

# Statistiques de base

Ce guide présente la marche à suivre pour effectuer les opérations de base prévues dans le document *Introduction à R – statistiques de base* (Leçon 1)

Les points traités sont les suivants :

- Statistiques descriptives
  - Moyenne arithmétique;
  - Médiane;
  - Variance;
  - Écart type;
  - Erreur type de la moyenne;
  - Quartiles;
  - Déciles;
- Fonctions quantiles;
- Densité de probabilité;
- Probabilités cumulatives.

# Statistiques descriptives

R Commander permet en quelques clics de souris le calcul de plusieurs statistiques descriptives.

Cette diapositive indiquera comment obtenir les descripteurs suivants :

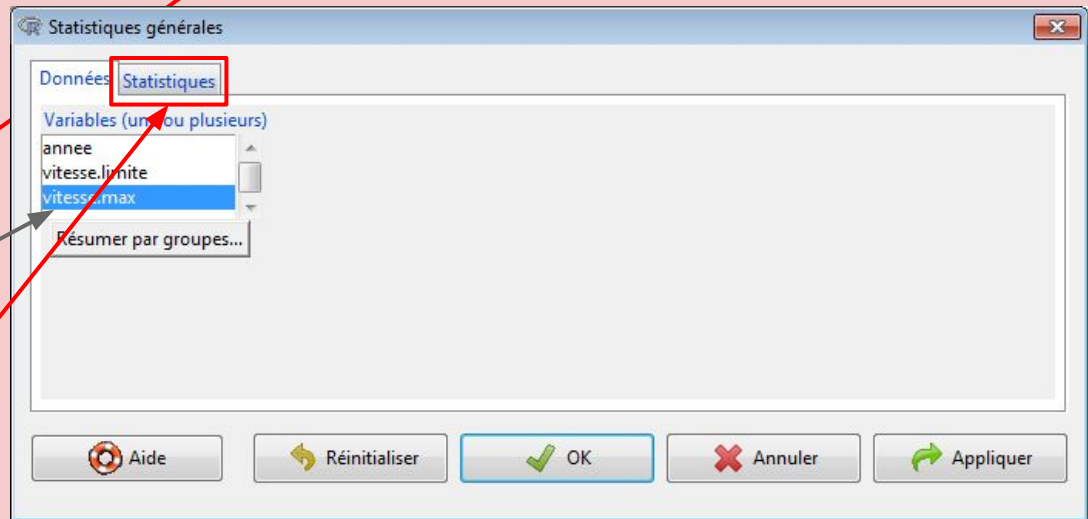
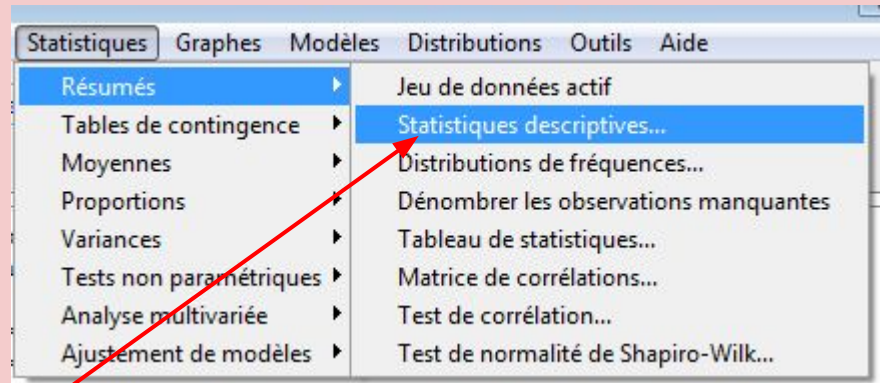
- Moyenne;
- Médiane;
- Écart type;
- Erreur type de la moyenne;
- Quartiles.

1) Téléchargez le fichier vitesse.txt en cliquant sur ce [lien](#). Importez ensuite ce fichier dans RCommander ;

2) Cliquez sur le menu Statistiques. Choisissez tout d'abord l'option Résumés et ensuite l'option Statistiques descriptives;

3) Dans la fenêtre qui apparaît, veuillez sélectionner la ou les variables pour lesquelles vous souhaitez connaître ces descripteurs (ici `vitesse.max`);

4) Cliquez sur l'onglet Statistiques afin de sélectionner les statistiques que vous souhaitez connaître;

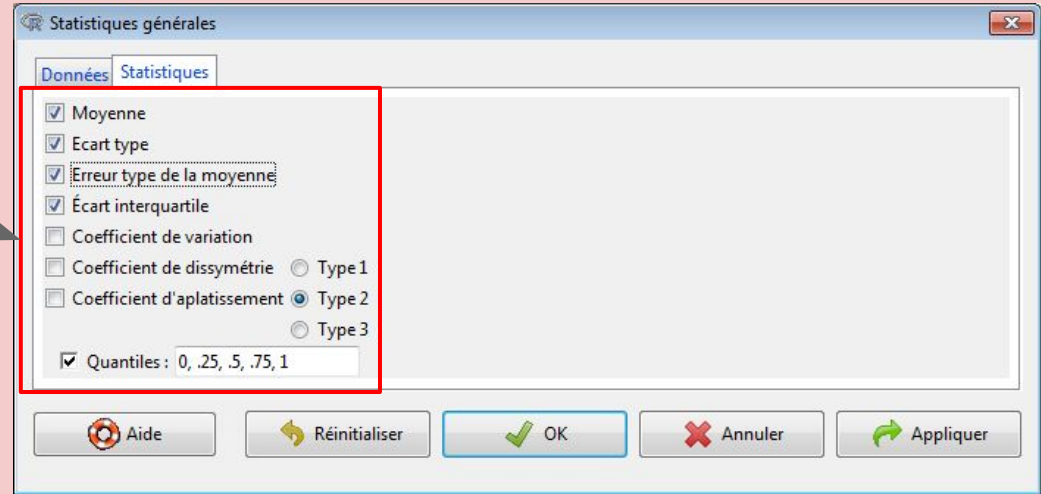


# Statistiques descriptives (suite)

5) Les statistiques disponibles sont présentées à gauche de la fenêtre

6) Après avoir cliqué sur le bouton OK, vous verrez apparaître dans la zone Sortie les descripteurs sélectionnés :

- mean : moyenne arithmétique;
- sd : Écart type;
- se (mean) : Erreur type de la moyenne;
- IQR : [Écart interquartile](#);
- 0 % : minimum;
- 25 % : premier quartile;
- 50 % : deuxième quartile (médiane);
- 75 % : troisième quartile;
- 100 % : maximum;
- n : nombre d'observations.



```
> numSummary(vitesse["vitesse.max"], statistics=c("mean", "sd", "se(mean)",  
+ "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))  
      mean      sd se(mean)      IQR 0% 25% 50%      75%    100%  n  
311.7474 99.39887 22.80366 120.135 19 267 324 387.135 435.31 19
```

# Statistiques descriptives (suite)

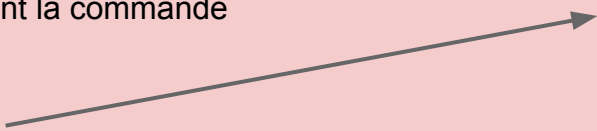
R Commander ne permet pas de calculer la variance et les déciles. Il faut donc faire appel aux fonctions de R.

L'utilisation de fonctions exige souvent de préciser la variable analysée. Pour ce faire, vous devrez indiquer cette variable en séparant le nom de l'objet et le nom de la variable par un symbole de dollar (\$). Par exemple, si vous souhaitez analyser la variable `vitesse.max` qui se trouve dans l'objet `vitesse`, vous devrez inscrire `vitesse$vitesse.max` dans la fonction.

Dans l'éventualité où l'objet contenait une seule colonne, il suffit de répéter le nom de l'objet. Le nom de la variable pourrait être `vitesse$vitesse`.

On calcule la variance en soumettant la commande suivante :


`var(vitesse$vitesse.max)`



```
> var(vitesse$vitesse.max)
[1] 9880.135
```

Les déciles quant à eux sont calculés en soumettant la commande qui suit.

`quantile(vitesse$vitesse.max, prob = c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1))`



```
> quantile(vitesse$vitesse.max, prob = c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1))
 10%   20%   30%   40%   50%   60%   70%   80%   90%  100%
220.100 253.400 277.400 308.200 324.000 333.800 371.592 396.110 416.022 435.310
```

Comme dans le cas des autres résultats, ces descripteurs apparaissent dans la zone Sortie.

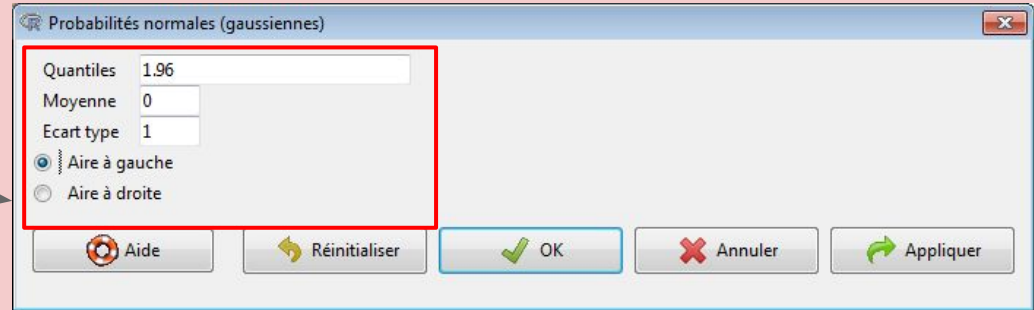
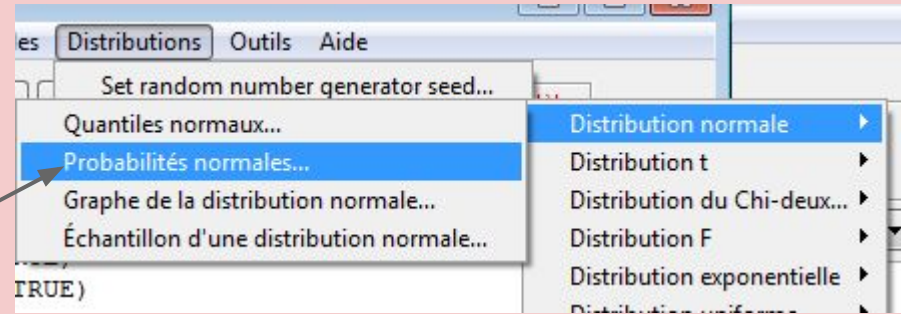
# Fonctions quantiles

R Commander permet de réaliser les fonctions de quantiles présentées à la section 2.2 du document cité précédemment..

On accède à la fonction `pnorm` de la façon suivante :

1) Cliquez sur le menu Distributions. Choisissez tout d'abord l'option Distributions continues, ensuite l'option Distribution normale et finalement l'option Probabilités normales;

2) À la fenêtre qui apparaît, saisissez la valeur seuil à la cellule Quantile (ici 1.96), la moyenne dans la cellule appropriée (ici 0) et finalement l'écart type (ici 1). Enfin, on sélectionne l'option Aire à gauche.



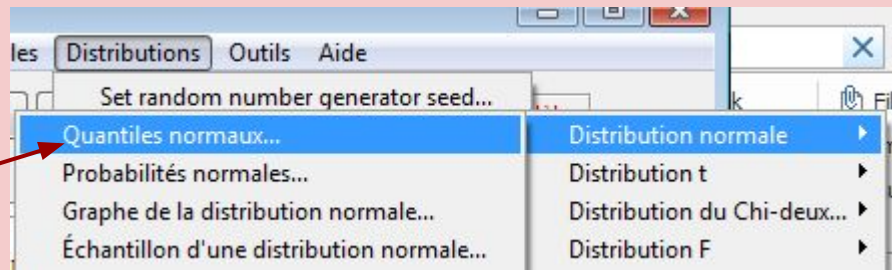
Après avoir cliqué sur OK, vous verrez apparaître dans la zone Sortie la réponse recherchée, soit 0,975.

```
> pnorm(c(1.96), mean=0, sd=1, lower.tail=TRUE)
[1] 0.9750021
```

# Fonctions quantiles (suite)

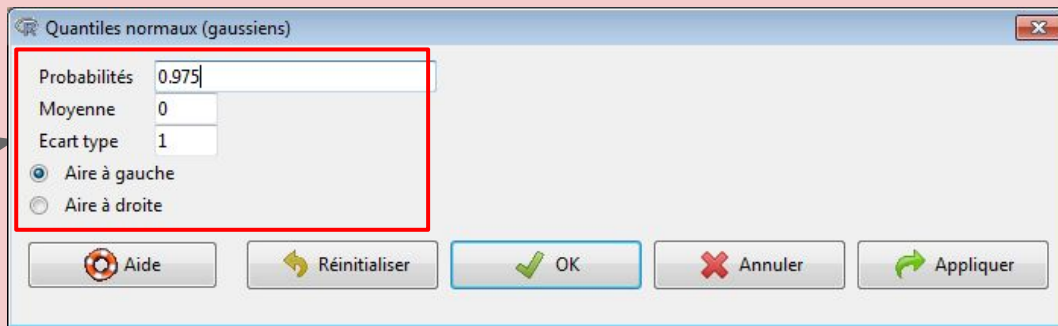
R Commander permet également d'effectuer la fonction `qnorm`. La marche à suivre est la suivante :

1) Cliquez sur le menu Distributions. Choisissez tout d'abord l'option Distributions continues, ensuite l'option Distribution normale et finalement l'option Quantiles normaux;

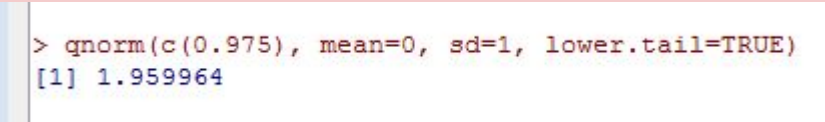


2) Remplissez ensuite les cellules de la fenêtre à l'aide des informations présentées dans la commande, c'est à dire :

- Probabilités : 0.975
- Moyenne : 0
- Écart type : 1



En cliquant sur OK, vous obtiendrez dans la zone Sortie la réponse, c.-à-d. 1,96.



```
> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

A screenshot of the R console window showing the execution of the `qnorm` function. The command is `> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)` and the output is `[1] 1.959964`. A grey arrow points from the text above to this console output.


# Densité de probabilité

La densité de probabilité d'une observation peut uniquement être calculée à l'aide de R en soumettant la commande suivante :

```
dnorm(x = 3.4, mean = 4.1, sd = 1.5)
```

où :

- `x` : la valeur pour laquelle on souhaite connaître la densité de probabilité;
- `mean` : moyenne de l'échantillon;
- `sd` : écart type de l'échantillon.



```
> dnorm(x = 3.4, mean = 4.1, sd = 1.5)
[1] 0.2385223
```

Comme on peut le constater à la page 4 des notes de cours, la densité de probabilité est de 0.23852.



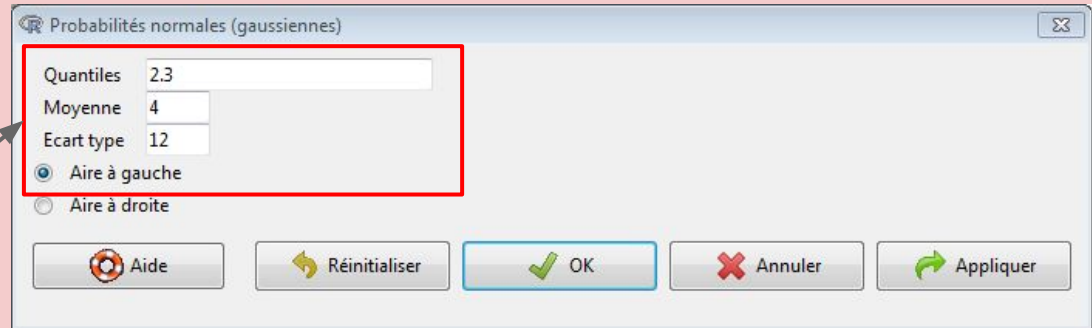
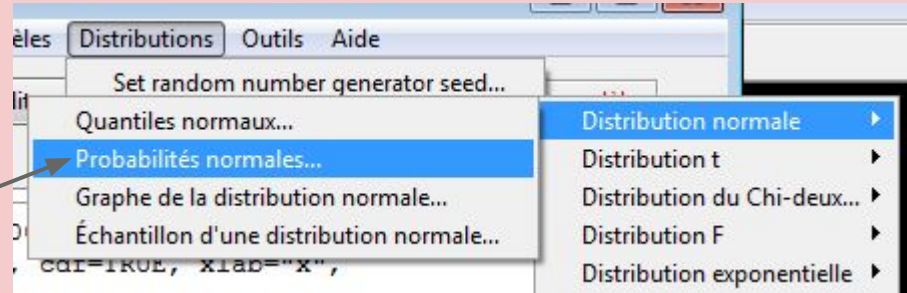
# Probabilités cumulatives

Les étapes qui suivent permettent de calculer la probabilité cumulative à partir d'un seuil prédéfini. Cette probabilité est calculée à partir de l'exemple présenté à la section 2.1 du document cité au début du guide..

1) Cliquez sur le menu Distributions. Choisissez tout d'abord l'option Distributions continues, ensuite l'option Distribution normale et finalement l'option Probabilités normales;

2) Dans la fenêtre qui apparaît, saisissez la valeur seuil à la cellule Quantile (ici 2.3), la moyenne dans la cellule appropriée (ici 4) et finalement l'écart type (ici 12). Enfin, on sélectionne l'option Aire à gauche étant donné que l'on souhaite connaître la probabilité cumulative à partir des valeurs inférieures à 2.3;

En cliquant sur OK, vous obtiendrez la probabilité cumulative de 0,44367



```
> pnorm(c(2.3), mean=4, sd=12, lower.tail=TRUE)
[1] 0.4436717
```