Propriétés d'une probabilité

- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$ et $P(\mathcal{S}) = 1$
- Si $A_1, A_2, A_3, \ldots$ sont mutuellement exclusifs, alors $P(A_1 \cup A_2 \cup A_3 \cdots) = \sum_i P(A_i)$.
- $P(A) = 1 - P\left(\overline{A}\right)$
- Si $A \subset B$, alors $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P\left(A \cup B \cup C\right) = P(A) + P(B) + P(C) - P\left(A \cap B\right) - P\left(B \cap C\right) - P\left(A \cap C\right) + P\left(A \cap B \cap C\right)$
- Si les n résultats de $\mathcal{S}$ sont équiprobables, alors $P(A) = n(A)/n$.

Techniques de dénombrement

- *Permutations* :
  - il existe $n!$ arrangements ordonnés de n éléments distincts ;
  - il existe $P_r^n = \dfrac{n!}{(n-r)!}$ façons de choisir r objets parmi n objets distincts en tenant compte de l'ordre de pige.
  - il existe $\dfrac{n!}{n_1! n_2! \cdots n_k!}$ façons de permuter n objets séparés en k groupes formés respectivement de $n_1, n_2, ..., n_k$ objets indiscernables.
- *Combinaisons* :
  - il existe $C_r^n = \dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$ façons de choisir r objets parmi n objets distincts sans tenir compte de l'ordre de pige.

Probabilités conditionnelles

- $P\left(A \mid B\right) = \dfrac{P\left(A \cap B\right)}{P(B)}$
- $P\left(A \cap B\right) \ = \ P(A)\, P\left(B \mid A\right) \ = \ P(B)\, P\left(A \mid B\right)$
- A et B sont indépendants si et seulement si $P(A \cap B) = P(A)\, P(B)$
- Si $B_1, B_2, \ldots, B_k$ forment une partition de $\mathcal{S}$, alors
  - $P(A) = \displaystyle\sum_{i=1}^{k} P\left(A \mid B_i\right)\ P\left(B_i\right)$ (théorème des probabilités totales)
  - $P\left(B_r \mid A\right) = \dfrac{P\left(A \mid B_r\right)\ P\left(B_r\right)}{P(A)} = \dfrac{P\left(A \mid B_r\right)\ P\left(B_r\right)}{\displaystyle\sum_{i=1}^{k} P\left(A \mid B_i\right)\ P\left(B_i\right)}$ (théorème de Bayes)

Variables aléatoires et distributions

Notion abordée	**Cas d'une variable discrète**	**Cas d'une variable continue**
Fonction de répartition	$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
Fonction de masse/densité	$p(x) = P(X = x)$	$f(x) = \frac{d}{dx}F(x)$
Probabilité $P(a < X \leq b)$	$F(b) - F(a) = \sum_{a<x\leq b} p(x)$	$F(b) - F(a) = \int_{a}^{b} f(x)dx$
Espérance de X : $E(X)$	$\mu = \sum_{i} x_i\, p(x_i)$	$\mu = \int_{-\infty}^{\infty} x\, f(x)dx$
Variance de X : $V(X)$	$\sigma^2 = \sum_{i} (x_i - \mu)^2\, p(x_i)$	$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)dx$

Notons que $V(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$.

Notion abordée	**Cas d'une variable discrète**	**Cas d'une variable continue**
Distribution conjointe	$p(x_1, x_2) = P(X_1 = x_1 \text{ et } X_2 = x_2)$	$f(x_1, x_2)$
Distribution marginale de X_1	$p_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2} p(x_1, x_2)$	$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2$
Dist. conditionnelle de X_1 sachant que $X_2 = b$	$p_{X_1\|b}(x_1) = P(X_1 = x_1 \| X_2 = b) = \frac{p(x_1, b)}{p_{X_2}(b)}$	$f_{X_1\|b}(x_1) = \frac{f(x_1, b)}{f_{X_2}(b)}$
Espérance conditionnelle de X_1 sachant que $X_2 = b$	$E(X_1\|b) = \sum_{x_1} x_1 p_{X_1\|b}(x_1)$	$E(X_1\|b) = \int_{-\infty}^{\infty} x_1 f_{X_1\|b}(x_1)dx_1$

- **Indépendance de deux variables aléatoires**
  - X_1 et X_2 sont indépendantes si et seulement si

$$\begin{aligned} p(x_1, x_2) &= p_{X_1}(x_1)p_{X_2}(x_2) \quad \text{pour tout } (x_1, x_2) \quad \text{(cas discret)} \\ f(x_1, x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \quad \text{pour tout } (x_1, x_2) \quad \text{(cas continu)} \end{aligned}$$

  - Si X_1 et X_2 sont indépendantes, alors leur covariance et leur corrélation sont nulles.

- **Covariance et corrélation**

  Covariance : $\text{Cov}(X_1, X_2) = E\left[(X_1 - \mu_1)\,(X_2 - \mu_2)\right] = E(X_1 X_2) - E(X_1)\,E(X_2)$

  Corrélation : $\rho = \dfrac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}}$

Lois discrètes usuelles

Lois discrètes	Fonction de masse	Espérance	Variance
Bernoulli (p)	$p(k) = p^k(1-p)^{1-k}, \quad \text{pour} \quad k = 0, 1$	p	$p(1-p)$
Binomiale (n, p)	$p(k) = \binom{n}{k} p^k(1-p)^{n-k}, \quad \text{pour} \quad k = 0, 1, \ldots, n$	np	$np(1-p)$
Géométrique (p)	$p(k) = (1-p)^{k-1}p, \quad \text{pour} \quad k = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Pascal (r, p)	$p(k) = \binom{k-1}{r-1} p^r(1-p)^{k-r}, \quad \text{pour} \quad k = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson (λ)	$p(k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad \text{pour} \quad k = 0, 1, 2, \ldots$	λ	λ

Lois continues usuelles

Lois discrètes	Fonction de masse	Espérance	Variance
Uniforme (α, β)	$f(x) = \frac{1}{\beta - \alpha}, \quad \text{pour} \quad \alpha \leq x \leq \beta$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - alpha)^2}{12}$
Exponentielle (λ)	$f(x) = \lambda \mathrm{e}^{-\lambda x}, \quad \text{pour} \quad x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma (r, λ)	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} \mathrm{e}^{-\lambda x}, \quad \text{pour} \quad x > 0$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$
Normale (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \text{pour} \quad -\infty < x < \infty$	μ	σ^2

Loi normale univariée : $X \sim N(\mu, \sigma^2)$

- Fonction de densité : $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$, pour $-\infty < x < \infty$
- Si $X \sim N(\mu, \sigma^2)$, alors $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$ (loi normale centrée réduite)

Théorème central limite

Si $X_1, X_2, ..., X_n$ sont des variables aléatoires
$\rightarrow$ indépendantes,
$\rightarrow$ identiquement distribuées (loi quelconque),
$\rightarrow$ de moyenne $\mu < \infty$,
$\rightarrow$ de variance $\sigma^2 < \infty$,
$\rightarrow$ et si n est assez grand,

alors

$$\sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2)$$

et

$$\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right).$$

- Si les X_i sont des variables aléatoires discrètes, on utilise une correction pour la continuité.

Loi normale bivariée : $(X_1,\ X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

- Fonction de densité conjointe :
  $f(x_1,x_2) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\},$
  pour $-\infty < x_1, x_2 < \infty$.

- Lois marginales :
  - $X_1 \sim N(\mu_1, \sigma_1^2)$ et $X_2 \sim N(\mu_2, \sigma_2^2)$

- Lois conditionnelles :
  - $X_1|x_2 \sim N\left(\mu_1 + \rho\,\sigma_1\left(\dfrac{x_2-\mu_2}{\sigma_2}\right),\ \sigma_1^2(1-\rho^2)\right)$
  - $X_2|x_1 \sim N\left(\mu_2 + \rho\,\sigma_2\left(\dfrac{x_1-\mu_1}{\sigma_1}\right),\ \sigma_2^2(1-\rho^2)\right)$

Statistique descriptive

- **Valeurs observées de la variable X**
  — Dans l'ordre de collecte : $x_1,\ x_2,\ ...,\ x_n$
  — En ordre croissant : $x_{(1)},\ x_{(2)}, ...,\ x_{(n)}$
- **Mesures de position**

Statistique	**Symbole**	**Formule**
Moyenne	$\overline{x}$	$\dfrac{1}{n}\sum_{i=1}^{n} x_i$
Médiane	$\widetilde{x} = Q2$	$q_{0,5} = x_{\left(\frac{n+1}{2}\right)}$
Mode	m	valeur la plus fréquente
Quantile d'ordre γ	q_γ	$x_{(\gamma\, n+0,5)}$
Premier quartile	$Q1$	$q_{0,25} = x_{(0,25\, n+0,5)}$
Troisième quartile	$Q3$	$q_{0,75} = x_{(0,5\, n+0,5)}$

- **Mesures de dispersion**

Statistique	**Symbole**	**Formule**
Variance	s^2	$\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
Écart-type	s	$\sqrt{s^2}$
Étendue	R	$x_{(n)} - x_{(1)}$
Coefficient de variation	CV	$\dfrac{s}{\overline{x}}$
Écart interquartile	EIQ	$Q3 - Q1$

Distributions échantillonnales

Nom de la loi	Notation	Espérance	Variance	Quantile* d'ordre $1-\alpha$	Exemple de statistique**
Normale centrée réduite	$N(0,1)$	0	1	z_α	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
Khi-carré avec k degrés de liberté	χ^2_k	k	$2k$	$\chi^2_{\alpha;k}$	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
Student avec k degrés de liberté	t_k	0	$\frac{k}{k-2}$	$t_{\alpha;k}$	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
Fisher avec u et v degrés de liberté	$F_{u,v}$	$\frac{v}{v-2}$	$\frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$	$F_{\alpha;u,v}$	$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$

* Quantile d'ordre $1-\alpha$: $P(\text{Variable}>\text{Quantile}) = \alpha$
** Basé sur des échantillons provenant d'une distribution $N(\mu,\, \sigma^2)$

Inférence à un échantillon :

Conditions	Paramètre	Estimation	Err.-type estimée	Intervalle de confiance de niveau $1-\alpha$	Distribution de la statistique de test	Région de rejet au seuil α
Échantillon aléatoire de taille n issu d'une loi $N(\mu,\sigma^2)$	μ	$\overline{x} = \frac{1}{n}\sum x_i$	$\frac{s}{\sqrt{n}}$	$\overline{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$H_0 : \mu = \mu_o$ vs $H_1 : \mu > \mu_o$ Rejet de H_0 si $\frac{\overline{x}-\mu_o}{s/\sqrt{n}} \geq t_{n-1,\alpha}$
Échantillon aléatoire de taille n issu d'une loi $N(\mu,\sigma^2)$	σ^2	$s^2 = \frac{\sum(x_i-\overline{x})^2}{n-1}$	$\frac{\sqrt{2}\, s^2}{\sqrt{n-1}}$	$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}},\, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right]$	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$H_0 : \sigma^2 = \sigma^2_o$ vs $H_1 : \sigma^2 > \sigma^2_o$ Rejet de H_0 si $\frac{(n-1)s^2}{\sigma^2_o} \geq \chi^2_{n-1,\alpha}$
Échantillon aléatoire de grande taille n issu d'une loi dichotomique	p	$\hat{p} = \frac{\text{nb de cas}}{n}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$	$H_0 : p = p_o$ vs $H_1 : p > p_o$ Rejet de H_0 si $\frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} \geq z_\alpha$

Tests d'hypothèses

$$
\begin{aligned}
\alpha &= P\left(\text{Rejeter } H_0 \mid H_0 \text{ est vraie}\right) &&= P\left(\text{erreur de } 1^{re} \text{ espèce}\right) = \text{Seuil du test} \\
\beta &= P\left(\text{Ne pas rejeter } H_0 \mid H_1 \text{ est vraie}\right) &&= P\left(\text{erreur de } 2^{e} \text{ espèce}\right) \\
1-\beta &= P\left(\text{Rejeter } H_0 \mid H_1 \text{ est vraie}\right) &&= \text{Puissance du test}
\end{aligned}
$$

Seuil observé (valeur P) : probabilité que la statistique du test prenne une valeur au moins aussi extrême que celle observée, si H_0 est vraie. (On rejette H_0 si la valeur P est inférieure à α.)

Inférence à deux échantillons indépendants :

Échantillons aléatoires indépendants de tailles n_1 et n_2 issus de lois...	Paramètre	Estimation	Erreur-type estimée	Intervalle de confiance de niveau $1-\alpha$	Distribution de la statistique de test	Région de rejet au seuil α
$N(\mu_1,\sigma^2)$ et $N(\mu_2,\sigma^2)$ (Variances égales)	$\mu_1-\mu_2$	$\overline{x}_1-\overline{x}_2$	$s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ $s_c^2=\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	$(\overline{x}_1-\overline{x}_2)\pm t_{k,\alpha/2}s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ $k=n_1+n_2-2$	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_k$ $k=n_1+n_2-2$	$H_0:\mu_1=\mu_2$ vs $H_1:\mu_1>\mu_2$ Rejet de H_0 si $\frac{\overline{x}_1-\overline{x}_2}{s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\geq t_{k,\alpha}$ (Test de Student)
$N(\mu_1,\sigma_1^2)$ et $N(\mu_2,\sigma_2^2)$ (Variances inégales)	$\mu_1-\mu_2$	$\overline{x}_1-\overline{x}_2$	$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	$(\overline{x}_1-\overline{x}_2)\pm t_{k,\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $k=\frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$	$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim t_k$ $k=\frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$	$H_0:\mu_1=\mu_2$ vs $H_1:\mu_1>\mu_2$ Rejet de H_0 si $\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\geq t_{k,\alpha}$ (Test de Welch)
$N(\mu_1,\sigma_1^2)$ et $N(\mu_2,\sigma_2^2)$	$\dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{S_1^2}{S_2^2}$	à omettre	$\left[\dfrac{1}{F_{k,\ell,\frac{\alpha}{2}}}\dfrac{s_1^2}{s_2^2},\ \dfrac{1}{F_{k,\ell,1-\frac{\alpha}{2}}}\dfrac{s_1^2}{s_2^2}\right]$ $k=n_1-1$ et $\ell=n_2-1$	$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F_{k,\ell}$ $k=n_1-1$ et $\ell=n_2-1$	$H_0:\sigma_1^2=\sigma_2^2$ vs $H_1:\sigma_1^2>\sigma_2^2$ Rejet de H_0 si $\frac{s_1^2}{s_2^2}\geq F_{k,\ell,\alpha}$ (Test de Fisher)
Binomiale$(1,p_1)$ et Binomiale$(1,p_2)$ (n_1 et n_2 grands)	p_1-p_2	$\hat{p}_1-\hat{p}_2$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\hat{p}=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$	$H_0:p_1=p_2$ vs $H_1:p_1>p_2$ Rejet de H_0 si $z_{obs}\geq z_\alpha$

Plan d'expérience complètement aléatoire à un facteur

a échantillons indépendants de taille n (donc $N = an$ observations), variances égales

Modèle : $Y_{ij} \;=\; \mu + \tau_i + E_{ij}$, où $E_{ij} \sim N(0, \sigma^2)$, $(i = 1, ..., a,\ j = 1, ..., n)$

Analyse de la variance : $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ (ou $H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$)

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique F_0	Valeur P
Traitements	$SS_{traitements}$	$a-1$	$MS_{traitements} = \frac{SS_{traitements}}{a-1}$	$f_0 = \frac{MS_{traitements}}{MS_E}$	$P(F > f_0)$
Erreur	SS_E	$N-a$	$MS_E = \frac{SS_E}{N-a}$		où $F \sim F_{a-1,N-a}$
Total	SS_T	$N-1$			

- $SS_{traitements} = \sum_{i=1}^{a} n(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a} \frac{Y_{i\bullet}^2}{n} - \frac{Y_{\bullet\bullet}^2}{N}$
- $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\bullet})^2 = (n-1)\sum_{i=1}^{a} S_i^2$
- $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$
- $e_{ij} = Y_{ij} - \overline{Y}_{i\bullet} =$ résidu

Plan d'expérience en blocs aléatoires complets

b blocs de a unités expérimentales recevant chacune un des a traitements (donc ab observations)

Modèle : $Y_{ij} \;=\; \mu + \tau_i + \beta_j + E_{ij}$, où $\beta_j \sim N(0, \sigma_\beta^2)$ et $E_{ij} \sim N(0, \sigma^2)$, $(i = 1, ..., a,\ j = 1, ..., b)$

Analyse de la variance : $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique F_0	Valeur P
Traitements	$SS_{traitements}$	$a-1$	$MS_{traitements} = \frac{SS_{traitements}}{a-1}$	$f_0 = \frac{MS_{traitements}}{MS_E}$	$P(F > f_0)$
Blocs	SS_{blocs}	$b-1$	$MS_{blocs} = \frac{SS_{blocs}}{b-1}$		où $F \sim F_{a-1,(a-1)(b-1)}$
Erreur	SS_E	$(a-1)(b-1)$	$MS_E = \frac{SS_E}{(a-1)(b-1)}$		
Total	SS_T	$ab-1$			

- $SS_{traitements} = \sum_{i=1}^{a} b(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a} \frac{Y_{i\bullet}^2}{b} - \frac{Y_{\bullet\bullet}^2}{ab}$
- $SS_{blocs} = a\sum_{j=1}^{b}(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 = \sum_{j=1}^{b} \frac{Y_{\bullet j}^2}{a} - \frac{Y_{\bullet\bullet}^2}{ab}$
- $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b} e_{ij}^2 = SS_T - SS_{traitements} - SS_{blocs}$
- $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$
- $e_{ij} = Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet} =$ résidu

Plan d'expérience complètement aléatoire à deux facteurs

$a \times b$ échantillons indépendants de taille n (donc abn observations), variances égales

Modèle : $Y_{ijk} \; = \; \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + E_{ijk}$, où $E_{ijk} \sim N(0, \sigma^2)$, $(i = 1, ..., a, \; j = 1, ..., b, \; k = 1, ..., n)$

Analyse de la variance : 1)H_0 : $(\tau\beta)_{ij} = 0$ pour tout (i, j)
2)H_0 : $\tau_1 = \tau_2 = \cdots = \tau_a = 0$
H_0 : $\beta_1 = \beta_2 = \cdots = \beta_b = 0$

Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique F_0	Valeur P
Facteur A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$f_{0A} = \frac{MS_A}{MS_E}$	$P(F_{a-1,ab(n-1)} > f_{0A})$
Facteur B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$f_{0B} = \frac{MS_B}{MS_E}$	$P(F_{b-1,ab(n-1)} > f_{0B})$
Interaction AB	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$f_{0AB} = \frac{MS_{AB}}{MS_E}$	$P(F_{(a-1)(b-1),ab(n-1)} > f_{0AB})$
Erreur	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$		
Total	SS_T	$abn-1$			

- $SS_A = bn \sum_{i=1}^{a} (\overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^{a} \frac{Y_{i\bullet\bullet}^2}{bn} - \frac{Y_{\bullet\bullet\bullet}^2}{abn}$
- $SS_B = an \sum_{j=1}^{b} (\overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet})^2 = \sum_{j=1}^{b} \frac{Y_{\bullet j\bullet}^2}{an} - \frac{Y_{\bullet\bullet\bullet}^2}{abn}$
- $SS_E = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} e_{ijk}^2 = SS_T - SS_{AB} - SS_A - SS_B$
- $SS_{AB} = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{\bullet\bullet\bullet})^2$
- $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \overline{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}^2 - \frac{Y_{\bullet\bullet\bullet}^2}{abn}$
- $e_{ijk} = Y_{ijk} - \overline{Y}_{ij\bullet}$ = résidu

Régression linéaire simple

Modèle : $Y_i = \beta_0 + \beta_1 X_i + E_i \qquad (i = 1, ..., n)$

où les Y_i sont indépendantes et suivent une loi $N(\beta_0 + \beta_1 X_i, \sigma^2)$.

Quantité à estimer	Estimateur	Intervalle de niveau $1-\alpha$
Pente β_1	$\hat\beta_1 = \dfrac{S_{XY}}{S_{XX}}$	$\hat\beta_1 \pm t_{\alpha/2;\,n-2}\sqrt{\dfrac{MS_E}{S_{XX}}}$
Ordonnée à l'origine β_0	$\hat\beta_0 = \overline{Y} - \hat\beta_1 \overline{X}$	$\hat\beta_0 \pm t_{\alpha/2;\,n-2}\sqrt{MS_E\left(\dfrac{1}{n} + \dfrac{\overline{X}^2}{S_{XX}}\right)}$
Variance σ^2	$\hat\sigma^2 = MS_E$	
Moyenne $E(Y\|x_0)$	$\hat Y_0 = \hat\beta_0 + \hat\beta_1 x_0$	$\hat Y_0 \pm t_{\alpha/2;\,n-2}\sqrt{MS_E\left(\dfrac{1}{n} + \dfrac{(x_0-\overline{X})^2}{S_{XX}}\right)}$
Observation future Y_0	$\hat Y_0 = \hat\beta_0 + \hat\beta_1 x_0$	$\hat Y_0 \pm t_{\alpha/2;\,n-2}\sqrt{MS_E\left(1 + \dfrac{1}{n} + \dfrac{(x_0-\overline{X})^2}{S_{XX}}\right)}$
Corrélation ρ	$r = \dfrac{S_{XY}}{(S_{XX}S_{YY})^{1/2}}$	$\left[\tanh\left(\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right), \tanh\left(\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)\right]$

Test de signification de la régression : $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

Source de variation	Sommes de carrés	Degrés de liberté	Moyenne des carrés	Statistique F_0	On rejette H_0 si...
Régression	SS_R	1	$MS_R = SS_R/1$	$F_0 = \dfrac{MS_R}{MS_E}$	$f_0 > F_{\alpha;1,n-2}$
Erreur	SS_E	$n-2$	$MS_E = SS_E/(n-2)$		
Total	SS_T	$n-1$			

- $S_{XY} = \sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^n X_i Y_i - n\,\overline{X}\,\overline{Y}$
- $S_{XX} = \sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - n\,\overline{X}^{\,2}$
- $S_{YY} = \sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^n Y_i^2 - n\,\overline{Y}^{\,2}$
- $R^2 = \dfrac{SS_R}{SS_T} = \hat\beta_1^2 \dfrac{S_{XX}}{S_{YY}} =$ coefficient de détermination
- $SS_R = \sum_{i=1}^n (\hat Y_i - \overline{Y})^2 = \hat\beta_1 S_{XY}$
- $SS_E = \sum_{i=1}^n (Y_i - \hat Y_i)^2 = \sum_{i=1}^n e_i^2$
- $SS_T = S_{YY}$
- $e_i = Y_i - \hat Y_i = Y_i - (\hat\beta_0 + \hat\beta_1\, X_i)$

Régression linéaire multiple

Modèle : $Y_i \quad = \quad \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + E_i \qquad (i = 1, ..., n)$

où les E_i sont indépendantes et suivent une loi $N(0, \sigma^2)$.

Forme matricielle : $\boldsymbol{Y} \quad = \quad \boldsymbol{X\beta} + \boldsymbol{E}$

Quantité à estimer	Estimateur	Intervalle de niveau $1-\alpha$
Variance σ^2	$\hat{\sigma}^2 = MS_E$	
Vecteur $\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$	
Coefficient β_j	$\hat{\beta}_j$	$\hat{\beta}_j \pm t_{\alpha/2;\, n-k-1}\sqrt{MS_E \cdot C_{jj}}$
Moyenne $E(Y\|\boldsymbol{x_0})$	$\hat{Y}_0 = \boldsymbol{x_0'\beta}$	$\hat{Y}_0 \pm t_{\alpha/2;\, n-k-1}\sqrt{MS_E\,(\boldsymbol{x_0'}(\boldsymbol{X'X})^{-1}\boldsymbol{x_0})}$
Observation future Y_0	$\hat{Y}_0 = \boldsymbol{x_0'\beta}$	$\hat{Y}_0 \pm t_{\alpha/2;\, n-k-1}\sqrt{MS_E\,(1+\boldsymbol{x_0'}(\boldsymbol{X'X})^{-1}\boldsymbol{x_0})}$

Test de signification de la régression : $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$
$H_1 : \beta_j \neq 0$ pour au moins un j

Source de variation	Sommes de carrés	Degrés de liberté	Moyenne des carrés	Statistique F_0	On rejette H_0 si...
Régression	SS_R	k	$MS_R = SS_R/k$	$F_0 = \dfrac{MS_R}{MS_E}$	$f_0 > F_{\alpha;k,n-k-1}$
Erreur	SS_E	$n-k-1$	$MS_E = SS_E/n-k-1$		
Total	SS_T	$n-1$			

- C_{jj} = élément (j,j) de $(\boldsymbol{X'X})^{-1}$
- $R^2 = \dfrac{SS_R}{SS_T} = 1 - \dfrac{SS_E}{SS_T}$
- $R^2_{aj} = 1 - \dfrac{SS_E/n-k-1}{SS_T/n-1}$
- $SS_R = \hat{\boldsymbol{\beta}}'\boldsymbol{X'Y} - \dfrac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$
- $SS_E = \boldsymbol{E'E} = \boldsymbol{Y'Y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X'Y}$
- $SS_T = \boldsymbol{Y'Y} - \dfrac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$