

BIJLAGE A BIJ MODEL BEREKENING STEEKPROEF METING VREEMDE FLESSEN

UITLEG AANPAK STEEKPROEF

Let op: het model berekening steekproef meting vreemde flessen betreft een digitaal (excel) bestand. Deze is te downloaden via www.bnrfles.nl/statiegeld

Probleem

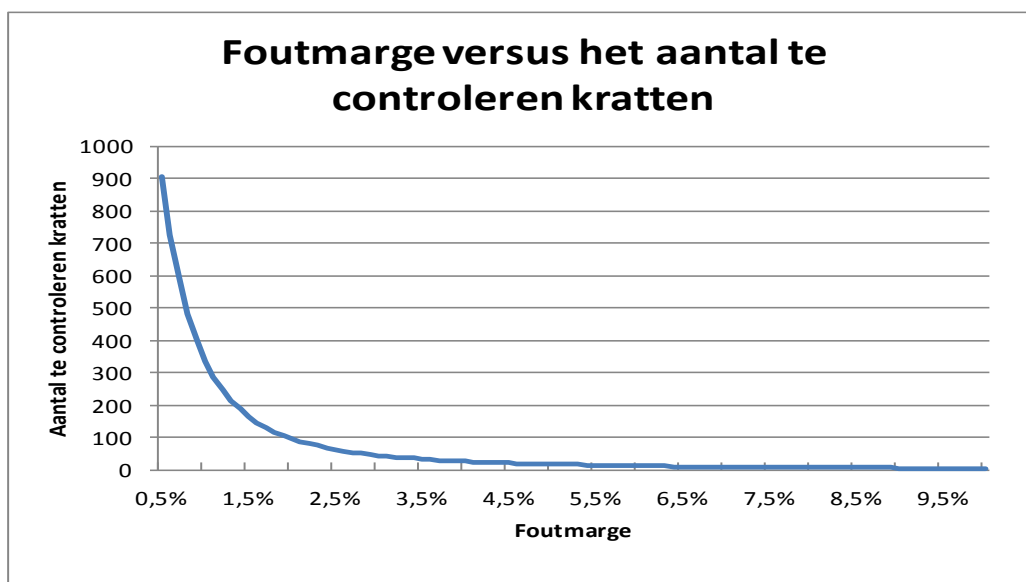
Dagelijks komen bij de bier brouw fabrieken vrachtwagens kratten met lege flesjes bezorgen. Deze kratjes komen bij de supermarkten vandaan. Er blijken echter nogal flesjes van andere merken in deze kratten te zitten (“vervuiling”). Het kan wenselijk zijn voor brouwers/merkeigenaren/importeurs om te meten hoeveel vreemde flessen zij ontvangen. Zij willen echter uiteraard niet alle afzonderlijke flesjes in de vrachtwagen tellen. Hoeveel kratten moeten zij daadwerkelijk tellen om toch een goed beeld te krijgen van het aantal foute flesjes in de vrachtwagen?

Oplossing

Hoe precies wij het percentage foute flesjes in de vrachtwagen kunnen schatten hangt af van de hoeveelheid flesjes die wij tellen. Als wij alle flesjes in de vrachtwagen zouden controleren zouden wij honderd procent zeker weten hoeveel flesjes daarvan van een ander merk zijn. Als wij maar één kratje zouden tellen kunnen wij weinig zeggen over het daadwerkelijke percentage foute flesjes.

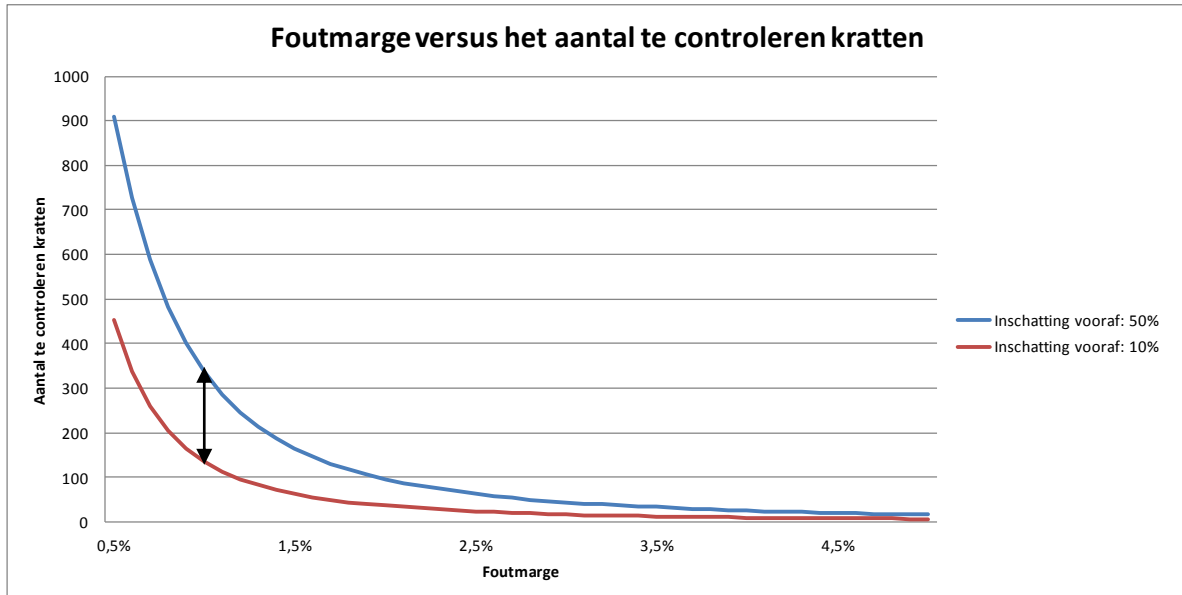
Hoe precies wij het percentage foute flesjes in een vrachtwagen kunnen schatten duiden wij aan met een foutmarge (afhankelijk van een betrouwbaarheidsinterval). Wij zouden bijvoorbeeld kunnen zeggen (met 95% zekerheid) dat het percentage foute flesjes ongeveer 4% is en zo goed als zeker tussen de 2% en de 6% valt. Op het moment dat we meer kratjes controleren zouden wij deze voorspelling preciezer kunnen maken. We zouden dan kunnen zeggen (met 95% zekerheid) dat het percentage flesjes ongeveer 4% is en zo goed als zeker tussen de 3% en de 6% valt.

Stel, we gaan er vanuit dat een normaal gevulde vrachtwagen pakweg 50.000 bierflesjes (2100 kratten) vervoert. Stel vervolgens veiligheidshalve dat de kans fiftyfifty is dat er een flesje fout (“vervuilt”) is of niet. Met behulp van het volgende figuur kun je dan zien hoeveel kratten er gecontroleerd moeten worden ten opzichte van de precisie van je inschatting (foutmarge).



Bovenstaande figuur laat zien dat je bij een volle vrachtwagen van 2100 kratjes je pakweg 350 kratjes moet controleren zodat je de vervuiling kunt voorspellen met een foutmarge van 1 procent.

Om een voorspelling te maken van het aantal foute flesjes moeten we van te voren ook een inschatting maken van hoeveel foute flesjes er ongeveer in de vrachtwagen aanwezig zouden zijn. Want op het moment dat je vooraf een goede inschatting maakt van het aantal foute flesjes hoef je minder kratjes te controleren voor dezelfde precisie, zie volgende figuur.



In bovenstaande figuur zien we dat het aantal te controleren kratjes fors daalt als vooraf richting gegeven kan worden aan de inschatting van het percentage foute flesjes, bij een gelijkblijvende precisie (foutmarge).

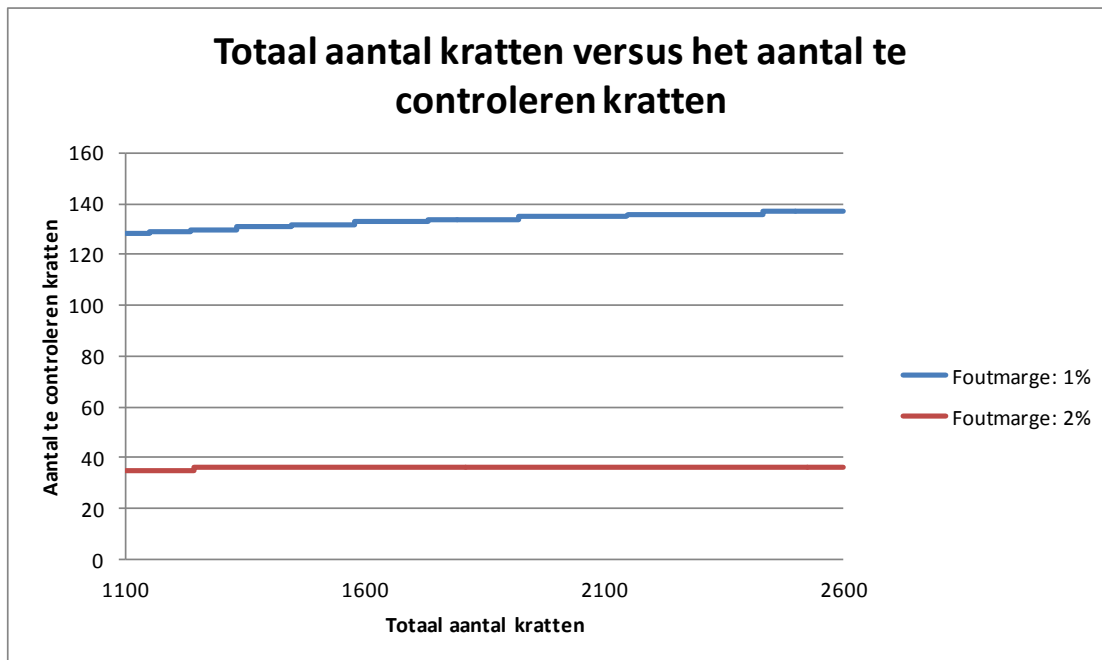
Partijen worden geadviseerd van te voren een gedegen inschatting te maken van het percentage “vervuilde” flesjes, waarbij het van belang is dat het percentage niet te laag wordt ingeschat, maar zeker ook niet te hoog. Een gedegen inschatting scheelt namelijk fors met het aantal in te controleren kratjes. Echter, weet dat je het risico loopt een minder accurate voorspellingen te doen indien het werkelijke aantal “vervuilde” flesjes ver boven je vooraf ingeschatte percentage zit. Op het moment dat van tevoren het aantal foute flesjes te laag wordt ingeschat zullen wij het aantal te tellen kratjes onderschatten (en neemt de *effectieve* precisie af).

Echter, op het moment dat van tevoren het aantal foute flesjes te hoog wordt ingeschat zullen wij het aantal te tellen kratjes overschatten (en neemt de *effectieve* precisie toe). Het is dus belangrijk een goede/veilige schatting te maken. Let hierbij op dat met meest veilige schatting, 50% (pure random) wordt bedoeld.

Op basis van historische gegevens van Nederlandse Brouwers blijkt dat het gewogen gemiddelde van ontvangst van vreemde flessen (in merkkratten en pinolenkratten samen) op ca. 5,6% ligt. We weten echter ook dat de ontvangst van vreemde flessen per krat sterk kan verschillen. Het verschil is het grootst tussen merkkratten en pinolenkratten, waarbij de gemiddelde ontvangst van het aantal vreemde flessen in merkkratten ruim onder de 5% ligt en het aantal vreemde flessen in pinolenkratten aanzienlijk hoger is.

Op basis van deze gegevens is voorzichtigheidshalve in het model uitgegaan van een ontvangst van vreemde flessen in merkkratten van 10% en pinolenkratten van 25%.

Het totaal aantal kratjes in vrachtwagen heeft echter maar marginaal invloed op het aantal te controleren kratjes. Intuïtief is dit moeilijk te begrijpen, maar dit komt omdat de mate van precisie en het van te voren ingeschatte percentage foute flesjes een veel grotere rol speelt. In het volgende figuur geven we een inzicht van het aantal te controleren kratjes ten opzichte van het totaal aantal kratten (met 24 flesjes) in een vrachtwagen. Daarbij zijn 2 grafieken afgebeeld met verschillende mate van precisie (foutmarge) en is onze inschatting vooraf dat het aantal foute flesjes niet meer is dan 10%. Hierbij nemen wij aan dat een gewone vrachtwagen die half gevuld is ongeveer 1100 kratjes (=26400 flesjes) vervoert en dat een bomvolle vrachtwagen 2600 kratjes kan vervoeren.



Let op! Bovenstaande figuur leert ons dat als een volle vrachtwagen voor de ene helft gevuld is met merkkratten A en de andere helft met merkkratten B, en je wilt voor ieder merk afzonderlijk het “vervuilings” percentage berekenen, dien je dus 2 keer 35 kratjes (in geval van 2% foutmarge) te controleren, en niet 2 keer 18 kratjes wat intuïtief logischer lijkt.

Laten we bovenstaande theorie uitleggen aan de hand van een voorbeeld:

- We hebben een vrachtwagen met 2100 merkkatjes
- We accepteren een foutmarge van 2%
- Op basis van ervaring schatten we vooraf in dat het aantal foute (“vervuilde”) flesjes niet meer is dan 5%, maar waarbij we in de berekening de veilige marge aanhouden en dus uitgaan van maximaal 10%

Het aantal te controleren kratjes is dan 36. En stel dat het getelde aantal foute flesjes uit deze steekproef van 36 kratjes 43 is, dan kunnen we met 95% zekerheid zeggen dat de “vervuiling” van de totale vrachtwagen tussen de 3,5 en 6,4 procent ligt. Ofwel, met 97,5% zekerheid kunnen we stellen dat minimaal 3,5 procent van de flesjes in de vrachtwagen vervuult zijn. Deze effectieve foutmarge (ook wel bandbreedte) ligt dus onder de 2% omdat je vooraf stelt dat het percentage “vervuiling” maximaal 10% is.

Appendix A: Achterliggende wiskunde

De formule voor het betrouwbaarheidsinterval is als volgt als wij percentages willen schatten:

$$p - z \cdot \sqrt{\frac{p(1-p)}{n}} < \pi < p + z \cdot \sqrt{\frac{p(1-p)}{n}}$$

Waar p het percentage is dat je vind in je steekproef, π het werkelijke percentage in de steekproef en n de grootte van je steekproef (het aantal te controleren flesjes). De foutmarge is de maximale fout tussen π en p . Door deze foutmarge vast te stellen op een bepaald percentage kunnen wij de steekproefgrootte vaststellen:

$$marge = z \cdot \sqrt{\frac{p_{prior}(1-p_{prior})}{n}}$$

Waar p_{prior} de vooraf ingeschatte kans is op het percentage vervuiling. Uitgaande van een betrouwbaarheid van 95% kunnen wij de optimale steekproef uitrekenen door bovenstaande formule om te schrijven

$$n = \frac{1.96^2 \cdot p_{prior}(1-p_{prior})}{marge^2}$$

De uitkomst van deze formule ronden wij altijd naar boven af. Wij moeten nu echter ook nog een correctiefactor bepalen omdat wij niet te maken hebben met een oneindige populatie maar met een eindige populatie deze correctie factor is als volgt:

$$correctiefactor = \frac{N-n}{N-1}$$

Waar N de grootte van de populatie is. De foutmarge voor een eindige populatie verandert vervolgens iets:

$$marge = z \cdot \sqrt{\frac{p_{prior}(1-p_{prior})}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Weer omschrijven naar n leidt tot de volgende formule:

$$n = \frac{1.96^2 \cdot p_{prior}(1-p_{prior})}{marge^2} \cdot \frac{N}{n+N-1}$$

Bron: Kish, L. (1965). *Survey Sampling*. Wiley. ISBN 0-471-48900-X.