

MODELISATION
STATISTIQUE &
APPLICATIONS

Températures, ensoleillement
et précipitations
à Lyon de 1956 à 2005

Etude statistique
sur les mois
de janvier et juillet

SOMMAIRE

Introduction (page 3)

1/ JUILLET (page 3)

1.1 Analyses univariées (page 3)

1.1.1 Températures maximales (Tx) : analyse approfondie (page 3)

1.1.2 Insolation (Ins) : analyse simple (page 7)

1.2 Analyses multivariées (page 10)

1.2.1 Température et insolation, quelle évolution ? (page 10)

1.2.2 Quel rapport entre température maximale et insolation ? (page 11)

1.2.3 Le modèle linéaire utilisé est-il correct et adapté? (page 13)

2/ JANVIER (page 15)

2.1 Analyses univariées et comparaisons avec juillet (page 15)

2.1.1 Températures maximales (Tx) : analyse simple (page 15)

2.1.2 Insolation (Ins) : analyse simple (page 18)

2.2 Analyses multivariées et comparaisons avec juillet (page 20)

2.2.1 Température et insolation, quelle évolution ? (page 20)

2.2.2 Quel rapport entre température maximale et insolation ? (page 22)

2.2.3 Le modèle linéaire utilisé est-il correct et adapté? (page 24)

2.2.4 Un modèle mieux adapté? (page 25)

3/ QUELQUES TESTS... (page 26)

3.1 Test sur la moyenne d'un échantillon gaussien (page 26)

3.2 Intervalle de confiance de la moyenne (page 27)

3.3 Test sur la variance d'un échantillon gaussien (page 27)

3.4 Intervalle de confiance de la variance (page 27)

3.5 Comparaison des variances de 2 échantillons (page 28)

3.6 Comparaison des moyennes de 2 échantillons (page 29)

Conclusion (page 30)

Annexes :

Les chiffres de Juillet (page 31)

Les chiffres de Janvier (page 32)

Introduction :

Ces données météorologiques nous ont été gracieusement envoyées par mail car indisponibles gratuitement. Un des webmaster du site <http://www.meteo-mc.fr/>, site de météo sur le massif central a répondu à notre demande en nous envoyant les données suivantes : Températures minimales, maximales et moyennes, précipitations et insolation pour les mois de juillet et janvier de 1956 à 2005 pour la station météorologique de Lyon/Bron.

Les moyennes sont établies logiquement de la manière suivantes : pour la moyenne des températures minimales (idem pour les maximales et moyennes) : Somme des températures minimales (ou maximales ou moyennes) de chaque jour divisée par le nombre de jours dans le mois c'est-à-dire 31. Pour les précipitations, c'est la somme des précipitations de chaque jour du mois, la neige étant comptabilisée pour les mois de janvier (elle est fondue pour obtenir son équivalent en eau). L'insolation (appelé insolation) est le nombre total d'heure de soleil dans le mois. Les unités sont le degré Celcius (°C) pour les températures, le millimètre (mm) ou litre par mètre carré (l/m²) pour les précipitations (ils sont tout les 2 équivalents : 1 mm = 1 l/m² = 1 litre de précipitation tombé par mètre carré) et l'heure (h) pour l'insolation.

Il a été choisi volontairement deux mois comportant 31 jours et opposés (le plus chaud de l'année et le plus froid) afin d'effectuer des comparaisons intéressantes.

1/ JUILLET :

1.1 Analyses univariées

1.1.1 Températures maximales (Tx) : analyse approfondie

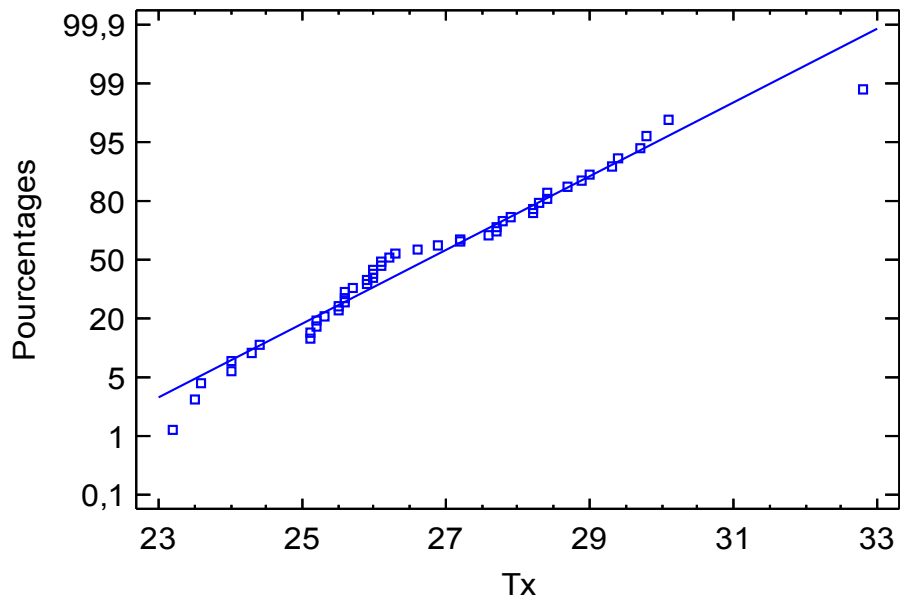
Intéressons nous d'abord aux températures maximales moyennes des 50 mois de juillet de 1956 à 2005.

Statistiques résumées :

Effectif	50
Moyenne	26,734
Médiane	26,15
Variance	3,98066
Ecart-type	1,99516
Coef. de variation	7,463%
Minimum	23,2
Maximum	32,8
Etendue	9,6
1er quartile	25,5
3ème quartile	28,2
Etendue inter-quartiles	2,7
Asymétrie standardisée	1,5328
Aplatissement standard.	0,481629

Les coefficients d'aplatissement standardisé et d'asymétrie standardisée sont particulièrement utiles car ils peuvent être utilisés pour déterminer si l'échantillon suit une loi normale. Des valeurs de ces statistiques en dehors de la plage -2 à +2 indiquent un écart significatif à la normalité, ce qui rend non valides les tests statistiques concernant la variance. Ici on remarque que ces deux coefficients sont positifs et inférieurs à deux. On peut donc dire que cet échantillon est gaussien et suit donc une loi normale.

Graphique de normalité



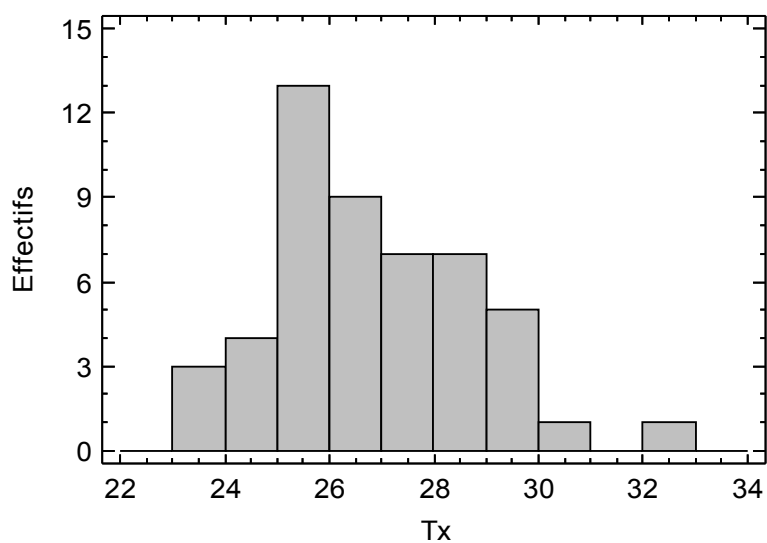
Ce graphique de normalité montre que cet échantillon suit une loi normale même si on remarque quelques valeurs faibles isolées et une valeur très élevée et isolée.

Diagramme tiges et feuilles : (unité = 0,1 10|2 représente 10,2)

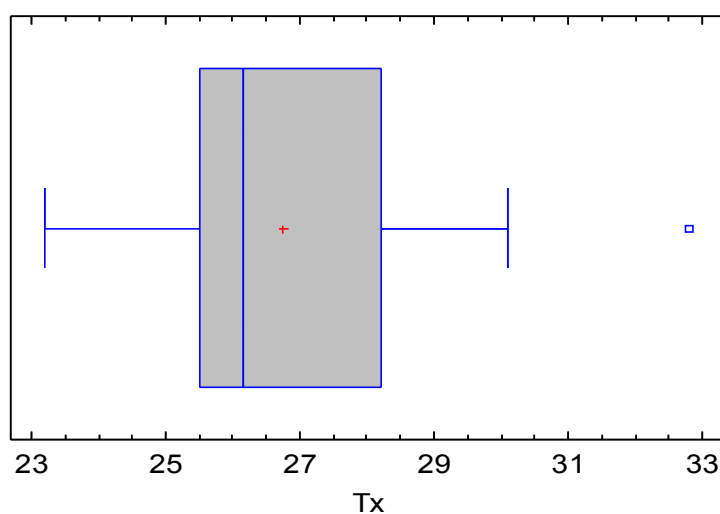
3	23 256
4	24 0034
13	25 1122355666799
9	26 000112369
7	27 2267789
7	28 2234479
5	29 03478
1	30 1
SUP	32,8

Voici le diagramme tiges et feuilles de cet échantillon, il représente aussi un tableau des effectifs. L'étendue des données a été divisée en 8 intervalles (appelés tiges), chacun représenté par une ligne du tableau. Les tiges ont comme libellés les deux premiers chiffres des valeurs des données se trouvant dans cet intervalle (donc les valeurs entières). Dans chaque ligne, les données individuelles sont représentées par un chiffre (appelé feuille) à droite de la barre verticale (et représentant la décimale). Cela construit un histogramme qui possède l'avantage que chaque donnée est visualisable. S'il y a des points très éloignés de la plupart des autres (appelés points extrêmes), ils sont placés sur des feuilles INF et SUP. Ici, il y a un point extrême SUP qui correspond à un mois de juillet anormalement chaud comparé aux autres car ayant une moyenne de températures maximales de 32.8°C soit plus de 6° au-dessus de la moyenne des 50 mois de juillet étudiés. La colonne de nombres à gauche donne les effectifs de chaque tige.

Histogramme



Boîte à moustaches



L'histogramme et la boîte à moustache mettent bien en valeur ce mois de juillet particulièrement chaud. Cette anomalie chaude s'explique sans doute tout simplement par une canicule persistante sur l'ensemble de mois. En regardant de plus près nos chiffres, on remarque qu'il s'agit de juillet 1983 et qu'il s'accompagne d'une forte insolation et de faibles précipitations, ce qui conforte cette hypothèse. A part ça on remarque une bonne concentration de températures maximales moyennes entre 25 et 26°, ce qui a pour conséquence de faire baisser la médiane en dessous de la moyenne arithmétique, cet écart est clairement visible sur la boîte à moustache.

Voyons les effectifs plus en détail :

Tableau des effectifs pour Tx

Classe	Limite infér.	Limite supér.	Point central	Effectif	Fréquence	Effectif cumulé	Fréquence cumulée
	< ou =	22,0		0	0,0000	0	0,0000
1	22,0	23,0	22,5	0	0,0000	0	0,0000
2	23,0	24,0	23,5	3	0,0600	3	0,0600
3	24,0	25,0	24,5	4	0,0800	7	0,1400
4	25,0	26,0	25,5	13	0,2600	20	0,4000
5	26,0	27,0	26,5	9	0,1800	29	0,5800
6	27,0	28,0	27,5	7	0,1400	36	0,7200
7	28,0	29,0	28,5	7	0,1400	43	0,8600
8	29,0	30,0	29,5	5	0,1000	48	0,9600
9	30,0	31,0	30,5	1	0,0200	49	0,9800
10	31,0	32,0	31,5	0	0,0000	49	0,9800
11	32,0	33,0	32,5	1	0,0200	50	1,0000
12	33,0	34,0	33,5	0	0,0000	50	1,0000
	>	34,0		0	0,0000	50	1,0000

Voici un tableau des effectifs obtenu en divisant l'étendue des Tx en intervalles de mêmes amplitudes et en comptant le nombre de données dans chaque intervalle. Les effectifs indiquent les nombres de données dans chaque intervalle, les fréquences indiquent les proportions dans chaque intervalle. Ce tableau nous permet par exemple de voir que l'intervalle [25°; 26°[contient 26% des valeurs soit environ 1/4 et que 96% des valeurs sont inférieures à 30°C, les mois de juillet « insupportables » sont donc plutôt rares à Lyon.

Voyons maintenant les quantiles :

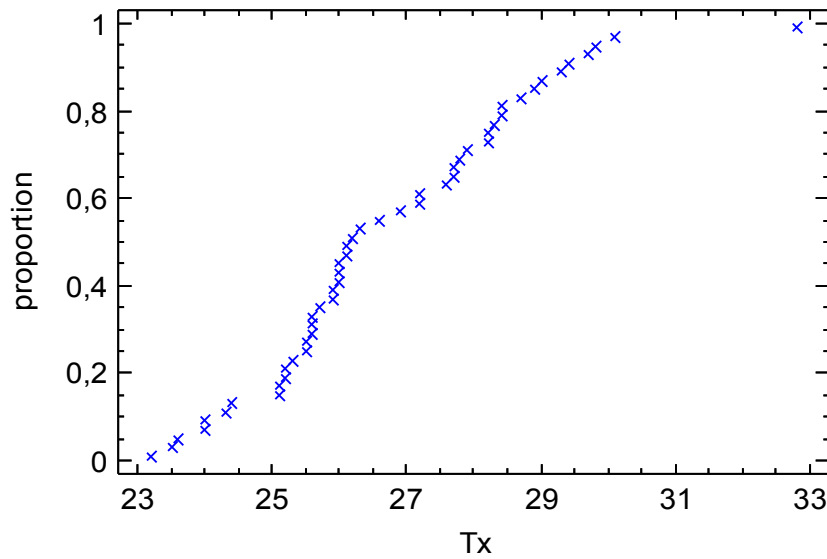
Quantiles pour Tx

	Quantiles	Limite inférieure	Limite supérieure
1,0%	23,2	20,7889	23,0058
5,0%	23,6	22,4317	24,1918
10,0%	24,15	23,2939	24,8382
25,0%	25,5	24,6976	25,9584
50,0%	26,15	26,167	27,301
75,0%	28,2	27,5096	28,7704
90,0%	29,35	28,6298	30,1741
95,0%	29,8	29,2762	31,0363
99,0%	32,8	30,4622	32,6791

Ce tableau donne les quantiles de l'échantillon des Tx ainsi que des limites de confiance pour ces quantiles. Les quantiles sont les valeurs au-dessous desquelles des pourcentages précisés des données se trouvent. Rappelons que 25% correspond au 1^{er} quartile, 75% au 3^{eme} quartile ou encore 50% à la médiane. Les limites normales de confiance sont calculées pour un niveau de confiance de 95%.

Voici le graphique correspondant où on peut voir encore une fois le mois de juillet 1983 se distinguer par son isolement :

Graphique des quantiles



Intervalles de confiance :

Intervalle de confiance à 95,0% pour la moyenne : 26,734 à +/- 0,567019 : [26,167; 27,301]

Intervalle de confiance à 95,0% pour la variance : [2.778; 6.181]

Voici les intervalles de confiance à 95,0 % (donc pour un risque de 5 %) pour la moyenne et la variance. L'interprétation classique de ces intervalles est que, pour un échantillonnage répété, ces intervalles contiennent la vraie moyenne ou la vraie variance de l'échantillon dont les données proviennent 95,0 % des fois. En termes pratiques, on peut dire qu'il y a 95,0 % de chance pour que la vraie moyenne des Tx soit comprise entre 26,167 et 27,301, alors que la vraie variance est comprise entre 2.778 et 6.181.

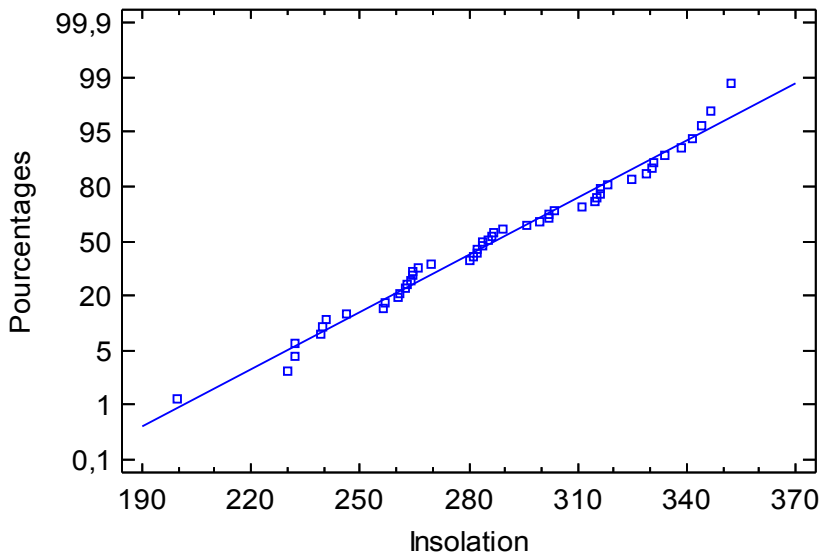
1.1.2 Insolation (Ins) : analyse simple

Statistiques résumées :

Effectif	50
Moyenne	287,144
Médiane	284,55
Variance	1301,41
Ecart-type	36,0751
Coef. de variation	12,5634%
Minimum	199,8
Maximum	352,0
Etendue	152,2
1er quartile	262,7
3ème quartile	316,0
Etendue inter-quartiles	53,3
Asymétrie standardisée	-0,306565
Aplatissement standard.	-0,919798

Ici on remarque que les coefficients d'asymétrie et d'aplatissement standardisés sont négatifs et supérieurs à -2, on en déduit que cet échantillon est gaussien et suit donc une loi normale.

Graphique de normalité



Ce graphique montre que cet échantillon suit une loi normale

Jouons avec les chiffres :

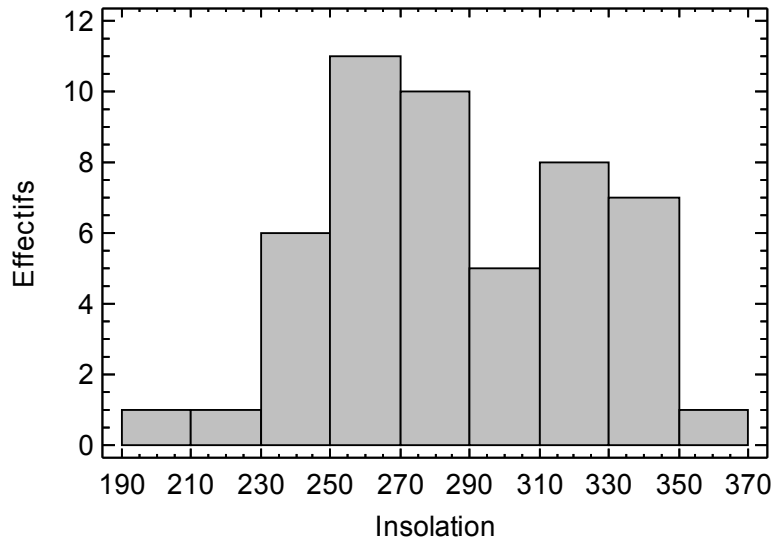
Nous disposons aussi des chiffres quotidiens pour le mois de juillet 2005. Rappelons que la journée où l'insolation totale quotidienne est potentiellement la plus élevée est le jour du solstice d'été : le 21 juin, ensuite les jours raccourcissent progressivement, donc au mois de juillet la durée d'insolation potentielle quotidienne diminue (lentement) au fil des jours. En les observant de plus près, on constate que l'ensoleillement quotidien maximum pour une journée du début du mois est d'environ 14.6h et d'environ 14h pour une journée de fin du mois. Considérons que la moyenne d'insolation maximale quotidienne au mois de juillet est de 14.3h par jour. Ainsi un mois de juillet sans aucun nuage connaîtrait une insolation totale de $14.3h \times 31 = 443.3h$. La moyenne arithmétique d'insolation des 50 derniers mois de juillet est de 287.1h soit un pourcentage d'insolation moyen d'environ 65 %. On peut donc dire qu'à Lyon au mois de juillet, le soleil est présent $\frac{2}{3}$ du temps et le ciel est couvert $\frac{1}{3}$ du temps. Le mois le plus ensoleillé a connu 352h d'insolation soit 79 % d'ensoleillement, alors que le mois le moins ensoleillé a connu 199.8h d'insolation soit 45 % d'ensoleillement.

Diagramme tiges et feuilles : (unité = 10,0 1|2 représente 120,0)

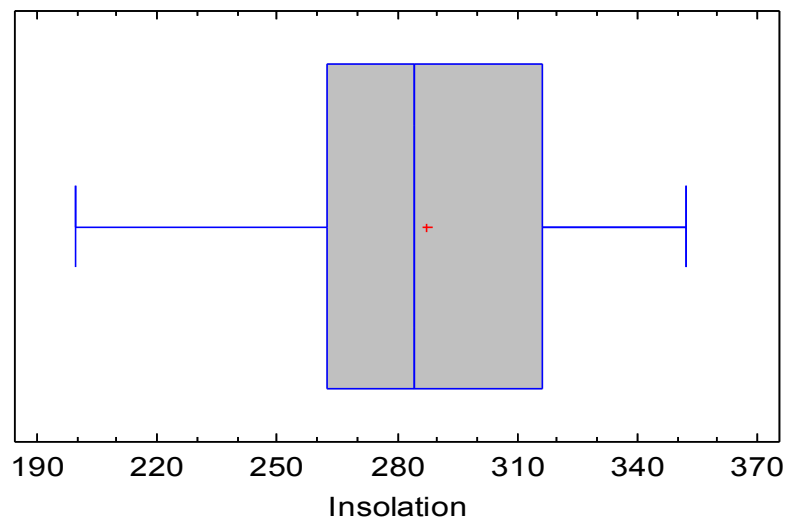
```

1      1|9
0      2|
5      2|23333
4      2|4455
9      2|666666666
12     2|8888888888899
9      3|000111111
6      3|223333
4      3|4445
    
```

Histogramme



Boîte à moustaches



L'histogramme montre 2 pics de fréquence, 1 vers 270h et l'autre vers 330h ce qui est assez difficile à expliquer. Ce qui est plus remarquable est peut-être le « trou » entre 290 et 310h, sans doute que sur un échantillon de plus grande taille ce « trou » aurait tendance à s'atténuer, voir à disparaître. La boîte à moustache ne montre, elle, rien de bien remarquable si ce n'est que les valeurs faibles sont assez rares et que juillet est donc plutôt bien ensoleillé à Lyon.

Intervalles de confiance :

Intervalle de confiance à 95,0% pour la moyenne : 287,144 à +/- 10,2524 : [276,892; 297,396]

Intervalle de confiance à 95,0% pour la variance : [908.10; 2020.89]

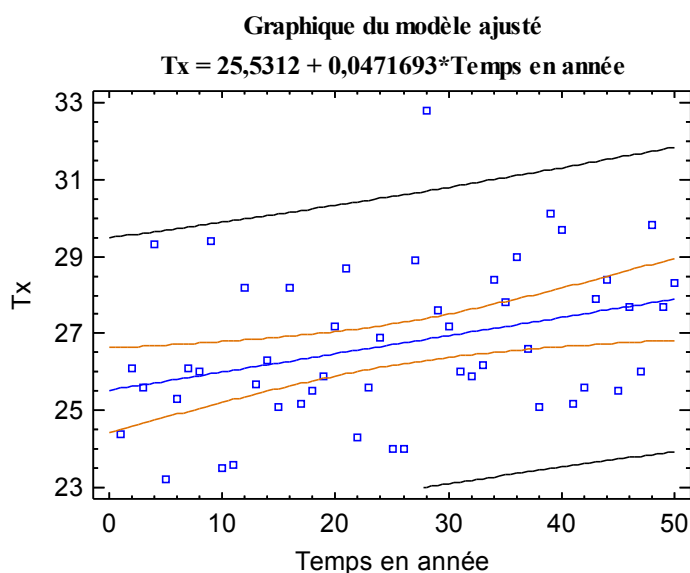
Voici les intervalles de confiance pour un risque de 5 % de la moyenne et de la variance.

1.2 Analyses multivariées :

1.2.1 Température et insolation, quelle évolution ?

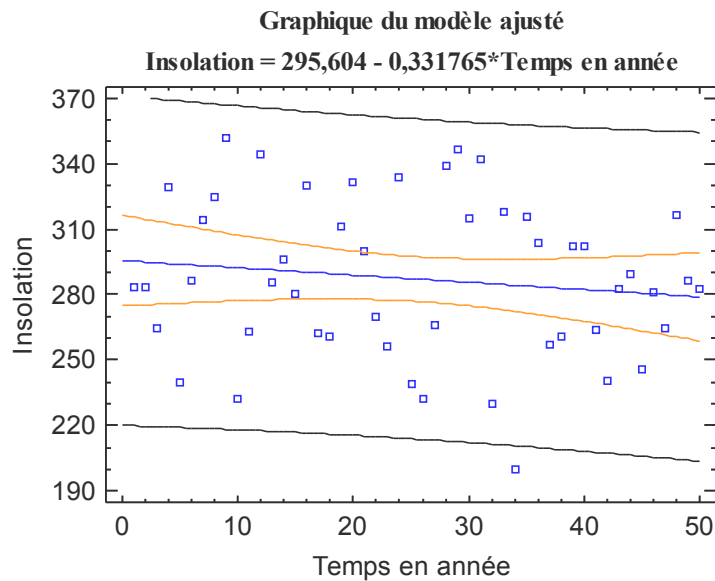
Ci-dessous se trouve le nuages de points des température max des 50 derniers mois de juillet (au temps $t = 0$ correspond juillet 1956 et $t = 50$ juillet 2005) ainsi que la droite de régression par moindres carrés des températures max en fonction du temps. Le coefficient de corrélation vaut 0,34 ce qui indique une relation assez faible entre les variables. Malgré tout ce graphique est assez intéressant et nous permet de constater sans surprise une **hausse** des températures maximales en juillet ces 50 dernières années finalement assez significative car de pente environ égale à 0,047, ce qui ferait une hausse de 4,7°C en 100 ans. Cette hausse est assez facile à expliquer car étant probablement due au réchauffement climatique que nous connaissons actuellement, même si cela reste à prouver.

L'équation du modèle ajusté est : $T_x = 25,5312 + 0,0471693 \times \text{Temps en année}$



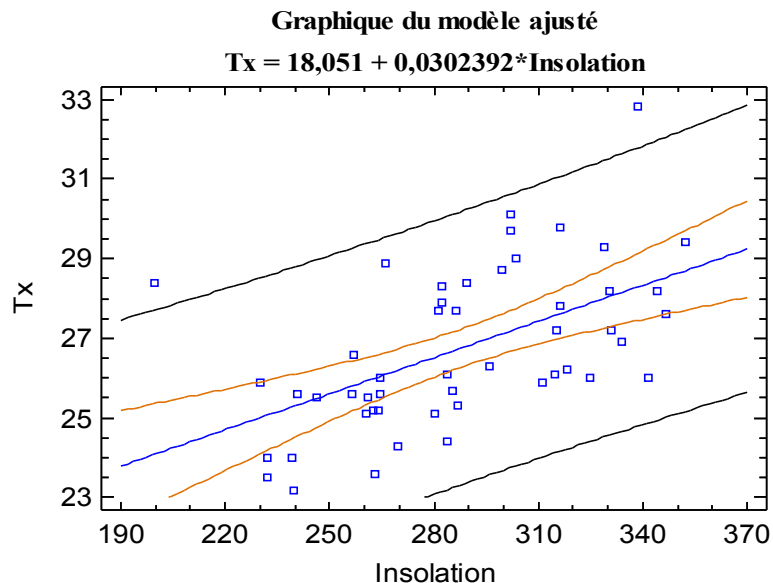
A la page suivante se trouve le nuages de points des insolation totales mensuelles des 50 derniers mois de juillet ainsi que la droite de régression par moindres carrés des insolation en fonction du temps. Le coefficient de corrélation vaut -0,13 ce qui indique une relation relativement faible entre les variables. Toutefois on peut tout de même remarquer qu'il est étonnant de constater une **baisse** de l'insolation sur ces 50 derniers mois de juillet, la pente est de -0,33, ce qui correspondrait à une diminution d'ensoleillement d'environ 33h en 100 ans soit environ 1h de soleil en moins par jour... Cette diminution semble à priori difficile à expliquer et nous ne nous hasarderons pas à émettre des hypothèses qui risqueraient d'être fausses. Contentons-nous donc de la constater, et de rappeler que le climat est tout sauf stable et qu'il se caractérise par un ensemble de variabilité au sein même d'autres variabilités (quotidiennes, hebdomadaires, mensuelles, décennales, centenales, millénaires...), et que ces mécanismes de variations recèlent encore beaucoup de secrets qui s'avèreront dur à percer et qui ne peuvent de plus qu'augmenter depuis que l'homme s'en trouve mêlé et en devient un des acteurs principaux.

L'équation du modèle ajusté est : $\text{Insolation} = 295,604 - 0,331765 \times \text{Temps en année}$



1.2.2 Quel rapport entre température maximale et insolation ?

Analysons maintenant les températures maximales et l'insolation totale des mois de juillet ensemble. Y'a-t-il un rapport entre les deux ? L'insolation a-t-elle une influence sur la moyenne des maximales du mois ? A priori on serait tenter de répondre oui, car une journée ensoleillée est généralement et logiquement plus chaude qu'une journée sans soleil notamment au mois de juillet où le soleil est plus haut dans le ciel et a donc une plus grande influence sur les températures. Toutefois il ne faut pas oublier que c'est avant tout la masse d'air et son origine qui sont responsables des températures que nous subissons ou que nous apprécions... Notons aussi que nous avons choisi de chercher une relation entre l'insolation et les températures maximales plutôt qu'avec les températures minimales ou moyennes car il nous semble évident que si l'ensoleillement a une influence sur les températures, elle serait bien plus marquée et visible sur les températures maximales, notamment en juillet...



Régression simple - Tx en fonction de Insolation

Variable à expliquer : Tx

Variable explicative : Insolation

Modèle linéaire : $Y = a + b \times X$

Coefficients

	<i>Estimation des moindres carrés</i>	<i>Erreur type</i>	<i>t</i>	<i>Probabilité</i>
Ordonnée	18,051	1,934	9,33349	0,0000
Pente	0,0302392	0,00668381	4,52425	0,0000

Analyse de variance

<i>Source</i>	<i>Somme des carrés</i>	<i>Ddl</i>	<i>Carré moyen</i>	<i>F</i>	<i>Probabilité</i>
Modèle	58,311	1	58,311	20,47	0,0000
Résidu	136,741	48	2,84877		
Total (Corr.)	195,052	49			

Coefficient de corrélation = 0,546764

R-carré = 29,8951 %

Estimation de l'écart-type du résidu = 1,68783

Erreur absolue moyenne = 1,34607

Test de Durbin-Watson = 1,66945 (P=0,1216)

Autocorrélation résiduelle d'ordre 1 = 0,136444

Voici les résultats de l'ajustement d'un modèle linéaire pour décrire la relation entre les températures maximales et l'insolation. L'équation du modèle ajusté est:

$$Tx = 18,051 + 0,0302392 * \text{Insolation}$$

En théorie, cette équation signifierait qu'en cas d'absence totale de soleil, la température maximale moyenne au mois de juillet serait d'environ 18°C ce qui n'est bien sur pas vraie en pratique. Comme la valeur de la probabilité dans le tableau de l'analyse de variance est inférieure à 0,05 (elle est même nulle), il y a une relation statistiquement significative entre les températures maximales et l'insolation au niveau de confiance de 95,0%. La statistique de R-carré indique que le modèle ajusté explique 29,9 % de la variabilité des températures maximales. Le coefficient de corrélation vaut 0,55 ce qui indique une relation modérément forte entre les variables. Ainsi on peut dire qu'au mois de juillet, l'insolation a une bonne influence sur les températures maximales et a tendance à tirer leur moyenne vers le haut, même si ce n'est pas systématiquement le cas comme on peut le voir sur le graphique puisque le mois de juillet ayant connu la plus faible insolation (juillet 1989 avec 199.8h d'ensoleillement) est aussi celui qui a connu une des moyenne de température maximale la plus élevée avec 28.4°C. Excepté cet exemple, beaucoup de mois ayant connu une faible insolation, s'accompagnent cependant de températures maximales faibles et en-dessous de la moyenne. La relation entre l'insolation et les températures maximales au mois de juillet est donc bien existante. L'explication évidente de cette relation étant celle adoptée en préambule à savoir qu'en juillet le soleil a autant d'influence par sa présence qui assure généralement une après-midi chaude et ensoleillée que par son absence plutôt synonyme du contraire même si des après-midi nuageuses et lourdes restent assez fréquentes par temps orageux. L'erreur-type d'estimation indique que l'écart-type des résidus est de 1,69. Cette valeur peut être utilisée pour construire des limites de prévision pour de nouvelles observations. L'erreur absolue moyenne de 1,35 est la valeur moyenne des résidus. La statistique de Durbin-Watson (DW) teste les résidus pour déterminer s'il y a une corrélation significative basée sur l'ordre dans lequel ils apparaissent dans le fichier des données. Comme la valeur de la probabilité est supérieure à 0,05, il n'y a pas d'autocorrélation sérielle des résidus au niveau de confiance de 95,0%.

Valeurs prévues

X	Y prévu	95,00%		95,00%	
		Limite inférieure	de prévision supérieure	Limite inférieure	de confiance supérieure
199,8	24,0928	20,47	27,7156	22,8247	25,3609
352,0	28,6952	25,1587	32,2317	27,7002	29,6902

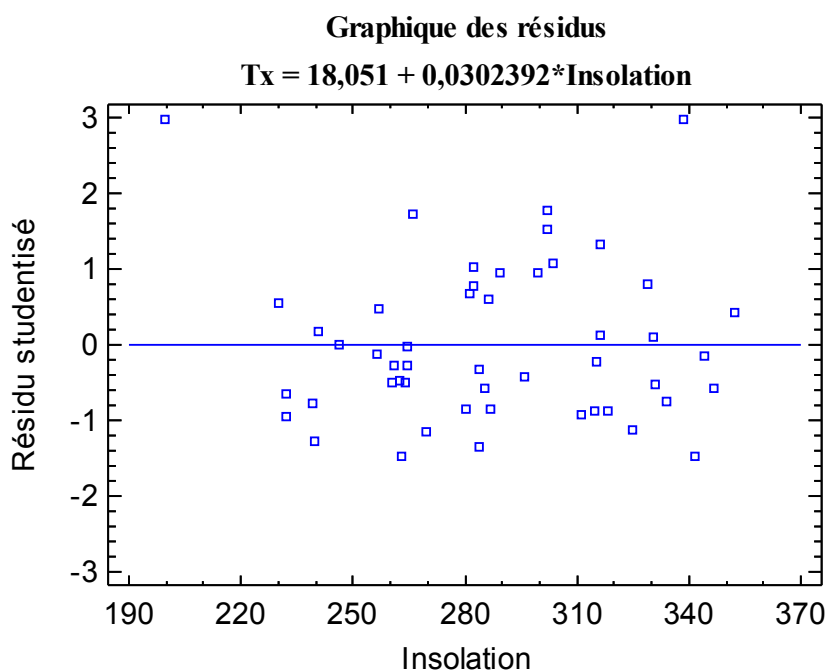
Ce tableau affiche les valeurs prévues de Tx en utilisant le modèle ajusté. En plus des meilleures prévisions, le tableau affiche les:

- (1) intervalles de prévision à 95,0% pour de nouvelles observations
- (2) intervalles de confiance à 95,0% pour la moyenne d'un ensemble d'observations

Les intervalles de prévision et de confiance correspondent aux limites internes et externes dans le graphique du modèle ajusté.

1.2.3 Le modèle linéaire utilisé est-il correct et adapté?

Le graphique des résidus ci-dessous montre bien que les résidus sont distribués de manière assez aléatoire autour de 0. Cela montre que notre modèle linéaire est correct et assez bien représentatif du rapport entre les températures maximales et l'insolation au mois de juillet.



Analyse de variance avec test du manque d'adéquation

Source	Somme des carrés	Ddl	Carré moyen	F	Probabilité
Modèle	58,311	1	58,311	20,47	0,0000
Résidu	136,741	48	2,84877		
Manque d'adéquation	135,091	45	3,00203	5,46	0,0927
Erreur pure	1,65	3	0,55		
Total (Corr.)	195,052	49			

Le test de manque d'adéquation permet de déterminer si le modèle sélectionné est adéquat pour décrire les données observées ou si un modèle plus compliqué doit être utilisé. Le test est effectué en comparant la variabilité des résidus du modèle actuel à la variabilité entre les

observations pour des valeurs répétées de la variable explicative X. Comme la valeur de la probabilité pour le test du manque d'adéquation dans le tableau est supérieure à 0,05, le modèle semble être adéquat pour les données observées au niveau de confiance de 95,0%.

Comparaison des modèles alternatifs

<i>Modèle</i>	<i>Corrélation</i>	<i>R-carré</i>
Réciproque Y, X carré	-0,5653	31,95%
Log Y, X carré	0,5603	31,40%
Réciproque Y	-0,5589	31,24%
Racine carrée Y, X carré	0,5574	31,07%
X carré	0,5542	30,71%
Réciproque Y, racine carrée X	-0,5535	30,63%
Exponentiel	0,5534	30,63%
Racine carrée Y	0,5502	30,28%
Log Y, racine carrée X	0,5477	30,00%
Y carré, X carré	0,5469	29,91%
Linéaire	0,5468	29,90%
Réciproque Y, log X	-0,5463	29,84%
Racine carrée Y et X	0,5444	29,64%
Racine carrée X	0,5408	29,25%
Multiplicatif	0,5403	29,19%
Y carré	0,5389	29,05%
Racine carrée Y, log X	0,5369	28,83%
Log X	0,5332	28,43%
Y carré, racine carrée X	0,5328	28,39%
Réciproque double	0,5265	27,72%
Y carré, log X	0,5250	27,56%
Courbe en S	-0,5201	27,05%
Racine carrée Y, réciproque X	-0,5165	26,68%
Réciproque X	-0,5127	26,29%
Y carré, réciproque X	-0,5043	25,43%

Ce tableau affiche les résultats de l'ajustement de plusieurs modèles curvilinéaires aux données. Parmi les modèles ajustés, le modèle réciproque Y et X carré a la plus grande valeur du R-carré avec 31,95 %. C'est 2,05 % supérieur au modèle linéaire ce qui est quand même assez faible. Aussi le modèle linéaire possède l'avantage d'être plus simple c'est pourquoi nous avons jugé bon de le conserver et l'estimons assez satisfaisant pour représenter le rapport entre températures maximales et insolation au cours des mois de juillet 1956 à 2005 à Lyon. Rappelons enfin que nous avons vu que les températures maximales avaient tendance à augmenter pendant que l'insolation diminuait sur ces cinquante mois juillet. Vu que nous venons de voir qu'il existe une relation entre les deux, on peut se demander si les températures augmenteraient encore plus vite en cas d'insolation stables ou en hausse ? Cette baisse surprenante et inexplicquée de l'insolation nous épargne-t-elle temporairement des mois de juillet caniculaires ? Réponse dans les années à venir...

2/ JANVIER :

2.1 Analyses univariées et comparaisons avec juillet :

2.1.1 Températures maximales (Tx) : analyse simple

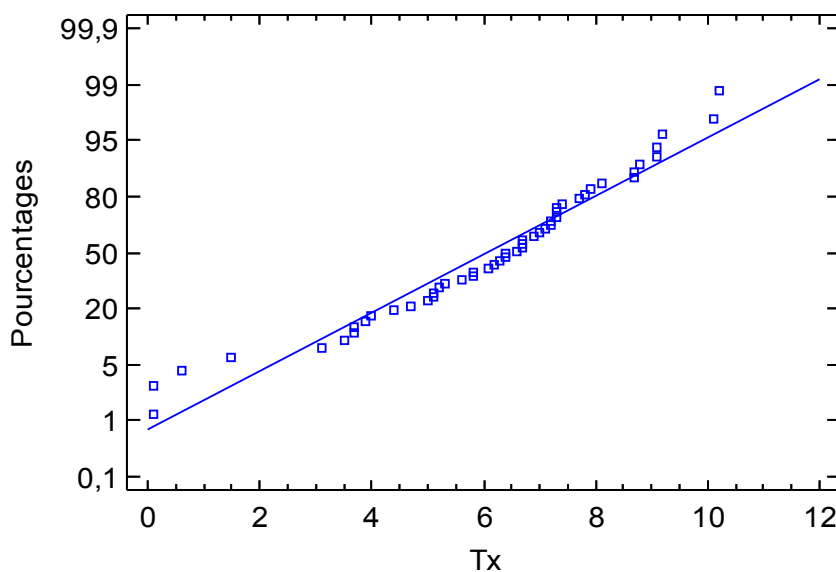
Intéressons nous d'abord aux températures maximales moyennes des 50 mois de janvier de 1956 à 2005.

Statistiques résumées :

Effectif	50
Moyenne	6,088
Médiane	6,5
Variance	5,6374
Ecart-type	2,37432
Coef. de variation	39,0%
Minimum	0,1
Maximum	10,2
Etendue	10,1
1er quartile	5,0
3ème quartile	7,4
Etendue inter-quartiles	2,4
Asymétrie std.	-2,21355
Aplatissement std.	0,815929

Ici on remarque que la valeur de l'aplatissement standardisé est dans la plage attendue pour des données suivant une loi normale (entre -2 et 2). En revanche le coefficient d'asymétrie standardisée est négatif et légèrement inférieur à -2, dans ce cas, la valeur de l'asymétrie standardisée n'est donc pas dans la plage attendue pour des données suivant une loi normale. Toutefois, elle en est assez proche, et comme l'échantillon est de grande taille (50), on peut appliquer les mêmes tests que sur un échantillon gaussiens.

Graphique de normalité



Ce graphique de normalité montre que le problème vient des valeurs de températures maximales faibles : elles sont trop rares ou trop faibles comparées aux autres, ce qui fait qu'elles sont isolées. Cela vient probablement du fait de 3-4 mois de janvier particuliers et ayant connu des températures très basses comparés aux autres. En effet, en hiver et plus particulièrement aux mois de décembre et janvier où le soleil est bas dans le ciel et a du mal à réchauffer l'atmosphère, on sait que lorsque de l'air froid s'installe en basse couche, il est particulièrement difficile à déloger et a même tendance à s'accroître. Ce phénomène vient du fait que l'air froid est plus lourd que l'air chaud et est donc bien plus tenace. De plus il existe des cas où après une invasion froide, de l'air plus doux revient, mais comme il est plus léger que l'air froid, il passe et s'installe par-dessus celui-ci... Ainsi à basse altitude, l'air froid reste installé et a même tendance à s'intensifier en raison de journées courtes et de nuits longues, alors qu'à moyenne altitude, de l'air plus doux est revenu et les températures sont plus clémentes. Le gradient de température, habituellement décroissant avec l'altitude (perte de 6.5°C tout les 1000 m en moyenne) est alors inversé, et il fait plus doux en moyenne montagne qu'en plaine ou vallée. Un tel phénomène est appelé inversion et peut persister parfois plusieurs semaines. Il se caractérise donc par un temps gris, brumeux et froid à basse altitude, alors qu'en moyenne montagne au dessus de la couche d'inversion, le temps est ensoleillé, doux et l'atmosphère est limpide. On peut ici penser que ce sont ces quelques mois de janvier, victime d'inversions persistantes qui sont mis en valeurs sur ce graphique en se caractérisant par des températures maximales bien en dessous des moyennes et cette explication laisserait présager (à confirmer plus tard) qu'ils s'accompagneraient donc d'insolations faibles. Les 4 mois de janvier froid en question sont : 1963 avec un moyenne de températures maximales de 0.1°C, 1964 avec 1.5°C, 1985 avec 0.6°C et 1987 avec 0.1°C. Ces dates sont très intéressante car il a été remarqué que les hivers rudes vont souvent par série, et surtout, ont tendance à se manifester en moyenne tous les 22 à 23 ans ce qui correspond tout à fait avec ces 4 dates. Cette récurrence d'apparitions de conditions rigoureuses en hiver semble trouver son explication dans les variations de l'activité solaire avec des cycles de 11 ans (appelé cycle de **Schwabe**) correspondant à un maximum et un minimum d'activité solaire. Et il est frappant que les séries d'hivers froids tombent systématiquement pendant que l'activité solaire est minimum. Précisons que le prochain minimum du cycle de Schwabe est prévu vers 2006-2007... On pourrait donc logiquement s'attendre à une série d'hiver froid, et il est très intéressant de constater que l'hiver 2005/2006 s'est caractérisé par un froid certes rarement intense mais très tenace, avec une série remarquable et rare de 5 mois ayant connu des températures inférieures aux normales (de novembre 2005 à mars 2006) et qu'il succède lui-même à deux hivers (2003/2004 et 2004/2005, 2002/2003 ayant été frais en janvier mais très doux le reste du temps) plutôt frais après une longue série d'hivers doux voir très doux à la fin des années 1990, début des années 2000. Cependant il faut rappeler que nous étudions ici les mois de janvier qui ne représente pas l'hiver en entier.

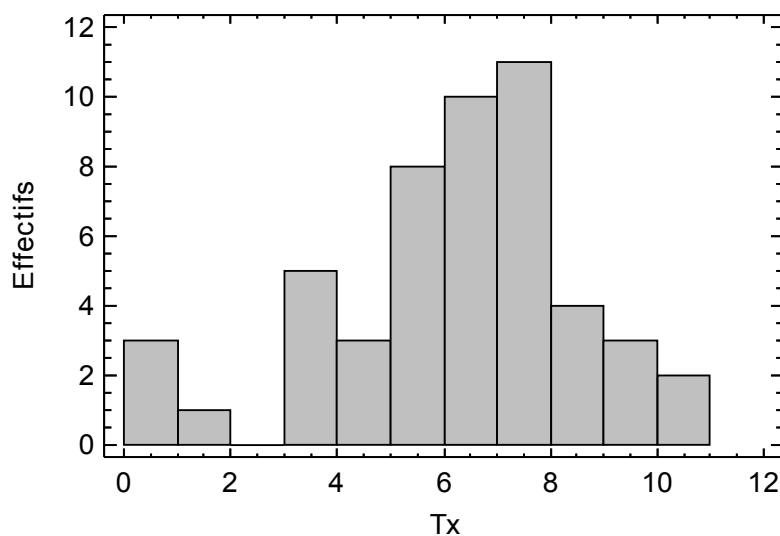
Revenons-en donc cependant à nos mois de janvier de 1956 à 2005 :

Diagramme tiges et feuilles : (unité = 0,1 1|2 représente 1,2)

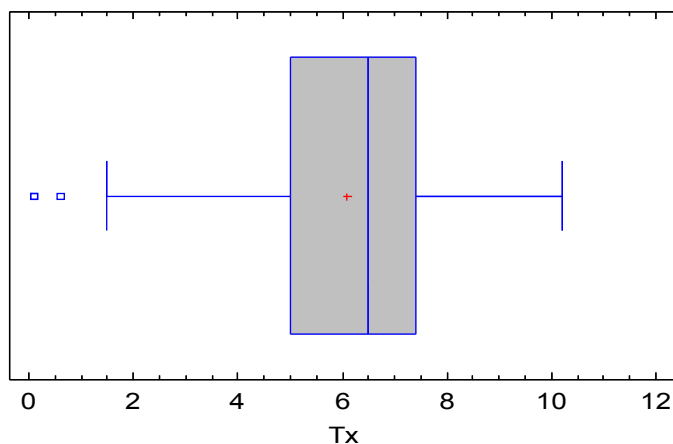
3	0 116
1	1 5
0	2
5	3 15779
3	4 047
8	5 01123688
10	6 1234467779
11	7 01223334789
4	8 1778
3	9 112
2	10 12

Voici le diagramme tiges et feuilles des effectifs dans lequel on voit apparaître très distinctement les 4 mois de janvier « particuliers » et plus froid que les autres.

Histogramme



Boîte à moustaches



L'histogramme et la boîte à moustache montrent aussi ces quelques mois de janvier isolés et ayant connu des températures maximales anormalement basses comparées aux autres.

Intervalle de confiance :

Intervalle de confiance à 95,0% pour la moyenne: 6,088 à +/- 0,674776 : [5,41322;6,76278]

Intervalle de confiance à 95,0% pour la variance: [3.93;8.75]

Voici les intervalles de confiance à 95,0 % (donc pour un risque de 5 %) pour la moyenne et la variance. Cependant ces deux intervalles supposent que la population dont l'échantillon est issu suit une loi normale. Alors que l'intervalle de confiance pour la moyenne est robuste et peu sensible à la non-normalité et on peut donc le considérer ici comme correct, l'intervalle de confiance pour la variance y est très sensible. Si les données ne suivent pas une loi normale, l'intervalle pour la variance peut être incorrect. Or on a vu ici que cet échantillon ne suit pas tout à fait une loi normale, donc il convient de rester très vigilant avec l'intervalle pour la variance affiché ici.

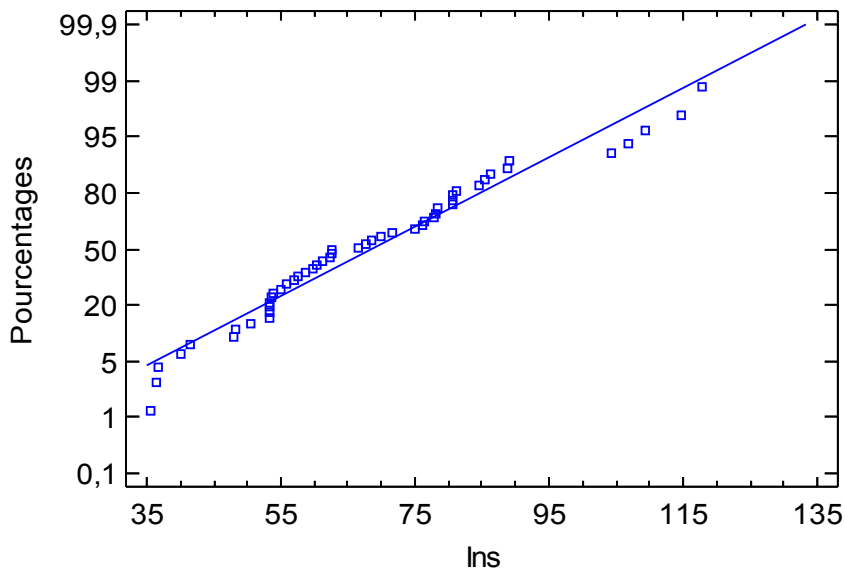
2.1.2 Insolation (Ins) : analyse simple

Statistiques résumées :

Effectif	50
Moyenne	68,546
Médiane	64,55
Variance	407,154
Ecart-type	20,178
Coef. de variation	29,4372%
Minimum	35,5
Maximum	117,9
Etendue	82,4
1er quartile	53,5
3ème quartile	80,7
Etendue inter-quartiles	27,2
Asymétrie std.	1,74505
Aplatissement std.	0,0450289

Les coefficients d'asymétrie et d'aplatissement standardisés sont bien dans la plage attendue (-2,2) pour un échantillon suivant une loi normale, on en déduit donc que cet échantillon est gaussien.

Graphique de normalité



Ce graphique montre que cet échantillon suit approximativement une loi normale malgré quelques valeurs élevées isolées de l'insolation pour 5 mois de janvier.

Jouons avec les chiffres :

Nous disposons aussi des chiffres quotidiens pour le mois de janvier 2005. Rappelons que la journée où l'insolation totale quotidienne est potentiellement la plus basse est le jour du solstice d'hiver : le 21 décembre, ensuite les jours rallongent progressivement, donc au mois de janvier la durée d'insolation potentielle quotidienne augmente au fil des jours. En les observant de plus près, on constate que l'ensoleillement quotidien maximum pour une journée du début du mois est d'environ 8h et d'environ 9.2h pour une journée de fin du mois.

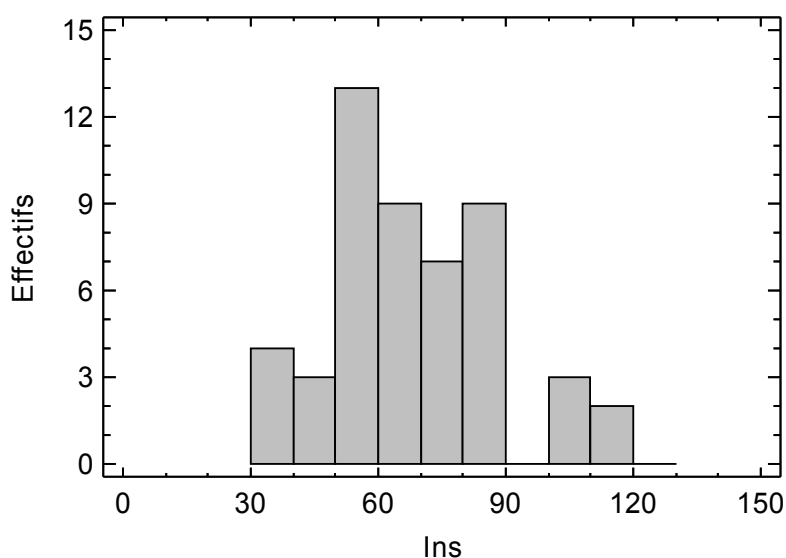
Considérons que la moyenne d'insolation maximale quotidienne au mois de janvier est de 8.6h par jour $((8+9.2) \div 2)$ (à comparer avec le mois de juillet : 14.3h par jour). Ainsi un mois de janvier sans aucun nuage connaîtrait une insolation totale de $8.6h \times 31=266.6$ h (rappel : 443.3h en juillet). La moyenne arithmétique d'insolation des 50 derniers mois de janvier est de 68.5h (287.1h en juillet) soit un pourcentage d'insolation moyen d'environ 25.7 % (65 % en juillet !). On peut donc dire qu'à Lyon au mois de janvier, le soleil est présent seulement 1/4 du temps et le ciel est couvert 3/4 du temps (rappel en juillet la proportion est de 2/3 de soleil pour 1/3 de « nuage »). Le mois de janvier le plus ensoleillé a connu 117.9h d'insolation soit environ 44 % d'ensoleillement ce qui est légèrement inférieur au mois de juillet le moins ensoleillé qui a connu 45 % d'insolation! En proportion, tout les mois de janvier ont donc connu un ensoleillement inférieur aux mois de juillet ce qui est assez remarquable et montre bien le faible ensoleillement des mois de janvier lyonnais contrastant avec l'ensoleillement important des mois de juillet. Le mois de janvier le moins ensoleillé a même connu 35.5 petites heures d'insolation soit environ 13 % d'ensoleillement...

Diagramme tiges et feuilles : (unité = 1,0 1|2 représente 12,0)

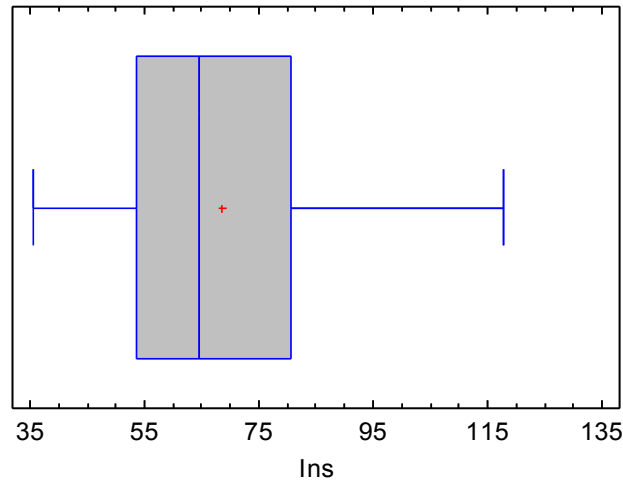
3	3 566
4	4 0178
13	5 0333333357789
9	6 012226789
7	7 1566788
9	8 000145689
0	9
3	10 469
2	11 47

Voici le diagramme tiges et feuilles des effectifs dans lequel on voit aussi apparaître les 5 mois de janvier particuliers ayant connu des insolutions élevées.

Histogramme



Boîte à moustaches



L’histogramme et la boîte à moustache montrent aussi ces quelques mois de janvier isolés et ayant connu des insolation bien supérieures aux autres.

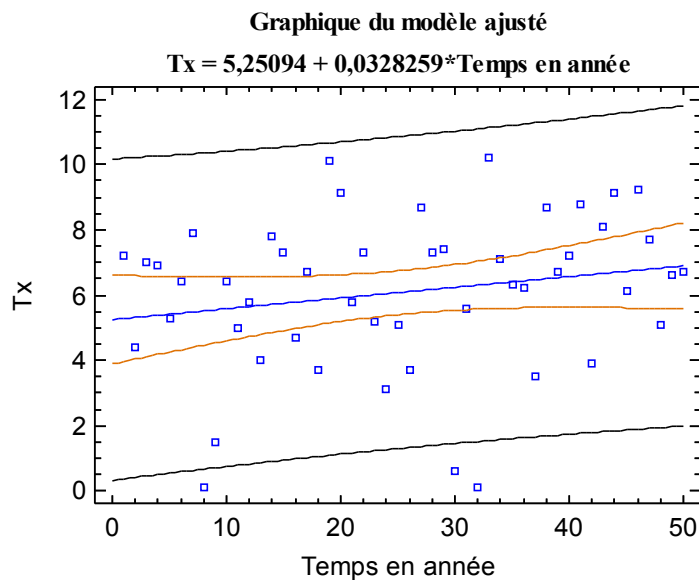
Intervalles de confiance :

Intervalle de confiance à 95,0% pour la moyenne: 68,546 à +/- 5,73455 : [62,8115;74,2805]
Intervalle de confiance à 95,0% pour la variance : [284.10;623.25]

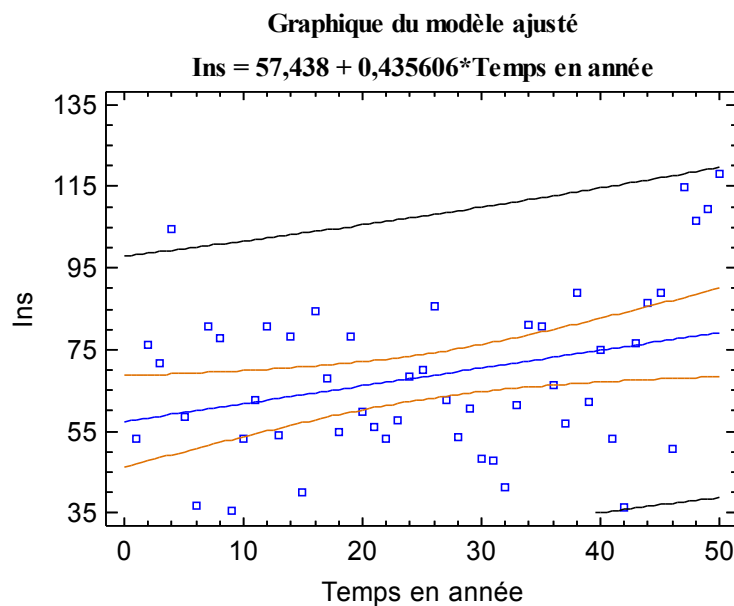
Voici les intervalles de confiance pour un risque de 5 % de la moyenne et de la variance.

2.2 Analyses multivariées et comparaisons avec juillet :

2.2.1 Température et insolation, quelle évolution ?



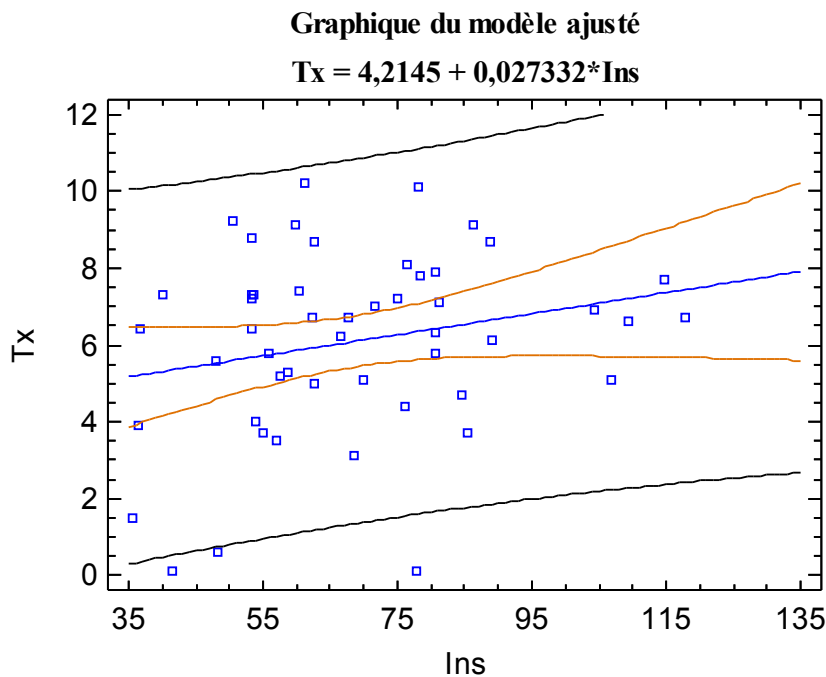
Voici le nuage de points des températures max de ces 50 mois de janvier (au temps $t=0$ correspond janvier 1956 et $t=50$ janvier 2005) ainsi que la droite de régression par moindres carrés des températures maximales en fonction du temps. Le coefficient de corrélation vaut 0,2 ce qui indique une relation assez faible entre les variables. Malgré tout et sans surprise on constate encore une **hausse** des températures maximales en fonction du temps de pente environ égale à 0.033, ce qui ferait une hausse de 3.3°C en 100 ans. Il est intéressant de constater que cette hausse des températures est plus lente qu'en juillet ce qui est difficilement explicable a priori. Par contre on remarque bien les 4 mois de janvier froids par série de 2 et espacés d'environ 22 ans, de même que les séries de janvier doux correspondant au maximum du cycle de Schwabe et qui apparaissent logiquement environ 11 ans après les janviers froids. Bien sur ce cycle n'explique pas tout, et des mois de janvier frais ou doux apparaissent par ailleurs sans d'autres explications que la variabilité naturelle du climat.



Ci-dessus se trouve le nuage de points des insolationes totales mensuelles de ces 50 mois de janvier ainsi que la droite de régression par moindres carrés des insolationes en fonction du temps. Le coefficient de corrélation vaut 0,31 ce qui indique une relation assez faible entre les variables. Toutefois il est ici plus étonnant de constater qu'on remarque une **hausse** assez forte de l'insolation sur ces 50 mois de janvier, la pente est d'environ 0.44, ce qui correspondrait à une augmentation d'ensoleillement d'environ 44h en 100 ans soit presque 1h30 de soleil en plus par jour... On remarque notamment que les 4 derniers mois de janvier sont quasiment les plus ensoleillés, ce qui est une fois de plus difficilement explicable. Beaucoup d'hypothèses sont envisageables, par exemple un changement de l'environnement entourant la station météorologique (rasage d'un immeuble...) permettant une meilleure insolation, mais la plus vraisemblable étant une série étonnamment consécutive de mois de janvier ensoleillés, la question qui se pose alors étant de savoir si cette série ainsi que cette hausse globale constatée sur le modèle linéaire est l'amorce d'un changement durable conduisant à des mois de janvier un peu plus ensoleillés ou si elle est simplement due au hasard (le hasard faisant parfois de belles choses...). Précisons pour être complet qu'une analyse réalisée par ailleurs montre que les précipitations sont en baisse au mois de janvier pendant la même période. Il semble donc que depuis 50 ans, les mois de janvier à Lyon aient tendance à être plus doux, plus ensoleillés et moins maussades, bref moins hivernaux... Cette tendance étant bien évidemment à confirmer dans les années et plutôt les décennies à venir.

2.2.2 Quel rapport entre température maximale et insolation ?

Analysons maintenant les températures maximales et l'insolation totale des mois de janvier ensemble. Y'a-t-il un rapport entre les deux ? L'insolation a-t-elle une influence sur la moyenne des maximales du mois ? La réponse est plus incertaine qu'au mois de juillet car on a vu que l'insolation est bien plus faible en heure et en pourcentage qu'au mois de juillet, et de plus un ensoleillement conséquent peut se produire lors des vagues de froid en hiver grâce au vent du nord, alors que les périodes de douceur s'accompagnent souvent d'un temps maussade et couvert... En se basant sur ces premiers éléments de réponse il semblerait donc que l'ensoleillement ait une influence bien moindre sur les températures maximales en janvier qu'en juillet, voir même inexistante. Cependant il ne faut pas oublier ce qu'on a expliqué plus haut, c'est-à-dire que les périodes d'inversion, froides à basse altitude et donc à Lyon, se caractérisent par un temps gris et brumeux, et donc une absence d'insolation... Ces phénomènes d'inversion vont-ils suffire à entraîner une relation entre température maximale et insolation en janvier ?



Régression simple - Tx en fonction de Ins

Variable à expliquer: Tx

Variable explicative: Ins

Modèle linéaire: $Y = a + b * X$

Coefficients

	Estimation des moindres carrés	Erreur type	t	Probabilité
Ordonnée	4,2145	1,17944	3,57329	0,0008
Pente	0,027332	0,0165195	1,65453	0,1045

Analyse de variance

Source	Somme des carrés	Ddl	Carré moyen	F	Probabilité
Modèle	14,9038	1	14,9038	2,74	0,1045
Résidu	261,329	48	5,44435		
Total (Corr.)	276,233	49			

Coefficient de corrélation = 0,232279
 R-carré = 5,39538 %
 Estimation de l'écart-type du résidu = 2,33331
 Erreur absolue moyenne = 1,79356
 Test de Durbin-Watson = 2,12083 (P=0,6493)
 Autocorrélation résiduelle d'ordre 1 = -0,065943

Voici les résultats de l'ajustement d'un modèle linéaire pour décrire la relation entre les températures maximales et l'insolation. L'équation du modèle ajusté est:

$$T_x = 4,2145 + 0,027332 \times \text{Ins}$$

En théorie, cette équation signifierait qu'en cas d'absence totale de soleil, la température maximale moyenne au mois de janvier serait d'environ 4.2°C ce qui n'est bien sur pas vraie en pratique. Comme la valeur de la probabilité dans le tableau de l'analyse de variance est supérieure ou égale à 0,05, il n'y a pas de relation statistiquement significative entre T_x et Ins . La statistique de R-carré indique que le modèle ajusté explique 5,4 % de la variabilité des températures max ce qui est très faible. Le coefficient de corrélation vaut 0,23, ce qui indique une relation faible entre les variables. Ainsi on peut dire qu'au mois de janvier, l'insolation a une influence très limitée sur les températures maximales. Sur le graphique du modèle ajusté, on remarque cependant que sur les 4 mois de janvier atypiques ayant connu des températures maximales très basses, trois s'accompagnent d'insolations très faibles et inférieures à la moyenne. Ces 3 mois de janvier pourraient donc avoir subi des inversions prolongées comme soupçonné plus haut. Les autres points sont inégalement répartis sur le graphique et on constate qu'insolation élevée n'est en janvier pas significative de températures maximales supérieures à la moyenne mais plutôt de températures maximales normales. De même comme attendu, les mois de janvier ayant connu les températures maximales les plus élevées ne sont pas ceux ayant connu les insolations les plus grandes, les périodes de douceur hivernales étant souvent accompagnées de temps maussade et donc faiblement ensoleillées. Ce qui confirme cette influence très faible voir inexistante de l'insolation sur la température maximale moyenne au mois de janvier à Lyon, les masses d'air jouant donc un rôle beaucoup plus important. L'erreur-type d'estimation indique que l'écart-type des résidus est de 2,3. Cette valeur peut être utilisée pour construire des limites de prévision pour de nouvelles observations. L'erreur absolue moyenne de 1,8 est la valeur moyenne des résidus, pour un bon modèle cette erreur moyenne doit être proche de 0 ce qui n'est pas le cas ici. La statistique de Durbin-Watson (DW) teste les résidus pour déterminer s'il y a une corrélation significative basée sur l'ordre dans lequel ils apparaissent dans le fichier des données. Comme la valeur de la probabilité est supérieure à 0,05, il n'y a pas d'autocorrélation sérielle des résidus au niveau de confiance de 95,0%.

Valeurs prévues

X	Y prévu	95,00%		95,00%	
		Limite inférieure	de prévision supérieure	Limite inférieure	de confiance supérieure
35,5	5,18479	0,321181	10,0484	3,90223	6,46734
117,9	7,43694	2,42325	12,4506	5,66849	9,2054

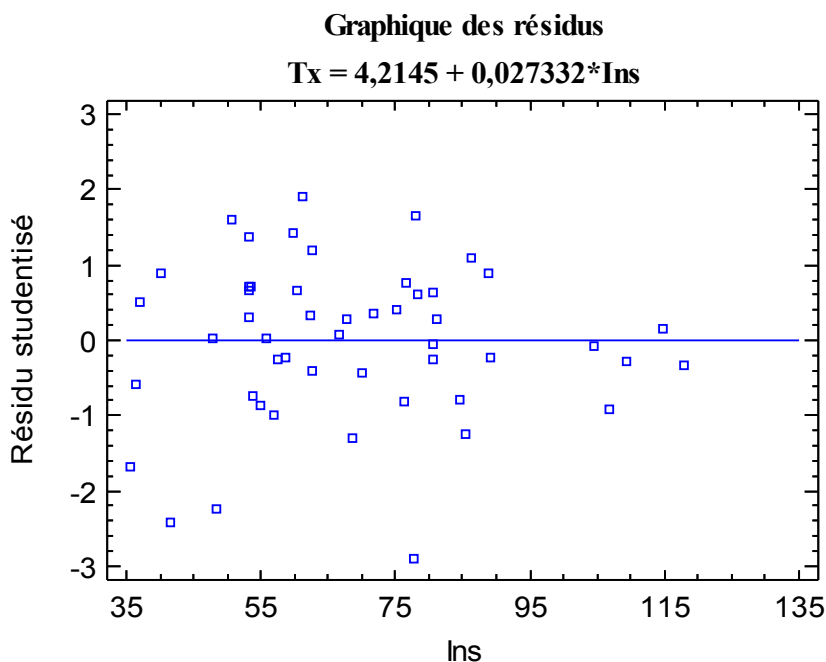
Ce tableau affiche les valeurs prévues de T_x en utilisant le modèle ajusté. En plus des meilleures prévisions, le tableau affiche les:

- (1) intervalles de prévision à 95,0% pour de nouvelles observations
- (2) intervalles de confiance à 95,0% pour la moyenne d'un ensemble d'observations

Rappelons que les intervalles de prévision et de confiance correspondent aux limites internes et externes dans le graphique du modèle ajusté.

2.2.3 Le modèle linéaire utilisé est-il correct et adapté?

Le graphique des résidus ci-dessous montre cependant que les résidus sembleraient à première vue être distribués de manière assez aléatoire autour de 0 ce qui n'est en réalité pas vraiment le cas. Cela montre que notre modèle linéaire n'est pas le plus adapté pour représenter l'éventuel rapport entre les températures maximales et l'insolation au mois de juillet.



Analyse de variance avec test du manque d'adéquation

Source	Somme des carrés	Ddl	Carré moyen	F	Probabilité
Modèle	14,9038	1	14,9038	2,74	0,1045
Résidu	261,329	48	5,44435		
Manque d'adéquation	258,264	45	5,7392	5,62	0,0893
Erreur pure	3,065	3	1,02167		
Total (Corr.)	276,233	49			

Le test de manque d'adéquation permet de déterminer si le modèle sélectionné est adéquat pour décrire les données observées ou si un modèle plus compliqué doit être utilisé. Le test est effectué en comparant la variabilité des résidus du modèle actuel à la variabilité entre les observations pour des valeurs répétées de la variable explicative X. Comme la valeur de la probabilité pour le test du manque d'adéquation dans le tableau est supérieure à 0,05, le modèle semble être étonnamment adéquat pour les données observées au niveau de confiance de 95,0%.

Comparaison des modèles alternatifs

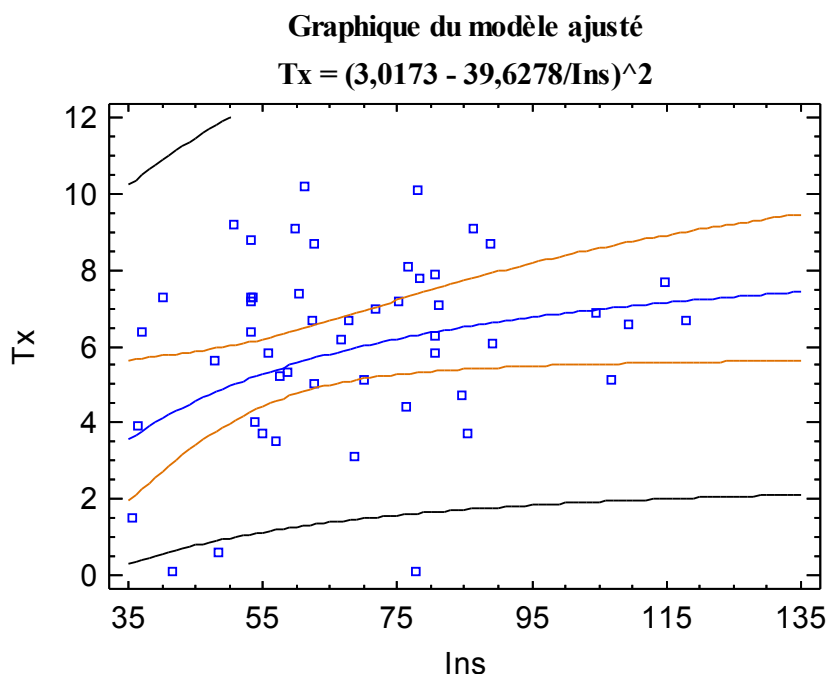
Modèle	Corrélation	R-carré
Racine carrée Y, réciproque X	-0,3031	9,19%
Réciproque X	-0,2955	8,73%
Racine carrée Y, log X	0,2757	7,60%
Log X	0,2663	7,09%
Courbe en S	-0,2634	6,94%
Racine carrée Y et X	0,2596	6,74%
Racine carrée X	0,2495	6,23%
Racine carrée Y	0,2430	5,91%
Multiplicatif	0,2412	5,82%

Y carré, réciproque X	-0,2412	5,82%
Linéaire	0,2323	5,40%
Log Y, racine carrée X	0,2280	5,20%
Exponentiel	0,2143	4,59%
Y carré, log X	0,2119	4,49%
Racine carrée Y, X carré	0,2110	4,45%
X carré	0,1991	3,96%
Y carré, racine carrée X	0,1955	3,82%
Log Y, X carré	0,1880	3,53%
Y carré	0,1788	3,20%
Y carré, X carré	0,1469	2,16%
Réciproque double	0,1443	2,08%
Réciproque Y, log X	-0,1325	1,76%
Réciproque Y, racine carrée X	-0,1256	1,58%
Réciproque Y	-0,1186	1,41%
Réciproque Y, X carré	-0,1060	1,12%

Ce tableau affiche les résultats de l'ajustement de plusieurs modèles curvilinéaires aux données. On constate tout d'abord que les valeurs du R-carré sont beaucoup plus faibles ici qu'au mois de juillet ce qui prouve une fois de plus l'influence bien plus réduite de l'ensoleillement sur les températures maximales en janvier qu'en juillet. Parmi les modèles ajustés, le modèle racine carrée Y et réciproque X a la plus grande valeur du R-carré avec environ 9,2 %. C'est 3,8 % supérieur au modèle linéaire actuellement sélectionné.

2.2.4 Un modèle mieux adapté?

Le modèle racine carrée Y et réciproque X est-il donc mieux adapté que le modèle linéaire? Voyons le graphique du modèle ajusté :



Ce graphique représente les résultats de l'ajustement d'un modèle racine carrée Y et réciproque X pour décrire la relation entre Tx et Ins.

L'équation du modèle ajusté est: $T_x = (3,0173 - 39,6278/Ins)^2$

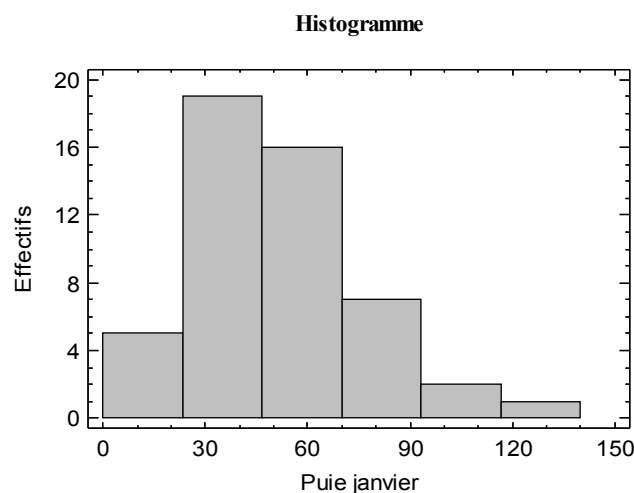
Comme la valeur de la probabilité dans le tableau de l'analyse de variance est égale à 0.03 et donc légèrement inférieure à 0,05, il y a une relation statistiquement assez significative entre Tx et Ins au niveau de confiance de 95,0% mais pas énorme non plus.

La statistique de R-carré indique que le modèle ajusté explique 9,2 % de la variabilité des Tx ce qui reste peu. Le coefficient de corrélation vaut -0,3 ce qui indique une relation relativement faible entre les variables. L'erreur-type d'estimation indique que l'écart-type des résidus est de 0,6. L'erreur absolue moyenne de 0,42 est la valeur moyenne des résidus, elle est ici plus proche de 0 que le modèle linéaire ce qui prouve que ce modèle est plus adapté et semble donc plus représentatif d'un éventuel rapport entre les températures maximales et l'insolation au mois de janvier. Mais cependant la valeur faible du R-carré (0.09) qui est beaucoup plus proche de 0 que de 1 indique que la relation entre température maximale et insolation au mois de janvier est peu évidente pour ne pas dire plus...

3/ QUELQUES TESTS...

3.1 Test sur la moyenne d'un échantillon gaussien

Intéressons nous maintenant aux précipitations des 50 derniers mois de janvier.
En voici l'histogramme :



Cet histogramme indique que cet échantillon suit à peu près une loi normale, et est donc gaussien, de plus les valeurs de l'asymétrie standardisée (1.86) et de l'aplatissement standardisé (0.99) sont dans les plages attendues pour des données suivant une loi normale. Tout va bien, on peut donc effectuer le test sur la moyenne!

Au vu de cet histogramme, il semble qu'on puisse estimer la moyenne vers 50 mm de précipitations par mois. Cette estimation est-elle bonne ? On effectue le test :

Prenons l'hypothèse H_0 : la moyenne théorique m vaut 50

Et l'hypothèse H_1 : la moyenne théorique m est différente de 50

On calcule : ↪ La moyenne empirique Moy de l'échantillon : $Moy = \sum (X_i) \div n = 52.396$

↪ La variance Var : $Var = (n \div n-1) \times ((1 \div n) \times \sum (X_i^2) - Moy^2) = 574.615$

↪ Le Tobs : $Tobs = (Moy - m) \div (\sqrt{Var \div n}) = 0.707$

On effectue un test de Student sur la moyenne pour un risque de niveau **5 %** :

On lit T_c dans la table de Student à $n-1$ degré de liberté soit 49 et à la colonne 0.975 qui correspond à un risque de 5 % ($1 - (5 \% \div 2)$) : cette valeur n'est pas dans le tableau de Student

dont nous disposons mais nous la supposons comprise entre 40 (2.021) et 60 (2.000) et on obtient : $T_c = 2.01$. Comme la valeur absolue de T_{obs} (0.707) est inférieure à la valeur de T_c (2.01) on accepte l'hypothèse H_0 pour un risque de 5 %. C'est-à-dire qu'on peut dire que la moyenne de cet échantillon est de 50 à un niveau de confiance de 95 %. En observant de près la table de Student on observe que l'hypothèse H_0 serait rejeté à partir de la colonne 0.7 où le T_c devient inférieur au T_{obs} , cela qui signifie qu'il faudrait prendre un risque de niveau 60 % pour rejeter notre hypothèse H_0 qui est donc une bonne estimation de la quantité moyenne de précipitations qu'il tombe au mois de janvier à Lyon.

3.2 Intervalle de confiance de la moyenne

Dans les analyses univariées de ce dossier, nous avons donné des intervalles de confiance de la moyenne à 95 %. Comment ont-ils été obtenus ?

Voici un exemple concret à l'aide de l'échantillon étudié précédemment.

On a vu que cet échantillon était gaussien, on peut donc donner un intervalle contenant la moyenne théorique avec une confiance de 95 %.

Pour ce niveau de confiance T_c vaut 2.01 (vu précédemment).

Or un intervalle de confiance de niveau 95% est $[Moy - T_c \sqrt{Var/n} ; Moy + T_c \sqrt{Var/n}]$

Ici on obtient $T_c \sqrt{Var/n} = 6.814$, donc l'intervalle de confiance est 52.396 ± 6.814 .

Un intervalle de confiance de la moyenne théorique des précipitations tombées à Lyon au cours des mois de janvier de 1956 à 2005 de niveau de confiance 95 % est donc [45.58;59.21]

3.3 Test sur la variance d'un échantillon gaussien

Effectuons maintenant un test sur la variance théorique de cet échantillon de risque 5 %.

Au vu de l'histogramme les valeurs de précipitation semblent être assez dispersées.

Prenons donc l'hypothèse H_0 : la variance théorique v vaut 200.

Et l'hypothèse H_1 : la variance théorique v est différente de 200

La variance empirique Var de cet échantillon vaut en fait 574.615.

On calcule X_{obs} : $X_{obs} = (n \times Var) \div v = 143.65375$

On lit dans la table du χ^2 à $n-1$ degré de liberté soit 49, cette valeur n'étant pas dans le tableau dont nous disposons, elle a été trouvée sur le site Internet suivant : <http://perso.wanadoo.fr/alain.pichereau/chi2test.html>

On a trouvé : $P(X_{49} \leq 31.5549) = 0.025$ et $P(X_{49} \leq 70.224) = 0.975$ (car le risque est de 5 %)

Donc pour accepter notre hypothèse H_0 il faudrait que X_{obs} soit dans l'intervalle

[31.5549 ; 70.224] ce qui n'est pas le cas. On en conclut que pour un risque de 5 %, l'hypothèse : la variance théorique des précipitations tombées aux cours des 50 derniers mois de janvier vaut 200 ne peut pas être acceptée.

Refaisons le test avec le même risque mais en prenant une variance théorique v de 500.

On calcule X_{obs} : $X_{obs} = (n \times Var) \div v = 57.4615$.

Cette valeur se trouve dans l'intervalle [31.5549;70.224], donc on en conclut qu'on peut dire que pour un risque de 5 %, la variance des précipitations tombées au mois de janvier vaut 500.

3.4 Intervalle de confiance de la variance

Dans les analyses univariées de ce dossier, nous avons donné des intervalles de confiance de la variance à 95 %. Comment ont-ils été obtenus ?

Voici un exemple concret toujours à l'aide du même échantillon.

On garde le même risque de 5 %, on peut calculer un intervalle de confiance de la variance car l'échantillon est gaussien.

On sait que X_{obs} appartient à $[a ; b]$ (on a calculé précédemment que pour un risque de 5 %, l'intervalle est $[a=31.5549 ; b=70.2224]$) avec une probabilité de 95 %.

Un intervalle de confiance de la variance de niveau de confiance 95 % est :

$$[(n \times \text{Var}) \div b ; (n \times \text{Var}) \div a] = [409.13 ; 910.5].$$

Donc un intervalle de la variance de la hauteur de précipitations tombées au cours des 50 derniers mois de janvier à Lyon est $[409.14 ; 910.5]$ pour un niveau de confiance de 95 %.

3.5 Comparaison des variances de 2 échantillons

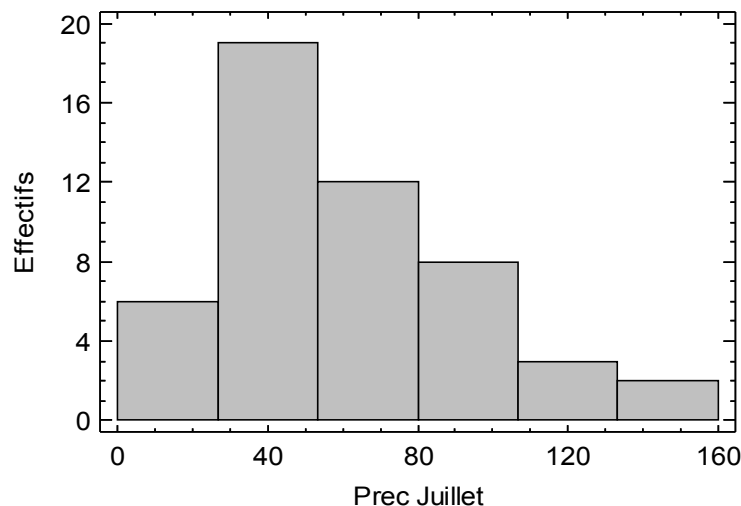
Nous allons maintenant comparer les variances de 2 échantillons, les précipitations des mois de janvier de 1956 à 2005 dont nous avons vu précédemment qu'il était gaussien et les précipitations des mois de juillet de la même période dont voici quelques informations :

Statistiques résumées pour Prec Juillet

Effectif	50
Moyenne	60,83
Médiane	53,9
Variance	1107,69
Minimum	10,9
Maximum	152,1
Etendue	141,2
Asymétrie std.	2,7966
Aplatissement std.	0,872316

Ici la valeur de l'aplatissement standardisé est dans la plage attendue pour des données suivant une loi normale mais pas la valeur de l'asymétrie standardisée. Toutefois deux informations vont nous « rassurer », tout d'abord l'histogramme de cet échantillon :

Histogramme



Cet histogramme montre bien que s'il ne suit pas tout à fait une loi de normale, notre échantillon n'en semble cependant pas si éloigné. Enfin rappelons que notre échantillon a pour taille 50, ce qui peut être considéré comme une grande valeur et nous permet donc d'appliquer les mêmes tests que pour un échantillon gaussien...

Passons donc au test de comparaisons des variances :

Variance janvier = $V_{ja} = 574.615$

Variance juillet = $V_{ju} = 1107.69$

On veut tester H_0 : les variances théoriques des échantillons sont égales pour un risque de 5 %

On calcule F_{obs} : $F_{obs} = V_{ju} \div V_{ja} = 1.93$

Si les deux échantillons ont la même variance alors F_{obs} suit la loi de Fisher-Snedecor de paramètre $(n-1, n-1)$ soit $(49, 49)$. On regarde donc dans la table de Fisher-Snedecor :

On lit F_2 dans la colonne 50 et la ligne 50 (qu'on suppose approximativement égales à 49) du côté 5 % $\div 2 = 2.5$ %, On trouve $F_2 = 1.752$, idem pour $F_1 = 1 \div 1.752 = 0.571$.

Or ici notre F_{obs} n'est pas compris dans l'intervalle $[0.571 ; 1.752]$, on en conclut donc que pour un risque 5 %, il faut rejeter notre hypothèse H_0 et que les variances théoriques des précipitations tombées au cours des mois de janvier et juillet à Lyon ne sont pas égales. Quand on regarde plus en détail on s'aperçoit que la variance des précipitations tombées au mois de juillet est bien supérieure à celle des précipitations tombées en janvier, ce qui explique la non-égalité du test des variances. Cet écart s'explique par le caractère différent des précipitations qu'il tombe au cours de ces deux mois : alors qu'au mois de janvier les précipitations sont dites plutôt stratiformes, c'est-à-dire qu'elles tombent sous des perturbations de grandes étendues et de manières régulières et bien réparties, les précipitations tombées au mois de juillet sont généralement dites convectives : elles se produisent sous des orages et lors d'instabilité et ne sont donc pas également réparties mais comportent un caractère beaucoup plus aléatoire et irrégulier. Ainsi d'une année sur l'autre, il est possible d'être plus ou moins épargné des dégradations orageuses principales responsables des précipitations en juillet, et les valeurs de ces précipitations sont donc bien plus variables qu'au mois de janvier où les précipitations comportent beaucoup moins ce caractère incertain. Cette différence de « nature » des précipitations tombées au cours de ces deux mois explique une variance bien plus forte en juillet qu'en janvier. Enfin ajoutons que la moyenne arithmétique des précipitations tombées en juillet est de 60.83 mm et que cette valeur peut facilement tomber au cours d'un ou deux orages violents alors que le reste du mois fut très beau, ainsi la valeur des précipitations tombées au cours d'un mois de juillet n'est que rarement représentative de l'impression (bonne ou mauvaise) laissée par ce mois, contrairement au mois de janvier où une valeur de précipitations mensuelle élevée est généralement synonyme d'un mois gris et pluvieux.

3.6 Comparaison des moyennes de 2 échantillons

Dans un premier temps, il est important de signaler que pour pouvoir effectuer un test de comparaison des moyennes de deux échantillons, il faut tout d'abord avoir effectué le test de comparaison des variances pratiqué ci-dessus et il faut que les variances des deux échantillons se soient avérées égales ce qui n'est malheureusement pas le cas pour nos deux échantillons...

La conclusion logique est que nos deux échantillons ne relèvent pas de la même loi pour le risque de 5 % calculé et que les moyennes ne sont donc pas comparables...

Nous allons toutefois quand même effectuer prudemment le test de comparaison des moyennes sur ces deux échantillons en gardant évidemment le même risque de 5 %, en oubliant pas que le F_{obs} calculé dans le test des variances (1.93) n'était pas si éloigné que ça de la plage dans laquelle il aurait du se trouver $[0.571 ; 1.752]$ pour accepter l'égalité des variances, et que si l'écart entre les variances de ces deux échantillons est significatif, il n'est pas abyssal. Procédons donc au test de comparaison des moyennes de ces deux échantillons :

Moy janvier = $M_{ja} = 52.396$ mm

Moy juillet = $M_{ju} = 60.83$ mm

Var janvier = $V_{ja} = 574.615$

Var juillet = $V_{ju} = 1107.69$

On calcule s , comme ici les deux échantillons sont de même taille $n = 50$ alors :

$$s = (V_{ja} + V_{ju}) \div n = 33.6461$$

On en déduit le T_{obs} : $T_{obs} = (M_{ja} - M_{ju}) \div (\sqrt{s}) = -1.454$

On lit la valeur T_c dans la table de la loi Student à 98 degrés de liberté ($2 \times 50 - 2$) dans la colonne 0.975 qui correspond à notre risque de 5 %.

On trouve une valeur approximative de T_c : $T_c = 1.984$

On compare ensuite la valeur absolue de T_{obs} (1.454) et T_c (1.984), on trouve $|T_{obs}| < T_c$ et on en conclut donc que pour un risque de 5 %, les moyennes théoriques des précipitations tombées en janvier et en juillet à Lyon peuvent être supposées égales.

On peut aussi obtenir un intervalle de confiance à 95 % de la différence D des moyennes théoriques : $D \in [M_{ja} - M_{ju} - T_c \sqrt{s} ; M_{ja} - M_{ju} + T_c \sqrt{s}] = -8.434 \pm 11.508 = [-19.942 ; 3.074]$

N'oublions pas de rappeler que le test sur les variances nous a montré que les variances de ces deux échantillons ne sont pas égales et qu'il convient donc de rester extrêmement prudent et mesuré avec le test d'égalité des moyennes qui n'aurait normalement pas du être effectué car l'inégalité des variances signifie que nos deux échantillons ne relèvent pas de la même loi. Cependant, on peut quand même remarquer que les valeurs moyennes des précipitations tombées en juillet et en janvier à Lyon sont relativement proches et que pour deux mois opposés de l'année (janvier gris et froid et juillet chaud et ensoleillé) il tombe en moyenne autant de précipitations, la raison principale due au « caractère » différent des précipitations tombées a déjà été expliqué plus haut, (pour résumer, il pleut moins souvent et moins longtemps mais plus fortement en juillet) et cela prouve l'absence de sécheresse estivale ou hivernale à Lyon contrairement aux régions méditerranéennes par exemple. A Lyon, les variations météorologiques saisonnières ou mensuelles se font beaucoup plus sur les températures que sur les précipitations.

Conclusion :

Pour résumer nous avons donc vu dans ce dossier que le réchauffement climatique actuellement en place est clairement identifiable à Lyon, avec une bonne hausse des températures maximales en juillet et une hausse plus modérées mais non négligeable de ces mêmes températures maximales en janvier sur la période allant de 1956 à 2005. Pour le mois de janvier, cette hausse des températures maximales s'accompagne dans le même temps d'une hausse de l'insolation (et d'une baisse des précipitations) qui semble donc nous conduire vers des mois de janvier de plus en plus cléments dans les années à venir. Le mois de juillet s'accompagne lui d'une baisse inexplicquée de l'insolation. Nous avons cependant déjà rappelé que le climat est en perpétuel évolution et que des variations (naturelles ou non) de ses divers paramètres sont bien sur tout à fait normales même si ces variations ne sont pas toujours expliquées et/ou explicables. En juillet, l'insolation a une influence significative sur les températures maximales, et on peut espérer avec une bonne confiance connaître des températures très agréables lorsque le ciel est dégagé car on a vu qu'une bonne insolation est généralement synonyme de températures maximales élevées, même si notre étude a été réalisée à l'échelle du mois et non de la journée et que des exceptions restent possibles. En revanche en janvier par contre, l'insolation du mois n'a pas vraiment d'influence sur les températures maximales, et une journée ensoleillée ne garantit absolument pas des températures douces. Enfin les deux mois connaissent des moyennes de précipitation relativement proches, ce point étant le seul paramètre météorologique les rapprochant l'un de l'autre, même si les précipitations qui tombent au mois de janvier et au mois de juillet sont de caractère très différent (orageux et instable contre régulier et uniforme), comme le prouve l'écart significatifs de leur variance.

Juillet	Tn	Tx	Tm	Ins	Pluie
1956	14,2	24,4	19,3	283,7	94,1
1957	14,3	26,1	20,2	283,7	46
1958	15,2	25,6	20,4	264,4	32,1
1959	15,9	29,3	22,6	328,9	19
1960	13,2	23,2	18,2	239,6	122,9
1961	14,1	25,3	19,7	286,7	76,3
1962	13,5	26,1	19,8	314,5	25
1963	15	26	20,5	324,7	76
1964	15	29,4	22,2	352,0	34,2
1965	13,7	23,5	18,6	232,1	152,1
1966	13,4	23,6	18,5	263,0	52,5
1967	15,6	28,2	21,9	344,1	28,6
1968	14,1	25,7	19,9	285,4	56,4
1969	15,3	26,3	20,8	295,9	21,6
1970	14,3	25,1	19,7	280,1	47,3
1971	15	28,2	21,6	330,3	30,2
1972	14	25,2	19,6	262,7	58,4
1973	14,7	25,5	20,1	261,2	101,8
1974	13,5	25,9	19,7	311,1	47,9
1975	14,6	27,2	20,9	331,2	45,8
1976	16,5	28,7	22,6	299,7	84
1977	14,3	24,3	19,3	269,5	149,7
1978	14,2	25,6	19,9	256,6	45,9
1979	14,5	26,9	20,7	334,1	31,7
1980	13,2	24	18,6	239,2	83,7
1981	14,2	24	19,1	232,1	83
1982	16,9	28,9	22,9	266,0	58,1
1983	19,4	32,8	26,1	338,7	10,9
1984	15,2	27,6	21,4	346,7	23,6
1985	16	27,2	21,6	315,1	80,2
1986	15,2	26	20,6	341,8	60,6
1987	16,3	25,9	21,1	229,9	48,5
1988	14,8	26,2	20,5	318,1	96
1989	17	28,4	22,7	199,8	37,7
1990	15,8	27,8	21,8	316,0	71,4
1991	17	29	23	303,4	36,1
1992	16	26,6	21,3	256,9	56,8
1993	14,7	25,1	19,9	260,5	109,8
1994	18,1	30,1	24,1	302,1	55,3
1995	18,1	29,7	23,9	301,9	19
1996	14,8	25,2	20	263,9	123,1
1997	15,6	25,6	20,6	240,7	59,2
1998	17,1	27,9	22,5	282,5	41,4
1999	17,2	28,4	22,8	289,6	71,2
2000	14,7	25,5	20,1	246,1	40
2001	17,1	27,7	22,4	281,2	71,4
2002	15,6	26	20,8	264,5	103,6
2003	18,2	29,8	24	316,5	51,6
2004	16,5	27,7	22,1	286,3	40,8
2005	17,7	28,3	23	282,5	29

Janvier	Tn	Tx	Tm	Ins	Pluie
1956	1,3	7,2	4,3	53,2	73,7
1957	-2,9	4,4	0,8	76,2	34,4
1958	0,8	7	3,9	71,7	37,8
1959	0,3	6,9	3,6	104,4	45,2
1960	-1,2	5,3	2,1	58,6	66,2
1961	0,5	6,4	3,5	36,8	64
1962	2,1	7,9	5,0	80,6	45,1
1963	-7	0,1	-3,5	77,8	41,5
1964	-2,7	1,5	-0,6	35,5	45,2
1965	1	6,4	3,7	53,3	54,6
1966	-1,6	5	1,7	62,5	47,5
1967	-0,5	5,8	2,7	80,7	41,2
1968	-2,2	4	0,9	53,9	79,3
1969	0,8	7,8	4,3	78,3	53,3
1970	1	7,3	4,2	40,0	73,2
1971	-4	4,7	0,4	84,5	43,5
1972	0,3	6,7	3,5	67,8	63,7
1973	-0,9	3,7	1,4	55,0	63,1
1974	2,8	10,1	6,5	78,1	33,8
1975	3,1	9,1	6,1	59,7	80,4
1976	0,1	5,8	3,0	55,9	29,9
1977	1,9	7,3	4,6	53,3	104
1978	-0,1	5,2	2,6	57,6	56,2
1979	-2,7	3,1	0,2	68,5	109,2
1980	-0,2	5,1	2,5	69,9	35,7
1981	-1,7	3,7	1,0	85,5	62,7
1982	2,3	8,7	5,5	62,6	57,9
1983	0,7	7,3	4,0	53,5	21,8
1984	1	7,4	4,2	60,4	63,3
1985	-5,8	0,6	-2,6	48,3	38,6
1986	0,2	5,6	2,9	47,9	63,6
1987	-4,1	0,1	-2,0	41,4	32,2
1988	4,2	10,2	7,2	61,3	68,4
1989	0,5	7,1	3,8	81,1	9,2
1990	-1,6	6,3	2,4	80,7	42
1991	0,7	6,2	3,5	66,5	21,2
1992	-1,8	3,5	0,9	57,0	25,8
1993	1	8,7	4,9	88,9	18,6
1994	1,1	6,7	3,9	62,2	121,8
1995	1,4	7,2	4,3	75,1	80,8
1996	2,5	8,8	5,7	53,3	38,2
1997	-1,2	3,9	1,4	36,5	72,4
1998	2,1	8,1	5,1	76,5	48,6
1999	1,7	9,1	5,4	86,3	69,8
2000	-0,1	6,1	3,0	89,0	9,6
2001	3,4	9,2	6,3	50,6	70,4
2002	-0,9	7,7	3,4	114,8	25,8
2003	-0,2	5,1	2,5	106,7	39,8
2004	1,3	6,6	4,0	109,5	64,2
2005	-0,1	6,7	3,3	117,9	31,4