Statistiek: Centrummaten

12/6/2013

dr. Brenda Casteleyn

Keu6
Coaching & Onderzoek

1. Theorie

1) Nominaal niveau:

Gebruik de Modus, dit is de meest frequente waarneming

2) Ordinaal niveau:

Mediaan: rangschik alle meetwaarden in oplopende volgorde en zoek wat de middelste waarde is. Of met een formule: de (N+1)/2 de waarneming.

Voorbeeld: In de volgende tabel zien we het scholingsniveau van 105 personen. De mediaan is (105+1)/2 =53 ste waarneming. De 53ste persoon bevindt zich in de klasse lager secundair. De mediaan bevindt zich dus in de klasse 'lager secundair onderwijs'.

Scholingsniveau	Frequentie
Lager	20
Lager secundair	35
Hoger secundair	10
Hogeschool	27
Universiteit	13
Totaal	105

3) Ordinaal (latent kwantitatief):

Middenkwartiel $\frac{Q_1 + Q_3}{2}$ met $Q_1 = \frac{(N+1)}{4}$ de waarneming en $Q_3 = \frac{3(N+1)}{4}$ de waarneming

Middenscharnier $\frac{S_1 + S_2}{2}$

met $S_1 = (\frac{(N+1)}{4} + 0{,}25)$ de waarneming en $S_2 = (\frac{3(N+1)}{4} - 0{,}25)$ de waarneming (N= even) en

met $S_1 = (\frac{(N+1)}{4} + 0{,}5)$ de waarneming en $S_2 = (\frac{3(N+1)}{4} - 0{,}5)$ de waarneming (N= oneven) en

Tukey's Trigemiddelde: $\frac{Q_1 + (2.Me) + Q_3}{4}$ of : $\frac{S_1 + (2.Me) + S_2}{4}$

4) Kwantitatief

Rekenkundig gemiddelde: som van de meetwaarden gedeeld door het aantal meetwaarden:

$x_{gem} = \sum_{i=1}^{N} \frac{x_i}{N}$

Voorbeeld: 5 mensen hebben als leeftijd 15; 20; 30, 40 en 50 jaar. Het gemiddelde is de som (15+20+30+40+50)/5 = 155/5 = 31 jaar.

Maar meestal hebben we veel meer mensen gemeten, zoals in onderstaande tabel. Dan maken we gebruik van het gewogen gemiddelde. De som bestaat dan uit de optelling van (leeftijd x frequentie) en gedeeld door het totaal aantal mensen. Dus in onderstaand voorbeeld: ((15.20) + (20.35) + (30.10) + (40.27) + (50.13))/(20+35+10+27+12) = 3030/105 = 29 jaar (afgerond)

Leeftijd	Frequentie	Leeftijd x frequentie
15	20	300
20	35	700
30	10	300
40	27	1080
50	13	650
Totaal	105	3030

$x_{gem} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$

Lineaire transformatie:

Dit betekent dat de meetwaarden door middel van een berekening naar een andere schaal zijn omgezet (bv. van euro's naar Belgische frank). De transformatie is lineair als de berekening volgende eigenschappen heeft: de oorspronkelijke waarde wordt met een getal vermenigvuldigd (bv. met 40 bij de omzetting van Euro naar Belgische frank) en daarna wordt er een getal bij opgesteld (nul in 't geval van omzetting van Euro naar frank).

$x_{frank} = 40.x_{euro}$

Als je 't gemiddelde al in de oorspronkelijke meetwaarde hebt berekend kun je het gemiddelde in de andere schaal vinden door gewoon dezelfde transformatieberekening op het gemiddelde te maken, dus het gemiddelde met hetzelfde getal te vermenigvuldigen en daarna op te tellen met hetzelfde. Als dus een gemiddelde 50 euro is, moet je deze 50 euro met 40 vermenigvuldigen en 0 bij optellen om het gemiddelde in Belgische frank te krijgen.

De algemene vorm van een lineaire transformtie is $x_{nieuw} = a.x_{oud} + b$

Niet-lineaire transformaties:

We maken hierbij het meetkundig gemiddelde van bv. $1/x_i$ (harmonisch); x^2 (kwadratisch) of log x_i (geometrisch). De uitkomst moeten we dan terug omzetten naar de oorspronkelijke grootte-orde (dus bij de harmonische de breuk omkeren; bij de kwadratische de vierkantswortel nemen en bij het geometrische antilog nemen).

Dus: <u>harmonisch gemiddelde</u>: gebruik formule voor rekenkundig gemiddelde maar vervang x_i door $1/x_i$ en draai op het einde van de berekening de breuk om.

Dit gebruik je als je het gemiddelde van verhoudingsgetallen moet berekenen. Bv. de gemiddelde snelheid van de heen- en terugreis.

Kwadratisch gemiddelde: gebruik formule voor rekenkundig gemiddelde maar vervang x_i door x_i^2 en trek op het einde de vierkantswortel uit de uitkomst.

Dit gebruik je als je meetwaarden zowel negatief als positief kunnen zijn. Bij het gewone gemiddelde zouden de positieven de negatieve elkaar opheffen. Om dat te vermijden (en dus alles positief te houden) gebruik je het kwadraat.   Bv. gemiddelde afwijkingen berekenen.

Meetkundig gemiddelde: gebruik formule voor rekenkundig gemiddelde maar vervang x_i door $\log x_i$ en neem op het einde het antilog van de uitkomst.

Dit gebruik je bij de berekening van groeifactoren (dit is de omzetting van een groeipercent, bv. bij een groei van 6% is de groeifactor 1,06[1]. Voorbeeld: berekening van gemiddelde intrestvoet.

2. Oefeningen

Examen Thijssen januari 2007 Vraag 4

Uit politierapporten over de criminaliteit in een Antwerpse probleemwijk blijkt dat in 1998 het aantal kleine misdrijven een toename van 10% kende ten opzichte van het voorgaand jaar. In 1999 nam de kleine misdrijven met één vierde toe. De 3 daaropvolgende jaren was er, mede wankzij een opwaardering van de buurt, telkens een daling van 1/5de. Door extreme spanningen verdriedubbelde het aantal kleine misdrijven in 2003. In 2004 nam het gelukkig terug met de helft af. In 2005 werden er precies evenveel kleine misdrijven geregistreerd als in 2004. Bereken de gemiddelde stijging van het aantal kleine misdrijven in deze buurt over de volledige waargenomen periode.

Examen Thijssen januari 2008 Vraag 5

Op 1 januari 2002 werd een nieuw onine forum opgestart. Eén jaar later waren er al 250 leden. Respectievelijk twee, drie en vier jaar later is dit ledenaantal aangegroeid tot 2500, 4600 en 7900. Bereken de gemiddelde groei van het ledenaantal en toon aan dat dit gemiddelde adequaat is.

Proefexamen Thijssen 2012 Semester 1 Vraag 2

In onderstaande frequentieverdeling is de eerste scharnier gelijk aan:

[1] Berekening groeifactor: iets groeit met 6% dus van 100 naar 106. Deel de laatste waarde (met de groei) door de beginwaarde (zonder de groei): 106/100 = 1.06.

x	1	2	3	4	5	6	7	8	9	10
f	1	2	2	5	1	5	9	8	3	3

A. 4,25
B. 4,5
C. 4
D. 5

Proefexamen Thijssen 2012 Semester 1 Vraag 3

Een automobielconstructeur maakt reclame voor een CO_2 vriendelijke wagen, waarvan wordt beweerd dat de CO_2 uitstoot maar 100 g/km bedraagt. Een consumentenorganisatie wil deze claim onderzoeken en laat 5 verschillende chauffeurs met deze wagen een traject van 100 km afleggen terwijl de CO_2 uitstoot wordt gemeten. De gemiddelde afwijking van de vooropgestelde norm bedraagt:

Chauffeur	Totale uitstoot (na 100 km) in g
A	9720
B	10327
C	11200
D	9856
E	10122

A. 510,52
B. 576,36
C. 245
D. 306,2

Proefexamen Thijssen 2012 Semester 1 Vraag 4

Een docent registreerde de daling van de aanwezigheid gedurende de eerste 6 weken van het academiejaar. In de tweede les waren er 5% minder studenten aanwezig dan in de eerste les. In de daaropvolgende weken daalde de aanwezigheid respectievelijk met 9%, 3%, 1% en 2% (telkens in vergelijking met de vorige les). Indien je weet dat er bij de eerste les 500 studenten aanwezig waren, hoeveel zaten er dan nog in de zesde lesweek,

A. ong. 480
B. ong. 428
C. ong. 391
D. ong. 408

3. Oplossingen

<u>Examen Thijssen januari 2007 Vraag 4</u>

Gegeven: groei aantal misdrijven: 1998: + 10%; 1999: +1/4; 2000: -1/5; 2001: -1/5; 2002: -1/5; 2003: x 3; 2004: -1/2; 2005: +0

Gevraagd: gemiddelde over heel de periode

Oplossing:

Omzetting naar groeifactoren:

1998: 110/100 = 1,10; 1999: 125/100 = 1.25; 2000: 80/100 = 0.8; 2001: 0.8; 2002:0.8; 2003: 300/100 = 3; 2004: 50/100 = 1/2; 2005: 100/100 = 1

Berekening via formule meetkundig gemiddelde

Berekening geometrisch gemiddelde: Antilog[(log (1,10) + log (1.25) + log (0.8) + log (0.8) + log (0.8) + log (3) + log (0.5)+ log (1))/8)]= Exp (ln (1.10*1.25*0.8*0.8*0.8*3*0.5*1)/8)

(in excel: gebruik ln voor log en exp voor antilog)

Berekening:exp(ln(1.056)/8) = Exp(0,054488185/8) = Exp(0,006811) = 1,006834

Controle: Als we het gemiddelde 8 keer toepassen als groei, dan krijgen we de groei ten opzichte van de eerste waarde: $1{,}006834^8$ is stijging van de criminaliteit tov het begin = 1.056

Wanneer we de groei berekenen met de gegeven waarden:

	Criminaliteit	actuele groei	gemiddelde groei
beginwaarde	100		
1998	100	1,1	1,006834271
1999	110	1,25	1,006834271
2000	137,5	0,8	1,006834271
2001	110	0,8	1,006834271
2002	88	0,8	1,006834271
2003	70,4	3	1,006834271
2004	211,2	0,5	1,006834271
2005	**105,6**	1	1,006834271
			1,056

We stellen vast dat 105.6 = 100. 1,056, ons gemiddelde is dus adequaat

Snelle methode: Om te testen of het gemiddelde adequaat is gebruiken we het gemiddelde als groeivoet en verheffen deze tot de macht, die gelijk is aan het aantal jaren. We kunnen deze formule ook gebruiken om het gemiddelde te berekenen: nl. bereken de groei van de laatste waarde ten opzichte van de eerste: 105,6/100 = 1,056. Het gemiddelde vinden we dan uit volgende formule:

$\sqrt[8]{1.056}$ = 1.006834271

Examen Thijssen januari 2008 Vraag 5

Gegeven: aantal leden forum: 2002: 0; 2003 250; 2004: 2500; 2005: 4600: 2006: 7900

Gevraagd: gemiddelde groei en toon aan dat dit gemiddelde adequaat is.

Oplossing: berekening groeifactoren: 2003: 250/1=250 ; 2004: 2500/250=10; 2005=1.84; 4600/2500 = 1.717391304.

Berekening via formule meetkundig gemiddelde

Bereken ln van het product van de groeivoeten en deel door n:

[(ln(250 x 10 x 1,84 x 1.717391304))/4] = 9,427722109

Om te controleren of dit gemiddelde adequaat is, ga je na of je na toepassing van het gemiddelde elk jaar op dezelfde uitkomst komt als wanneer je de groeivoet van elk jaar zou toepassen. We stellen vast dat $9{,}427722109^4$ = 7900, wat klopt: ten opzichte van 2002 is het aantal leden met 7900 gestegen.

	aantal leden	groeivoet	gemiddelde
2002	0		
2003	250	250	9,427722109
2004	2500	10	9,427722109
2005	4600	1,84	9,427722109
2006	7900	1,717391304	9,427722109
			7900

Snelle manier:

Groei ten opzichte van het begin: 7900/1 = 7900. Berekening gemiddelde: $\sqrt[4]{7900}$ = 9,4277

Groei in percentages: (9,4299 - 1)* 100 = 842.77%

Adequaat? $1(9{,}4277^4)$ = 7900

Proefexamen Thijssen 2012 Semester 1 Vraag 2

Gegeven:

x	1	2	3	4	5	6	7	8	9	10
f	1	2	2	5	1	5	9	8	3	3

Gevraagd: 1ste scharnier?

Bepaal N: = 1+2+2+5+1+5+9+8+3+3 = 39. N is oneven dus de formule voor S_1 =($\frac{(N+1)}{4}$+ 0,5)
Vul in: S_1 = (40/4 + 0.5) = 10,5de waarneming. Deze waarneming zit tussen de x=4 en x=5, net in het midden. Dus het scharnier zit bij x = 4.5

- ➔ Antwoord B

<u>Proefexamen Thijssen 2012 Semester 1 Vraag 3</u>

Gegeven: norm = 100 g/km; tabel met gemeten waarden per 100 km.

Gevraagd: gemiddelde afwijking

Oplossing: bereken per chauffeur de afwijking tov de norm (= 100 g/km of 10 000g/100 km)

Chauffeur	Totale uitstoot (na 100 km) in g	Afwijking tov 10 000 g
A	9720	280
B	10327	-327
C	11200	-1200
D	9856	144
E	10122	-122

Om gemiddelde te nemen van afwijkingen gebruiken we kwadratische gemiddelde. Daarvoor berekenen we het kwadraat van elke afwijking. Al deze kwadraten tellen we op en we delen deze som door n. Ten slotte nemen we de vierkantswortel van het eindresultaat.

Chauffeur	Totale uitstoot (na 100 km) in g	Afwijking tov 10 000 g	Kwadraten van afwijkingen
A	9720	280	78400
B	10327	-327	106929
C	11200	-1200	1440000
D	9856	144	20736
E	10122	-122	14884
		som kwadraten	1660949
		som/n	332189,8
		vierkantswortel (som/n)	576,3590895

We krijgen dan als gemiddelde: 576,36

- ➔ Antwoord B

Proefexamen Thijssen 2012 Semester 1 Vraag 4

Gegeven: daling afwezigheden:

Les 1: 500 studenten
Les 2: -5%
Les 3: -9%
Les 4: -3%
Les 5: -1%
Les 6: -2%

Gevraagd: Aantal studenten in zesde lesweek.

Oplossing: Bepaal aan de hand van groeicijfers het aantal studenten voor elke les:

Het groeicijfer berekenen we door het eindcijfer (dus na de groei) te delen door het begincijfer; voor les 2 geeft dat; 475/500 (475 is het begincijfer - 5%)=0.95. Maar het is gemakkelijker met 100 als vertrekcijfer te beginnen. Dan komen we op hetzelfde groeicijfer uit: 95/100 (5% minder dan 100 is 95) = 0.95. Zo berekenen we het groeicijfers van elke les. Daarna kunnen we telkens het groeicijfer vermenigvuldigen met het aantal studenten van de voorgaande les (zie tabel); Op les 6 hebben we dan 407 studenten.

	Formule groeicijfer	Groeicijfer	Formule aantal	Aantal studenten
Les 1: 500				
Les 2: -5%	95/100	0,95	0,95 x 500	475
Les 3: -9%	91/100	0,91	0,91 x 475	432
Les 4: -3%	97/100	0,97	0,97 x 419	419
Les 5: -1%	91/100	0,99	0,99 x 415	415
Les 6: -2%	98/100	0,98	0,98 * 406	407

➔ Antwoord D