Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

Rappels #4.2 (statistique & tests)

Soit un modèle paramétrique, $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$.
Un test d'hypothèses composites sur un paramètre réel θ a pour objectif de décider entre deux hypothèses $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta \setminus \Theta_0$ où $\Theta_0 \subset \Theta$, à partir de données.

- Lorsque $\Theta = \mathbb{R}$, $\Theta_0 = \theta_0$ et donc $\Theta \setminus \Theta_0 = \{\theta \neq \theta_0\}$, on parle de test bilatéral.
- Lorsque $\Theta = \mathbb{R}$, $\Theta_0 = \{\theta \leq \theta_0\}$ et donc $\Theta \setminus \Theta_0 = \{\theta > \theta_0\}$, on parle de test unilatéral à droite.
- Lorsque $\Theta = \mathbb{R}$, $\Theta_0 = \{\theta \geq \theta_0\}$ et donc $\Theta \setminus \Theta_0 = \{\theta < \theta_0\}$, on parle de test unilatéral à gauche.

Puisqu'il nous faut décider entre les deux hypothèses au vu d'observation, la décision sera entachée d'erreur

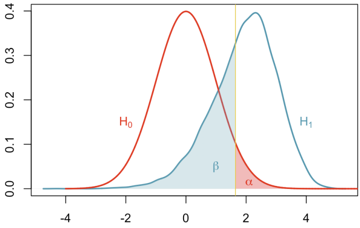
	État de la nature	
	H_0 est vraie	H_1 est vraie
Ne pas rejeter H_0	décision correcte	**erreur de type II**
Rejet de H_0	**erreur de type I**	décision correcte

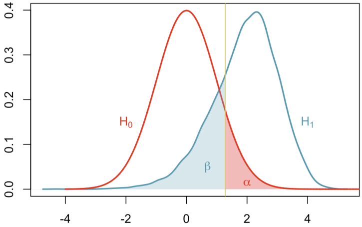
- Erreur de type I: $\mathbb{P}($rejeter $H_0 \mid H_0$ est vraie $) = \alpha$.
- Erreur de type II: $\mathbb{P}($ne pas rejeter $H_0 \mid H_0$ est fausse $) = \beta$.
- Les risques d'erreurs dépendent de θ. $\alpha = \alpha(\theta), \theta \in \Theta_0$; $\beta = \beta(\theta), \theta \in \Theta \setminus \Theta_0$.
- pour les tests unilatéraux, il est fréquent de remplacer l'hypothèse nulle par $H_0 : \theta = \theta_0$, si le risque d'erreur de première espèce est maximisé pour la valeur $\theta = \theta_0$.

Type I and Type II Error Balance

Traditionally we try to set Type I error probability as 5% or 1%, as in there is only a 5 or 1 in 100 chance that the variation that we are seeing is due to chance.

E.g. on student's height, $H_0 : \mu_M = \mu_F + 10\text{cm}$





Tests for Proportion (one sample)

Here $Y_i \sim \mathcal{B}(\theta)$, with $\theta = \mathbb{E}(Y) = \mathbb{P}(Y = 1)$. Set $\widehat{\theta} = \overline{y}$

```
> y = c(1,1,0,1,0,1,0,0,1,0,1)
> mean(y)
[1] 0.5454545
> n = length(y)
```

We want to test $H_0 : \theta = \theta^\star$, e.g. $\theta^\star = 1/2$.

```
> phat = mean(y)
> pstar = 1/2
> alpha = 5/100
```

Gaussian confidence interval

$$\left[\widehat{\theta} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}\right]$$

```
> phat+qnorm(c(.025,.975))*sqrt(phat*(1-phat)/n)
[1] 0.2512024 0.8397067
```

Z-test

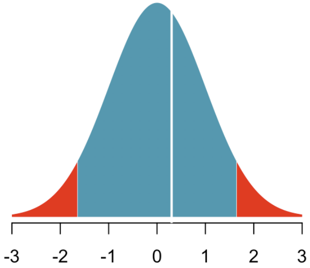
$$z = \sqrt{n}\frac{\widehat{\theta} - \theta^\star}{\sqrt{\theta^\star(1-\theta^\star)}}$$

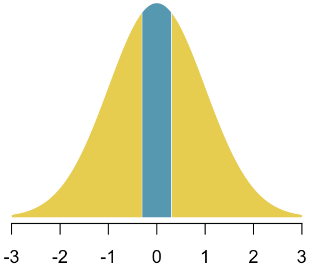
so we reject H_0 at level α if

$$z \notin \left[\Phi^{-1}\left(\frac{\alpha}{2}\right), \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

```
> (z = sqrt(n)*(phat-pstar)/(sqrt(
    pstar*(1-pstar))))
[1] 0.3015113
> abs(z)>qnorm(1-alpha/2)
[1] FALSE
```

so here, we do not reject H_0





Z-test

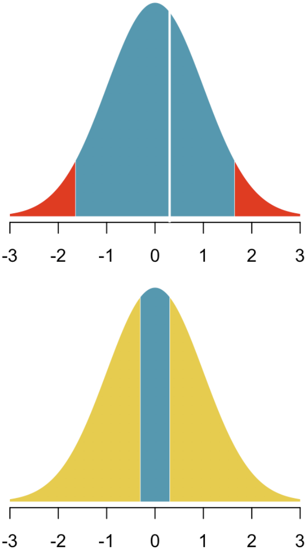
$$z = \sqrt{n}\frac{\widehat{\theta} - \theta^\star}{\sqrt{\theta^\star(1-\theta^\star)}}$$

We can also use the p-value, i.e., if $Z \sim \mathcal{N}(0,1)$

$$p = \mathbb{P}[|Z| > |z|] \leq \alpha$$

```
> 2*(1-pnorm(abs(z)))
[1] 0.7630246
```

so here, $p > \alpha$ so we do not reject H_0.

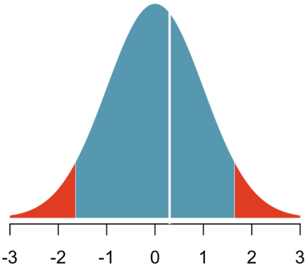


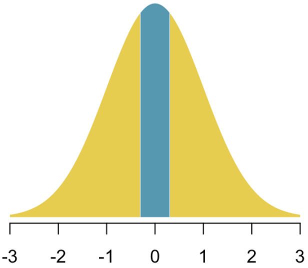
t-test

Since proportion is also the mean, use

$$t = \sqrt{n-1}\frac{\widehat{\theta}-\theta^\star}{\sqrt{\widehat{\theta}(1-\widehat{\theta})}}$$

```
> (t = sqrt(n-1)*(phat-pstar)/(sqrt(
    phat*(1-phat))))
[1] 0.2886751
> t.test(y,mu = 1/2)
data:  y
t = 0.28868, p-value = 0.7787
alternative hypothesis: true mean is
     not equal to 0.5
95 percent confidence interval:
 0.1946137 0.8962954
sample estimates:
mean of x
0.5454545
```





Test for Proportion (Two Samples)

	randomized	
	size	number
treatment	200,000	57
control	200,000	142
no consent	350,000	92

Jonas Salk's polio vaccine, 1954

H_0 : vaccine has no effect

$H_0 : p_t = p_c$

$X_\cdot$: number of observed cases in a group, $X_\cdot \sim \mathcal{N}(np_\cdot, np_\cdot(1-p_\cdot))$
here $p_\cdot$ small, $\text{Var}(X_\cdot) \simeq np_\cdot$

X_t: number of observed cases in the treatment group

X_c: number of observed cases in the control group

Two groups have similar size (n), $X_c - X_t \sim \mathcal{N}(\star, n(p_c + p_t))$

$$z = \frac{142 - 57}{\sqrt{142 + 57}} \simeq 6.1$$

(very unlikely under H_0, since Z should follow a $\mathcal{N}(0,1)$)

Acceptation / Rejection Regions

Consider n coin flipping. We observed 55% tails. Is the coin biased?

Not biased (H_0) means $p = p_0 = 50\%$

Under H_0, $\overline{x} \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right)$, i.e. if H_0 is true

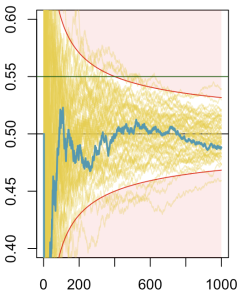
with 95% chance, $\overline{x} \in \left[\frac{1}{2} \pm \frac{1}{\sqrt{n}}\right]$

hence, 55% belongs to that interval if $\frac{1}{\sqrt{n}} \geq 5\%$

i.e. $n \leq 400$

Equivalently, $z = 2\sqrt{n}\left(\overline{x} - \frac{1}{2}\right) \sim \mathcal{N}(0,1)$

We reject H_0 if $|z| > 2$ (or 1.96).



p-value

Consider n coin flipping. We observed 55% tails. Is the coin biased?

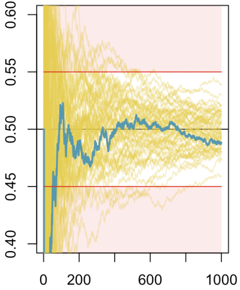
Conversely, we can compute

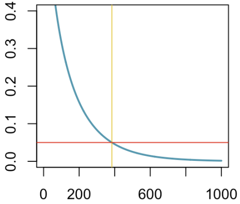
$$\mathbb{P}(|\overline{X}| > 55\%) \text{ when } \overline{X} \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right)$$

called p-value.

We reject H_0 if $p < 5\%$.

If we could replicate experiments of this sample size, how often will we see a statistic this extreme, assuming that H_0 is true ?





The "$p < 5\%$" Dogma

"if p is between 10% and 90% there is certainly no reason to suspect the hypothesis tested. If it is below 2% it is strongly indicated that the hypothesis fails to account for the whole of facts [...] We shall not often be astray if we draw a conventional line at 5%"
Ronald Fisher

see La guerre des étoiles, *p*-value and statistical practice or It's time to talk about ditching statistical significance