Pleuvra-t-il demain en Australie ?

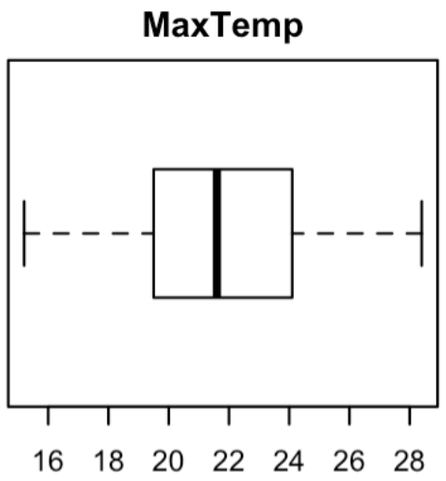
Loan Godard
Élève ingénieur à l’ENSIIE

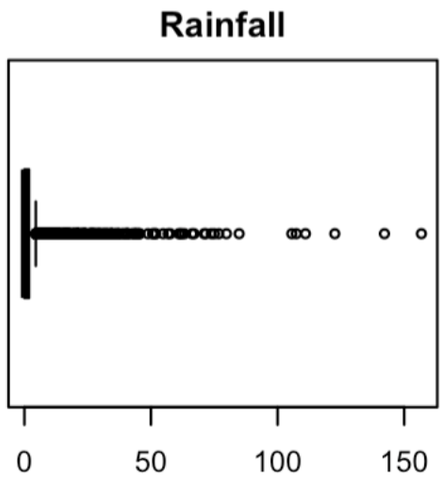
Les données

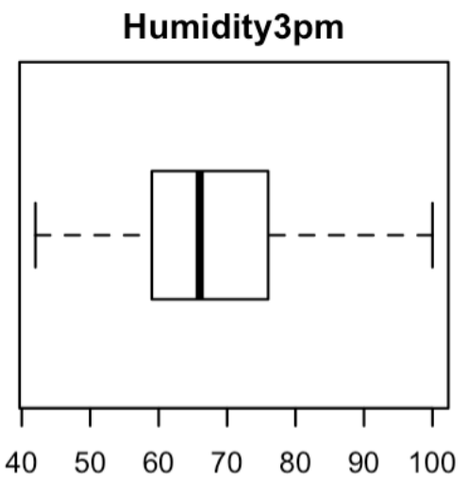
Date	Location	MinTemp	MaxTemp	Rainfall	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
2008-12-01	Albury	13.4	22.9	0.6	W	44	W	WNW	20	24	71	22	1007.7	1007.1	16.9	21.8	0	0
2008-12-02	Albury	7.4	25.1	0.0	WNW	44	NNW	WSW	4	22	44	25	1010.6	1007.8	17.2	24.3	0	0
2008-12-03	Albury	12.9	25.7	0.0	WSW	46	W	WSW	19	26	38	30	1007.6	1008.7	21.0	23.2	0	0
2008-12-04	Albury	9.2	28.0	0.0	NE	24	SE	E	11	9	45	16	1017.6	1012.8	18.1	26.5	0	0
2008-12-05	Albury	17.5	32.3	1.0	W	41	ENE	NW	7	20	82	33	1010.8	1006.0	17.8	29.7	0	0
2008-12-06	Albury	14.6	29.7	0.2	WNW	56	W	W	19	24	55	23	1009.2	1005.4	20.6	28.9	0	0
2008-12-07	Albury	14.3	25.0	0.0	W	50	SW	W	20	24	49	19	1009.6	1008.2	18.1	24.6	0	0
2008-12-08	Albury	7.7	26.7	0.0	W	35	SSE	W	6	17	48	19	1013.4	1010.1	16.3	25.5	0	0
2008-12-09	Albury	9.7	31.9	0.0	NNW	80	SE	NW	7	28	42	9	1008.9	1003.6	18.3	30.2	0	1
2008-12-10	Albury	13.1	30.1	1.4	W	28	S	SSE	15	11	58	27	1007.0	1005.7	20.1	28.2	1	0
2008-12-11	Albury	13.4	30.4	0.0	N	30	SSE	ESE	17	6	48	22	1011.8	1008.7	20.4	28.8	0	1
2008-12-12	Albury	15.9	21.7	2.2	NNE	31	NE	ENE	15	13	89	91	1010.5	1004.2	15.9	17.0	1	1
2008-12-13	Albury	15.9	18.6	15.6	W	61	NNW	NNW	28	28	76	93	994.3	993.0	17.4	15.8	1	1
2008-12-14	Albury	12.6	21.0	3.6	SW	44	W	SSW	24	20	65	43	1001.2	1001.8	15.8	19.8	1	0
2008-12-16	Albury	9.8	27.7	*NA*	WNW	50	*NA*	WNW	*NA*	22	50	28	1013.4	1010.3	17.3	26.2	0	0
2008-12-17	Albury	14.1	20.9	0.0	ENE	22	SSW	E	11	9	69	82	1012.2	1010.4	17.2	18.1	0	1
2008-12-18	Albury	13.5	22.9	16.8	W	63	N	WNW	6	20	80	65	1005.8	1002.2	18.0	21.5	1	1
2008-12-19	Albury	11.2	22.5	10.6	SSE	43	WSW	SW	24	17	47	32	1009.4	1009.7	15.5	21.0	1	0
2008-12-20	Albury	9.8	25.6	0.0	SSE	26	SE	NNW	17	6	45	26	1019.2	1017.1	15.8	23.2	0	0
2008-12-21	Albury	11.5	29.3	0.0	S	24	SE	SE	9	9	56	28	1019.3	1014.8	19.1	27.3	0	0
2008-12-22	Albury	17.1	33.0	0.0	NE	43	NE	N	17	22	38	28	1013.6	1008.1	24.5	31.6	0	0
2008-12-23	Albury	20.5	31.8	0.0	WNW	41	W	W	19	20	54	24	1007.8	1005.7	23.8	30.8	0	0
2008-12-24	Albury	15.3	30.9	0.0	N	33	ESE	NW	6	13	55	23	1011.0	1008.2	20.9	29.0	0	0
2008-12-25	Albury	12.6	32.4	0.0	W	43	E	W	4	19	49	17	1012.9	1010.1	21.5	31.2	0	0
2008-12-26	Albury	16.2	33.9	0.0	WSW	35	SE	WSW	9	13	45	19	1010.9	1007.6	23.2	33.0	0	0
2008-12-27	Albury	16.9	33.0	0.0	WSW	57	*NA*	W	0	26	41	28	1006.8	1003.6	26.6	31.2	0	0
2008-12-28	Albury	20.1	32.7	0.0	WNW	48	N	WNW	13	30	56	15	1005.2	1001.7	24.6	32.1	0	0
2008-12-29	Albury	19.7	27.2	0.0	WNW	46	NW	WSW	19	30	49	22	1004.8	1004.2	21.6	26.1	0	1
2008-12-30	Albury	12.5	24.2	1.2	WNW	50	WSW	SW	11	22	78	70	1005.6	1003.4	12.5	18.2	1	0

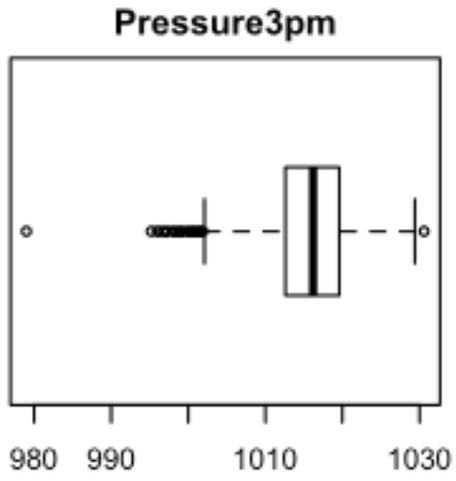
- 22 Variables étudiées
- 6 Qualitatives, 16 Quantitatives

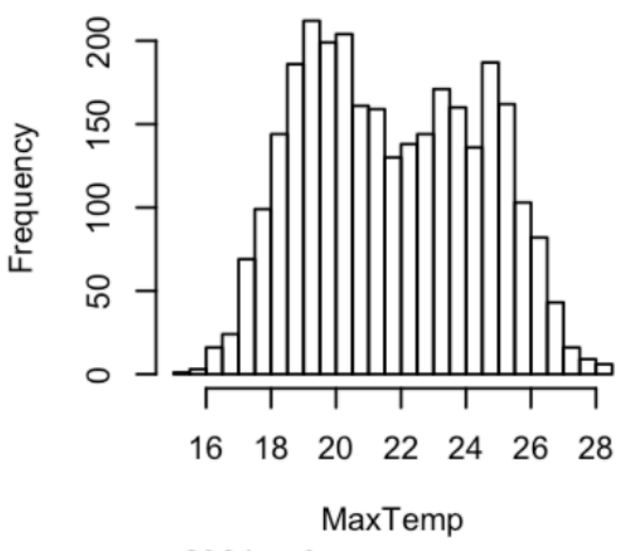
Un étude descriptive

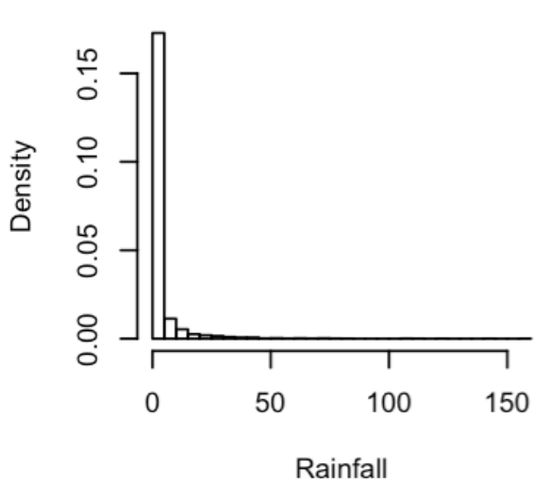


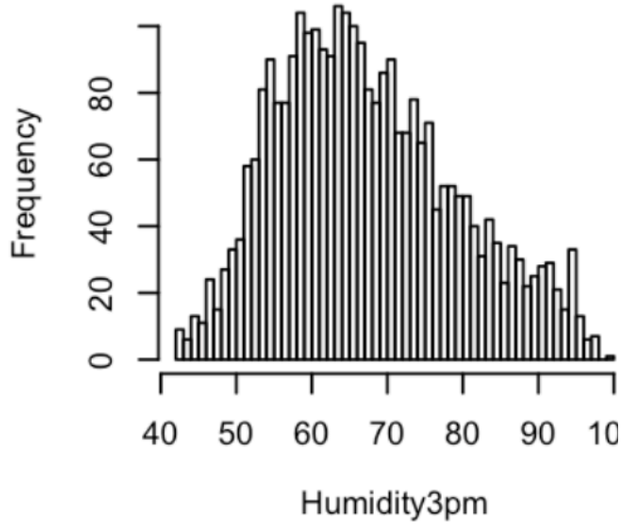


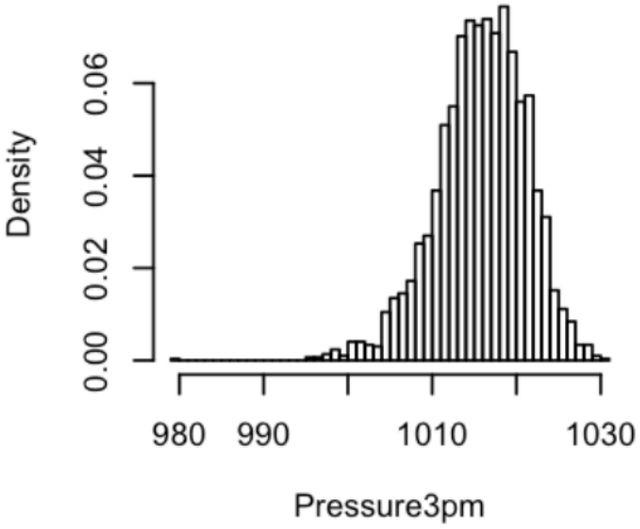




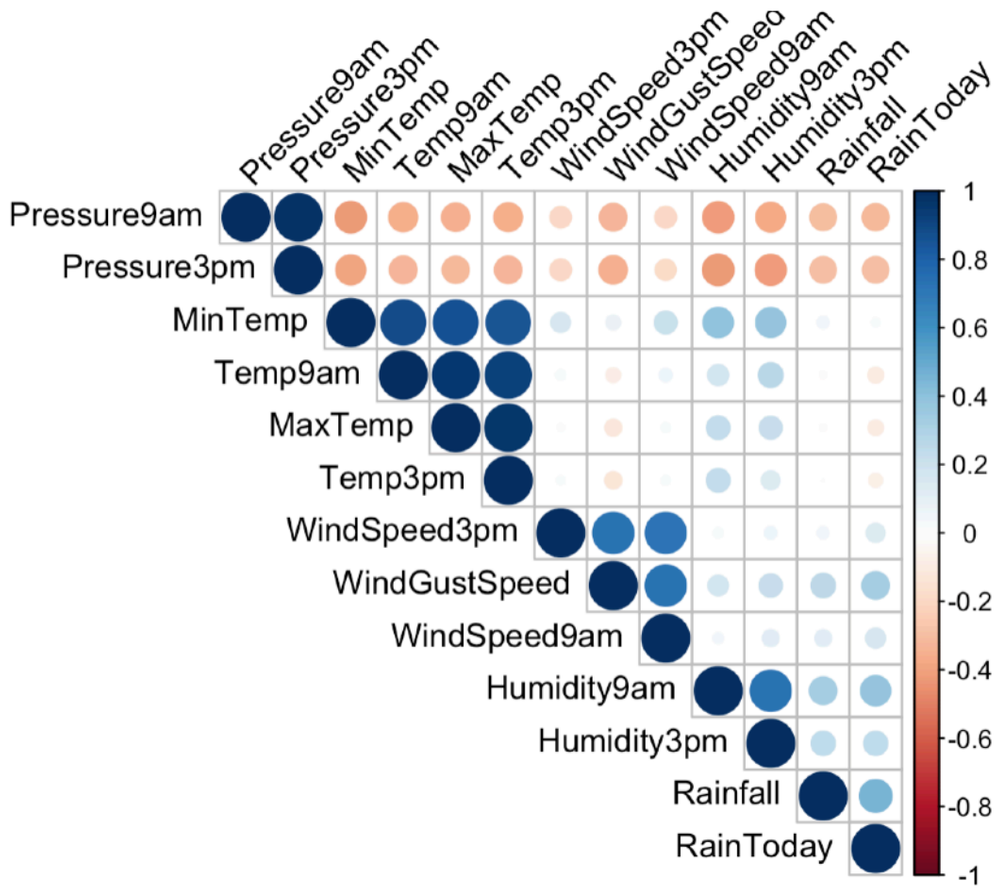






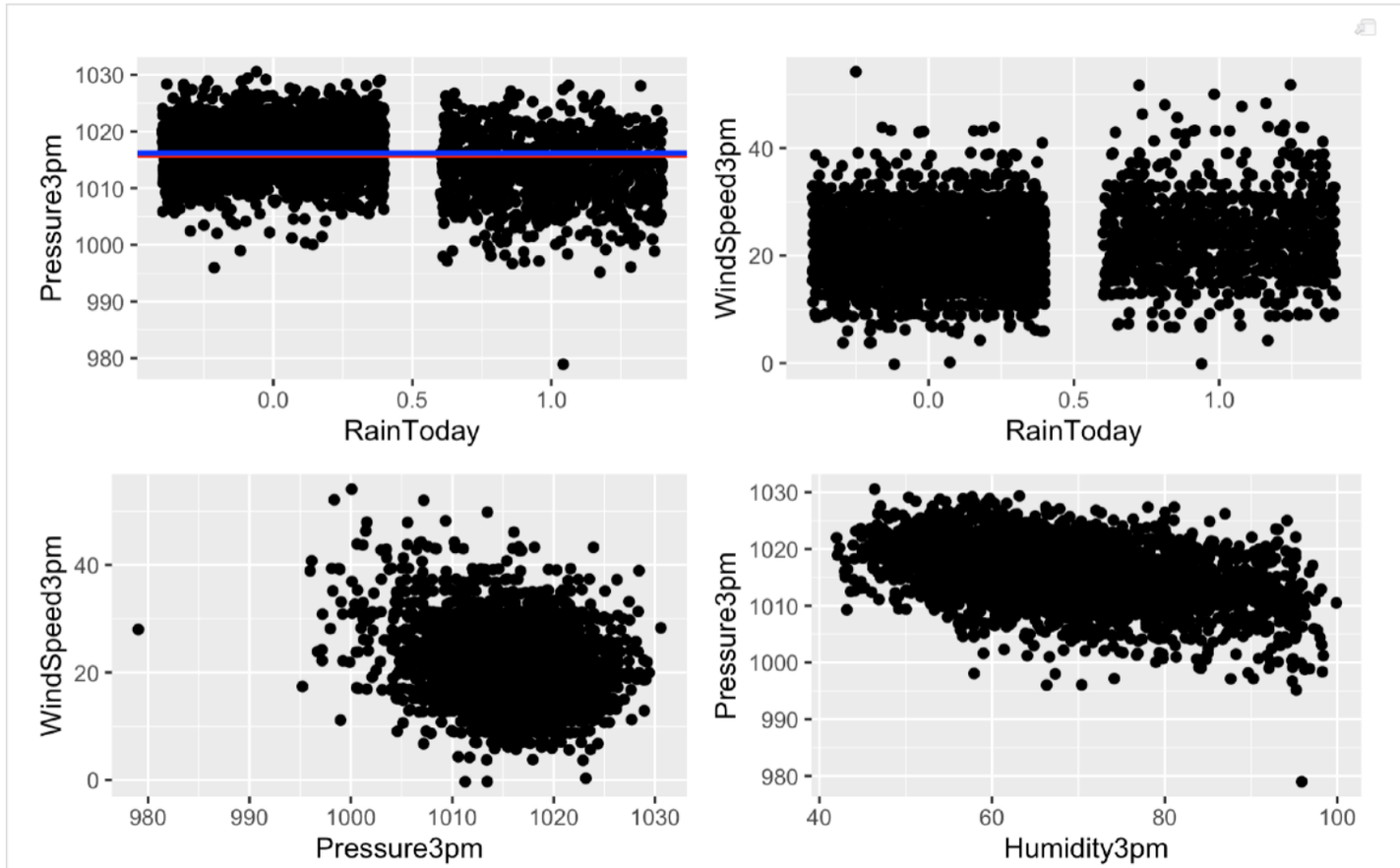


Une étude bivariée

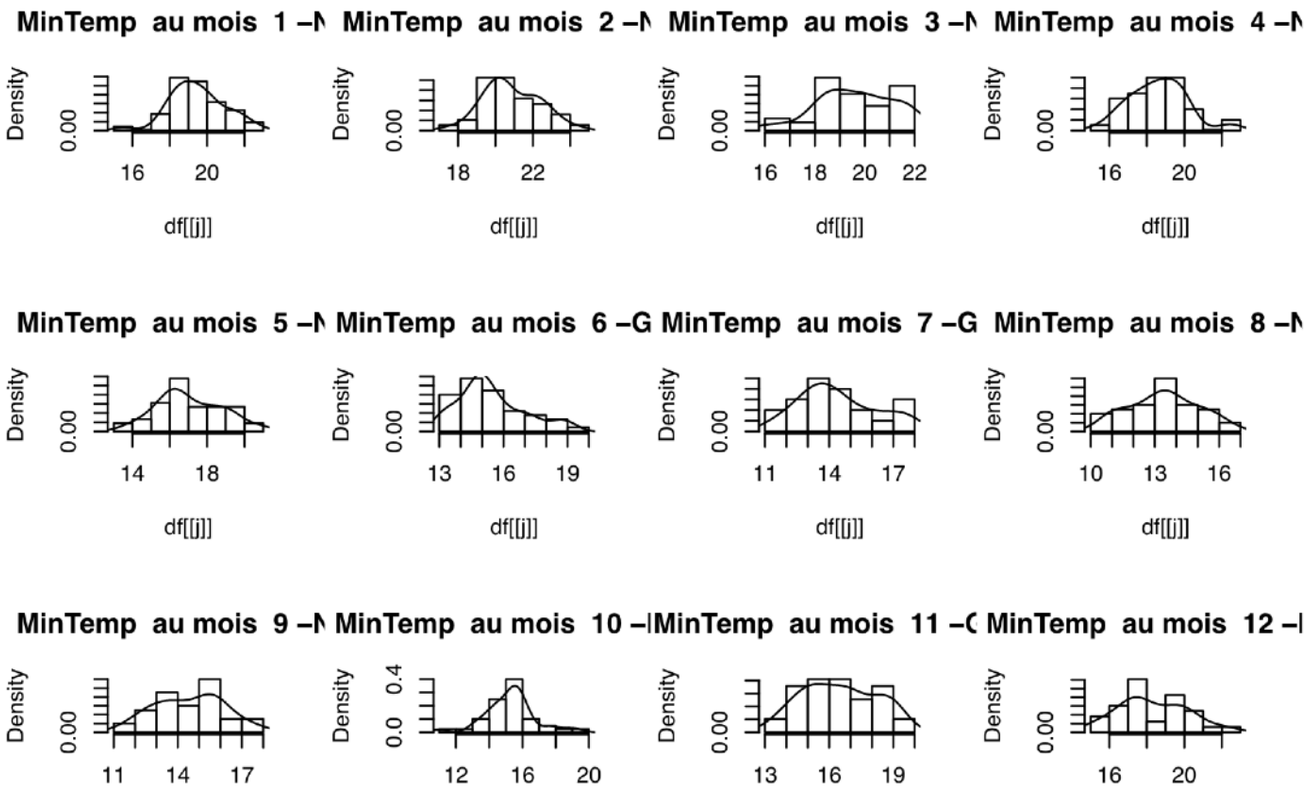


Matrice de corrélation

Étude bivariée de certaines variables intéressantes

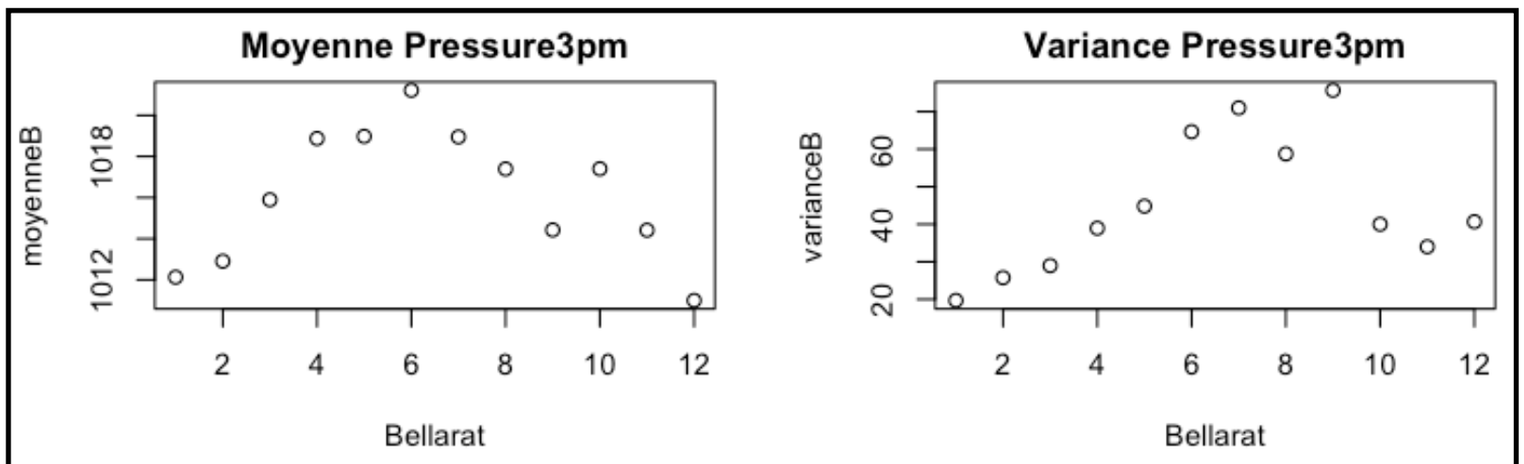


Étude de jours aléatoire, ordonnés par mois



	Gaussienne	Normale
MinTemp		✅
MaxTemp		✅
RainFall	✅	
WindGustSpeed	✅	
WindSpeed		✅
Humidity		✅
Pressure		✅
Temp		✅

Modélisation des lois



Évolution moyenne et variance au cours du temps

Régression logistique

```
Call:  glm(formula = dataAlbury$RainTomorrow ~ dataAlbury[, 3] + dataAlbury[,
    4] + dataAlbury[, 7] + dataAlbury[, 10] + dataAlbury[, 11] +
    dataAlbury[, 12] + dataAlbury[, 13] + dataAlbury[, 14] +
    dataAlbury[, 15] + dataAlbury[, 16] + dataAlbury[, 17], family = "binomial",
    data = dataAlbury)

Coefficients:
     (Intercept)   dataAlbury[, 3]   dataAlbury[, 4]   dataAlbury[, 7]  dataAlbury[, 10]  dataAlbury[, 11]
       87.495523          0.121990         -0.134381          0.055246         -0.002787         -0.007478
dataAlbury[, 12]  dataAlbury[, 13]  dataAlbury[, 14]  dataAlbury[, 15]  dataAlbury[, 16]  dataAlbury[, 17]
       -0.004519          0.086402          0.433297         -0.527152          0.036482          0.019868

Degrees of Freedom: 3010 Total (i.e. Null);  2999 Residual
Null Deviance:      3057
Residual Deviance: 1720         AIC: 1744
```

Estimation des Bi

```
Call:  glm(formula = dataBallarat$RainTomorrow ~ dataBallarat[, 3] +
    dataBallarat[, 4] + dataBallarat[, 7] + dataBallarat[, 10] +
    dataBallarat[, 11] + dataBallarat[, 12] + dataBallarat[,
    13] + dataBallarat[, 14] + dataBallarat[, 15] + dataBallarat[,
    16] + dataBallarat[, 17], family = "binomial", data = dataBallarat)

Coefficients:
       (Intercept)   dataBallarat[, 3]   dataBallarat[, 4]   dataBallarat[, 7]  dataBallarat[, 10]
         109.401185            0.013973           -0.057543            0.055215           -0.024986
dataBallarat[, 11]  dataBallarat[, 12]  dataBallarat[, 13]  dataBallarat[, 14]  dataBallarat[, 15]
          -0.008068           -0.009782            0.066986            0.168822           -0.282904
dataBallarat[, 16]  dataBallarat[, 17]
           0.101863           -0.030281

Degrees of Freedom: 3027 Total (i.e. Null);  3016 Residual
Null Deviance:      3457
Residual Deviance: 2147         AIC: 2171
```

Cible

$$Y = f(X) = \mathbf{P}(Y = 1|X) = \frac{e^{\sum_i B_i X_i}}{1 + e^{\sum_i B_i X_i}}$$

Données

À Estimer