

Een voorbeeld van een hypothesetoets

Een speelgoedfabrikant koopt een machine om aan de lopende band speelgoedhaasjes te maken. Bij aankoop van de machine staat in het contract dat minstens 80% van de haasjes die uit de machine komen goed zijn, dus hoogstens 20% uitval.

De beheerder van het machinepark twijfelt na enige tijd of de machine wel goed is. Hij denkt dat meer dan 20% van de haasjes er niet goed uitkomen. Vele haasjes missen een oor en vele hebben er zelfs drie.

Hij zou een steekproef willen nemen maar weet niet goed hoe hij dat moet aanpakken, dus hij haalt er iemand bij die wiskunde B1 in zijn pakket heeft gehad en vraagt hoe dit aan te pakken.

Die besluit om een steekproef te gaan nemen van 1000 haasjes die uit de machine komen. Voordat hij dat doet gaat hij wat rekenen, een model maken.

Hij denkt te maken te hebben met een binomiale verdeling $\text{Bin}(n,p)$, X aantal goede haasjes.

- a. Is dit correct? Zo ja wat is n en wat is p ?

Ervan uitgaande dat de machine goed is, dus $p=0,8$, bekijkt hij de kansverdeling die bij de steekproef van 1000 hoort.

- b. Maak een schets van het bijbehorende histogram
c. Bepaal ook $E(X)$ en $SD(X)$ (rond $SD(X)$ af op 2 decimalen)

Nu wil hij de grens g zoeken waarbij $P(X=0) + P(X=1) + P(X=2) + \dots + P(X=g)$ ongeveer 5% is. Dan kan hij als straks het aantal goede haasjes in dit gebied ligt, $aantal \leq g$, zeggen dat de kans daarop slechts 5% is, dus dat de gestelde 80% te hoog zal zijn gesteld.

Ho vindt hij die g . Daartoe gaat hij de verdeling benaderen door een normale verdeling.

- d. Teken in histogram van b) een vloeiende klokvormige kromme. De stochast daarbij noem je Z , die dan normaal is verdeeld volgens $\text{Norm}(\mu, \sigma)$
Bereken dat gemiddelde μ en de spreiding σ .
e. Met de functie $\text{invNorm}(a,b,c)$ kan hij g vinden. De letters zijn a : het percentage dat je wilt hebben, dus 0,05 hier; b is gemiddelde dus μ en c de spreiding σ .
De functie invNorm berekent de grens als je oppervlakte weet. Bepaal de grens.

Nu kan hij de steekproef uitvoeren en als het resultaat in het kritieke gebied terecht komt ($0,1,2,\dots,g$) dan verwerpt hij de veronderstelling dat minstens 80% goed is. Als de uitkomst boven de g uitkomt kan hij die gestelde 80% niet verwerpen.

- f. De gevonden g is een reëel getal, tussen twee gehele getallen $k \leq g \leq l$ in. Bereken met binomcdf ook de kans $P(X \leq k)$ en $P(X \leq l)$. Ligt de 5% ertussen?

In hoofdstuk S3 wordt dit model geformaliseerd:

$H_0: p=0,80$ (lees; de nulhypothese is dat $p=80\%$)

$H_1: p<0,8$ (de alternatieve hypothese is dat het percentage goede haasjes kleiner is)

$N=1000$ de steekproefgrootte is 1000

$\alpha=0,05$ de onnauwkeurigheid drempel is 5%

Met bovenstaande kun je kritieke gebied (verwerp-gebied) bepalen en daarna steekproef uitvoeren.

Antwoorden: $N=1000$, $p=0,8$, $\mu=800$, $\sigma=12,65$, $\text{Norm}(800;12.65)$

$g=779,2$, $P(X \leq 779) = 0,054$ en $P(X \leq 780) = 0,063$