangs E2

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2 = n_1 n_2 - U_1$$

$U_1+U_2 = n_1 n_2$
Petits effectifs: $U_{obs} = \min(U_1,U_2)$
Grand effectifs (n>20):

$$Z_{obs} = \frac{W_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \sim N(0,1) \text{ sous } H_0$$

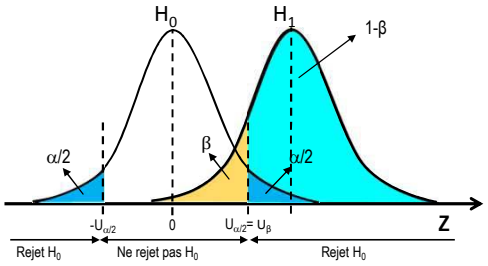
Analyse de puissance

Mention science de la vie – L2

Test de comparaison de 2 moyennes – analyse de puissance

Déterminer le nombre de sujets?

- Bénéfice minimum cliniquement intéressant avec le nouveau traitement ($\Delta=\mu_A-\mu_B$)
- Variabilité de la réponse (variance dans les deux groupes)
- Risque α ($P(|U|>U_\alpha \mid H_0)$)
- Risque β ($P(U<U_\beta \mid H_1)$)
- Puissance $1-\beta$ ($P(U>U_\beta \mid H_1)$)



Hypothese:
H_0: $\mu_A=\mu_B$
H_1: $\mu_A\neq\mu_B$

$$U=\frac{\bar{X}_A-\bar{X}_B-(\mu_A-\mu_B)}{\sqrt{\frac{\sigma^2_A}{n_A}+\frac{\sigma^2_B}{n_B}}}\sim N(0,1)$$

$$sous\ H_0:\ U=\frac{\bar{X}_A-\bar{X}_B}{\sqrt{\frac{\sigma^2_A}{n_A}+\frac{\sigma^2_B}{n_B}}}\sim N(0,1)$$

$$sous\ H_1:\ U=\frac{\bar{X}_A-\bar{X}_B}{\sqrt{\frac{\sigma^2_A}{n_A}+\frac{\sigma^2_B}{n_B}}}\sim N\left(\frac{\mu_A-\mu_B}{\sqrt{\frac{\sigma^2_A}{n_A}+\frac{\sigma^2_B}{n_B}}},1\right)$$

$\alpha = P(U > U_{\alpha/2} \mid H_0)$ => $U_{\alpha/2} = U_\beta$
$\beta = P(U < U_\beta \mid H_1)$

Cohen's d: $d=\frac{\mu_A-\mu_B}{s_p}$; $s^2{}_p=\frac{(n_A-1)s^2{}_A+(n_B-1)s^2{}_B}{n_A+n_B-2}$

Test de comparaison de 2 moyennes – analyse de puissance

Comment déterminer le nombre de sujets?

- Différence minimum pour être cliniquement intéressant ($\Delta=\mu_A-\mu_B$)
- Variabilité de la réponse (variance dans les deux groupes)
- Risque α (U_α = qnorm(1-a/2,0,1) = 1.96 (a=5%)
- Puissance $1-\beta$ ($U_{2\beta}$ = -qnorm(b,0,1) = 1.28 (b=10%)

⇒ **Nombre de sujets n par groupe**

$$n=[u_\alpha+u_{2\beta}]^2\frac{\sigma_A{}^2+\sigma_B{}^2}{[\mu_A-\mu_B]^2}$$

n est le nombre par échantillon (groupe)
U_α et $U_{2\beta}$ basé sur un tableau Z~N(0,1) bilatérale.
Formule vrai pour Z-test,
mais pas pour T-test (sous-estimation)

R code (ADU dataset):

```
> s1 <- var(adu1)^0.5
> s2 <- var(adu2)^0.5
> n <- (Ua+U2b)^2*(s1^2+s2^2)/(mean(adu1)-mean(adu2))^2
> s3 <- ((length(adu1)-1)*(s1^2+s2^2)/(2*length(adu1)-2))^0.5
> cohenD <- (mean(adu1)-mean(adu2))/s3
```

Analyse de puissance – Taille d'effet

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{\mu_1 - \mu_2}{s} \qquad s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$$

$$w = \sqrt{\sum_{i=1}^{N} \frac{(p_{0i} - p_{1i})^2}{p_{0i}}}$$

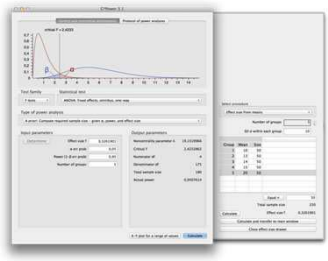
$$r_{xy} = \frac{Cov(x,y)}{\hat{\sigma}_x \hat{\sigma}_y}$$

	Z-test	χ^2 test	Pearson correlation
Effect size	Cohen's d	W	r
Small	0.2	0.1	0.1
Medium	0.5	0.3	0.3
High	0.8	0.5	0.5

« Cohen a suggéré que d = 0,2 soit considéré comme une «petite» taille d'effet, 0,5 représente une taille d'effet «moyenne» et 0,8 une «grande» taille d'effet. **Cela signifie** que si les moyennes de deux groupes ne **diffèrent pas de 0,2 écart-type** ou plus, la différence est négligeable, même si elle est statistiquement significative »

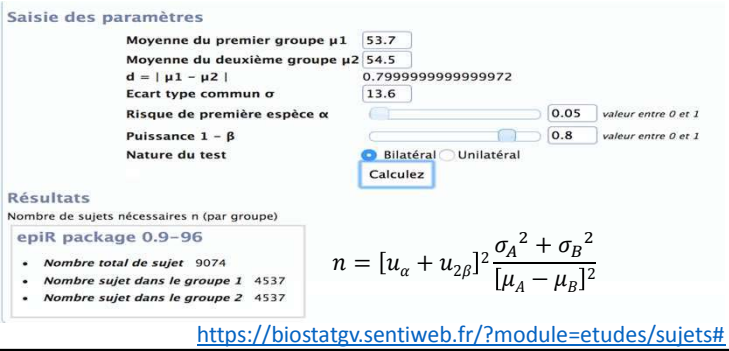
Analyse de puissance - outils

G*POWER



http://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html

BiostaTGV
Etudes cliniques



$$n = [u_\alpha + u_{2\beta}]^2 \frac{{\sigma_A}^2 + {\sigma_B}^2}{[\mu_A - \mu_B]^2}$$

https://biostatgv.sentiweb.fr/?module=etudes/sujets#

Contrôle continue

L'objectif est de résoudre un exercice qui consistera à choisir et à appliquer le bon test statistique à un problème biologique donné. vous recevrez une question ainsi que les données correspondantes. Vous devrez résoudre le problème sans utiliser des notes. Vous utiliserez R-Studio sur les ordinateurs pour effectuer l'analyse, mais votre rapport devrait être rédigé à la main. Ce que j'attends de vous et à quoi vous devriez donc vous préparer:

1. Importation de données
2. Visualisation des données
3. Choix du test statistique
4. Hypothèse
5. Énumérer et valider les conditions (les hypothèses et conditions intermédiaires doivent être décrites)
6. Effectuer le test
7. Conclusion (statistique et biologique)

Les examens des dernières années et les données correspondantes sont accessible sur Moodle si vous souhaitez vous exercer. Comme vous le verrez, la différence entre l'examen et le "contrôle continue" est que, pour l'examen, les tests sont pré-calculés (votre travail consiste à choisir les bons analyses et à les interpréter). Pour le "contrôle continue", vous devrez également écrire le code vous-même. Par conséquent, assurez-vous que vous avez mémorisé les fonctions R essentiels. Bien que vous puissiez utiliser la fonction d'aide intégrée de R (pas d'Internet), il vous sera utile de mémoriser les éléments essentiels pour gagner du temps.

Contrôle continue continue

Code R: Je m'attends à ce que vous connaissiez les commandes suivantes et sachiez quand les utiliser (toutes les fonctions sont décrites dans les diapositives de mon cours):

Importer des données:
data <- read.table(file.choose(), header = TRUE)

Afficher les données:
barplot(données)
boxplot(données)
mosaicplot(matrice)
qqnorm(variable)
plot(x,y) ou plot(y~x)
À noter! Dans mes diapositives, j'ai utilisé un certain nombre de paramètres supplémentaires pour rendre les plots plus jolies, mais il vous est seulement demandé d'utiliser les paramètres critiques, les données.

Tests:
shapiro.test(variable) # Normalité (je ne m'attends pas à ce que vous utilisiez les autres tests de normalité mentionnés dans le cours, mais vous devez savoir qu'ils existent)
var.test(variable1, variable2) # Homoscédasticité
t.test(variable1, variable2)
chisq.test(matrice)

Au-delà, vous devez étudier les conditions des différents tests pour pouvoir identifier la variable, définir l'hypothèse et tirer une conclusion. Ce sont normalement des éléments que vous avez vus dans 2V314 et que je vous ai encore présentés la semaine dernière.
Pour info, vous pouvez utiliser vos propres ordinateurs pour le CC.