

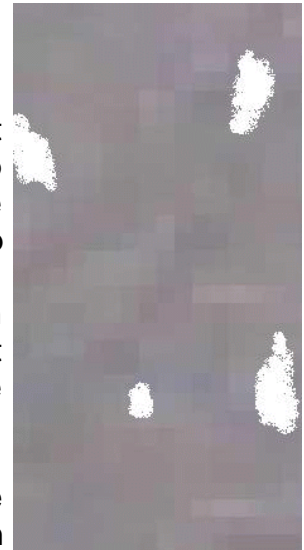
Statistiektips: Wanneer *gemiddelde*, en wanneer *mediaan* gebruiken?

Voorbeeld 1:

Ooit moest ik een softwareprogramma schrijven voor het analyseren van medische diagnosestrips. Op de (poreuze papieren) strip werd een druppel aangebracht van een lichaamsvloeistof die we hier niet nader hoeven te omschrijven. Daardoor werden sommige stroken op de strip donkerder grijs tot zwart. De grijstint verraadde welke ziekte of genetische afwijking de patiënt had. (Op het zicht lijkt dit wat op een zwangerschapstest.)

Om de juiste gradatie van grijs nauwkeuriger vast te stellen, werden de strips met een skanner ingelezen. Het bepalen van de grijstint kwam er dan op neer van het gemiddelde te berekenen van alle pixelwaarden (0=zwart, tot 255=wit) die te zien waren op een strook. Nu was er één probleem: dikwijls waren er "vlekjes" op de strips, stukjes die wit bleven omdat ze niet goed ontwikkelden. Als de grijswaarde van de meeste pixels bv. schommelde rond de 80, en 10% van de pixels was wit gebleven, dan ging het gemiddelde daardoor drastisch omhoog, waardoor de diagnose foutief zou zijn.

Oplossing: neem de mediaan i.p.v. het gemiddelde! Zolang er minder dan 50% witte pixels zijn, blijft de mediaan immers in de buurt van 80.

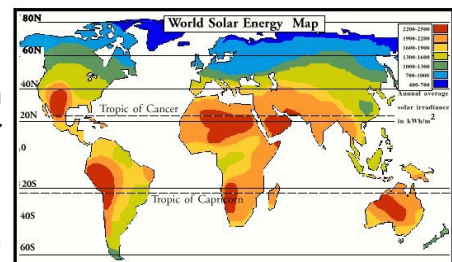


teststrook met vlekken

Voorbeeld 2:

Ik zoek een goede plaats om een zonne-energiecentrale te bouwen. Ik beschik over zonneschijngegevens van diverse plaatsen, verzameld over vele jaren. Ik kan per plaats de mediaan berekenen of het gemiddelde. Waarmee ben ik het meest?

Als ik de mediaan bereken, dan heb ik geen idee van hoeveel ik kan verdienen in de superzonnige jaren, of hoeveel ik kan verliezen in supersombere jaren. Als ik daarentegen het gemiddelde bereken, kan ik tamelijk nauwkeurig schatten hoeveel mijn lange-termijnopbrengst zal zijn (tenzij het klimaat verandert natuurlijk).



Besluit:

Als uitschieters in de gegevens niet van belang zijn, dan is de *mediaan* waarschijnlijk te verkiezen. Als alle gegevens even belangrijk zijn, ook de extreme waarden, dan is het *gemiddelde* normaal gezien te verkiezen.

Oefeningen:

a) In een klas van 15 leerlingen hebben er 11 voor hun toets 10 op 10. 4 leerlingen hebben geen klop gedaan en hebben dus 0/10.

Welke centrummaat (mediaan/gemiddelde) geeft het best de begrijpelijkheid van de leerstof weer?

b) Ik wil een handelszaak in karnavalsartikelen overnemen.

Welke centrummaat in verband met de maandelijkse verkoopscijfers interesseert mij het meest?