

VWO 6, Wiskunde A

Computertoets WOON-bestand

Wat is het WOON-databestand?

Het databestand wonen is een selectie uit het WOONbestand van 2012, bestaat uit 117 duizend records en vertegenwoordigt ongeveer 7,1 miljoen huishoudens in een zelfstandige woning. Het geeft informatie over onder andere de samenstelling van het huishouden, de tevredenheid met de woning en woonomgeving, buurtcontacten en het vóórkomen van rommel op straat of overlast.

Beschrijving van variabelen in WOON

Variabelen in bestand WOON

Het bestand bevat onderstaande variabelen:

1 *Hweegwon_n* Huishoudensgewicht (doen we niets mee in deze PO)

2 *G4_3* Wel/niet vier grote steden (Amsterdam, Rotterdam, Den Haag, Utrecht)

met

1 G4

2 G27 (Groningen, Leeuwarden, Emmen, Almelo, Deventer, Enschede, Hengelo, Zwolle, Arnhem, Nijmegen, Amersfoort, Lelystad, Alkmaar, Haarlem, Zaanstad, Dordrecht, Leiden, Schiedam, Breda, Eindhoven, Helmond, Den Bosch, Tilburg, Heerlen, Maastricht, Venlo en Sittard-Geleen)

3 rest

3 *SamHH5* Samenstelling van het huishouden

met

1 Eenpersoonshuishouden

2 Paar of niet gezinshuishouden

3 Paar + kind(eren)

4 Eenoudergezin plus kind(eren)

4 *Huko* Huishouden is huurder of eigenaar koopwoning

met

1 Eigenaar koopwoning

2 Huurder huurwoning

5 *TWoning* Tevredenheid met huidige woning

met

1 Zeer tevreden

2 Tevreden

- 3 Niet tevreden, maar ook niet ontevreden
- 4 Ontevreden
- 5 Zeer ontevreden

6 TWoonOmg Tevredenheid met huidige woonomgeving

met

- 1 Zeer tevreden
- 2 Tevreden
- 3 Niet tevreden, maar ook niet ontevreden
- 4 Ontevreden
- 5 Zeer ontevreden

7 MensKen Mensen kennen elkaar in deze buurt nauwelijks

met

- 1 Helemaal mee eens
- 2 Mee eens
- 3 Niet mee eens, maar ook niet mee oneens
- 4 Oneens
- 5 Helemaal mee oneens

8 ConBuur1 Ik heb veel contact met mijn direct burens

met

- 1 Helemaal mee eens
- 2 Mee eens
- 3 Niet mee eens, maar ook niet mee oneens
- 4 Oneens
- 5 Helemaal mee oneens

9 ConBuur2 Ik heb veel contact met andere buurtbewoners

met

- 1 Helemaal mee eens
- 2 Mee eens
- 3 Niet mee eens, maar ook niet mee oneens
- 4 Oneens
- 5 Helemaal mee oneens

10 OBeklad Voorkomen bekladding van muren en gebouwen

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit

11 ORommel Voorkomen rommel op straat

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit

12 OHPoep Voorkomen hondenpoep op straat

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit

13 OVerniel Voorkomen vernieling van telefooncellen, bus- of tramhokje

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit
- 4 N.v.t., geen telefooncellen, bus- of tramhokjes in de buurt

14 OJong Overlast door jongeren

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit

15 OVerkeer Last van het verkeer

met

- 1 Vaak
- 2 Soms
- 3 (Bijna) nooit

16 Vorm Een-/meergezins huidige woning

met

- 1 Eengezinswoning
- 2 Meergezinswoning

17 TypWon Een-/meergezins en type eengezinswoning

met

- 1 Eengezinswoning, vrijstaand
- 2 Eengezinswoning, 2-onder-1-kap
- 3 Eengezinswoning, hoekwoning
- 5 Eengezinswoning, tussenwoning/overig
- 6 Etagewoning

18 BJrWon Bouwjaar woning

met

- 1 1945 of eerder (vooroorlogse woning)
- 2 1946 tot en met 1959
- 3 1960 tot en met 1970
- 4 1971 tot en met 1980
- 5 1981 tot en met 1990
- 6 1991 tot en met 2000
- 7 2001 of later

19 WozWaarde WOZ-waarde woning

20 VromHH5 Besteedbaar inkomen gecorrigeerd voor woonlasten

met

1 Tot € 17.600

2 € 17.600 tot € 25.600

3 € 25.600 tot € 35.800

4 € 35.800 tot € 49.300

5 € 49.300 en hoger

Opdracht 1 (intro) (3 ptn)

Open het bestand WOON

We kijken eerst naar de WOZ waarden van de huizen.

De gemeente stelt deze de waarde van uw woning vast of ander onroerend goed vast. De gemeente doet dit op basis van de Wet waardering onroerende zaken (WOZ). De WOZ-waarde bepaalt de hoogte van een aantal belastingen en gemeentelijke heffingen, zoals de onroerende zaakbelasting en het rioolrecht.

In het bestand WOON komen een aantal huizen met een hoge WOZ waarde.

We willen de WOZ waarden van verschillende groepen vergelijken en daarom kijken we enkel naar woningen met een WOZ waarde van minder dan 700000.

Filter de woningen met WOZ van 700000 of meer er uit.

Via Data en selectiefilter kun je de woningen die een WOZ waarde hebben die kleiner is dan 700000.

Doe dit en controleer of je nu onderstaande tabel kan krijgen:

WozWaarde	Freq.	Perc.
0-<50000	175	0,15
50000-<100000	3469	3,01
100000-<150000	18791	16,32
150000-<200000	28319	24,60
200000-<250000	24303	21,11
250000-<300000	14499	12,59
300000-<350000	8574	7,45
350000-<400000	5705	4,96
400000-<450000	3951	3,43
450000-<500000	2759	2,40
500000-<550000	1930	1,68
550000-<600000	1232	1,07
600000-<650000	819	0,71

650000-<700000	605	0,53
Totaal	115131	100%

Leg uit waarom de keuze gemaakt wordt van “enkel naar woningen met een WOZ waarde van minder dan 700000”.

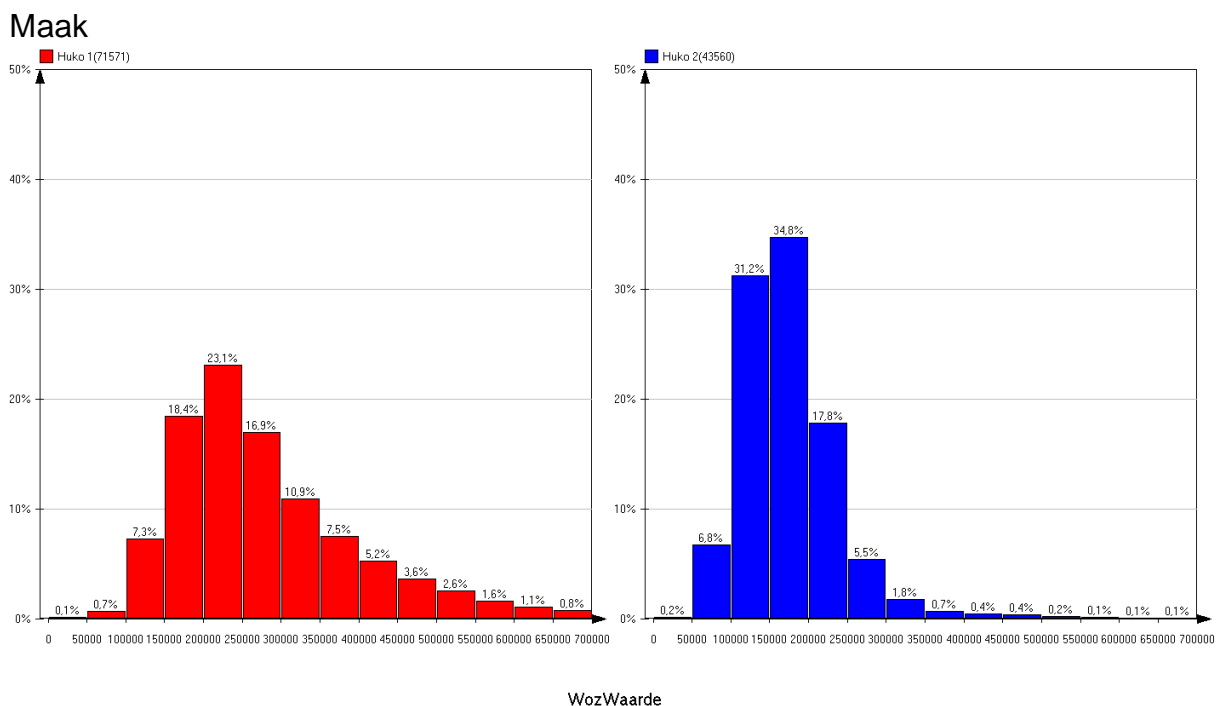
Laat voor de rest van deze PO dit filter staan.

Opdracht 2 (8 ptn)

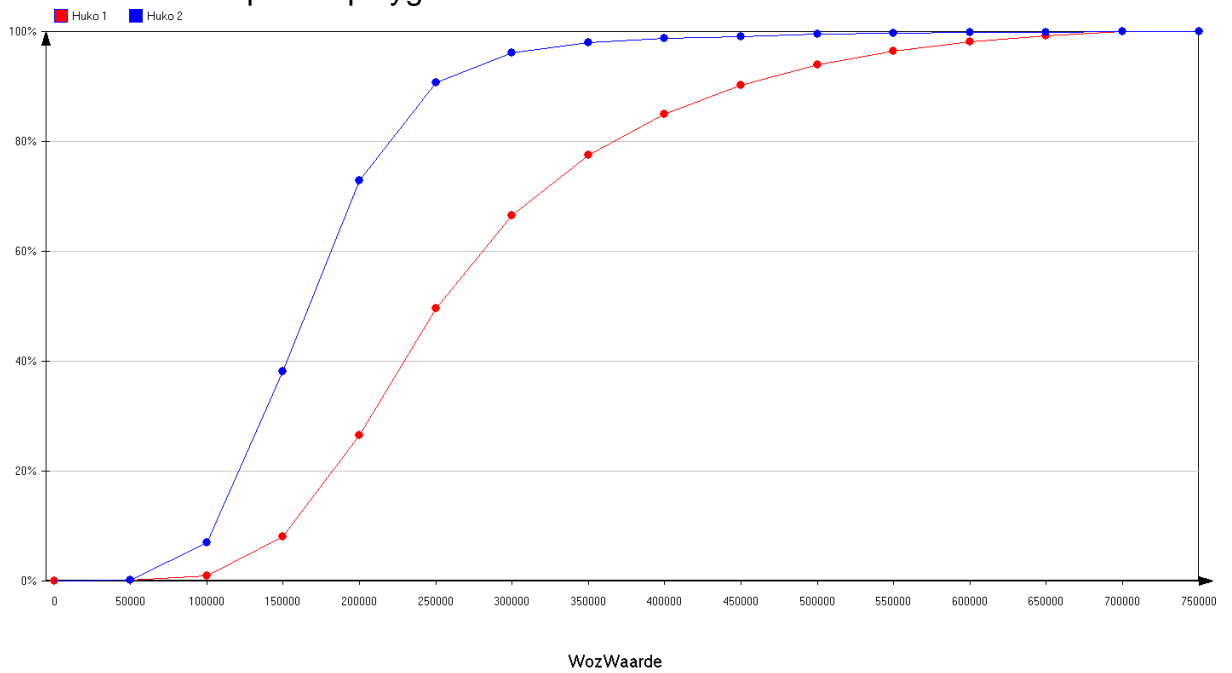
In deze opdracht kijken we naar verschillen in WOZ waarden van koopwoningen en huurwoningen.

Om de verschillen koopwoningen (Huko 1) en huurwoningen (Huko 2) te bekijken zijn verschillende representaties gemaakt met VUStat. Geef bij onderstaande representaties welke verschillen er tussen de twee groepen zichtbaar zijn in deze representaties (bijv. de WOZ waarde van de een is groter dan die van de ander; of de spreiding van de een is groter dan de ander, of de verdeling van een heeft een andere vorm dan die van de ander, enz.). Geen berekeningen maar wel een toelichting bij je antwoorden.

a) staafdiagram

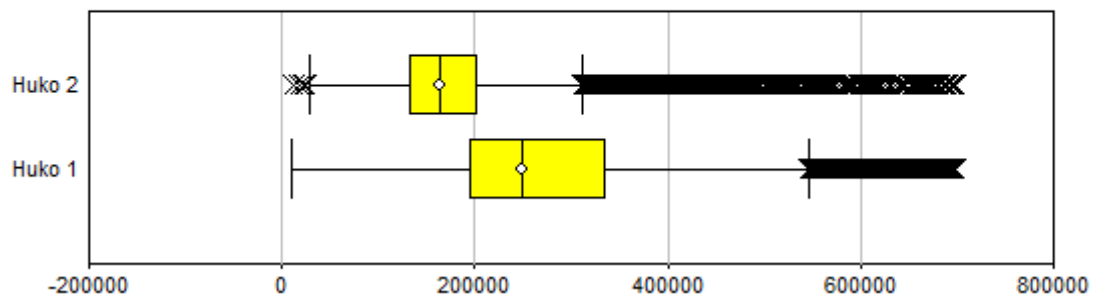


b) Cumulatieve frequentie polygonen:



c) Boxplots:

Variabele: WozWaarde



d) Kentallen

Huko	Huko 1	Huko 2
Aantal waarnemingen	71571	43560
Gemiddelde	278353,6	174631,2
Mediaan	250000	164000,0
Modus	235000	140000

Minimum	10000	12000
Maximum	699000	699000
Stand.Afw. steekproef	116554,19	66411,80
Stand_Afw.populat ie	116553,38	66411,04

Opdracht 3

(6 ptn)

Gebruik nu de vuistregels van het formuleblad om kwantitatief de verschillen tussen beide groepen uit opdracht 2 (koopwoningen en huurwoningen) te berekenen. Gebruik bij de representaties van opdrachten 2b,c,d een passende vuistregel van het formuleblad om de grootte van de verschillen te bepalen.

Opdracht 4 (12 ptn)

Doe nu zelf een dergelijk onderzoek naar de verschillen tussen de WOZ waarde van de G27 steden en de rest van het land. Maak verschillende representaties en bespreek per representatie de verschillen tussen de G27 steden en de rest van het land. Gebruik, indien mogelijk, passende vuistregels van het formuleblad om de grootte van de verschillen te bepalen.

Opdracht 5 (4 ptn)

Geef een 95% betrouwbaarheidsinterval van het percentage huishoudens dat tevreden is met hun huidige woningomgeving.

Opdracht 6 (6 ptn)

“Nog geen 20% van alle huizen in Nederland is van voor de oorlog (een vooroorlogse woning)”. Onderzoek deze uitspraak door middel van toetsen van hypothesen.

Opdracht 7 (12 ptn)

“Hoekwoningen hebben een hogere WOZ waarde dan tussenwoningen”. Om dit te onderzoeken bekijken we de statistische significantie en de praktische significantie.

- Leg in eigen woorden uit wat het verschil is tussen statistische significantie en praktische significantie.
- Onderzoek of de WOZ waarde voor hoekwoningen statistisch significant hoger is dan voor tussenwoningen.

c) Onderzoek of de praktisch significantie hoger is.

Opdracht 8 (4 ptn)

“De tevredenheid met de woning neemt toe als de WOZ waarde hoger is”.
Onderzoek deze uitspraak. Licht je antwoord toe.

Opdracht 9 (5 ptn)

Eengezinswoningen zijn woningen waar 1 gezin woont, meergezinswoningen zijn woningen waar meerdere huishoudens wonen. Deze meergezinswoningen zullen in het algemeen veel groter zijn en daarom duurder. Daarom mag je verwachten dat eengezinswoningen vaker koopwoningen zijn en meergezinswoningen vaker huurwoningen. Maar hoe groot is dit verschil eigenlijk?

Onderzoek met een 2x2 tabel en een passende vuistregel de grootte van het verschil tussen eengezins- en meergezinswoningen met betrekking tot koop- en huurwoningen.

Opdracht 10 (max 10 ptn)

Formuleer zelf een interessante vraag in deze context en beantwoord hem.

Hierbij kun je bijvoorbeeld denken aan

- In sommige buurten lijken de mensen elkaar nauwelijks te kennen. Iemand vermoedt dat dit vooral een probleem is in wijken in grote steden waarin relatief veel oude huurwoningen staan. Ga na of dit vermoeden door de gegevens in dit bestand worden ondersteund.
- Komen verschillende vormen van overlast vaak allemaal tegelijk voor op eenzelfde plek of komen specifieke vormen overlast op de ene plek voor en andere vormen juist weer op een andere plek?
-

Formuleblad statistiek VWO

- 1) Als X en Y twee onafhankelijke toevalsvariabelen dan geldt: $E(X+Y) = E(X) + E(Y)$ en $SD(X+Y) = \sqrt{(SD(X))^2 + (SD(Y))^2}$
- 2) Als $S = X_1+X_2+\dots+X_n$ en X_1, X_2, \dots onafhankelijk zijn en allemaal dezelfde kansverdeling hebben, dan is T normaal verdeeld en $E(S) = n \cdot E(X_1)$ en $SD(S) = \sqrt{n} \cdot SD(X_1)$
- 3) Als $G = (X_1+X_2+\dots+X_n)/n$ en X_1, X_2, \dots zijn onafhankelijk dan $E(G) = E(X_1)$ en $SD(G) = \frac{SD(X_1)}{\sqrt{n}}$
- 4) Als X binomiaal verdeeld met n en p dan $E(X) = n \cdot p$ en $SD(X) = \sqrt{n \cdot p \cdot (1-p)}$

Verwachtingswaarde en standaardafwijking

Betrouwbaarheidsintervallen

Het 95%-betrouwbaarheidsinterval voor de populatieproportie is

$P \pm 2 \cdot \sqrt{\frac{P(1-P)}{n}}$, met P de steekproefproportie en n de steekproefomvang.

Het 95%-betrouwbaarheidsinterval voor het populatiegemiddelde is

$\bar{X} \pm 2 \cdot \frac{S}{\sqrt{n}}$, met \bar{X} het steekproefgemiddelde, n de steekproefomvang en S de steekproefstandaardafwijking.

Verschil tussen twee populaties toetsen

Het verschil tussen twee populatiegemiddelden toetsen (μ_1 en μ_2) (beide populaties bij benadering normaal verdeeld), n_1, n_2 omvang steekproeven

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Steekproefgemiddelden \bar{X}_1 en \bar{X}_2 en (steekproef)standaardafwijkingen S_1 en S_2 ,

dan is toetsingsgrootheid $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ standaardnormaal verdeeld

d.w.z. met gemiddelde 0 en standaardafwijking 1

Het verschil tussen twee populatieproporties (p_1 en p_2) toetsen; n_1, n_2 omvang steekproeven

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

Steekproefproporties P_1 en P_2 ,

dan is toetsingsgrootte $Z = \frac{(P_1 - P_2) - (p_1 - p_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$ standaardnormaal verdeeld

d.w.z. met gemiddelde 0 en standaardafwijking 1

Vuistregels bij de grootte van het verschil tussen twee groepen

2x2 kruistabel $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, met $phi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$

- als $phi < -0,4$ of $phi > 0,4$, dan zeggen we “het verschil is groot”,
- als $-0,4 < phi < -0,2$ of $0,2 < phi < 0,4$, dan zeggen we “het verschil is middelmatig”,
- als $-0,2 < phi < 0,2$, dan zeggen we “het verschil is gering”.

maximale verschil in cumulatief percentage ($\max V_{cp}$) (met steekproefomvang $n > 100$)

- als $\max V_{cp} > 40$, dan zeggen we “het verschil is groot”,
- als $20 < \max V_{cp} \leq 40$, dan zeggen we “het verschil is middelmatig”,
- als $\max V_{cp} \leq 20$, dan zeggen we “het verschil is gering”.

effectgrootte $E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$

met \bar{X}_1 en \bar{X}_2 de steekproefgemiddelden ($\bar{X}_1 \geq \bar{X}_2$), S_1 en S_2 de steekproefstandaardafwijkingen

- als $E > 0,8$, dan zeggen we “het verschil is groot”,
- als $0,4 < E \leq 0,8$, dan zeggen we “het verschil is middelmatig”,
- als $E \leq 0,4$, dan zeggen we “het verschil is gering”.

twee boxplots vergelijken

- als de boxen elkaar niet overlappen, dan zeggen we “het verschil is groot”, anders
- als de boxen elkaar wel overlappen en de mediaan van de ene boxplot buiten de box van de andere boxplot ligt, dan zeggen we “het verschil is middelmatig”, en
- in alle andere gevallen zeggen we “het verschil is gering”.

Correlatiecoefficient R

- Als $R \leq -0,7$ of $R \geq 0,7$ dan is er sprake van sterke samenhang
- Als $-0,7 \leq R \leq -0,3$ of $0,3 \leq R \leq 0,7$ dan is er sprake van middelmatige samenhang
- Als $-0,3 \leq R \leq 0,3$ dan is er sprake van zwakke samenhang

Berekening van regressielijn van Y op X ($y = a \cdot x + b$)

$a = R \cdot \frac{S_y}{S_x}$ en b via $b = \bar{y} - a \cdot \bar{x}$, waarbij S_x en S_y de standaardafwijkingen