

Statistiek: Vorm van de verdeling

1/4/2014

dr. Brenda Casteleyn



Keu6
Coaching & Onderzoek

1. Theorie

Enkel de theorie die nodig is voor de oefeningen is hierin opgenomen.

Scheefheid of asymmetrie

Indien de meetwaarden links van de mediaan meer spreiding vertonen dan rechts van de mediaan spreken we over een linksscheve verdeling, in het omgekeerde geval spreken we van een rechtsscheve verdeling.

Scheefheidscoëfficiënt van Yule

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Q_1)} \text{ of } \frac{(Q_3 + Q_1 - 2Me)}{(Q_3 - Q_1)} \text{ of } \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

De noemer is hier toegevoegd om de maat relatief te maken. De uitkomst varieert dus tussen -1 en 1.

Voor de berekening van de scheefheidscoëfficiënt van Fisher hebben we wat kennis over momenten en binomium van Newton nodig, dat komt dus nu eerst aan bod.

Uitleg over momenten

Het r-de moment definiëren we als volgt:

$m_r' = \frac{\sum_{i=1}^n x_i^r}{n} = \frac{x_1^r + x_2^r + x_3^r + \dots + x_n^r}{n}$ Het is dus de som van alle waarnemingen die tot de r-de macht verheven werden en gedeeld door het totaal aantal waarnemingen.

Merk dus op: het eerste moment (r=1) is gewoon het rekenkundig gemiddelde; het tweede moment (r=2) is het gemiddelde van de kwadraten.

Het r-de moment *rond een bepaald getal A* (gecentreerd moment) definiëren we als volgt:

$$m_r = \frac{\sum_{i=1}^n (x_i - A)^r}{n} = \frac{(x_1 - A)^r + (x_2 - A)^r + (x_3 - A)^r + \dots + (x_n - A)^r}{n}$$

Als je voor A het gemiddelde¹ neemt, wordt dit:

$$\mu_r = \frac{\sum_{i=1}^n (x_i - \mu)^r}{n} = \frac{(x_1 - \mu)^r + (x_2 - \mu)^r + (x_3 - \mu)^r + \dots + (x_n - \mu)^r}{n}$$

Als je voor de waarde A het gemiddelde gebruikt en r = 2 stelt, krijg je de formule voor variantie VAR(x)

¹ Ik maak hier bij de formules geen onderscheid tussen populatie of steekproef.

Als je voor A niet het gemiddelde maar 0 neemt, krijg je terug het r-de moment, dus dat moment kan je ook het r-de moment rond 0 noemen.

Als we met gegroepeerde data werken worden de formules:

$$\text{Moment rond 0: } m_r' = \frac{\sum_{i=1}^n f_i x_i^r}{n} = \frac{f_1 x_1^r + f_2 x_2^r + f_3 x_3^r + \dots + f_n x_n^r}{n}$$

$$\text{Moment rond } \mu = \mu_r = \frac{\sum_{i=1}^n f_i (x_i - \mu)^r}{n} = \frac{f_1 (x_1 - \mu)^r + f_2 (x_2 - \mu)^r + f_3 (x_3 - \mu)^r + \dots + f_n (x_n - \mu)^r}{n}$$

Verder is nog van belang om te zien dat je momenten rond een getal (gecentreerde momenten) kan ontbinden naar momenten rond nul. In het boek van Thijssen staat dit uitgelegd op blz. 114 voor het tweede moment rond het gemiddelde (maw de variantie). Door uitwerking van het kwadraat kan je het moment opsplitsen in verschillende termen, zodat je uiteindelijk een som van momenten rond nul vindt.

Als je hetzelfde wil doen voor momenten met r groter dan 2, moet je telkens de machten uitwerken en heb je het binomium van Newton nodig.

Binomium van Newton

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

waarbij
$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

Voorbeeld:

$$\begin{aligned} (a + b)^5 &= \sum_{i=0}^5 \binom{5}{i} a^i b^{5-i} \\ &= \binom{5}{0} a^0 b^5 + \binom{5}{1} a^1 b^4 + \binom{5}{2} a^2 b^3 + \binom{5}{3} a^3 b^2 + \binom{5}{4} a^4 b^1 + \binom{5}{5} a^5 b^0 \end{aligned}$$

Om de coëfficiënten $\binom{n}{i}$ te berekenen kan je de formule gebruiken of de driehoek van Pascale. (zie ook nuttige rekenregels in boek van Thijssen blz. 151-152)

$$\begin{aligned} &= 1 \cdot a^0 \cdot b^5 + 5 \cdot a \cdot b^4 + 10 \cdot a^2 b^3 + 10 \cdot a^3 b^2 + 5 \cdot a^4 b^1 + 1 \cdot a^5 b^0 \\ &= b^5 + 5 \cdot a \cdot b^4 + 10 \cdot a^2 b^3 + 10 \cdot a^3 b^2 + 5 \cdot a^4 b + a^5 \end{aligned}$$

Noteer: als je de 3de term wil, dan is dat $\binom{5}{2} a^2 b^3$ dus bij deze term is $i = 2$

Als je dus bv. de 520ste term wil, is k dan gelijk aan 519. Dat komt omdat de eerste term staat voor i=0.

Scheefheidscoëfficiënt van Fisher

Definitie: deze coëfficiënt is gelijk aan: $\Upsilon_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\sqrt{\mu_2}^3} = \frac{\mu_3}{\mu_2 \sqrt{\mu_2}}$

Om het derde moment rond het gemiddelde te berekenen, gebruiken we het Binomium van Newton om de formule om te zetten van een moment rond het gemiddelde naar de som van momenten rond nul (zie Thijssen blz. 153). De formule wordt nu herleid tot:

$$\Upsilon_1 = \frac{\mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3}{\mu_2\sqrt{\mu_2}} = \frac{\mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3}{\sigma^3}$$

2. Oefeningen

Proefexamen Thijssen 2008 Vraag 4

In de eerste 6 maanden na de opening van een nieuwe zaak wordt elke dag het aantal klanten genoteerd. De helft van de dagen kwam er nooit meer dan 45 klanten langs en in 75% van de gevallen waren er steeds meer dan 22 klanten. Het middenkwartiel bedraagt 38,5 klanten en deze waarde is net iets kleiner dan het gemiddeld aantal klanten van 42,8. Er was echter ook een topdag waarop 95 klanten langskwamen. Is hier sprake van een symmetrische verdeling?

Staaf je antwoord aan de hand van een geschikt kengetal.

Proefexamen Thijssen 2012 Vraag 5

Welke van volgende formules is GEEN berekening van de scheefheidscoëfficiënt van Yule?

- A. $\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$
- B. $\frac{(Q_3 + Q_1 - 2Me)}{(Q_3 - Q_1)}$
- C. $\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Q_1)}$
- D. $\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) - (Me - Q_1)}$

Proefexamen Thijssen 2012 Vraag 5

Geef de 13de term van $[x - (y \cdot \log_2 131072)]^{17}$ en werk deze term zoveel mogelijk uit

- A. $6188 \cdot x^{13} \cdot 17y^5$
- B. $2380 \cdot x^5 \cdot 17y^{12}$

- C. $6188.17^{12}x^5.y^{12}$
D. $2380.17.x^{12}.y^5$

Fictieve oefening

Bereken de scheefheidscoëfficiënt van Fisher Υ_1 voor de verdeling met volgende 4 waarnemingen: 1, 2, 4 en 8. Dus $n = 4$ en gemiddelde is 3,75

3. Oplossingen

Proefexamen Thijssen 2008 Vraag 4

Gegeven:

Mediaan = 45

Gemiddelde = 42,8

Eerste kwartiel (Q_1) = 22

Middenkwartiel ($(Q_3+Q_1)/2$) = 38.5

Oplossing:

$$Q_3 = 38,5 \cdot 2 - Q_1 = 38,5 \cdot 2 - 22 = 55$$

Nu kunnen we de scheefheidscoëfficiënt van Yule berekenen:

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Q_1)} = \frac{(55 - 45) - (45 - 22)}{(55 - 22)} = -0,39 = \text{linksscheef.}$$

Proefexamen Thijssen 2012 Vraag 5

Gevraagd: Welke van volgende formules is GEEN berekening van de scheefheidscoëfficiënt van Yule?

Oplossing: dit is een rechtstreeks uit de theorie af te leiden

➔ Antwoord D

Proefexamen Thijssen 2012 Vraag 5

Geef de 13de term van $[x - (y \cdot \log_2 131072)]^{17}$ en werk deze term zoveel mogelijk uit

Oplossing: De 13de term is de term waarvoor $k = 12$:

$$(a + b)^{17} = \sum_{i=0}^{17} \binom{17}{k} a^k b^{17-k} \quad \text{waarbij } b = x \text{ en } a = -(y \cdot \log_2 131072)$$

(noot: wanneer we x gelijkstellen aan a , krijgen we voor b een negatief getal, en komt er in de uitkomst ook een negatief getal, geen van de opties voorziet een negatief getal)

We berekenen het $\log_2 131072$, dit is 17 dus $a = -17 \cdot y$

De 13de term is dan $= \binom{17}{12} (-17y)^{12} x^5$

$$= \frac{17!}{12!5!} 17^{12} y^{12} x^5$$

$$= \frac{17 \cdot 16 \cdot 15 \cdot 14 \cdot 13}{5 \cdot 4 \cdot 3 \cdot 2} 17^{12} y^{12} x^5$$

$$= 6188 17^{12} y^{12} x^5$$

→ Antwoord C

Fictieve oefening

Bereken de scheefheidscoëfficiënt van Fisher Υ_1 voor de verdeling met volgende 4 waarnemingen: 1, 2, 4 en 8. Dus $n = 4$ en gemiddelde is 3,75

Volgnummer	x_i	x_i^2	x_i^3	$x_i - x_{\text{gem}}$	$(x_i - x_{\text{gem}})^2$
1	1	1	1	-2,75	7,5625
2	2	4	8	-1,75	3,0625
3	4	16	64	0,25	0,0625
4	8	64	512	4,25	18,0625
Totaal	15	85	585		28,75
: 4	3,75	21,25	146,25		7,1875
	$= \mu_1'$	$= \mu_2'$	$= \mu_3'$		$= \mu_2$

We gebruiken volgende formule:

$$\Upsilon_1 = \frac{\mu_3' - 3 \mu_1' \mu_2' + 2(\mu_1')^3}{\mu_2' \sqrt{\mu_2'}} \text{ of } = \frac{\mu_3' - 3 \mu_1' \mu_2' + 2(\mu_1')^3}{\sigma^3}$$

Het eerste moment ten opzichte van nul vinden we door alle waarden van x op te tellen en te delen door n . μ_1' is dus 3,75 en komt overeen met het gemiddelde

Het tweede moment ten opzichte van nul vinden we door alle kwadraten van x op te tellen en te delen door n . μ_2' is dus 21,25

Het derde moment ten opzichte van nul vinden we door alle derdemachten van x op te tellen en te delen door n . μ_3' is dus 146,25

De tweede momenten in de noemer zijn momenten ten opzichte van het gemiddelde (dus de gekwadrateerde afwijkingen ten opzichte van het gemiddelde optellen en dit totaal delen door n). $\mu_2 = 7,1875$

Nu kunnen we de waarden voor Υ_1 invullen in de formule:

$$\Upsilon_1 = \frac{\mu_3' - 3 \mu_1' \mu_2' + 2(\mu_1')^3}{\mu_2' \sqrt{\mu_2'}} = \frac{146,25 - 3(3,75 \cdot 21,25) + 2(3,75)^3}{7,1875 \sqrt{7,1875}} = \frac{12,65625}{19,27} = 0,66$$

Je kan voor de noemer ook σ^3 berekenen en de tweede formule gebruiken. Voor de berekening van σ moet je de som van de gekwadrateerde afwijkingen ten opzichte van het gemiddelde delen door n en dan de vierkantswortel nemen. Dat geeft $\sqrt{7,1875} = 2,68095$. Dus $\sigma = 2,68095$ en σ^3 is dan 19,27 wat overeenkomt met de waarde uit de andere formule.