

STATISTIQUES INFÉRENTIELLES

Introduction

Nous allons observer et tester les valeurs données concernant le nombre de panneau photovoltaïque (SolWatt) dans chaque commune belge. Le principe de la statistique inférentielle est d'estimer en utilisant des tests sur un échantillon si une population correspond ou non à une hypothèse donnée.

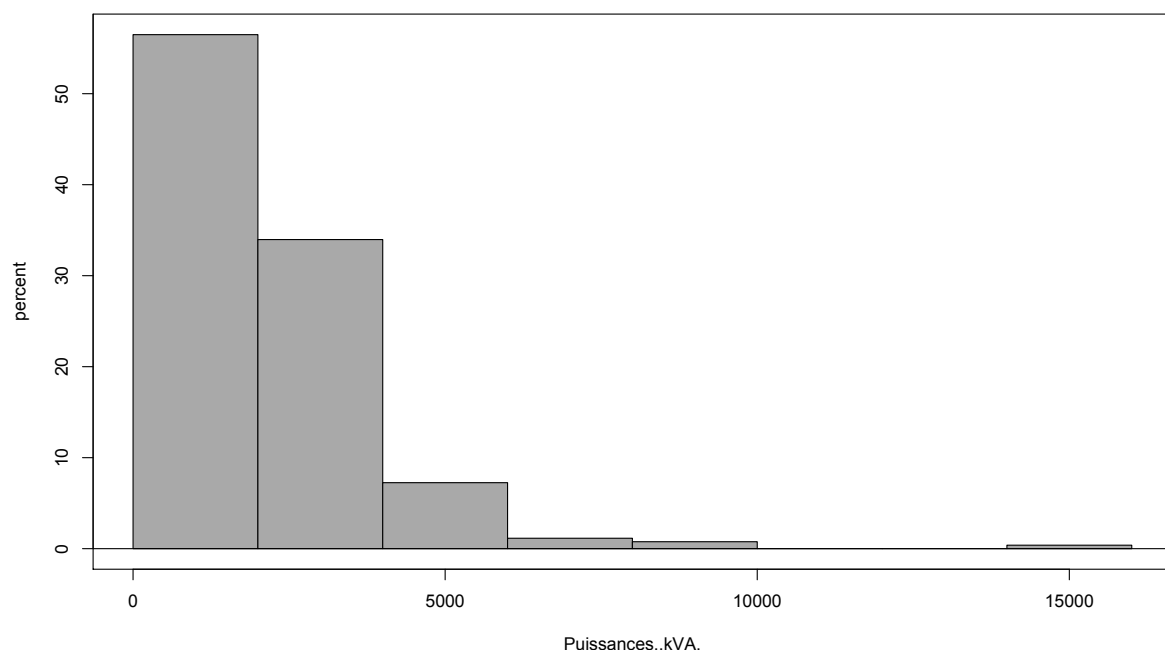
« La *p*-valeur est la probabilité d'obtenir la même valeur (ou une valeur encore plus extrême) du test si l'hypothèse H_0 était vraie. Contrairement à ce qui est parfois écrit, la valeur-*p* n'est pas la probabilité que l'hypothèse nulle soit vraie. La procédure généralement employée consiste à comparer la valeur-*p* à un seuil préalablement défini (traditionnellement 5 %). Si la valeur-*p* est inférieure à ce seuil, on rejette l'hypothèse nulle en faveur de l'hypothèse alternative, et le résultat du test est déclaré « statistiquement significatif ». Dans le cas contraire, si la valeur-*p* est supérieure au seuil, on ne rejette pas l'hypothèse nulle, et on ne peut rien conclure quant aux hypothèses formulées. »

Test de catégorie 1

- Nos données retiré via R :

Moyenne	sq	Ecart inter-cartil	Coefficient viariation	Quantil 0%	Quantil 25%	Quantil 50%	Quantil 75%	Quantil 100%	N
2138.2	1570.332	1711.899	0.7344178	295.289	1099.506	1783.863	2811.405	14452.2	262

Histogramme



Test de normalité

Etant donné que notre $n > 30$ nous n'effectuerons pas cette vérification.

Test de valeur:

On considère que l'offre de Solwatt peut être encouragée à partir du moment où la puissance des photovoltaïque d'une commune atteint 2000 kVA. On a une moyenne de 2138 dans notre échantillon alors peut-on considérer que c'est significativement plus grand que 2000?

En effet, tant que je reste en dessous de 2000 ça va bien. Si j'ai 2010 on peut encore considérer que c'est ok car proche de 2000 mais 2138 est-ce vraiment significatif? L'écart que nous observons entre 2138 et 2000 est-il dû à notre échantillon ou est-ce une réelle certitude de dire que nous dépassons les 2000 kVA?

$H_0 : \mu = 2000$

$H_1 : \mu > 2000$

Test-Univarié One Sample T-Test

Data: Puissances..kVA.

$t = 1.4245$, $df = 261$, $p\text{-value} = 0.07775$ (7%)

Alternative hypothesis: true mean is greater than 2000

95 percent confidence

interval: 1978.055 Inf

sample estimates:

mean of x

2138.2

Au vu de ce test nous pouvons partir sur la conclusion que nous ne rejetons pas H_0 car la p-valeur n'est pas au-delà de 5%. Nous avons donc 7% de chance de trouver un échantillon de ce genre là. Dans ce cas, on pourra pas dire significativement dire que notre moyenne est toujours au-delà de 2000 kVA.

Test de catégorie 2

Introduction

La puissance des photovoltaïque est-elle en moyenne la même au Luxembourg qu'au Brabant Wallon ?

Exposé des données :

1. Luxembourg

Moyenne	sq	Ecart inter-cartil	Coefficient viariation	Quantil 0%	Quantil 25%	Quantil 50%	Quantil 75%	Quantil 100%	N
1543.867	962.1751	1325.269	0.6232243	295.289	825.3015	1223.032	2150.57	4536.85	44

2. Brabant Wallon

Moyenne	sq	Ecart inter-cartil	Coefficient viariation	Quantil 0%	Quantil 25%	Quantil 50%	Quantil 75%	Quantil 100%	N
1981.07	1008.11251	1014.779	0.5088725	509.875	1333.314	1811.507	2348.093	4721.011	27

Test de normalité

- Luxembourg

Pas nécessaire car $n > 30$

- Brabant Wallon

Test effectué en ligne de commande dans R via Shapiro-Wilk :

data: rnorm(100, mean = 1981.07, sd = 1008.11251) W = 0.98687, p-value = 0.4289

Notre P-Valeur n'est pas significative et donc l'échantillon suit une loi normale.

F test to compare two variances

data: Puissances..kVA. by NA..1

F = 1.0978, num df = 26, denom df = 43, p-value = 0.7692

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.5612284 2.2823139

sample estimates:

ratio of variances 1.097765

On considère que notre test des variances est vérifié car la p-value est au dessus de 5%. On peut donc passer au test des moyennes

Test de valeur

H_0 : moyenne Brabant = moyenne Luxembourg

H_1 : moyenne Brabant > moyenne Luxembourg

Welch Two Sample t-test

data: Puissances..kVA. by province

t = 1.8048, df = 53.15, p-value = 0.07677

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-48.63922 923.04580

sample estimates:

mean in group Brabant mean in group Luxembourg

1981.070

1543.867

notre p-valeur est de 7% (donc > 5%) on pourra dire que notre H_0 n'est pas rejeté. Donc on ne peut pas significativement dire que nos moyennes sont équivalentes.