**SAMENVATTING JOY OF STATS**

Statistiek is, volgens de Zweedse bioloog Hans Rosling, bij uitstek de manier om ons te vertellen of die dingen die we denken en vermoeden ook echt waar zijn. Ze kunnen ons de wereld duiden. Daarnaast geven de mogelijkheden die moderne computers ons leveren de kans om wetenschap op een totaal andere manier te gaan uitvoeren.

Rosling geeft een aantal voorbeelden waaruit de kracht van statistiek duidelijk word, zo heeft de SFPD haar database met misdaden opengesteld voor publiek. Hiermee is een aantal statistici aan de slag gegaan en hebben een kaart geplot met waar in de stad misdaden worden gepleegd. Dit wordt inmiddels weer door de politie gebruikt om efficiënter haar werk te doen.

Terug naar de geschiedenis van de statistiek, deze begon 2 eeuwen terug om grip te krijgen op hoe de bevolking groeide. Men dacht dat zweden 20 miljoen inwoners had, maar dit bleek na de tellingen slechts 2 miljoen te zijn. Dit maakte het land strategisch veel minder sterk dan dat men eigenlijk dacht. Het was dan ook van belang dat de vijand deze informatie niet in haar bezit kreeg. Statistiek bleek verder te gaan dan een lijst met cijfers, de informatie was van strategisch en economisch belang.

Cijfers alleen zeggen niks, je moet ze analyseren. Dit begon met het gemiddelde bepalen en door de eeuwen heen zijn uit al die analyses verschillende belangrijke meetwaarden gebleken: variatie en distributie zijn minstens zo belangrijk als gemiddelden, de normale verdeling komt in allerlei vormen terug van de lengte tot mensen tot de sterkte van beton. Evenals de Poisson verdeling die veelal gebruikt word om onsamenhangende bewegingen te verklaren.

Terug naar de geschiedenis en het belang van data en vooral de analyse daarin: Nightingale verzamelde data van overleden soldaten, hieruit bleek dat de meeste soldaten stierven door ziekten die ze opliepen in het ziekenhuis, niet aan de verwondingen die ze tijdens de strijd opliepen.

Door het vertalen van cijfers naar beeld zoals Nightingale deed is het mogelijk veel grotere datasets te kunnen gebruiken en vooral begrijpelijk te kunnen maken.

De ontwikkeling van data-analyse ging verder met de ontdekking dat je correlaties wiskundig kon bepalen de eerste grote correlatie die werd aangetoond: roken en longkanker.

Data is tegenwoordig de zuurstof van de wetenschap die zich voornamelijk bezighoud met het verklaren van de wereld om ons heen. Data en statistiek zijn hier uitstekende tools voor. Zo gebruikt Google gebruikt het hele internet om twee talige documenten te vinden waaruit vervolgens een kansberekening kan worden gemaakt wat een woord of zinsnede in een andere taal betekend. Met deze ontwikkeling wordt duidelijk: computers, data en statistiek kunnen leidend worden in wetenschappelijk onderzoek.

Thijs de Jong  
2094349