

Séance de travaux dirigés 7

Démarche statistique

Exercice 1 : Comparaison de mesures de la teneur en protéine

Pour évaluer la teneur en protéine de blé à l'échelle de la parcelle, deux méthodes sont disponibles:

- Méthode *bottillon* : mesure la teneur en protéine sur un échantillon prélevé au champ, que l'on appelle un bottillon ;
- Méthode *moissonneuse* : calcul automatique de la quantité de protéine au moment de la moisson.

L'avantage de la méthode *moissonneuse* est que le calcul est exact ; l'inconvénient est que l'on ne connaît la teneur en protéine qu'une fois que le champ est fauché. Or il est intéressant de ne faucher que lorsque la teneur en protéine est suffisamment élevée. D'où l'utilisation de la méthode *bottillon*.

Dans cet exercice, on cherche à vérifier que la méthode *bottillon* est fiable, par comparaison avec la méthode *moissonneuse*, considérée comme exacte. On dispose pour cela de 188 paires de mesures réalisées dans des conditions différentes (plusieurs variétés, différentes doses et dates d'apport d'azote).

La figure 1 présente les données.

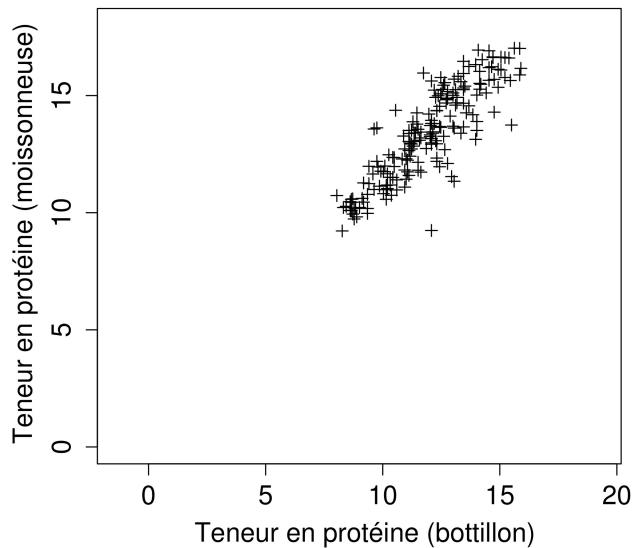


Figure 1: Teneur en protéine (en pourcentage) selon deux méthodes de calcul

1. Définir un modèle permettant de répondre à la question posée en introduction de l'exercice.

Le traitement de ces données par le modèle défini ci-dessus a conduit au tableau 1.

2. Donner une estimation du coefficient de corrélation entre les mesures du taux de protéines par les deux méthodes.
3. Donner une estimation des paramètres du modèle de la question 1. Calculer un intervalle de confiance de niveau de confiance 95 % de ces paramètres.
4. Quelles hypothèses sur les paramètres du modèle correspondent à l'assertion : les deux procédés donnent les mêmes résultats ?
5. Construire le test de cette assertion pour un risque maximal inférieur à 5% de décider à tort que les deux procédés ne donnent pas les mêmes résultats.
6. Conclure à partir du tableau 2.

```

Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.60731   0.46782   5.573 8.67e-08
x           0.90174   0.03877  23.260 < 2e-16

Residual standard error: 1.006 on 186 degrees of freedom
Multiple R-Squared:  0.7442,    Adjusted R-squared:  0.7428
F-statistic: 541 on 1 and 186 DF,  p-value: < 2.2e-16

```

Table 1: Teneur en protéine: résultats de la régression (sortie R)

R script

```

> proba = seq(from=0.850,to=0.999,by=0.001)
> proba

[1] 0.850 0.851 0.852 0.853 0.854 0.855 0.856 0.857 0.858 0.859 0.860 0.861 0.862
[14] 0.863 0.864 0.865 0.866 0.867 0.868 0.869 0.870 0.871 0.872 0.873 0.874 0.875
[27] 0.876 0.877 0.878 0.879 0.880 0.881 0.882 0.883 0.884 0.885 0.886 0.887 0.888
[40] 0.889 0.890 0.891 0.892 0.893 0.894 0.895 0.896 0.897 0.898 0.899 0.900 0.901
[53] 0.902 0.903 0.904 0.905 0.906 0.907 0.908 0.909 0.910 0.911 0.912 0.913 0.914
[66] 0.915 0.916 0.917 0.918 0.919 0.920 0.921 0.922 0.923 0.924 0.925 0.926 0.927
[79] 0.928 0.929 0.930 0.931 0.932 0.933 0.934 0.935 0.936 0.937 0.938 0.939 0.940
[92] 0.941 0.942 0.943 0.944 0.945 0.946 0.947 0.948 0.949 0.950 0.951 0.952 0.953
[105] 0.954 0.955 0.956 0.957 0.958 0.959 0.960 0.961 0.962 0.963 0.964 0.965 0.966
[118] 0.967 0.968 0.969 0.970 0.971 0.972 0.973 0.974 0.975 0.976 0.977 0.978 0.979
[131] 0.980 0.981 0.982 0.983 0.984 0.985 0.986 0.987 0.988 0.989 0.990 0.991 0.992
[144] 0.993 0.994 0.995 0.996 0.997 0.998 0.999

> round(qt(proba,df=186),digits=3)
> # Quantiles de la loi de Student

[1] 1.039 1.044 1.048 1.052 1.057 1.061 1.066 1.070 1.074 1.079 1.083 1.088 1.093
[14] 1.097 1.102 1.106 1.111 1.116 1.120 1.125 1.130 1.135 1.139 1.144 1.149 1.154
[27] 1.159 1.164 1.169 1.174 1.179 1.184 1.189 1.194 1.199 1.204 1.210 1.215 1.220
[40] 1.225 1.231 1.236 1.241 1.247 1.252 1.258 1.263 1.269 1.275 1.280 1.286 1.292
[53] 1.298 1.304 1.309 1.315 1.321 1.327 1.333 1.340 1.346 1.352 1.358 1.365 1.371
[66] 1.378 1.384 1.391 1.397 1.404 1.411 1.418 1.424 1.431 1.438 1.446 1.453 1.460
[79] 1.467 1.475 1.482 1.490 1.497 1.505 1.513 1.521 1.529 1.537 1.545 1.554 1.562
[92] 1.570 1.579 1.588 1.597 1.606 1.615 1.624 1.634 1.643 1.653 1.663 1.673 1.683
[105] 1.694 1.704 1.715 1.726 1.737 1.749 1.760 1.772 1.784 1.797 1.809 1.822 1.836
[118] 1.849 1.863 1.878 1.892 1.907 1.923 1.939 1.956 1.973 1.991 2.009 2.028 2.048
[131] 2.068 2.090 2.112 2.136 2.161 2.187 2.215 2.244 2.276 2.310 2.347 2.387 2.431
[144] 2.481 2.537 2.603 2.681 2.780 2.914 3.135

```

Table 2: Quantiles $t_p^{(186)}$ de la loi de Student \mathcal{T}_{186} d'ordre $0.850 \leq p \leq 0.999$.

Exercice 2 : Influence du déboisement sur la relation entre température de l'eau et température de l'air

On cherche à étudier l'influence du déboisement d'une rivière sur la relation entre la température de l'eau et la température de l'air. On dispose pour cela de mesures conjointes de la température d'eau dans la Flume et de la température de l'air effectuées avant le déboisement et après le déboisement. L'unité expérimentale est une décennie et les mesures de températures sont en fait des maxima sur une décennie.

Le tableau des données contient 56 lignes correspondant aux 28 décades d'observation avant le déboisement et aux 28 décades d'observation après le déboisement.

Le tableau 3 donne les tests séparés des différences des températures moyennes de l'air et de l'eau entre les deux périodes, celle qui précède le déboisement et celle qui lui succède.

```
> mod = lm(EAU~DEBOIST,data=dta)
> anova(mod)
```

Analysis of Variance Table

Response: EAU

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
DEBOIST	??	0.02	0.0172	7e-04	0.9789
Residuals	??	1318.50	24.4167		

```
> mod = lm(AIR~DEBOIST,data=dta)
> anova(mod)
```

Analysis of Variance Table

Response: AIR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
DEBOIST	??	1.43	1.4287	0.0623	0.8038
Residuals	??	1237.80	22.9222		

Table 3: Tests séparés des différences des températures moyennes de l'air et de l'eau entre les deux périodes (avant et après le déboisement).

1. Quels sont les nombres de degrés de liberté remplacés par ?? dans le tableau 3 ?

2. Doit-on considérer que la différence des températures moyennes de l'eau entre les deux périodes est significative ? Même question pour la température de l'air ?

A partir de ces données, le graphique de la figure 2 représente le lien entre les températures mesurées dans l'eau et dans l'air. La forme de ce nuage de points suggère que les variations de la température de l'eau sont proportionnelles aux variations de la température de l'air. Dans la suite, le coefficient de proportionnalité est appelé coefficient de transfert de chaleur entre l'air et l'eau.

3. A partir du graphique de la figure 2, proposer un modèle du lien entre la température de l'eau et la température de l'air permettant de tester un effet du déboisement sur le coefficient de transfert. Vous donnerez explicitement :

- la variable à expliquer,
- les variables explicatives,
- les paramètres des effets.

4. Montrez que le test de l'égalité des coefficients de transfert avant et après le déboisement peut être vu comme un test de comparaison de deux modèles emboîtés.

Le tableau 4 donne les tables d'analyse de la variance du modèle de la question 3, ajusté sur les données dont on dispose et du sous-modèle obtenu en supposant égaux les coefficients de transfert avant et après le déboisement.

5. A partir du tableau 4, calculez la statistique de Fisher du test de comparaison de modèle introduit dans la question 4 (remplacé par ?? dans le tableau). Peut-on dire que le déboisement a modifié de manière significative le coefficient de transfert ?

Le tableau 5 donne les résultats de l'ajustement du modèle de la question 3 sur les données dont on dispose.

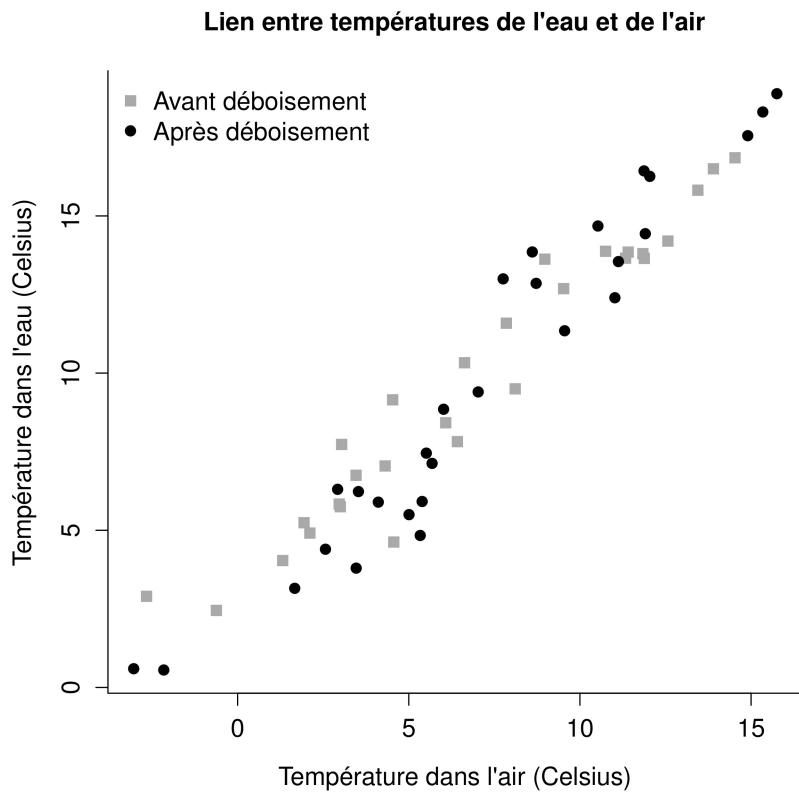


Figure 2: Températures dans l'air et dans l'eau, avant et après déboisement

R script

```
> mod0 = lm(EAU~AIR+DEBOIST,data=dta)
> anova(mod0)
```

Analysis of Variance Table

```
Response: EAU
          Df  Sum Sq Mean Sq F value Pr(>F)
AIR         1 1224.22 1224.22 696.313 <2e-16
DEBOIST     1    1.12    1.12   0.636 0.4287
Residuals  53   93.18   1.76
```

```
> mod = lm(EAU~AIR*DEBOIST,data=dta)
> anova(mod)
```

Analysis of Variance Table

```
Response: EAU
          Df  Sum Sq Mean Sq F value Pr(>F)
AIR         1 1224.22 1224.22 765.5109 < 2e-16
DEBOIST     1    1.12    1.12   0.6993 0.40686
AIR:DEBOIST 1      ??      ??       ?? 0.01548
Residuals  52   83.16   1.60
```

Table 4: Tables d'analyse de la variance du modèle de la question 3 (en bas) et du sous-modèle obtenu sous l'hypothèse d'égalité des coefficients de transfert avant et après le déboisement (en haut).

```
> mod = lm(EAU~AIR*DEBOIST,data=dta)
> summary(mod)

Call:
lm(formula = EAU ~ AIR * DEBOIST, data = dta)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.0004	-0.5763	-0.2083	0.8624	2.6446

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.95448	0.43399	4.504	3.81e-05
AIR	1.08259	0.05018	21.576	< 2e-16
DEBOISTAVANT	1.55301	0.60977	2.547	0.0139
AIR:DEBOISTAVANT	-0.18003	0.07191	-2.503	0.0155

Residual standard error: 1.265 on ?? degrees of freedom
 Multiple R-squared: 0.9369, Adjusted R-squared: 0.9333
 F-statistic: 257.5 on 3 and ?? DF, p-value: < 2.2e-16

Table 5: Ajustement du modèle liant la température de l'eau à celle de l'air et intégrant l'effet du déboisement.

6. Quel est le nombre de degrés de liberté remplacé par ?? dans le tableau 5 ?
7. A partir du tableau 5, quelle est la valeur prédictée par le modèle de la température de l'eau dans une décennie post-déboisement au cours de laquelle la température de l'air est de 10 degrés Celsius ? Même question dans une période pré-déboisement ?
8. Quel impact a le déboisement sur le coefficient de transfert (le transfert de chaleur de l'air vers l'eau est-il plus rapide avant ou après le déboisement) ?

