

DATA VERWERKING ZONDER PERIODIEKE AFHANKELIJKHEID

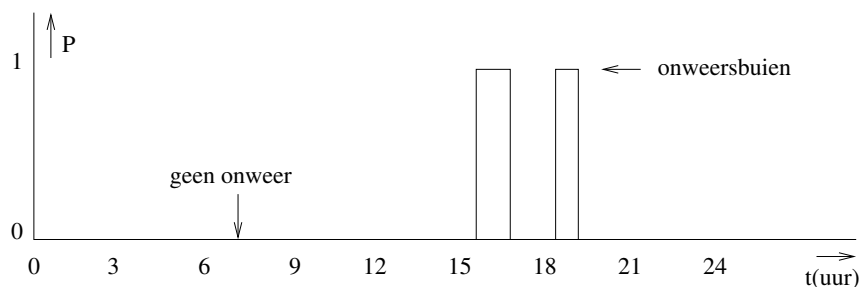
N.G. SCHULTHEISS

1. INLEIDING

Deze module volgt op de module “Data verwerking met periodieke afhankelijkheid”, waarin werd uitgelegd hoe je een spreadsheet met gegevens van de detectoren kunt maken. Hier werd uitgelegd hoe conclusies te trekken zijn, als er sprake is van een periodieke afhankelijkheid. In het geval van de afhankelijkheid van weersverschijnselen zoals bliksem, zal dit niet periodiek zijn. Het is namelijk niet zo dat er om een bepaalde tijd een bliksemflits wordt gegenereerd.

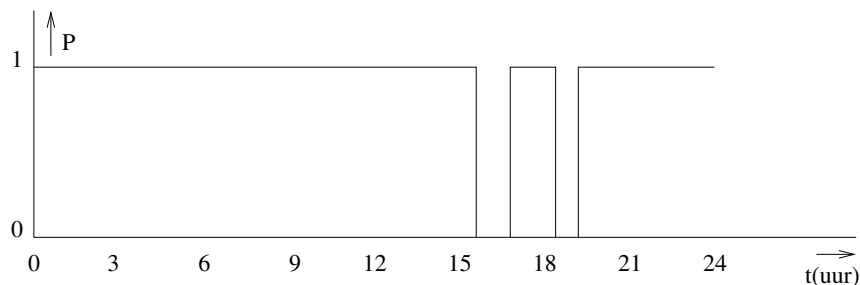
2. KANSEN

Om iets over de afhankelijkheid van een verschijnsel te kunnen zeggen, bekijken we eerst de kans (probability: $P()$) dat het verschijnsel heeft plaatsgevonden. Een dergelijk grafiek is te zien in figuur 2.1.



FIGUUR 2.1. De kans op onweer

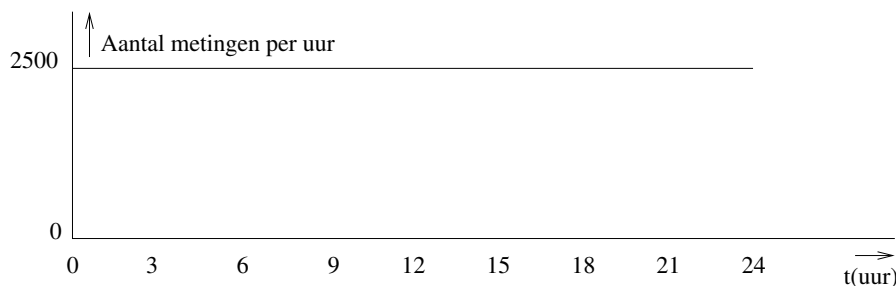
We moeten eigenlijk twee kansen vergelijken, de kans dat er onweer is en de kans dat er geen onweer is. De kans dat er geen onweer is, is te zien in figuur 2.2.



FIGUUR 2.2. De kans dat er geen onweer is.

Opdracht 1: Leg uit dat op ieder tijdstip geldt: $P(\text{onweer}) + P(\text{geenOnweer}) = 1$.

Laten we eerst aannemen dat de kosmische straling niet afhangt van het weer. In grafiek 2.3 is de kosmische straling daarom constant genomen. Om te kijken of de kosmische straling nu afhangt van het onweer, vermenigvuldigen we op alle tijdstippen de kans op onweer met het aantal coïncidenties. Als we alle de uitkomsten voor de hele dag optellen krijgen we een getal. Hetzelfde is te doen met de kans dat er geen onweer is. Nu krijgen we weer een getal. Als we deze getallen vergelijken, lijkt het of de kans op kosmische straling bij mooi weer groter is.



FIGUUR 2.3. Constante kosmische straling

Opdracht 2: Toon aan dat dit resultaat te corrigeren is door in beide gevallen te delen door het oppervlak onder de kans-grafieken.

3. EEN PROCEDURE VOOR EEN SPREADSHEET

We kunnen dit op de volgende manier in een spreadsheet verwerken:

- Laad een .csv bestand van “<http://data.hisparc.nl>” in een spreadsheetprogramma.
- Bereken de kans voor ieder uur dat er onweer was en zet deze waarden in een “onweerkolom”. Is er in een uur een kwartier onweer geweest, dan geldt: $P(\text{onweer})=0,25$ (of 0.25 als het spreadsheetprogramma niet in het nederlands is geïnstalleerd).
- De kans dat er in dit geval geen onweer is, is te berekenen met: $P(\text{geenOnweer})=1-P(\text{onweer})$. Deze waarden staan in de kolom “geenOnweer”.
- Bereken de som voor de kolommen “onweer” en “geenOnweer”.
- Bereken op ieder tijdstip $\text{aantalCoïncidenties} \cdot P(\text{onweer})$ en $\text{aantalCoïncidenties} \cdot P(\text{geenOnweer})$ en zet deze waarden in twee kolommen.
- Bereken de som van $\text{aantalCoïncidenties} \cdot P(\text{onweer})$ en deel dit door de som van de kolom “onweer”, dit is “deKansOpOnweer”.
- Bereken de som van $\text{aantalCoïncidenties} \cdot P(\text{geenOnweer})$ en deel dit door de som van de kolom “geenOnweer”, dit is “geenKansOpOnweer”.
- We kunnen nu “deKansOpOnweer” en “geenKansOpOnweer” vergelijken. De grootste wint.

Met deze aanpak zijn er veel soortgelijke problemen op te lossen. We kunnen bijvoorbeeld kijken of er een hemellichaam boven de detector is. Op een vergelijkbare wijze is een grafiek te maken met de kans dat een hemellichaam boven de detector is en de kans dat dit hemellichaam niet boven de detector is.

Opdracht 3: Station 98 geeft de Nikhef weerinformatie, deze is te zien in figuur 3.1. Op het Nikhef zijn ook diverse detectoren van kosmische straling te vinden. Leg uit wat het probleem is, als we willen onderzoeken of de kosmische straling van de luchtdruk afhangt.

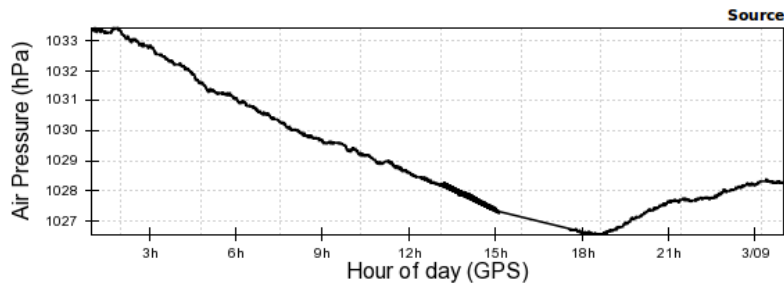
Als we op “source” klikken kunnen we hier ook een .csv bestand downloaden. Dit bestand werkt met een tijdstempel: het aantal seconden vanaf 1 januari 1970. Aangezien een uur 3600s heeft, kunnen we hier ook een tabel met de weergegevens per minuut / uur van maken.

4. CORRELATIE VAN TWEE GRAFIEKEN

Als we twee grafieken willen correleren, moeten we dit iets slimmer aanpakken. We gaan uit van de luchtdrukafhankelijk van de kosmische straling.

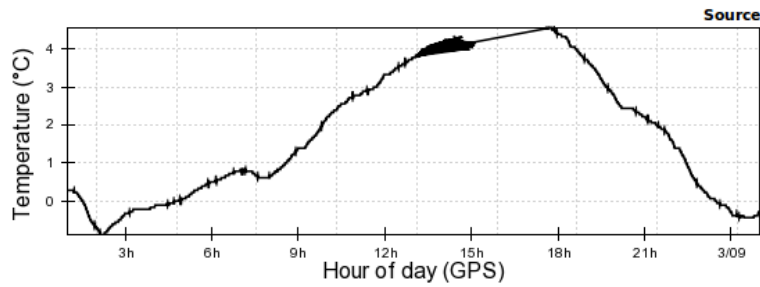
- Haal de gegevens van de kosmische straling op en zet deze in een spreadsheet.
- Haal de gegevens van het weer op en zet deze in een spreadsheet.
- Bepaal het gemiddelde van de luchtdruk over de dag.
- Bepaal hoeveel de luchtdruk op ieder moment afwijkt van het gemiddelde en zet deze waarden in een nieuwe kolom.
- Vermenigvuldig de afwijking met de kosmische straling en zet dit resultaat in een extra kolom.
- Tel alle waarden van de extra kolom op. Is het resultaat positief dan geeft een hogere luchtdruk meer kosmische straling. Is het resultaat negatief dan geeft een lagere luchtdruk meer kosmische straling.

Barometer data



March 2010						
Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				
2005 2006 2010						
Mar						
Station List						

Temperature data



FIGUUR 3.1. De weerdata voor 8 maart 2010

Opdracht 4: *Vergelijk de procedure hierboven met de procedure uit hoofdstuk 3. Toon aan dat de procedure van hoofdstuk 4 hetzelfde resultaat geeft als de procedure uit hoofdstuk 4.*