

# Formulaire de statistiques

## I. Statistiques descriptives :

**Moyenne arithmétique :** (population:  $m_x = \mu$ ) (échantillon =  $\bar{x} = M_x$ )

$$\frac{1}{n} \sum_{i=1}^n x_i$$

Excel FR =MOYENNE(série)  
Excel NL =GEMIDDELDE(série)  
Excel EN =AVERAGE(série)

**Somme des carrés des écarts :**

$$SCE = \sum_{i=1}^n (x_i - M_x)^2$$

Excel FR =SOMME.CARRES.ECARTS(série)  
Excel NL =DEV.KWAD (série)  
Excel EN =DEVSQ(série)

**Variance ou carré moyen des écarts d'un échantillon :**

$$S_{x;n}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - M_x)^2 = \frac{SCE}{n}$$

Excel FR =VAR.P(série)  
Excel NL, EN =VARP(série)

**Estimation de la variance d'une population :**

$$S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - M_x)^2 = \frac{SCE}{n-1}$$

Excel FR =VAR(série)  
Excel NL, EN =VAR(série)

**Écart-type de l'échantillon :**

$$S_{x;n} = \sqrt{S_{x;n}^2}$$

Excel FR =ECARTYPEP(série)  
Excel NL, EN =STDEVP(série)

**Écart-type de la population :**

$$S_{x;n-1} = \sqrt{S_{x;n-1}^2}$$

Excel FR =ECARTYPE(série)  
Excel NL, EN =STDEV(série)

**Coefficient de variation :**

$$CV = \frac{S_x}{M_x}$$

**Somme des produits des écarts :**

$$SPE = \sum_{i=1}^n [(x_i - M_x)(y_i - M_y)]$$

Excel FR =SOMME((zone des X-M<sub>x</sub>)\*(zone des Y-M<sub>y</sub>))  
Excel NL =SOM((zone des X-M<sub>x</sub>)\*(zone des Y-M<sub>y</sub>))  
Excel EN =SUM((zone des X-M<sub>x</sub>)\*(zone des Y-M<sub>y</sub>))  
Calcul matriciel : Mac: pomme+enter ; PC: ctrl+shift+enter

**Covariance ou produit moyen des écarts :**

$$S_{x,y} = \frac{SPE}{n}$$

Excel FR =COVARIANCE(série)  
Excel NL =COVARIANTIE(série)  
Excel EN =COVAR(série)

**Coefficient de détermination :** Note: formule Excel valable uniquement pour un modèle linéaire  $Y_i = B_0 + B_1 X_i$

$$R^2 = \frac{SCE_{régression}}{SCE_{totale}}$$

Excel FR =COEFFICIENT.DETERMINATION(série)  
Excel NL =R.KWADRAAT(série)  
Excel EN =SRQ(série)

**Coefficient de corrélation : cas particulier du modèle linéaire :  $r = \sqrt{R^2}$**

$$r = \frac{S_{x,y}}{S_x \cdot S_y} = \frac{SPE}{n \cdot S_x \cdot S_y}$$

Excel FR =COEFFICIENT.CORRELATION(série)  
Excel NL =CORRELATIE(série)  
Excel EN =CORREL(série)

**Droite des moindres carrés :  $Y_i = B_0 + B_1 \cdot X_i$**

$$B_1 = \frac{SPE}{SCE_x}$$

Excel FR =PENTE(série)  
Excel NL =RICHTING(série)  
Excel EN =SLOPE(série)

$$B_0 = M_y - B_1 \cdot M_x$$

Excel FR =ORDONNEE.ORIGINE(série)  
Excel NL =SNIJPUNT(série)  
Excel EN =INTERCEPT(série)

## II. Probabilités :

|                                       |   |
|---------------------------------------|---|
| <b>Loi des probabilités totales</b>   | $P(A \cup B) = P(A) + P(B) - P(A \cap B)$             |
| <b>Loi des probabilités composées</b> | $P(A \cap B) = P(A) \cdot P(B A) = P(B) \cdot P(A B)$ |
| <b>Recomposition de P(A)</b>          | $P(A) = P(A \cap B) + P(A \cap B^*)$                  |
| <b>Événements incompatibles</b>       | $P(A \cap B) = 0$                                     |
| <b>Événements indépendants</b>        | $P(A B) = P(A)$ ou $P(B A) = P(B)$                    |

## III. Les variables aléatoires discontinues :

| <b>Variable aléatoire Binomiale</b> X va Bi (n ; π) |  |
|---|--|
| <b>Espérance de X</b>                               | $E(X) = n \cdot \pi$   |
| <b>Variance de X</b>                                | $var(X) = n \cdot \pi \cdot (1 - \pi)$   |
| <b>Nombre de combinaisons</b>                       | $C_n^x = \frac{n!}{x!(n-x)!}$ Excel FR, EN =FACT(n)/(FACT(x)*FACT(n-x))<br>Excel NL =FACULTEIT(n)/(FACULTEIT(x)*FACULTEIT(n-x))  |
| <b>Probabilité</b>                                  | $P(X = x_i) = C_n^x \cdot \pi^x \cdot (1 - \pi)^{n-x}$ Excel FR =LOI.BINOMIALE(x;n;π;cumulatif)<br>Excel NL =BINOMIALE.VERD(x;n;π;cumulatif)<br>Excel EN =BINOMDIST(x;n;π;condition)<br>Pour P(X=xi) cumulatif= FR :FAUX; NL: VERVALSING; EN: FALSE.<br>Pour P(X≤xi) cumulatif= FR :VRAI; NL:WARE; EN: TRUE. |

| <b>Variable aléatoire de Poisson</b> X va Po (μ) |   |
|--|---|
| <b>Probabilité</b>                               | $P(X = x) = \frac{e^{-\mu} \cdot \mu^x}{x!}$ Excel FR =LOI.POISSON(x;μ;cumulatif)<br>Excel NL, EN =POISSON(x;μ;cumulatif)<br>Pour P(X=xi) cumulatif= FR :FAUX; NL: VERVALSING; EN: FALSE.<br>Pour P(X≤xi) cumulatif= FR :VRAI; NL:WARE; EN: TRUE. |
| <b>Espérances</b>                                | $E(X) = \sigma_x^2 = \mu$   |

### IV. Les variables aléatoires continues :

| Variable aléatoire Normale   |   |
|--|---|
| Réduction d'une variable lorsqu'elle représente un individu pour X v.a. N ( $\mu; \sigma^2$ ) en Z v.a. N (0;1)                                      | $Z = \frac{X - \mu}{\sigma}$                  |
| Réduction d'une variable lorsqu'elle représente la valeur moyenne d'un échantillon de n individus pour X va N( $\mu; \sigma^2/n$ ) en Z v.a. N (0;1) | $Z = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}}$ |

### V. Inférence statistique :

| Test $\alpha$ - $\beta$ |  |
|-------------------------|--|
| Estimation de n         | $n \geq \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \cdot \sigma^2}{(\mu_0 - \mu_1)^2}$ |

| A: comparaison d'une moyenne à un standard - test de conformité |  |  |
|---|--|--|
| $\sigma^2$ connue   | $Z_{obs} = \frac{M_{x\,obs} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}}$ | <b>Obtenir le <math>Z_{théorique}</math> :</b><br>Table de Student, pour $n = \infty$<br><b>ou</b><br>Excel FR =LOI.NORMALE.STANDARD.INVERSE(probabilité)<br>Excel NL =STAND.NORM.INV(probabilité)<br>Excel EN =NORMSINV(probabilité)<br>avec probabilité = $P(Z \leq Z_{théorique})$  |
| $\sigma^2$ inconnue<br>On prend donc $S^2$ comme estimateur     | $t_{obs} = \frac{M_{x\,obs} - \mu_{H_0}}{\frac{S}{\sqrt{n}}}$      | <b>Obtenir le <math>t_{théorique}</math> :</b><br>Table de Student, pour $k = n-1$ degrés de libertés (dl)<br><b>ou</b><br>Excel FR =LOI.STUDENT.INVERSE(probabilité;k)<br>Excel NL =T.INV(probabilité;k)<br>Excel EN =TINV(probabilité;k)<br>avec probabilité = $P(t \leq t_{théorique})$ et $k =$ degrés de libertés = $n-1$ |

| B: comparaison de variances issues d'échantillons indépendants  |  |
|---|--|
| <b>Pour deux variances :</b><br><br>$F_{obs} = \frac{S_{max}^2}{S_{min}^2}$   | <b>Calculer le <math>F_{obs}</math> :</b><br>Excel FR, NL, EN =MAX(série des variances)/MIN(série des variances)<br><b>Obtenir le <math>F_{théorique}</math> :</b><br>Table de Fisher, pour $k = (n-1)$ degrés de libertés (dl) de $S_{max}^2$ et $r = (n-1)$ dl de $S_{min}^2$<br><b>ou</b><br>Excel FR =INVERSE.LOIF(alpha;k;r)<br>Excel NL =F.INVERSE(alpha;k;r)<br>Excel EN =FINV(alpha;k;r)<br>avec alpha = 1 – confiance, en valeur décimale (ex: alpha = 0,05 si confiance à 95%) |
| <b>Pour deux variances ou plus, issues d'échantillons de même taille :</b><br><br>$H_{obs} = \frac{S_{max}^2}{S_{min}^2}$ | <b>Calculer le <math>H_{obs}</math> :</b><br>Excel FR, NL, EN =MAX(série des variances)/MIN(série des variances)<br><b>Obtenir le <math>H_{théorique}</math> :</b><br>Table de Hartley, pour $n_a$ variances comparées et $dl = (n_i-1)$<br>Pas de formule en Excel.   |

| <b>C : comparaison de proportions</b>  |  |
|--|--|
| $\chi^2_{obs} = \sum_{i=1}^k \left( \frac{(f_{i_{obs}} - f_{i_{théor}})^2}{f_{i_{théor}}} \right)$ | <p><b>Obtenir le <math>\chi^2_{théorique}</math> :</b><br/> <math>\chi^2_{théorique} = \chi^2_{(k-1) \cdot (r-1); 1-\alpha}</math><br/>                     Table de <math>\chi^2</math>, pour (k-1).(r-1) dl et une probabilité 1-<math>\alpha</math><br/> <b>ou</b><br/>                     Excel FR =KHIDEUX.INVERSE(alpha;dl)<br/>                     Excel NL =CHI.KWADRAAT.INV(alpha;dl)<br/>                     Excel EN =CHIINV(alpha;dl)<br/> <b>Attention</b> dans la formule Excel c'est alpha qu'il faut signaler !</p> |

## VI. ANOVA I :

| Source de variabilité | SCE  | dl                 | CM                   | F <sub>obs</sub>    | F <sub>table</sub>   |
|-----------------------|--|--------------------|----------------------|---------------------|--|
| <b>Totale</b>         | Excel FR =SOMME.CARRES.ECARTS(tous les individus)<br>Excel NL =DEV.KWAD(tous les individus)<br>Excel EN =DEVSQ(tous les individus)   | N - 1              |                      |                     |  |
| <b>Factorielle</b>    | Excel FR = n <sub>i</sub> * SOMME.CARRES.ECARTS(toutes les M)<br>Excel NL = n <sub>i</sub> *DEV.KWAD(toutes les M)<br>Excel EN = n <sub>i</sub> *DEVSQ(toutes les M)                                       | n <sub>a</sub> - 1 | $\frac{SCE_F}{dl_F}$ | $\frac{CM_F}{CM_R}$ | $F_{dl_F; dl_R; 1-\alpha}$<br><br>avec<br>k = dl <sub>F</sub><br>r = dl <sub>R</sub> |
| <b>Résiduelle</b>     | Excel FR = (n <sub>i</sub> -1) * SOMME(toutes les S <sup>2</sup> )<br>Excel NL = (n <sub>i</sub> -1) * SOM(toutes les S <sup>2</sup> )<br>Excel EN = (n <sub>i</sub> -1) * SUM(toutes les S <sup>2</sup> ) | N - n <sub>a</sub> | $\frac{SCE_R}{dl_R}$ |                     |  |

avec

*n<sub>i</sub>* = nombre d'individus par échantillon

*n<sub>a</sub>* = nombre d'échantillons

*N* = nombre total d'individus

Note: le F<sub>théorique</sub> peut aussi être trouvé avec l'instruction Excel suivante:

Excel FR =INVERSE.LOIF(alpha;dl<sub>F</sub>;dl<sub>R</sub>)

Excel NL =F.INVERSE(alpha;dl<sub>F</sub>;dl<sub>R</sub>)

Excel EN =FINV(alpha;dl<sub>F</sub>;dl<sub>R</sub>)

avec alpha = 1 - confiance, en valeur décimale (ex: alpha = 0,05 si confiance à 95%)

### Compléments ANOVA I aléatoire :

|                                      |  |  |
|--------------------------------------|--|--|
| <b>Variance entre échantillons</b>   | $\sigma_a^2 = \frac{E(CM_F) - E(CM_R)}{n_i}$                                 |  |
| <b>Variance entre répliques</b>      | $\sigma^2 = E(CM_R)$   |  |
| <b>Intervalle de confiance</b>       | formule simplifiée   | formule complète   |
|                                      | $M_x \pm t_{n_a-1; 1-\alpha/2} \cdot \sqrt{\frac{CM_F}{N}}$                  | $M_x \pm t_{n_a-1; 1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n_a \cdot n_i} + \frac{\sigma_a^2}{n_a}}$ |
| <b>Nombre de répliques optimum</b>   | $n_i = \sqrt{\frac{c_a \cdot \sigma^2}{c \cdot \sigma_a^2}}$                 |  |
| <b>Nombre d'échantillons optimum</b> | $n_a = \frac{16 \left( \frac{\sigma^2}{n_i} + \sigma_a^2 \right)}{\Delta^2}$ |  |

**VII. ANOVA I et Régression linéaire :**

| Source de variabilité | SCE                     | dl              | CM                             | F observé               | F table                        |
|-----------------------|-------------------------|-----------------|--------------------------------|-------------------------|--------------------------------|
| <b>Totale</b>         | <i>voir ANOVA 1</i>     |                 |                                |                         |                                |
| <b>Résiduelle</b>     |                         |                 |                                |                         |                                |
| <b>Factorielle</b>    |                         |                 |                                |                         |                                |
| <b>Linéaire</b>       | $= \frac{SPE^2}{SCE_x}$ | 1               | $= \frac{SCE_{lin}}{dl_{lin}}$ | $\frac{CM_{lin}}{CM_R}$ | $F_{dl_{lin}; dl_R; 1-\alpha}$ |
| <b>Non linéaire</b>   | $= SCE_F - SCE_{lin}$   | $n_F - n_{lin}$ | $= \frac{SCE_{NL}}{dl_{NL}}$   | $\frac{CM_{NL}}{CM_R}$  | $F_{dl_{NL}; dl_R; 1-\alpha}$  |

Avec

|                        |  |
|------------------------|--|
| <b>SPE</b>             | Excel FR =SOMME((zone des X-M <sub>x</sub> )*(zone des Y-M <sub>y</sub> ))<br>Excel NL =SOM((zone des X-M <sub>x</sub> )*(zone des Y-M <sub>y</sub> ))<br>Excel EN =SUM((zone des X-M <sub>x</sub> )*(zone des Y-M <sub>y</sub> ))<br>Calcul matriciel : Mac: pomme+enter ; PC: ctrl+shift+enter |
| <b>SCE<sub>x</sub></b> | Excel FR =n <sub>i</sub> * SOMME.CARRES.ECARTS(série de x)<br>Excel NL =n <sub>i</sub> * DEV.KWAD(série de x)<br>Excel EN =n <sub>i</sub> * DEVSQ(série de x)  |

**VIII. ANOVA II croisée fixe :** Où B et C sont les deux critères fixes

| Source de variabilité | SCE   | dl  | CM                         | F observé              | F table                       |
|-----------------------|---|---|----------------------------|------------------------|-------------------------------|
| <b>Totale</b>         | <i>voir ANOVA 1</i>   |   |                            |                        |                               |
| <b>Résiduelle</b>     |   |   |                            |                        |                               |
| <b>Factorielle</b>    |   |   |                            |                        |                               |
| <b>B</b>              | Excel FR =n <sub>i,echB</sub> * SOMME.CARRES.ECARTS(moyennes de B)<br>Excel NL =n <sub>i,echB</sub> * DEV.KWAD(moyennes de B)<br>Excel EN =n <sub>i,echB</sub> * DEVSQ(moyennes de B) | n <sub>catéB</sub> -1                             | $\frac{SCE_B}{dl_B}$       | $\frac{CM_B}{CM_R}$    | $F_{dl_B; dl_R; 1-\alpha}$    |
| <b>C</b>              | Excel FR =n <sub>i,echC</sub> * SOMME.CARRES.ECARTS(moyennes de C)<br>Excel NL =n <sub>i,echC</sub> * DEV.KWAD(moyennes de C)<br>Excel EN =n <sub>i,echC</sub> * DEVSQ(moyennes de C) | n <sub>catéC</sub> -1                             | $\frac{SCE_C}{dl_C}$       | $\frac{CM_C}{CM_R}$    | $F_{dl_C; dl_R; 1-\alpha}$    |
| <b>Interaction BC</b> | =SCE <sub>F</sub> -SCE <sub>B</sub> -SCE <sub>C</sub>   | dl <sub>F</sub> -dl <sub>B</sub> -dl <sub>C</sub> | $\frac{SCE_{BC}}{dl_{BC}}$ | $\frac{CM_{BC}}{CM_R}$ | $F_{dl_{BC}; dl_R; 1-\alpha}$ |

### IX. Contrastes de Scheffé :

| Contraste –Différence Observée  |   |
|---|---|
| $L = \left  \sum_{i=1}^{n_a} (c_i \cdot M_{x_i}) \right $   | Excel FR =ABS(SOMMEPROD(série des c ; série des moyennes))<br>Excel NL =ABS(SOMPRODUKT(série des c ; série des moyennes))<br>Excel EN =ABS(SUMPRODUCT(série des c ; série des moyennes))  |
| Plus petite différence significative: PPDS :  |   |
| $PPDS = \sqrt{\left[ F_{n_a-1; N-n_a; 1-\alpha} \cdot (n_a - 1) \cdot \frac{CM_R}{n_i} \cdot \sum_{i=1}^{n_a} c_i^2 \right]}$ |   |
| où $\sum_{i=1}^{n_a} c_i^2$ se calcule avec   | Excel FR =SOMME.CARRES(série des c)<br>Excel NL =KWADRATENSOM(série des c)<br>Excel EN =SUMSQ(série des c)  |
| PPDS =  | Excel FR =RACINE(INVERSE.LOIF(alpha,dl <sub>F</sub> ,dl <sub>R</sub> )*(na-1)*(CMR/ni)*SOMME.CARRES(série des c))<br>Excel NL =WORTEL(F.INVERSE(alpha,dl <sub>F</sub> ,dl <sub>R</sub> )*(na-1)*(CMR/ni)*KWADRATENSOM(série des c))<br>Excel EN =SQRT(FINV(alpha,dl <sub>F</sub> ,dl <sub>R</sub> )*(na-1)*(CMR/ni)*SUMSQ(série des c))<br>avec dl <sub>F</sub> = na-1 et dl <sub>R</sub> =N-na |

**X. ANOVA : contrastes et modèles**

Écrire le modèle en notant toutes les sources de variabilité.

Critères fixes en minuscules, critères aléatoires en majuscules et hiérarchisation entre parenthèses.

$$X_{(ijk)l} = \mu + a_i + B_{(i)j} + c_k + ac_{ik} + Bc_{(i)jk} + E_{(ijk)l}$$

*Règle 1 : Construction d'une table avec autant de lignes qu'il y a de sources de variabilité.*

*Adjonction d'une colonne à gauche des sources de variabilité.*

$\delta^2$  : pour les critères fixes

$\sigma^2$  : pour les critères aléatoires

*Adjonction d'autant de colonnes à droite des sources de variabilité qu'il y a d'indices (de facteurs) dans le modèle (i, j, k...).*

Si l'indice :

- n'est pas repris dans le membre en tête de ligne : mettre la valeur de l'indice en question
- est repris dans le membre en tête de ligne, se rapporte à un critère fixe (minuscule) et n'est pas entre parenthèses : mettre la valeur 0.
- ne répond pas aux conditions précédentes : mettre la valeur 1.

***Règle 2 :*** Adjonction d'autant de colonnes à droite des indices qu'il y a de sources de variabilité dans le modèle.

Considérer uniquement les sources de variabilité qui ont au moins tous les indices repris en tête de ligne.

***Règle 3 :*** Pondération des sources de variabilité introduites dans la table à l'étape précédente.

Masquer les colonnes correspondant aux indices qui ne sont pas entre parenthèses  
Pondérer chaque terme par le produit des indices non masqués.

***Règle 4 :*** Détermination des degrés de liberté de chaque v.a. de Fisher

Effectuer le produit de la valeur maximale de tous les indices représentés en tête de ligne, après avoir retiré 1 à ceux qui ne sont pas entre parenthèses.

***Conclusions***