Calcul des probabilités et quantiles sous Excel, R et Python

Les fonctions de distribution de probabilité utilisés couramment en inférence statistique (intervalles de confiance et tests) : loi normale, loi de Student, loi du Khi-2, loi de Fisher.

Ricco Rakotomalala

Calcul des quantiles et des probabilités critiques

Les calculs des quantiles et des valeurs des fonctions de répartition des principales distributions dérivées de la loi normale sont nécessaires en statistique inférentielle, lors de la construction des intervalles de confiance, lors de la mise en œuvre des tests d'hypothèses.

Des fonctions disponibles dans différents outils nous permettent d'obtenir les valeurs et nous affranchissent des tables statistiques.

- Via les fonctions statistiques sous Excel (de nouvelles fonctions sont arrivées avec Excel 2010)
- Via les fonctions du package « stats » de R (directement accessibles)
- Via les fonctions du package « scipy » de Python que l'on doit importer au préalable

```
import scipy.stats as stats
```

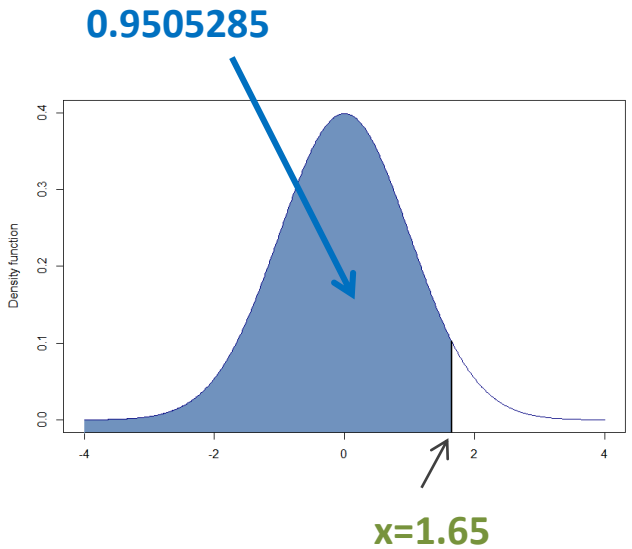
LOI NORMALE

Fonction de répartition de **la loi normale centrée (μ = 0) et réduite (σ = 1)**.

La probabilité d'être inférieure à une valeur **x = 1.65** est égale à **0.9505285**

Fonction de densité de la loi normale

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



EXCEL

« VRAI » pour qu'on ait la fonction de répartition et non la fonction de densité (FAUX). Paramètre obligatoire.

LOI.NORMALE.N(1.65; 0 ; 1 ; VRAI)

(μ = 0) et (σ = 1). Paramètres obligatoires.

LOI.NORMALE.STANDARD.N(1.65;VRAI) — Directement la loi centrée et réduite.

R

pnorm(1.65, mean = 0, sd = 1, lower.tail = TRUE)

Loi normale centrée (μ = 0) et réduite (σ = 1). Paramètres optionnels.

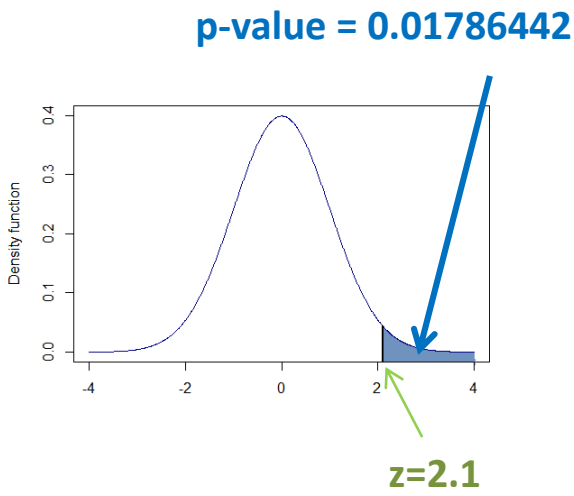
C'est bien la surface entre $]-\infty ; q\,]$ qui est calculée. Paramètre optionnel.

Python

stats.norm.cdf(1.65, loc = 0, scale = 1)

Loi normale centrée (μ = 0) et réduite (σ = 1). Paramètres optionnels.

Calcul de la p-value de **la loi normale centrée et réduite** dans un test unilatéral à droite. La probabilité d'être supérieur à une statistique de test **z = 2.1** est égale à 0.01786442



EXCEL

1-LOI.NORMALE.STANDARD.N(2.1;VRAI)

R

1 - pnorm(2.1)

pnorm(2.1, lower.tail = FALSE)

C'est la surface entre $[z\ ;\ +\infty[$ qui est calculée.

Python

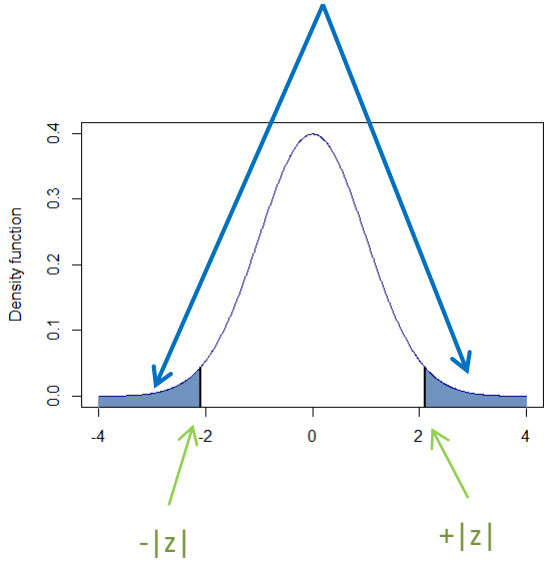
1 - stats.norm.cdf(2.1)

stats.norm.sf(2.1)

sf = 1 - cdf

Calcul de la p-value de **la loi normale centrée et réduite** dans un test bilatéral. La probabilité d'être supérieur à une statistique de test z = 2.1 en valeur absolue est égale à 0.03572884

p-value = 2 * 0.01786442 = 0.03572884



EXCEL

2*(1-LOI.NORMALE.STANDARD.N(2.1;VRAI))

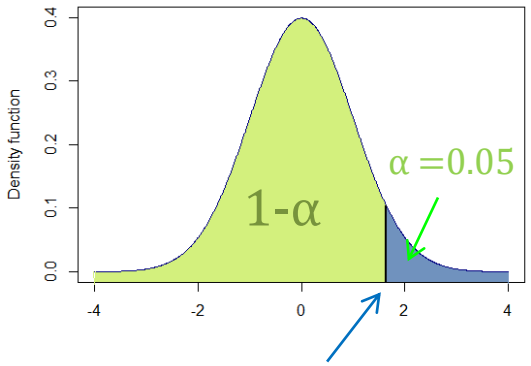
R

```
2 * pnorm(2.1, lower.tail = FALSE)
```

Python

```
2 * (1 - stats.norm.cdf(2.1))
```

Calcul du quantile (q) de **la loi normale centrée et réduite** à partir d'une probabilité $(1 - \alpha) = 0.95$



q = 1.644854

EXCEL

LOI.NORMALE.INVERSE.N(0.95;0;1)

LOI.NORMALE.STANDARD.INVERSE.N(0.95)

R

```
qnorm(0.95,mean=0,sd = 1,lower.tail=TRUE)
qnorm(0.05,mean=0,sd=1,lower.tail=FALSE)
```

Python

```
stats.norm.ppf(0.95, loc =0, scale = 1)
```

Génération d'un nombre aléatoire suivant une loi $\mathcal{N}(\mu=0,\sigma=1)$ (centrée et réduite)

ALEA() est une fonction EXCEL qui génère une valeur aléatoire comprise en [0 ; 1[suivant une distribution uniforme.

EXCEL

```
LOI.NORMALE.STANDARD.INVERSE.N(ALEA())
```

R

```
rnorm(n=1,mean=0,sd = 1)
```

Nombre de valeurs aléatoires à renvoyer. Si (n > 1), nous obtenons un vecteur en sortie de la fonction. Paramètre obligatoire.

Python

Initialisation du générateur. Si random_state = entier, nous obtiendrons toujours la même séquence de valeurs aléatoires. Paramètre optionnel.

```
stats.norm.rvs(loc=0,scale=1, size=1, random_state = none)
```

Nombre de valeurs aléatoires à renvoyer. Si (size > 1), nous obtenons un vecteur en sortie de la fonction. Paramètre optionnel.

Approximations de la fonction de répartition de la loi normale centrée et réduite. Des formules « simples » pour x > 0

$$\Phi_1(x) = 1 - \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(\frac{0.4361836}{1+0.33267x} + \frac{-0.1201676}{(1+0.33267x)^2} + \frac{0.9772980}{(1+0.33267x)^3}\right)$$

(https://fr.wikipedia.org/wiki/Loi_normale)

$$\Phi_2(x) = 0.5 + \frac{1}{2}\left\{1 - \frac{1}{30}\left[7e^{-\frac{x^2}{2}} + 16e^{-x^2(2-\sqrt{2})} + \left(7 + \frac{1}{4}\pi x^2\right)e^{-x^2}\right]\right\}^{\frac{1}{2}}$$

(http://mathworld.wolfram.com/NormalDistributionFunction.html)

$\Phi_1(1.65) = 0.9494966$

$\Phi_2(1.65) = 0.9505364$

(Excel, R et Python ➔ 0.9505285)

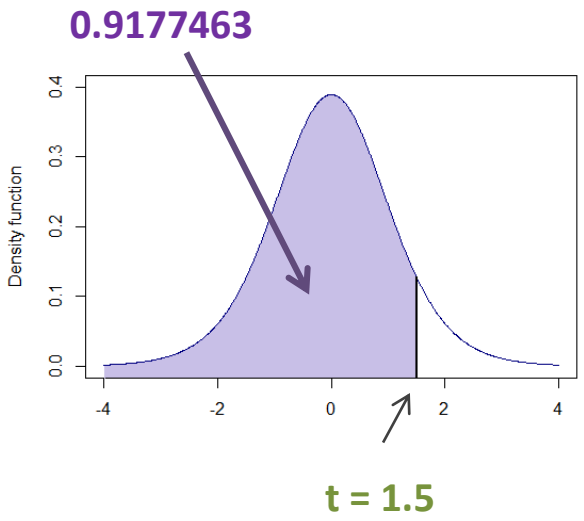
LOI DE STUDENT

Fonction de répartition de la loi de Student à k (k > 0) degrés de liberté.

Probabilité d'avoir une plus petite valeur que t = 1.5 avec k = 10.

Fonction de densité de la loi de Student

$$f_k(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$



EXCEL

« VRAI » pour qu'on ait la fonction de répartition et non la fonction de densité (FAUX). Paramètre obligatoire.

LOI.STUDENT.N(1.5;10;VRAI)

1 - LOI.STUDENT.DROITE(1.5;10)

On peut aussi passer par la probabilité d'excéder t = 1.5

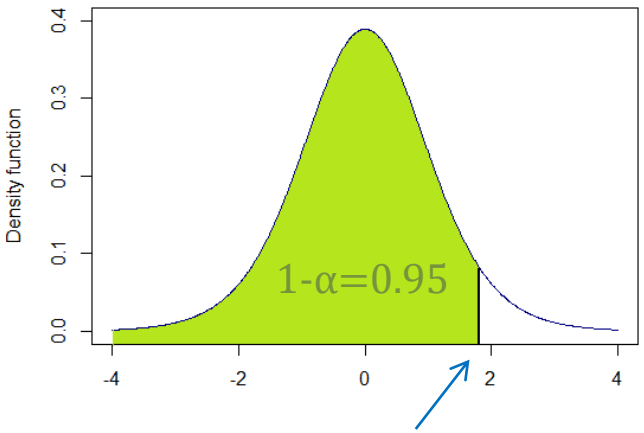
R

```
pt(1.5,df=10,lower.tail=TRUE)
1 - pt(1.5,df=10,lower.tail=FALSE)
```

Python

```
stats.t.cdf(1.5,df=10)
```

Quantile (q) de la loi de Student à k = 10 degrés de liberté pour une probabilité (1 – α) = 0.95



q = 1.812461

EXCEL

LOI.STUDENT.INVERSE.N(0.95;10)

R

```
qt(0.95,df=10,lower.tail=TRUE)
qt(0.05,df=10,lower.tail=FALSE)
```

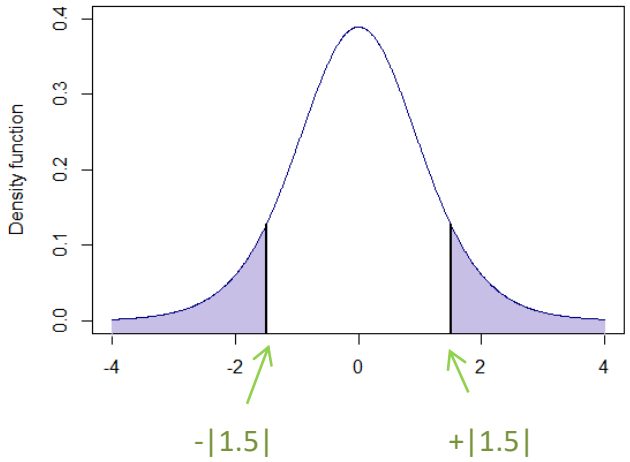
Python

```
stats.t.ppf(0.95,df=10)
```

Probabilité et quantile bilatéraux pour la loi de Student sous EXCEL.

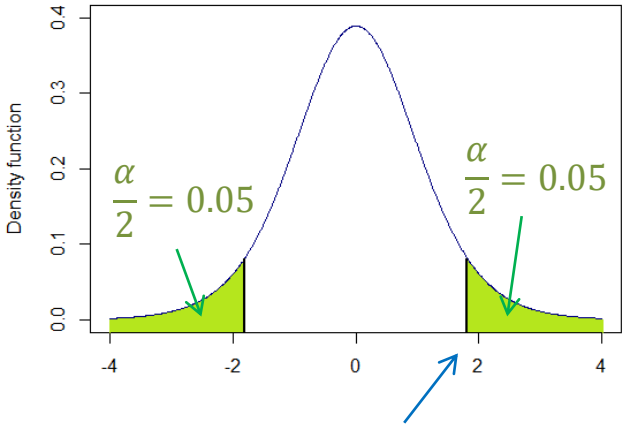
Excel propose deux fonctions spécifiques.

p-value = 2 * 0.08225366 = 0.16450733



ABS() fonction « valeur absolue ». Indispensable si la statistique de test prend une valeur négative.

LOI.STUDENT.BILATERALE(ABS(**1.5**);10)



q = 1.812461

LOI.STUDENT.INVERSE.BILATERALE(0.1;10)

$\alpha = 0.1$

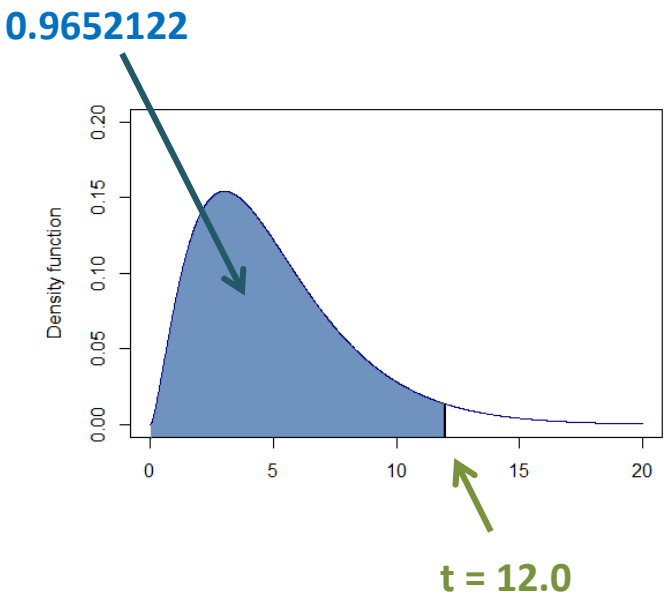
LOI DU KHI-2

Fonction de répartition de la loi du KHI-2 à k (k > 0) degrés de liberté.

Probabilité d'avoir une plus petite valeur que t = 12.0 avec k = 5.

Fonction de densité de la loi du χ^2

$$f_k(t) = \frac{1}{2^{\frac{k}{2}}\,\Gamma\left(\frac{k}{2}\right)} t^{\frac{k}{2}-1} e^{-\frac{t}{2}}$$



« VRAI » pour qu'on ait la fonction de répartition et non la fonction de densité (FAUX). Paramètre obligatoire.

EXCEL

LOI.KHI2.N(12.0;5;VRAI)

1 - LOI.KHI2.DROITE(12.0;5)

On peut aussi passer par la probabilité d'excéder t = 12.0

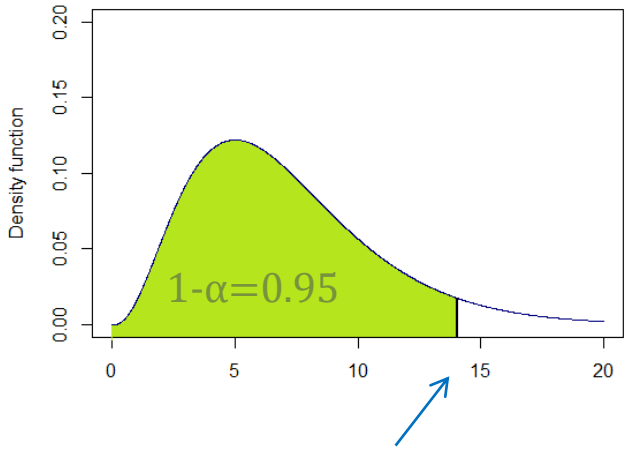
R

```
pchisq(12.0,df=5)
1 - pchisq(12.0,df=5,lower.tail=FALSE)
```

Python

```
stats.chi2.cdf(12.0,df=5)
```

Quantile (q) de la loi du KHI-2 à k = 7 degrés de liberté pour une probabilité (1 – α) = 0.95



q = 14.06714

EXCEL

LOI.KHIDEUX.INVERSE (0.95;7)

LOI.KHIDEUX.INVERSE.DROITE (0.05;7)

R

```
qchsiq(0.95,df=7)
qchisq(0.05,df=7,lower.tail=FALSE)
```

Python

```
stats.chi2.ppf(0.95, df=7)
```

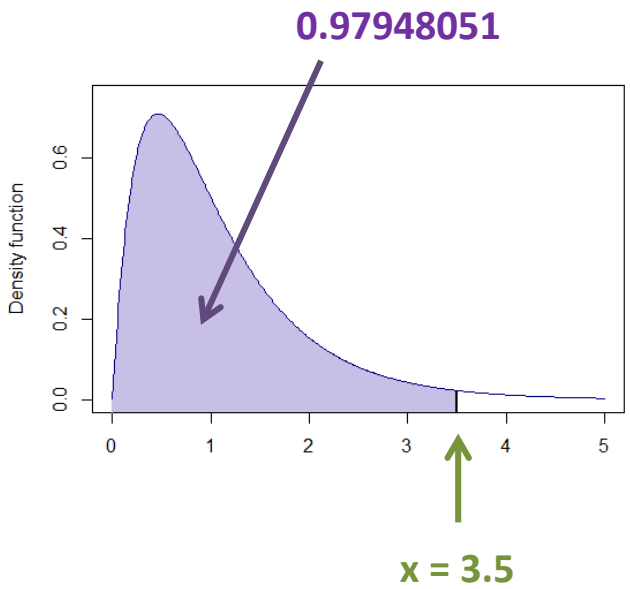
LOI DE FISHER

Fonction de répartition de la loi de Fisher à d1 (d1 > 0) et d2 (d2 > 0) degrés de liberté. Probabilité d'avoir une plus petite valeur que x = 3.5 avec d1 = 4, d2 = 26.

Fonction de densité de la loi de Fisher

$$f(x) = \frac{\left(\frac{d_1 x}{d_1 x + d_2}\right)^{\frac{d_1}{2}} \left(1 - \frac{d_1 x}{d_1 x + d_2}\right)^{\frac{d_2}{2}}}{x\, \mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$

B() est la fonction bêta



« VRAI » pour qu'on ait la fonction de répartition et non la fonction de densité (FAUX). Paramètre obligatoire.

EXCEL

LOI.F.N(3.5;4;26;VRAI)

1 - LOI.F.DROITE(3.5;5;26)

On peut aussi passer par la probabilité d'excéder x = 3.5

R

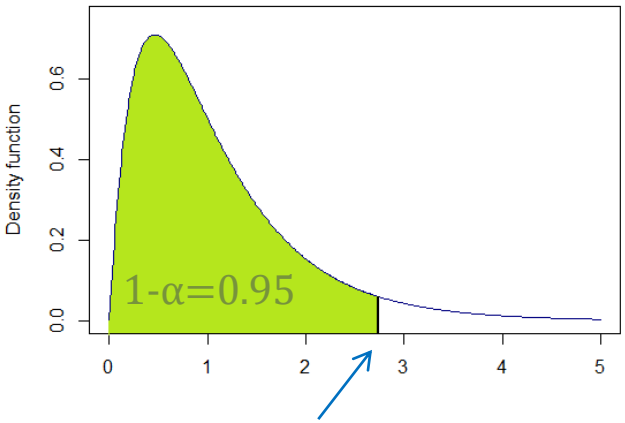
pf(3.5,df1=4,df2=26)

1 - pf(3.5,df1=4,df2=26,lower.tail=FALSE)

Python

stats.f.cdf(3.5,dfn=4,dfd=26)

Quantile (q) de la loi de Fisher à (d1 = 4, d2 = 26) degrés de liberté pour une probabilité (1 – α) = 0.95



q = 2.742594

EXCEL

INVERSE.LOI.F.N(0.95;4;26)

INVERSE.LOI.F.DROITE(0.05;4;26)

R

```
qf(0.95,df1=4,df2=26)
qf(0.05,df1=4,df2=26,lower.tail=FALSE)
```

Python

```
stats.f.ppf(0.95,dfn=4,dfd=26)
```

Références

Références

Scipy.org – Statistical functions (scipy.stats)

https://docs.scipy.org/doc/scipy/reference/stats.html

Support Microsoft – Fonctions Excel (par catégorie) – Fonctions statistiques

https://support.office.com/fr-fr/article/Fonctions-Excel-par-cat%C3%A9gorie-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb

R Tutorial – Basic Probability Distributions

http://www.cyclismo.org/tutorial/R/probability.html