

VASTGOED ECONOMIE / 19

HBO /
University of Applied Sciences

Raymond Reinhardt
3R Business Development

raymond.reinhardt@3r-bdc.com
©3R

*o.m. gebaseerd o.a. op:
Marktonderzoek van R. Kooiker,
M. Broekhoff, e.a.; Statistiek van
J. Pompen.*



1

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Statistiek**: wetenschap die gericht is op waarnemen, bestuderen en analyseren van massaverschijnselen.

245.000	98.000	167.000	442.000	85.000
367.000	422.000	143.000	168.000	92.000
162.000	754.000	194.000	147.000	169.000
635.000	517.000	241.000	219.000	183.000
190.000	229.000	175.000	95.000	111.000

25 waarnemingen: **totale frequentie** = 25
(Tabel-)bandbreedte = **variatiebreedte** =
754.000 - 85.000 = 669.000

3R

2

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Klassenverdeling**, hier met **klassenbreedte 100.000**:

0	< X ≤	100.000	4 x	} Absolute frequentie
100.000	< X ≤	200.000	11 x	
200.000	< X ≤	300.000	4 x	
300.000	< X ≤	400.000	1 x	
400.000	< X ≤	500.000	2 x	
500.000	< X ≤	600.000	1 x	
600.000	< X ≤	700.000	1 x	
700.000	< X ≤	800.000	1 x	
			<u>25 x</u>	

3R

3

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Verdere frequenties:

0	< X ≤	100.000	4	16%	4	16%
100.000	< X ≤	200.000	11	44%	15	60%
200.000	< X ≤	300.000	4	16%	19	76%
300.000	< X ≤	400.000	1	4%	20	80%
400.000	< X ≤	500.000	2	8%	22	88%
500.000	< X ≤	600.000	1	4%	23	92%
600.000	< X ≤	700.000	1	4%	24	96%
700.000	< X ≤	800.000	1	4%	25	100%
			25	100%		

absol. freq.
 relat. freq.
 cumul. absol. freq.
 cumul. relat. freq.

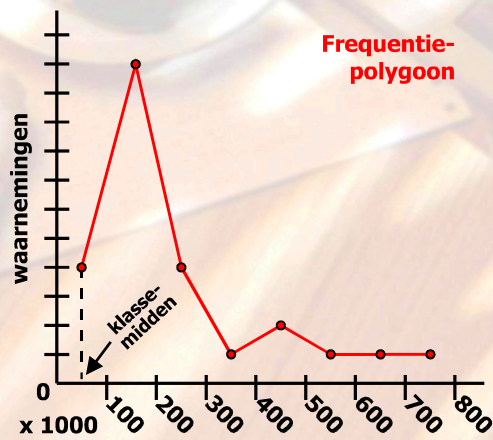
3R

4

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Grafieken:



3R

5

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Grafieken:



3R

6

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Grafieken:

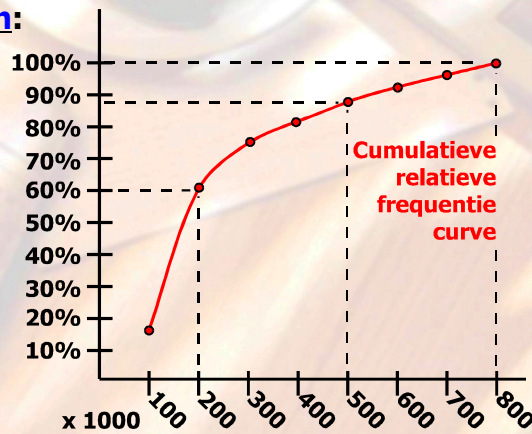


3R

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Grafieken:

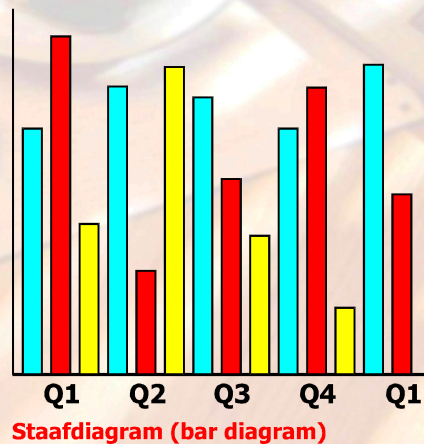


3R

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● Grafieken:

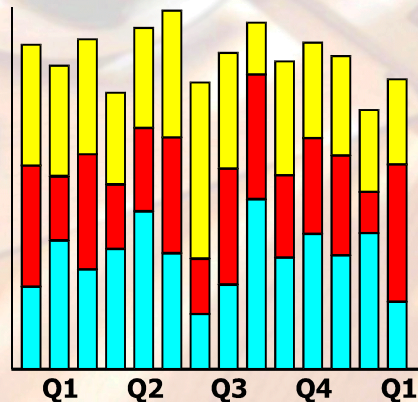


3R

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- Grafieken:



Stapeldiagram (stacked bar diagram)

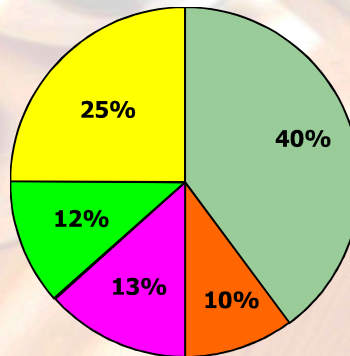
3R

10

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- Grafieken:



Cirkeldiagram (pie diagram)

3R

11

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- Het werken met indexcijfers:

Om ontwikkelingen in de markt te kunnen analyseren en de ontwikkelingen in tal van verschillende gegevens met elkaar te vergelijken, wordt vaak gewerkt met indexcijfers.

Berekeningsmethodiek:

$$\text{Indexcijfer jaar Y} = \frac{\text{Gegeven voor jaar Y}}{\text{Gegeven voor basisjaar B}} \times 100$$

$$\text{Indexcijfer aangepaste reeks A} = \frac{\text{Oorspr. indexcijfer A}}{\text{Indexcijfer A voor basisjaar B}} \times 100$$

3R

12

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● **Opgave 1:**

Jaar	2005	2006	2007	2008	2009	2010
PIC ₁	93,6	100,0	103,2		108,4	
Prijs		454,1		481,3		510,8
PIC ₂				100,0		

PIC₁ = prijsindexcijfers waarbij basisjaar = 2006
PIC₂ = prijsindexcijfers waarbij basisjaar = 2008
Vul de ontbrekende indexen en prijzen in.
[tip: eerst prijzen en PIC₁, dan PIC₂]

3R

13

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● **Opgave 1, oplossing:**

Jaar	2005	2006	2007	2008	2009	2010
PIC ₁	93,6	100,0	103,2		108,4	
Prijs		454,1		481,3		510,8
PIC ₂				100,0		

3R

14

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

● **Opgave 2:**

Het verbruik van een zeker type energiedrank was in 2006, in geld uitgedrukt, 200 miljoen euro, en in 2012 was dit 360 miljoen euro.

Het indexcijfer van het verbruik in stuks in dezelfde periode steeg van 125 in 2006 tot 180 in 2012 (index 2003 = 100).

Bereken hoeveel procent de prijs van dit type energiedrank over de periode 2006-2012 blijktbaar is gestegen?

[Tip: beschouw de indexcijfers van verbruik als afzet-verhoudingsgetallen.]

3R

15

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Opgave 2, oplossing:**

+25%

3R

16

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Centrummaten:** met één enkel getal een typerend beeld aangeven m.b.t. de totale massa; geeft aan rond welk getal waarnemingen zijn gegroepeerd:
 - **Modus:** de waarde van de meest voorkomende variabele;
Modale klasse: meest voorkomende klasse (hoogste frequentie, bij gelijke klassebreedten), of de klasse met de hoogste frequentiedichtheid (bij ongelijke klassebreedten).
 - **Mediaan:** de middelste waarneming nadat alle waarnemingen naar grootte zijn geordend. (Bij oneven aantal: middelste waarneming, bij even aantal: gemiddelde van 2 middelste waarnemingen.)

3R

17

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Centrummaten:**
 - **Rekenkundig gemiddelde:**
 - **ongewogen:** waarden waarnemingen / aantal
 - **gewogen:** op basis van wegingsfactoren:

Dag:	Dag- prijs:		Kwantiteit:	Totaalprijs:
Ma	1,75	x	30.000 kg	= 52.500
Di	2,23	x	22.000 kg	= 49.060
Wo	1,94	x	37.000 kg	= 71.780
Do	2,42	x	20.000 kg	= 48.400
Vr	2,07	x	41.000 kg	= 84.870
	10,41		150.000 kg	306.610

ongewogen
gemiddelde
2,08

gewogen
gemiddelde
2,04

3R

18

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Opgave 3:**
Een autoverkoper verkoopt in een zekere periode:
1 auto van 25.000 euro
3 auto's van 35.000 euro
2 auto's van 40.000 euro, en
1 auto van 70.000 euro.

A/ wat was in deze periode de ongewogen, rekenkundig gemiddelde prijs van verkochte auto's?

B/ wat was in deze periode de gewogen, rekenkundig gemiddelde prijs van verkochte auto's?

3R

19

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Opgave 3, oplossing:** A/ € 42.500; B/ € 40.000

3R

20

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Spreidingsmaten:** voor het verkrijgen van een indruk van de opbouw van waarnemingen.
 - **Spreiding:** de mate waarin individuele waarnemingen afwijken van de centrale tendentie (een centrummaat):
 - **variatiebreedte:** verschil tussen grootste en kleinste waarneming (de 'bandbreedte')
 - **gemiddelde afwijking (deviatie d)** >>
 - **standaardafwijking (standaarddeviatie S)** >>

3R

21

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Spreidingsmaten:**

- **Deviatie:** som van de absolute waarden der afwijkingen van de waarnemingen t.o.v. het rekenkundig gemiddelde, gedeeld door het aantal waarnemingen:

Voorbeeld:

Waarnemingen: 5, 7, 12, 14, 37, 45

Gemiddelde der waarnemingen: $120 / 6 = 20$

Afwijkingen: $|5 - 20| + |7 - 20| + |12 - 20| + |14 - 20| + |37 - 20| + |45 - 20| = 84$

Gemiddelde afwijking: $84 / 6 = 14 = \text{deviatie } d$
(Vergelijk: 9, 16, 21, 22, 24, 28: deviatie = 5)

3R

22

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Spreidingsmaten:**

- **Standaarddeviatie** = wortel uit de variantie, het kwadratisch gemiddelde (van een getallenreeks)

$$\text{variantie} = s^2 \Rightarrow s = \sqrt{\text{var}}$$

$$\text{variantie} = \frac{\sum (x_{1,2,\dots} - \Delta)^2}{N}$$

$$s = \sqrt{\frac{\sum (x_{1,2,\dots} - \Delta)^2}{N}}$$

$x_{1,2,\dots}$ = waarnemingen
 Δ = gemiddelde
 N = aantal waarn.

3R

23

PRAKTISCHE STATISTIEK

Statistiek basisbegrippen

- **Spreidingsmaten:**

- **Standaarddeviatie:**

Voorbeeld:

Waarnemingen: 5, 7, 12, 14, 37, 45 (x_1, x_2, \dots enz.)

Gemiddelde: $120 / 6 = 20$, $N = 6$

Variantie: $[(5 - 20)^2 + (7 - 20)^2 + (12 - 20)^2 + (14 - 20)^2 + (37 - 20)^2 + (45 - 20)^2] / 6 = 1408 / 6 = 234,67$

Standaarddeviatie = $\sqrt{234,67} = 15,3188$

Nb: waarnemingen: 9, 16, 21, 22, 24, 28 levert:

Standaarddeviatie = $\sqrt{37,00} = 6,0828$

3R

24

PRAKTISCHE STATISTIEK

Steekproeven

- **Steekproef, -populatie, -kader:**
 - **Steekproef:** een selectie van een subgroep van elementen (bv. individuen) uit een grotere verzameling van elementen (de populatie).
 - **Steekproefpopulatie:** de gehele groep van elementen (of individuen) waarover de onderzoeker d.m.v. steekproefonderzoek uitspraken wil doen. Voorbeelden: autobezitters, huishoudens, bedrijven.
 - **Steekproefkader:** een steekproefpopulatie, uitgedrukt in een lijst van elementen (of individuen), zoals een adressenlijst, een telefoonboek, enz.

3R

25

PRAKTISCHE STATISTIEK

Steekproeven

- **Steekproefvereisten:**
 - de populatie moet **homogeen** zijn
 - de steekproefelementen moeten **aselect gekozen** zijn (ieder element uit de populatie moet een even grote kans hebben om gekozen te worden)
 - de steekproef moet **voldoende groot** zijn (hetgeen afhangt van de gewenste betrouwbaarheid en nauwkeurigheid)
 - de steekproef moet **betrouwbaar** zijn (bij herhaling steeds vergelijkbare uitkomsten)
 - de steekproef moet **valide** zijn (meten wat gemeten dient te worden)
 - de steekproef moet **nauwkeurig** zijn

3R

26

PRAKTISCHE STATISTIEK

Steekproeven

- **Methoden van steekproeftrekking, o.a.:**
 - **Volkomen a-selecte steekproef**
= 100% representatieve steekproef.
Kan uitsluitend wanneer men beschikt over de gehele onderzoekspopulatie = steekproefkader.
Zodoende: vrijwel nooit volkomen a-select.
 - **Systematische steekproef (intervalsgewijs)**
= methode waarbij gehele adressenbestanden op systematische wijze worden doorlopen (iedere zoveelste element).

3R

27

PRAKTISCHE STATISTIEK

Steekproeven

- **Methoden van steekproeftrekking, o.a.:**
 - **Proportioneel gestratificeerde steekproef** waarbij gehele onderzoekspopulatie in een aantal deelpopulaties (strata = lagen) worden verdeeld, waarna uit elk van deze strata, *naar evenredigheid qua strata-verdeling*, afzonderlijke steekproeven worden getrokken.
 - **Disproportioneel gestratificeerde steekproef** waarbij de verhoudingen van de strata in de steekproef ongelijk zijn aan de verhoudingen van de strata in de populatie.

3R

28

PRAKTISCHE STATISTIEK

Steekproeven

- **Methoden van steekproeftrekking, o.a.:**
 - **Geclusterde steekproef (clustersteekproef)** waarbij steekproefkader wordt gesplitst in groepen (clusters) waarna een a-selecte steekproef van clusters wordt getrokken. Vervolgens wordt ieder element uit ieder getrokken cluster ondervraagd (**1-traps clustersteekproef**). Ook mogelijk: na trekking van een aantal clusters wordt een a-selecte steekproef van elementen uit ieder getrokken cluster getrokken (**2-traps clustersteekproef**).

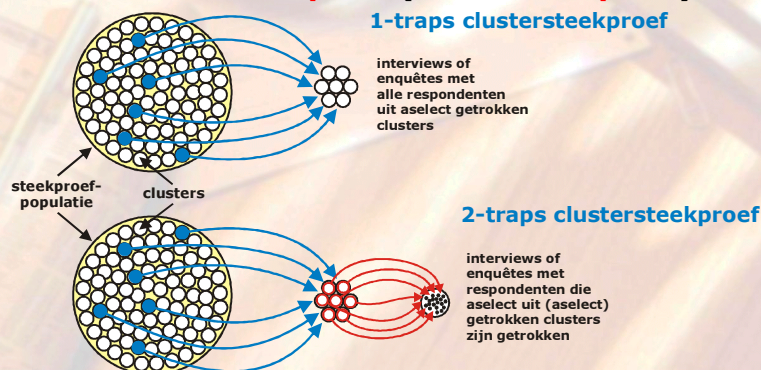
3R

29

PRAKTISCHE STATISTIEK

Steekproeven

- **Methoden van steekproeftrekking, o.a.:**
 - **Geclusterde steekproef (clustersteekproef)**



3R

30

PRAKTISCHE STATISTIEK

Steekproeven

- **Methoden van steekproeftrekking**, o.a.:
 - **Random walk methode (startadressenmethode)**
Manier om respondenten te selecteren voor een steekproef. Volgens toeval wordt een straat met huisnummer getrokken, bijv. 1e huisnummer plus 15, bij eerstvolgende straat rechtsaf, enz.
 - **Quotasteekproef**
Niet-aselecte steekproef waarbij de populatie op basis van belangrijk geachte variabelen wordt opgedeeld in een aantal strata. De interviewer moet dan een aantal personen in een bepaalde stratum ondervragen.

3R

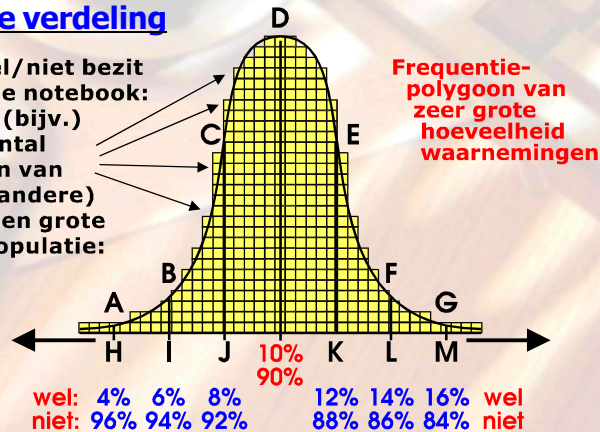
31

PRAKTISCHE STATISTIEK

Steekproeven

- **De normale verdeling**

verdeling wel/niet bezit van een Apple notebook: op basis van (bijv.) een groot aantal steekproeven van steeds **225** (andere) mensen uit een grote homogene populatie:



3R

32

PRAKTISCHE STATISTIEK

Steekproeven

- **De normale verdeling: de kromme van Gauss:**

- Uit het voorbeeld blijkt dat de meeste uitkomsten zich rondom 10% bevinden.

Wanneer uit de populatie (=N) steeds een groot aantal onafhankelijke waarnemingen (of in ons voorbeeld: steekproeven) wordt getrokken, vormen de uitkomsten samen een **symmetrische, klokvormige figuur**, waarin **modus, mediaan** en **rekenkundig gemiddelde** van de uitkomsten steeds samenvallen.

3R

33

PRAKTISCHE STATISTIEK

Steekproeven

- **Steekproef-standaarddeviatie: standaardfout**
 - Duidelijk is ook dat niet alle uitkomsten van de steekproeven gelijk zijn aan de meest voorkomende waarde, en is hier sprake van een zekere *variabiliteit* in deze uitkomsten.

Indien men algemene uitspraken wil doen over de gehele populatie, zal hier rekening moeten worden gehouden met een vorm van deviatie: de zgn. standaardfout s .

>>

3R

34

PRAKTISCHE STATISTIEK

Steekproeven

- **Steekproef-standaarddeviatie: standaardfout**
 - Standaardfout bij steekproeven:

$$s = \sqrt{\frac{p \times q}{n}}$$

waarbij p = % wel, q = % niet, n = aantal waarnemingen in steekproef; in ons voorbeeld, uitgaande van de **centrale tendentie 10%**:

$$s = \sqrt{\frac{10 \times 90}{225}} = 2(\%)$$

3R

35

PRAKTISCHE STATISTIEK

Steekproeven

- **Standaardfout en betrouwbaarheid van uitspraak:**
 - Wis- en sterrenkundige *Carl Friedrich Gauss* (1777-1855) ontdekte aan de hand van tal van sterrenkundige observaties o.a. de volgende patronen die ook in ons voorbeeld (10%) gelden:
 - **68,3%** van alle waarnemingen ligt tussen: $10\% \pm 1 \times s$ (nl. oppervlak JCDEK), ofwel: tussen 8% en 12%, en ook:
 - **95,4%** van alle waarnemingen ligt tussen: $10\% \pm 2 \times s$ (nl. oppervlak IBCDEFL), ofwel: tussen 6% en 14%, en ook:
 - **99,7%** van alle waarnemingen ligt tussen: $10\% \pm 3 \times s$ (nl. oppervlak HABCDEFGM), ofwel: tussen 4% en 16%.

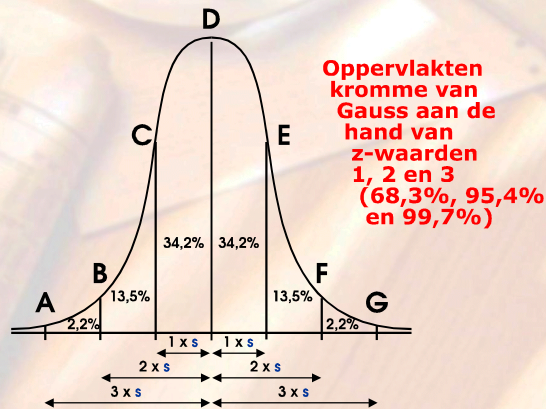
3R

36

PRAKTISCHE STATISTIEK

Steekproeven

- **Standaardfout en betrouwbaarheid van uitspraak:**



37

PRAKTISCHE STATISTIEK

Steekproeven

- **Standaardfout en betrouwbaarheid van uitspraak:**

- Wat betreft het bezit van een Apple-notebook, kunnen we nu dus stellen, dat:
 - met een betrouwbaarheid van bijv. **68,3%**, tussen **10% +/- (1 x 2%)**, ofwel **tussen 8% en 12% van de populatie** dit bezit;
- en ook:
 - met een betrouwbaarheid van bijv. **95,4%**, tussen **10% +/- (2 x 2%)**, ofwel **tussen 6% en 14% van de populatie** dit bezit;
- en ook:
 - met een betrouwbaarheid van bijv. **99,7%**, tussen **10% +/- (3 x 2%)**, ofwel **tussen 4% en 16% van de populatie** dit bezit.

3R

38

PRAKTISCHE STATISTIEK

Steekproeven

- **Standaardfout en betrouwbaarheid van uitspraak:**

- De factoren **1, 2** en **3** zijn zgn. **z-waarden** (*excentriciteitswaarden*). Per betrouwbaarheids-% gelden andere (vaste) z-waarden, waaronder:

- [betrouwbaarheid **0,0%** : z-waarde **0**]
- betrouwbaarheid **68,3%** : z-waarde **1**
- betrouwbaarheid **90,0%** : z-waarde **1,68**
- betrouwbaarheid **95,0%** : z-waarde **1,96**
- betrouwbaarheid **95,4%** : z-waarde **2**
- betrouwbaarheid **99,7%** : z-waarde **3**

3R

39

PRAKTISCHE STATISTIEK

Steekproeven

- **Betrouwbaarheid en precisie van uitspraak:**
 - Omdat een steekproef in zekere mate indicatief zou moeten zijn voor de gehele populatie, moet rekening worden gehouden met een zekere marge die de 'precisie' van de uitspraak bepaalt:

nauwkeurigheidsmarge m (%) =
z-waarde \times standaardfout, ofwel: $m = z \times s$

- In ons voorbeeld, bij $s = 2\%$ en een betrouwbaarheid van (bv.) $99,7\%$ ($z = 3$), wordt de marge: $z \times s = 3 \times 2 = 6\%$. Bij deze betrouwbaarheid kan dus gesteld worden dat tussen $10\% - 6\%$ ($= 4\%$) en $10\% + 6\%$ ($= 16\%$) van de populatie een Apple notebook bezit. De bandbreedte (hier $4\% \leftrightarrow 16\%$) heet een **betrouwbaarheidsinterval**.

3R

40

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 4:**
In een steekproefonderzoek naar het dvd-speler-bezit onder Rotterdammers bleek 40% van de 400 ondervraagden zo'n apparaat te bezitten. Tussen welke grenzen ligt dit percentage in de populatie, uitgaande van een betrouwbaarheid van $95,4\%$?

3R

41

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 4, oplossing:** $35,1\% \ll \gg 44,9\%$

3R

42

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 5:**
De aanbieder van dvd-spelers wil nu een onderzoek doen naar het % dvd-speler-bezitters in een andere grote plaats. Als eis wordt gesteld dat de nauwkeurigheid (de marge) maximaal 5% mag zijn bij een betrouwbaarheid van 95,4%. Hoe groot moet de steekproef zijn om aan deze eisen te kunnen voldoen?

[Hint: omdat hier vooraf geen steekproefuitkomst bekend is, zal hier voor de zekerheid moeten worden uitgegaan van de maximale steekproefstandaardfout!]

3R

43

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 5, oplossing:** 400

3R

44

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 6:**
In 2 aan elkaar grenzende gemeenten (gemeente A, 8.000 huishoudens; gemeente B, 10.000 huishoudens) is onlangs steekproefonderzoek onder de huishoudens gedaan naar het gebruik van groene energie.
In gemeente A gaven 120 van de ondervraagde 300 huishoudens aan groene energie te gebruiken.
In gemeente B bleken 275 van de ondervraagde 500 huishoudens dit ook te doen.
Is hier wel of niet sprake van een significant verschil tussen deze uitkomsten? Toon dit d.m.v. berekening aan waarbij een betrouwbaarheidseis wordt gesteld van 99,7%.

3R

45

PRAKTISCHE STATISTIEK

Steekproeven

- **Opgave 6, oplossing:** betrouwbaarheidsintervallen overlappen elkaar; geen significant verschil.

3R

46

PRAKTISCHE STATISTIEK

Steekproeven

- **Relatie betrouwbaarheid / nauwkeurigheid:**
 - hoe groter de betrouwbaarheid ('zekerheid') van uitspraak, des te minder nauwkeurig (minder precies) deze is(!)
 - Voorbeeld: steekproefuitkomst 40% onder 600 ondervraagden. Bereken de marges bij betrouwbaarheid 99,7% ($z=3$) resp. 68,3% ($z=1$).

$$\begin{array}{l|l} m = 3 \times \sqrt{\frac{40 \times 60}{600}} & m = 1 \times \sqrt{\frac{40 \times 60}{600}} \\ m = 3 \times \sqrt{4} = 6,0\% & m = 1 \times \sqrt{4} = 2,0\% \end{array}$$

3R

47

PRAKTISCHE STATISTIEK

Steekproeven

- **Correctie bij relatief grote steekproeven:**
Indien een steekproef de gehele populatie zou betreffen ($n = N$), dan is niet echt sprake meer van een steekproef maar van een gehele meting, en kan er geen standaardfout s bestaan: $s = 0$!
De formule voor de standaardfout houdt hier geen rekening mee, zodat met een correctiefactor moet worden gerekend wanneer $n \geq 10\% N$:

$$s = \sqrt{\frac{p \times q}{n}} \times \sqrt{\frac{N - n}{N - 1}}$$

3R

48

PRAKTISCHE STATISTIEK

Steekproeven

- **Preciese bepaling benodigde steekproefomvang** :

$$n = \frac{2500 \times N}{\left[\left(\frac{m}{z} \right)^2 \times (N - 1) + 2500 \right]}$$

m = toegestane marge
 n = steekproefomvang
 N = steekpr.-populatie

Voorbeeld: populatie $N = 5000$; $m = 3\%$;
betrouwbaarheid = 90% ($\rightarrow z = 1,68$).
Benodigde steekproefomvang ? :

$$n = \frac{2500 \times 5000}{\left[\left(\frac{3}{1,68} \right)^2 \times (5000 - 1) + 2500 \right]} = \frac{12500000}{18440,69} = 678$$

3R