

Mesure de la force de la liaison entre deux variables quantitatives : le coefficient de corrélation

- Deux variables sont dites dépendantes quand la connaissance de l'une donne une indication sur la valeur de l'autre
- La notion de « force » d'une liaison
- Causalité et liaison

- Le coefficient de corrélation
- $r = 0 \rightarrow$ indépendance (?)
- $r = \pm 1 \rightarrow$ dépendance totale : $Y = a X + b$

Traduit la force de la liaison

$$r = \frac{(x_1 y_1 + \dots + x_n y_n) - n m_X m_Y}{(n-1) s_X s_Y}$$

Standardisation
($-1 \leq r \leq 1$)

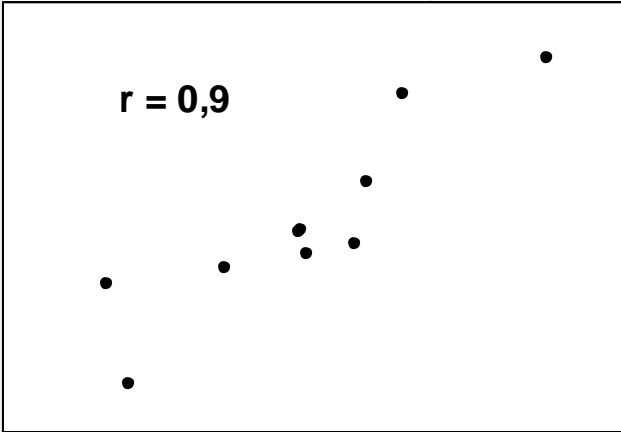
- Quel sens donner à un coefficient de corrélation ?
- Pourcentage de variance partagé ?
- Illustration :

Illustration

Introduction à la statistique avec R > La corrélation

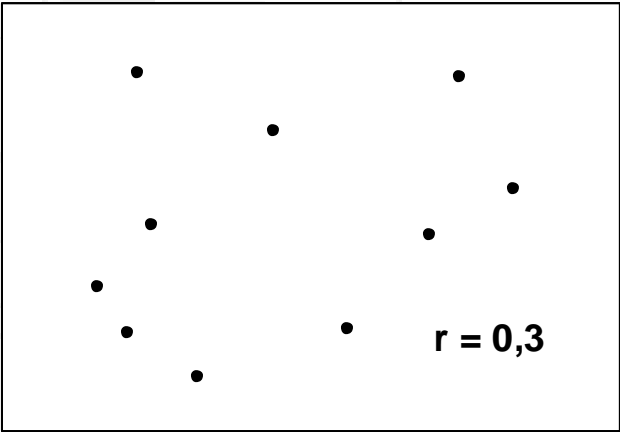


$r = 0,9$



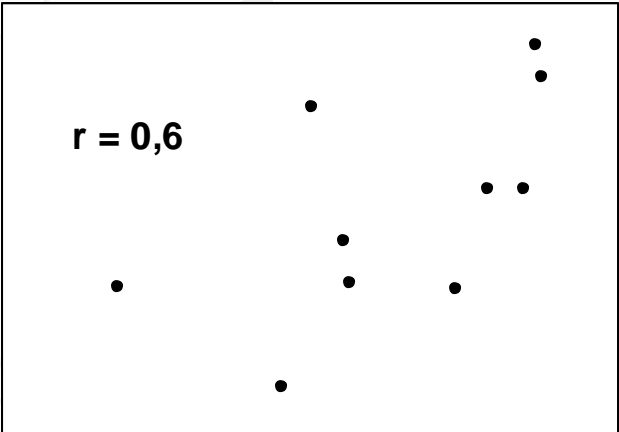
A scatter plot showing a strong positive linear correlation with a correlation coefficient of $r = 0,9$. The data points are tightly clustered along a diagonal line from the bottom-left to the top-right.

$r = 0,3$



A scatter plot showing a weak positive linear correlation with a correlation coefficient of $r = 0,3$. The data points are widely scattered, with only a slight upward trend visible.

$r = 0,6$



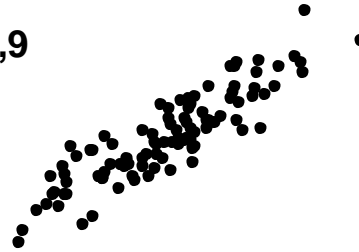
A scatter plot showing a moderate positive linear correlation with a correlation coefficient of $r = 0,6$. The data points are moderately clustered along a diagonal line from the bottom-left to the top-right.

Illustration

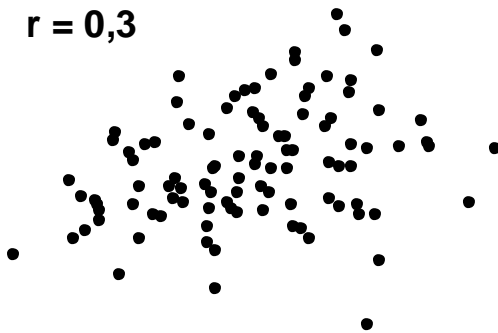
Introduction à la statistique avec R > La corrélation



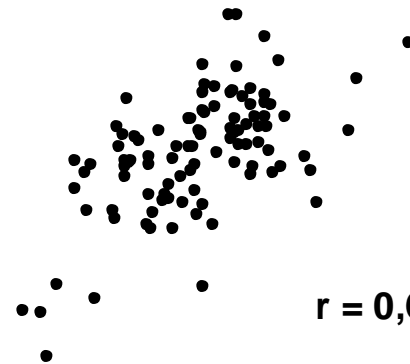
$r = 0,9$



$r = 0,3$



$r = 0,6$

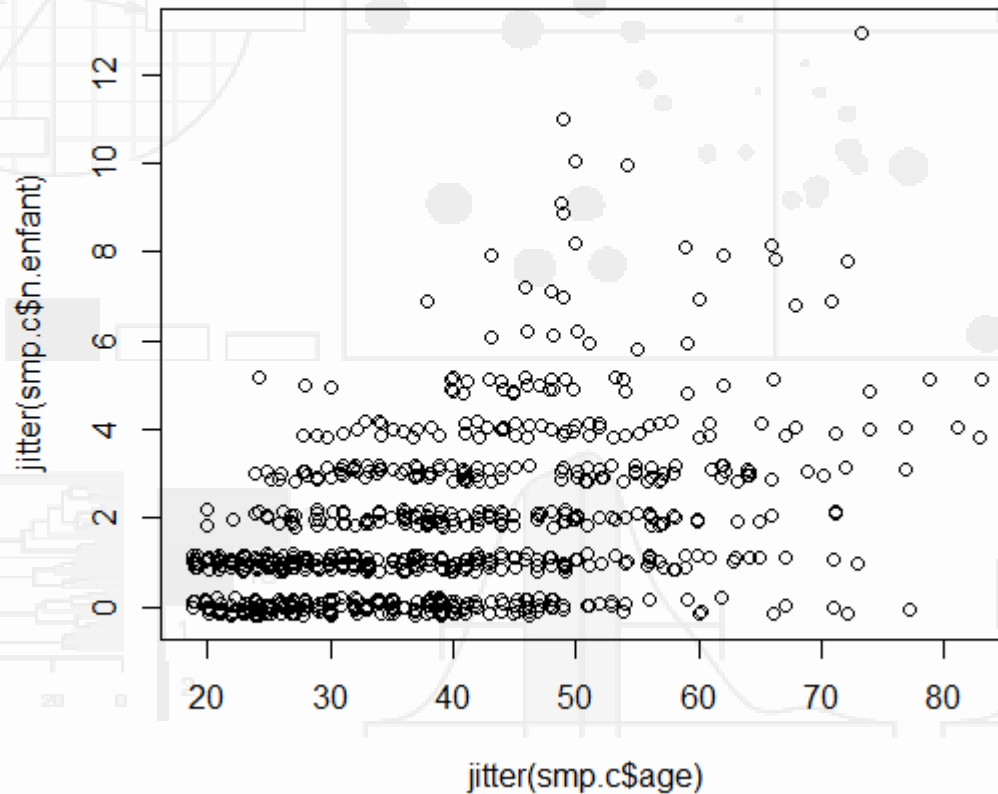


Exemple

Introduction à la statistique avec R > La corrélation



```
> plot(jitter(smp.c$age),  
       jitter(smp.c$n.enfant))
```



Exemple

Introduction à la statistique avec R > La corrélation



```
R
Fichier Edition Voir Misc Packages Fenêtres Aide
[Icons]
R R Console
> str(smp.c)
'data.frame': 799 obs. of 9 variables:
 $ age      : int  31 49 50 47 23 34 24 52 42 45 ...
 $ prof     : Factor w/ 7 levels "agriculteur",...: 3 NA 7 6
 $ dep.cons : int  0 0 0 0 1 0 1 0 1 0 ...
 $ scz.cons : int  0 0 0 0 0 0 0 0 0 0 ...
 $ grav.cons: int  1 2 2 1 2 1 5 1 5 5 ...
 $ n.enfant : int  2 7 2 0 1 3 5 2 1 2 ...
 $ rs       : int  2 2 2 2 2 1 3 2 3 2 ...
 $ ed       : int  1 2 3 2 2 2 3 2 3 2 ...
 $ dr       : int  1 1 2 2 2 1 2 2 1 2 ...
> cor(smp.c$age, smp.c$n.enfant, use="complete.obs")
[1] 0.4326039
> |
```


- Le coefficient de corrélation ne répond pas à toutes les questions
- Les relations quadratiques (en « U »)
- La concordance

Conclusion

Introduction à la statistique avec R > La corrélation



```
plot(jitter(smp.c$age), jitter(smp.c$n.enfant))  
cor(smp.c$age, smp.c$n.enfant, na.rm="complete.obs")
```