
Huiswerkopdracht 2014-2015

Statistiek - Huiswerkopdracht

01-12-2014

Opdracht 1

A. Gemiddelde: $(1,85 + 2,07 + 1,9 + 1,94 + 1,82 + 2,06 + 2,11 + 1,77 + 1,89 + 1,99) \div 10 = 1,94$

Mediaan: Eerst de prijzen per liter rangorden dan de middelste waarde nemen, dus:

1,77 1,82 1,85 1,89 1,90 1,94 1,99 2,06 2,07 2,11

Er zijn 10 waarden dus de mediaan ligt tussen 5^e en 6^e waarde: $(1,90 + 1,94) \div 2 = \text{€}1,92$

Standaardafwijking: $s^2 = ((1,85 - 1,94)^2 + (2,07 - 1,94)^2 + (1,90 - 1,94)^2 + (1,94 - 1,94)^2 + (1,82 - 1,94)^2 + (2,06 - 1,94)^2 + (2,11 - 1,94)^2 + (1,77 - 1,94)^2 + (1,89 - 1,94)^2 + (1,99 - 1,94)^2) : (10 - 1) = 0,1313333333$

$$s = \sqrt{s^2}$$

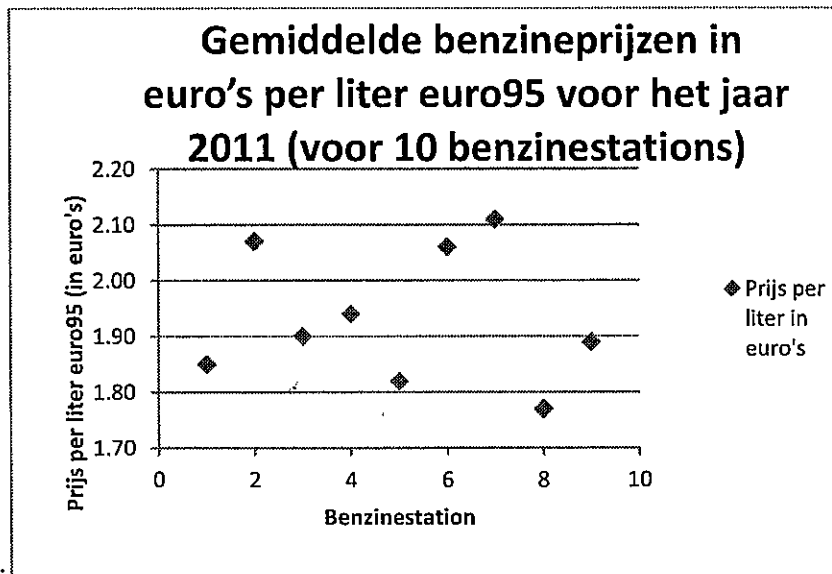
$$\sigma = \sqrt{0,1313333} = 0,1146$$

30^e percentiel: $P = \text{gem} \pm Z * SD$

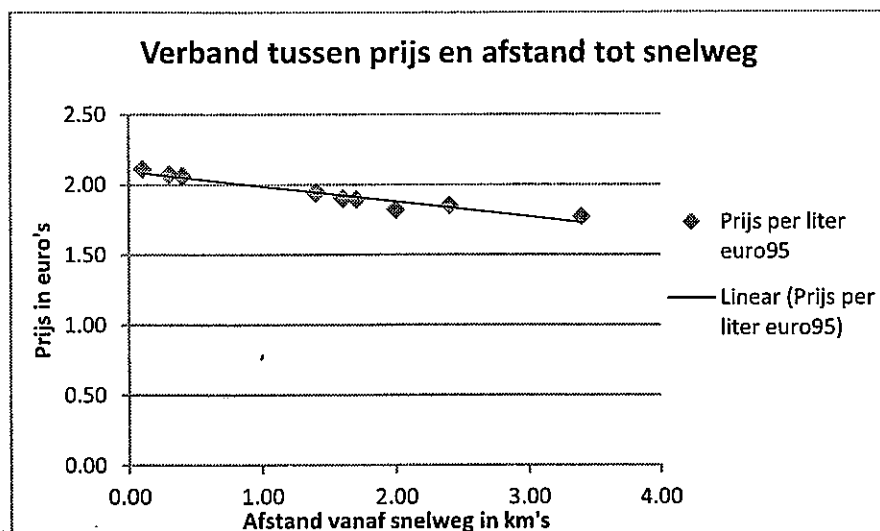
In Excel: =PERCENTIEL(A2:A11;0,3)= €1,878

$Lp = (10 + 1) \times (30 \div 100) = 3,3$. Hier ligt het 30^e percentiel dus op 0,3 afstand tussen de 3^e en 4^e waarde (1.85 en 1.89). Het 30^e percentiel is dus:

$$0,3 \times (1,89 - 1,85) + 1,85 = \text{€}1,862$$



B. Grafiek:



C. Grafiek:

D. De lineaire regressielijn laat zien dat de prijs per liter in euro's afneemt naarmate de afstand tot de snelweg groter wordt. Dit is in het echt ook zo omdat de exploitatiekosten langs de snelweg hoger zijn.

E. De prijs per liter euro 95 bij de snelweg is: €2,11 per liter.

De prijs per liter euro 95 één kilometer van de snelweg af is: €1,99 per liter.

Er zal dus in totaal: $500 \times (2,11 - 1,99) = €60$ worden bespaard.

F. Het gemiddelde wordt : $\frac{1.85+2.07+1.90+1.94+1.82+2.06+2.11+1.77+1.89+1.99+3.14}{11} = €2,05$

De mediaan: 1.77 1.82 1.85 1.89 1.90 1.94 1.99 2.06 2.07 2.11 3.14
:1,94

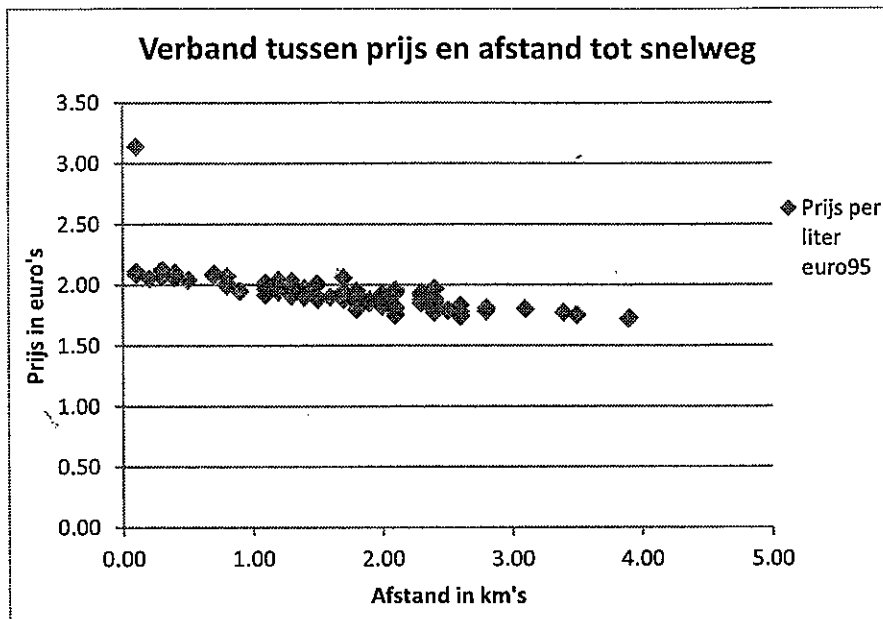
Het gemiddelde is niet heel veel veranderd ten opzichte van het antwoord bij vraag a (slechts 9cent). Dit komt doordat de steekproef iets groter is geworden en er nu door 11 wordt gedeeld in plaats van 10.

$$G. P(X \leq 9) = 1 - (P(X = 10) + P(X = 11)) = 1 - ((11 \text{ NCR } 10) \times (0,2)^{10} \times (0,8)^1) + ((11 \text{ NCR } 11) \times (0,2)^{11} \times (0,8)^0)$$

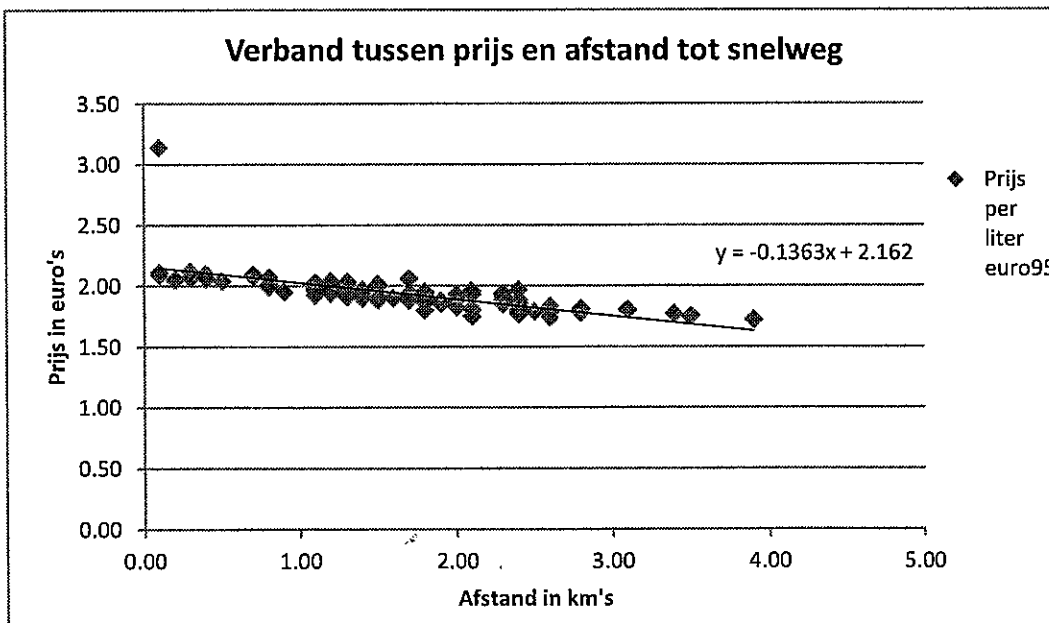
De kans dat er 9 of minder stations slachtoffer worden van een overval is dus:

$$P(X \leq 9) = 0,999999078$$

ii A. Grafiek:



B. Grafiek:

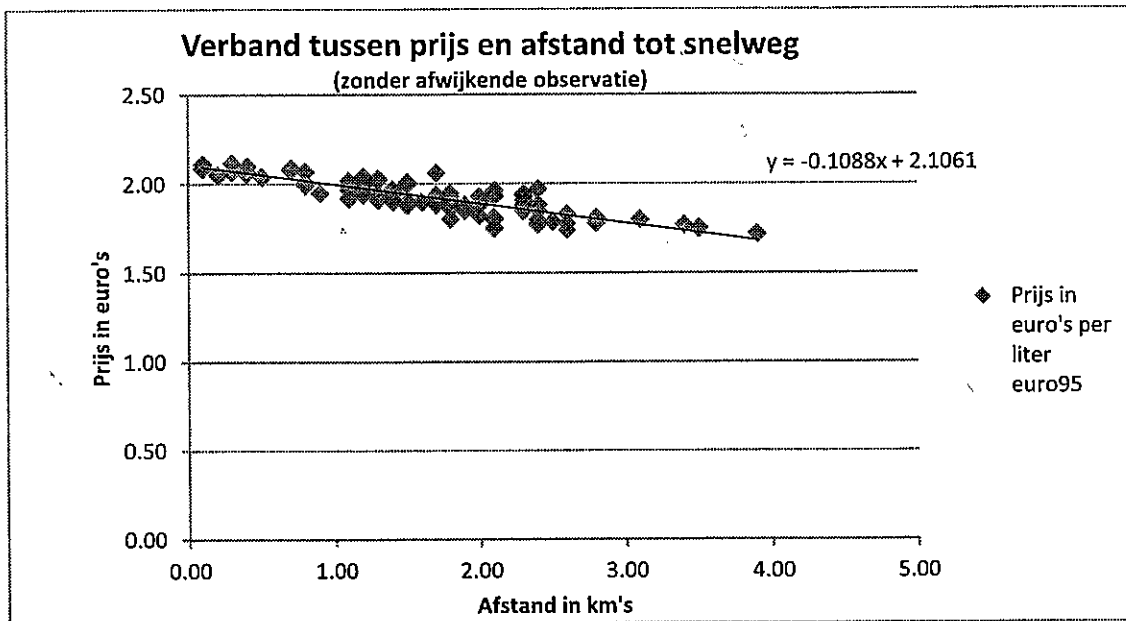


De formule van de trendlijn van de grote observatie is: $y = -0,1363x + 2,162$

De formule van de trendlijn van de kleine observatie is: $y = -0,1077x + 2,0935$

Je ziet in de formule een duidelijk verschil maar op het oog is er weinig verandering in de trendlijn zichtbaar. Het verschil is dus niet aanzienlijk groot. Beide trendlijnen laten een bijna gelijke lineair dalende trend zien.

C. Grafiek:



Opdracht 2

A. De kans dat project B slaagt is 10% dus: $P(B) = 0,1$.

Als project B slaagt is de kans 0,8 dat project A dan ook slaagt, de kans dat project A slaagt als project B niet slaagt is slechts 0,3.

De kans dat project A succesvol is is dus:

$$P(A) = (0,1 \times 0,8) + (0,9 \times 0,3) = 0,35$$

B. $P(A \text{ succesvol en } B \text{ succesvol}) = 0,8$

C. Het verschil tussen $P(A)$ en het antwoord bij vraag b ontstaat omdat $P(A)$ afhankelijk is van $P(B)$ en het antwoord bij vraag b is een kans die gegeven is wanneer al bekend is dat project B succesvol is geweest. Dus daar is $P(A)$ niet meer afhankelijk van B en dit zorgt voor een andere kans.

$$D. P(A \text{ of } B \text{ of allebei}) = P(A) + P(B) + P(A \text{ en } B) = 0,1 + (0,1 \times 0,8) + (0,9 \times 0,3) = 0,45$$

Opdracht 3

$$A. P(110 \leq X \leq 130) = P(X \leq 130) - P(X \leq 110)$$

$$P\left(\frac{130 - 140}{25}\right) = P(Z < -0,4)$$

$$P\left(\frac{110 - 140}{25}\right) = P(Z < -1,2)$$

$$P(X \leq 130) = 0,3446$$

$$P(X \leq 110) = 0,1151$$

$$P(110 \leq X \leq 130) = 0,3446 - 0,1151 = 0,2295$$

B. De 20,9% met de beste score mogen door naar de 2^e ronde. Dat betekent dat 79,1% lager scoort dan de minimale waarde om door te gaan naar ronde twee.

Z waarde hierbij is: 0,84

$$\sigma = 25 \quad \mu = 140$$

$$\frac{(x - 140)}{25} = 0,84$$

$$x = (0,84 \times 25) + 140$$

$$x = 161$$

De minimale score om toegelaten te worden tot ronde 2 is dus 161

Extra oefenopgaven Statistiek BDK 2013-2014

Meerkeuzevragen

1. Welke van onderstaande beweringen is juist?
 - a. Als de populatiecorrelatiecoëfficiënt tussen X en Y positief is, dan wordt een toename in Y (gedeeltelijk) veroorzaakt door een toename in X .
 - b. In een steekproef bestaande uit 100 waarnemingen getrokken uit een standaard normaal verdeelde populatie liggen tenminste 75 waarnemingen tussen -2 en 2.
 - c. De variatiebreedte (*range*) kan worden gebruikt om de variabiliteit van nominale data te beschrijven.
 - d. De t -verdeling is voor het eerst beschreven door John Tukey.
2. Voor een steekproef bestaande uit n observatieparen (x_i, y_i) (waarbij i loopt van 1 tot en met n) wordt een regressielijn geschat met behulp van de kleinste kwadratenmethode (*least squares method*). De variabele Y is de afhankelijke variabele en de variabele X is de onafhankelijke variabele. Het steekproefgemiddelde aangaande X is \bar{x} en het steekproefgemiddelde aangaande Y is \bar{y} . Wat kun je zeggen over de lokatie van het punt (\bar{x}, \bar{y}) ?
 - a. Het punt (\bar{x}, \bar{y}) ligt op de regressielijn.
 - b. Als de steekproefcorrelatiecoëfficiënt r_{xy} negatief is, dan ligt het punt (\bar{x}, \bar{y}) onder de regressielijn. Als r_{xy} positief is, dan ligt het punt (\bar{x}, \bar{y}) boven de regressielijn.
 - c. Hoe beter de regressielijn de data beschrijft, des te kleiner de afstand tussen het punt (\bar{x}, \bar{y}) en de regressielijn.
 - d. Zonder extra informatie kun je niets zeggen over het verband tussen de lokatie van het punt (\bar{x}, \bar{y}) en de regressielijn.
3. Vier studenten berekenen de steekproefcorrelatiecoëfficiënt tussen de toevalsvariabelen X en Y op basis van dezelfde steekproef. Slechts één student maakt geen fouten. Wie is dat?
 - a. Student A: $s_x = 1,61$, $s_y = 3,14$ en $s_{xy} = 5,77$ levert $r_{xy} = 1,14$.

- b. Student B: $s_x = 2,32$, $s_y = 3,14$ en $s_{xy} = -5,77$ levert $r_{xy} = 0,79$.
- c. Student C: $s_x = 1,61$, $s_y = -3,14$ en $s_{xy} = 2,23$ levert $r_{xy} = -0,44$.
- d. Student D: $s_x = 2,32$, $s_y = 3,14$ en $s_{xy} = -2,23$ levert $r_{xy} = -0,31$.

4. Beschouw de volgende steekproef:

observatie	1	2	3
x_i	1	3	4
y_i	4	4	2

Wat is de helling van de kleinste kwadratenlijn $\hat{y} = b_0 + b_1x$ behorend bij deze steekproef?

- a. Die helling is -0,57.
- b. Die helling is -0,23.
- c. Die helling is 0,23.
- d. Die helling is 0,57.

Open vraag

1. Een wetenschapper onderzoekt in welke mate intelligentie wordt overgedragen van moeders op dochters. Hij meet het IQ van 7 moeders en hun dochters. Dit levert de volgende steekproef op:

observatie	1	2	3	4	5	6	7
IQ moeder (x_i)	86	106	111	84	108	115	90
IQ dochter (y_i)	90	119	127	105	111	111	107

- a. Teken het spreidingsdiagram (*scatter diagram*) van deze data.
- b. Bereken de steekproefcorrelatiecoëfficiënt.
- c. Bereken de kleinste kwadratenschatting voor het lineaire regressiemodel

$$y = \beta_0 + \beta_1x + \epsilon,$$

waarbij het IQ van de moeder de onafhankelijke variabele is en het IQ van de dochter de afhankelijke variabele is.