

## Hoofdstuk 7

### Toetsen van hypothesen

Toetsen van hypothesen is, o.a. in de medische en chemische wereld, een veel gebruikte statistische techniek. Het wordt vaak gebruikt om een *gevestigde norm* eventueel te weerleggen, bijvoorbeeld om aan te tonen dat een nieuw geneesmiddel een betere kans biedt op genezing dan het klassieke geneesmiddel.

Het is een techniek die toelaat conclusies te trekken over een hele populatie gebaseerd op steekproefgegevens, conclusies waarvan we niet zeker zijn maar waarvan we het risico op een verkeerde uitspraak onder controle hebben. Het is de bedoeling om deze techniek toe te lichten met twee voorbeelden en hierbij aandacht te besteden aan de typische statistische denkwijze.

#### 7.1 Toets voor het gemiddelde van een normale verdeling

##### 7.1.1 De standaardafwijking is gekend

Een gezaghebbend tijdschrift publiceert dat het geboortegewicht in Vlaanderen normaal verdeeld is met een gemiddelde van 3.3 kg en een standaardafwijking van 0.55 kg.

Een gynaecoloog heeft de indruk dat het gemiddelde geboortegewicht in zijn kliniek groter is dan 3.3 kg. Om zijn hypothese te toetsen houdt hij gegevens bij i.v.m. dertig kinderen, zie lijst GK1 pagina 116.

De gynaecoloog berekent het gemiddelde van zijn steekproef en bekomt als resultaat  $\bar{x} = 3.483$  kg. Hij vraagt zich af of hij met dit resultaat mag besluiten dat het gemiddelde geboortegewicht in zijn kliniek groter is dan 3.3 kg.

Om zijn veronderstelling te toetsen, start hij met het opstellen van een kansmodel voor de populatie waaruit de steekproef werd getrokken. De arts gaat ervan uit dat de geboortegewichten in zijn kliniek normaal verdeeld zijn met een standaardafwijking gelijk aan deze van gans Vlaanderen,  $\sigma = 0.55$  kg. Een wat onrealistische veronderstelling die we in de volgende paragraaf zullen weglaten.

Vervolgens formuleert hij twee hypothesen omtrent de parameter  $\mu$  van de populatie: de *nullhypothese*,  $H_0$  (de norm die hij graag zou weerleggen), en een *alternatieve hypothese*,  $H_1$  (zijn vermoeden dat hij graag zou willen aantonen of statistisch bewijzen). De alternatieve hypothese is dat het gemiddelde geboortegewicht groter is dan 3.3 kg. We noteren de hypothesen als volgt :

$$\begin{cases} H_0 : \mu = 3.3 \text{ kg} \\ H_1 : \mu > 3.3 \text{ kg} \end{cases}$$

We toetsen  $H_0 : \mu = 3.3$  versus  $H_1 : \mu > 3.3$ .

De hypotheses-toets verloopt verder als volgt. We veronderstellen in de redenering dat de nullhypothese  $H_0$  waar is. Indien het gemiddelde geboortegewicht  $\bar{x}$  van de steekproef van de gynaecoloog *te groot* is, verwerpen we de nullhypothese  $H_0$  en aanvaarden we de alternatieve hypothese  $H_1$ .

Wat betekent *te groot* ?

Zou je uit een populatie met een gemiddelde van 3.3 kg gemakkelijk een steekproefgemiddelde van 3.483 kg of zelfs meer kunnen verkrijgen?

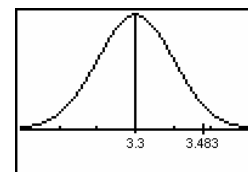
Het verkregen steekproefgemiddelde  $\bar{x} = 3.483$  kg is een door het toeval verkregen waarde van de stochast  $\bar{X} = \frac{1}{30}(X_1 + \dots + X_{30})$ , waarbij elke  $X_i$  verdeeld is zoals de populatiestochast  $X$ .

Om een uitspraak te doen over het al dan niet te groot zijn van de geobserveerde waarde  $\bar{x} = 3.483$ , bekijken we de ligging van  $\bar{x}$  op de grafiek van de dichtheidsfunctie van  $\bar{X}$ . Indien we veronderstellen dat  $H_0$  waar is, geldt :

$$X \sim N(3.3, 0.55) \text{ en } \bar{X} \sim N\left(3.3, \frac{0.55}{\sqrt{30}}\right) \text{ met } \frac{0.55}{\sqrt{30}} \approx 0.1.$$

De stochast  $\bar{X}$  is de *toetsingsgrootheid*, dit is een functie van de steekproef waarmee we een uitspraak gaan doen over de populatieparameter  $\mu$ .

De waarde  $\bar{x} = 3.483$  bevindt zich erg rechts.  $\bar{x}$  ligt zelfs zo ver naar rechts dat de kans (indien  $H_0$  waar is) om zo'n geobserveerde waarde te krijgen of groter, zeer klein is.

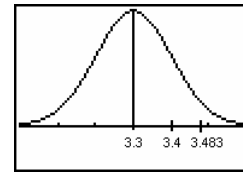


Dichtheidsfunctie  $\bar{X}$

Een dergelijke geobserveerde waarde onder de nullhypothese kan dus moeilijk verklaard worden door louter toeval.

We zeggen dat de geobserveerde waarde  $\bar{x}$  *statistisch significant* is en verwerpen de nulhypothese.

Stel even dat de geobserveerde waarde  $\bar{x}$  van de gynaecoloog 3.4 kg was. Op de grafiek van de dichtheidsfunctie van  $\bar{X}$  zien we dat in dit geval de kans (indien  $H_0$  waar is) om een geobserveerde waarde van 3.4 kg of meer te verkrijgen, al heel wat groter is dan voor  $\bar{x} = 3.483$  kg. Het is niet meer zo evident om de nulhypothese te verwerpen.



Dichtheidsfunctie  $\bar{X}$

Hoe verder de geobserveerde waarde  $\bar{x}$  zich bevindt in de positieve richting, hoe meer we geneigd zijn om de nulhypothese te verwerpen.

Een karakteristiek die gebruikt wordt om weer te geven hoe sterk de geobserveerde waarde afwijkt van de nulhypothese, is de **p**-waarde (*probability value* of *overschrijdingskans*).

De **p**-waarde is de kans, indien de nulhypothese waar is, om een waarde te verkrijgen van de toetsingsgrootheid die minstens even extreem is als de geobserveerde waarde. Met extreem bedoelen we waarden die nog meer zouden wijzen in de richting van de alternatieve hypothese.

Voor de geobserveerde waarde van de gynaecoloog geeft dit :

```
normalcdf(3.483,
10^99,3.3,0.55√
(30))
.0341954071
```

$$P(\bar{X} \geq 3.483) = 0.034 = 3.4\%$$

Deze kans is klein en vormt een goed bewijs tegen de nulhypothese. Slechts in 34 steekproeven op 1000 zal een dergelijke gebeurtenis optreden onder de nulhypothese.

Hoe kleiner de **p**-waarde hoe meer het aanneembaar is de nulhypothese te verwerpen. Hoe klein moet de **p**-waarde zijn om de nulhypothese te verwerpen ?

Dit hangt af van de situatie en de belangrijkheid of de gevolgen van de beslissing. Om te beslissen hanteert men de volgende regel. Men kiest vooraf een *significantieniveau*  $\alpha$  : we verwerpen  $H_0$  bij een **p**-waarde  $\leq \alpha$  (we zeggen dat de steekproefdata statistisch significant zijn op het  $\alpha$ -niveau) en aanvaarden  $H_0$  bij een **p**-waarde groter dan  $\alpha$ .

Om historische redenen, gebonden aan de statistische tabellen, kiest men vaak  $\alpha = 0.05 = 5\%$  of  $\alpha = 0.01 = 1\%$ .

De gynaecoloog had gekozen voor  $\alpha = 0.05$ , op basis van de **p**-waarde = 0.034 verwerpt hij dus de nulhypothese en aanvaardt de alternatieve hypothese  $\mu > 3.3$ . De statisticus beperkt zich vaak tot het rapporteren van de **p**-waarde en laat de beslissing meestal over aan zijn opdrachtgever die zijn eigen significantieniveau op voorhand bepaalde.

De **p**-waarde geeft meer informatie dan het significantieniveau dat alleen dient om de grens te leggen tussen aanvaarden en verwerpen van de nulhypothese. Zo zijn de steekproefdata van de gynaecoloog significant op het 5% niveau, ook op het 4% of het 3.5% niveau, maar niet op het 3% niveau.

De **p**-waarde is bijgevolg het kleinste significantieniveau waarvoor de steekproefdata significant zijn.

Wanneer men enkel beschikt over statistische tabellen van de standaard normale verdeling is men verplicht een andere steekproefgrootte te nemen dan  $\bar{X}$ , nl.

$$Z = \frac{\bar{X} - 3.3}{0.55/\sqrt{30}} \sim N(0,1).$$

We spreken dan over de Z-test. We berekenen de **p**-waarde in dit geval als volgt :

$$\begin{aligned} P\left(\frac{\bar{X} - 3.3}{0.55/\sqrt{30}} \geq \frac{\bar{x} - 3.3}{0.55/\sqrt{30}}\right) &= P\left(Z \geq \frac{3.483 - 3.3}{0.55/\sqrt{30}}\right) \\ &= P(Z \geq 1.822) = 0.034 \end{aligned}$$

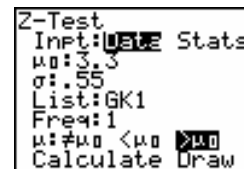
Het uitvoeren van deze toets doe je met de **TI-83** met het commando **STAT<TESTS> 1:Z-Test**.

De hieronder afgebeelde data (in kg) van de gynaecoloog plaatsen we in de lijst **GK1**.

3.54	3.67	3.13	3.76	3.99	4.13	3.22	1.63	2.77	3.36
3.49	3.22	3.04	3.76	3.86	3.58	4.08	3.58	3.22	3.13
2.72	3.22	3.4	4.31	3.76	3.36	4.58	3.13	4.08	3.76

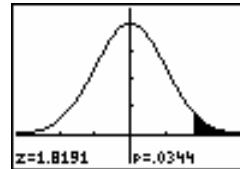
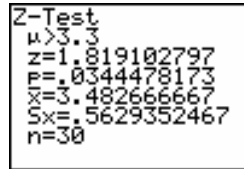
Vul dan het **Z-Test**-venster in zoals hiernaast is afgebeeld.

Indien de resultaten van de steekproef in een lijst gegeven zijn, selecteer je voor het item **Inpt** de optie **Data** en als je enkel statistische kengetallen kent van de steekproef, bv.  $\bar{x} = 3.4$ , selecteer je de optie **Stats**.



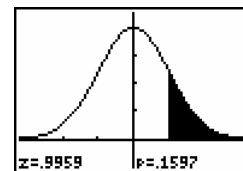
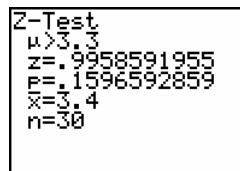
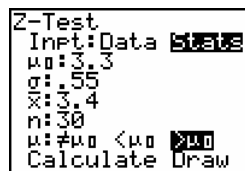
Plaats de cursor op **Calculate** en druk op **ENTER**. Het resultaat vind je links op de onderstaande figuur.

Activeer opnieuw het **Z-Test**-venster, selecteer **Draw** en druk op **ENTER**. Het resultaat vind je rechts op de onderstaande figuur.



$z$  is de geobserveerde waarde van de toetsingsgrootheid  $Z$  met  $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ .

Voor het geval  $\bar{x} = 3.4$  kg vind je hieronder wat schermafdrukken van de **TI-83** met o.a. de resultaten van de **Z-Test**.

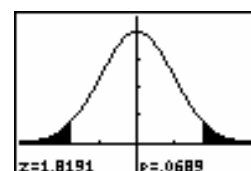
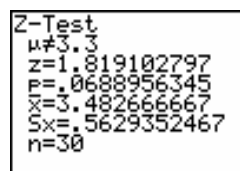
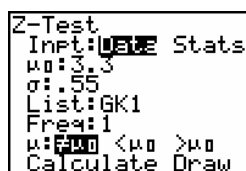


De **p**-waarde bij  $\bar{x} = 3.4$  kg of  $z = 0.9958$  is 0.1597. Dit is gevoelig groter dan  $0.05 = 5\%$ , zodat we de nulhypothese niet kunnen verwerpen op een 5% significantieniveau. We aanvaarden de nulhypothese, wat niet betekent dat we ze hebben “bewezen”. Sommige auteurs spreken derhalve liever over het niet verwerpen van de nulhypothese i.p.v. het aanvaarden van de nulhypothese.

De juist uitgevoerde toets noemt men een *éénzijdige toets* omdat men op voorhand vermoedt in welke richting een afwijking van  $H_0$  verwacht wordt.

Indien men dit niet op voorhand vermoedt, nemen we als alternatieve hypothese dat de populatie verschilt van wat er in de nulhypothese gesteld wordt. We toetsen  $H_0: \mu = 3.3$  versus  $H_1: \mu \neq 3.3$ . Dit noemen we een *tweezijdige toets*.

Het uitvoeren van deze toets verloopt analoog. Het **Z-Test**-venster vul je in dit geval in zoals hier linksonder. Het resultaat van deze toets is :



Merk op dat in dit geval de **p**-waarde het dubbel is van de **p**-waarde van de éézijdige toets. Dit is een vrij logisch verband daar we voor een tweezijdige toets een even extreme afwijking (d.w.z. met eenzelfde kans) in beide richtingen beschouwen. Zowel te grote als te kleine waarden zijn nu verdacht.

Voor de tweezijdige toets is  $p = 0.068 > 0.05$ . M.a.w. we verwerpen de nulhypothese niet op het 5% significantieniveau.

Dit laatste kan nogal verwarrend overkomen daar de nulhypothese bij de éézijdige toets verworpen wordt en in het geval van een tweezijdige niet. Welk besluit moeten we nemen ?

Indien we à priori geen vermoeden hebben omtrent het gemiddelde voeren we een tweezijdige toets uit a.h.v. een steekproef ( $H_0: \mu = \mu_0$  versus  $H_1: \mu \neq \mu_0$ ).

Indien we de geobserveerde waarde  $\bar{x}$  uit de steekproef zouden gebruiken om de alternatieve hypothese te veranderen (bv.  $H_0: \mu = \mu_0$  versus  $H_1: \mu > \mu_0$ ) om dan een éézijdige toets uit te voeren met dezelfde steekproef, zou dit statistisch oneerlijk zijn t.o.v. de onwetendheid waarmee we gestart zijn.

### 7.1.2 De standaardafwijking is niet gekend

In de vorige paragraaf hebben we ondersteld dat de variantie van de populatie gekend is, wat in de realiteit natuurlijk niet het geval is. Wanneer de variantie niet gekend is, wordt deze vervangen door de steekproefvariantie.

De algemene formule hiervoor is ( $n =$  steekproefgrootte) :  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

De stochastische veranderlijke  $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$  heeft, in de veronderstelling dat  $H_0$

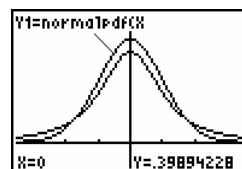
waar is en de populatie normaal verdeeld is, een Student  $t$ -verdeling of kortweg een  $t$ -verdeling met  $(n - 1)$  vrijheidsgraden.

Een  $t$ -verdeling is een symmetrische continue kansverdeling waarbij de dichtheidsfunctie afhankelijk is van één parameter. Deze parameter is het aantal vrijheidsgraden van de  $t$ -verdeling. De dichtheidsfunctie is een goede benadering voor de dichtheidsfunctie van de standaard normale verdeling indien het aantal vrijheidsgraden groot genoeg is. De verwachtingwaarde,  $E(T)$ , van een  $t$ -verdeling is altijd gelijk aan nul.

```

Plot1 Plot2 Plot3
\Y1=normalpdf(X
\Y2=t.pdf(X,2)
\Y3=
\Y4=
\Y5=
\Y6=
\Y7=

```



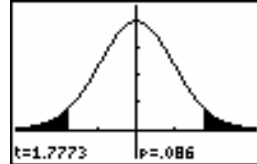
We gebruiken in deze situatie de stochast  $T$  als toetsingsgrootheid en spreken over een T-test.

Met de **TI-83** voeren we de toets als volgt uit : **STAT<TESTS> 2:T-Test.**

Voor de toets  $H_0 : \mu = 3.3$  versus  $H_1 : \mu \neq 3.3$  vullen we het venster in zoals hieronder met als resultaat:

```
T-Test
Inpt: DATA Stats
μ₀: 3.3
List: GK1
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw
```

```
T-Test
μ≠3.3
t=1.777303063
p=.0860090586
x̄=3.482666667
Sx=.5629352467
n=30
```

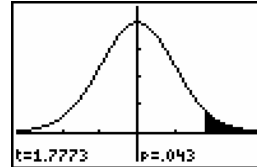


De geobserveerde waarde  $t$  wordt als volgt berekend :  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ .

Voor de toets  $H_0 : \mu = 3.3$  versus  $H_1 : \mu > 3.3$  vinden we het onderstaande resultaat.

```
T-Test
Inpt: DATA Stats
μ₀: 3.3
List: GK1
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw
```

```
T-Test
μ>3.3
t=1.777303063
p=.0430045293
x̄=3.482666667
Sx=.5629352467
n=30
```



## 7.2 Toetsen zonder p-waarde

Indien we niet beschikken over een rekentoestel met voldoende statistische mogelijkheden of over statistische software, is het berekenen van de **p**-waarde omslachtig. Het uitvoeren van een hypothesetoets kan ook op de volgende manier.

Als voorbeeld toetsen we op een 5% significantieniveau.

Voor het voorbeeld uit paragraaf 7.1 geldt dat  $\bar{X} \sim N(0.33, \frac{0.55}{\sqrt{30}})$ .

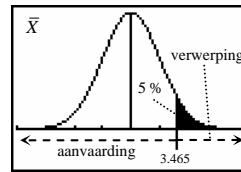
Eerst bepalen we de waarde  $x_0$  waarvoor geldt dat :  $P(\bar{X} \leq x_0) = 1 - 0.05 = 0.95$ .

De **TI-83** berekent deze waarde met het commando :

**2nd[DISTR]<DISTR> 3:invNorm( .**

Deze waarde noemen we het kritisch punt voor ons toetsingsprobleem. Dit punt verdeelt de mogelijke waarden voor  $\bar{X}$  in twee gebieden : een verwerpingsgebied en een aanvaardingsgebied.

```
invNorm(0.95, 3.3
, .55/sqrt(30))
3.465169296
```

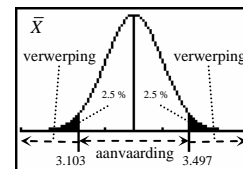


Voor de hypothesetoets  $H_0 : \mu = 3.3$  versus  $H_1 : \mu > 3.3$  vergelijken we de geobserveerde waarde  $\bar{x}$  met het kritisch punt  $x_0$ . Indien  $\bar{x} \geq x_0$  verwerpen we de nulhypothese, zo niet aanvaarden we de nulhypothese.

Voor de data van de gynaecoloog geldt dat  $\bar{x} = 3.483 > x_0 = 3.465$ . M.a.w. we verwerpen de nulhypothese.

Merk op dat voor een geobserveerde  $\bar{x}$ -waarde  $\geq 3.465$  gelegen in het verwerpingsgebied, de corresponderende **p**-waarde  $\leq 0.05 = 5\%$  is. In overeenstemming met vorige paragraaf verwerpen we dan ook de nulhypothese.

In het geval van een tweezijdige toets vinden we op een 5% significantieniveau de volgende verwerpings- en aanvaardingsgebieden.



### 7.3 Type I en Type II fouten

Bij het uitvoeren van een hypothesetoets kunnen de volgende fouten optreden. Indien we  $H_0$  verwerpen wanneer ze juist is, spreken we van een fout van type I en wanneer we  $H_0$  aanvaarden wanneer ze verkeerd is van een fout van type II.

Schematisch geeft dit :

werkelijkheid	$H_0$ is juist	$H_0$ is verkeerd
uitspraak		
verwerp $H_0$	type I fout	juiste uitspraak
aanvaard $H_0$	juiste uitspraak	type II fout



Het significantieniveau bij het uitvoeren van een hypothesetoets bepaalt een beslissingsregel voor het verwerpen van de nulhypothese.

In het geval van een toets op een 5% significantieniveau is de kans om in het verwerpsgebied te komen 5%. M.a.w. de kans om de nulhypothese te verwerpen terwijl ze waar is, is 5 %.

Het significantieniveau van een toets is gelijk aan de kans op het maken van een fout van type I. Het is vanzelfsprekend dat we de kans op een fout van type I zo klein mogelijk willen houden. Jammer genoeg is dit ten koste van de kans op een type II fout die hierdoor groter wordt (we gaan hier niet verder op in).

In vele gevallen is het maken van een type I fout erger dan het maken van een fout van type II. Denk maar aan een rechtszaak waarbij de nulhypothese is dat de persoon onschuldig is.

Indien we toetsen op een 5% significantieniveau is de kans op een type I fout 5 op 100. M.a.w. er is een betrouwbaarheid van 95 % dat de juiste beslissing wordt genomen indien  $H_0$  correct is.

## 7.4 Toets omtrent een populatieproportie

*In een rapport van het ministerie van Landbouw wordt gesteld dat inenting van dieren met een bepaald geneesmiddel van de firma BCP ter preventie van de ziekte dioxin, bij 50 % van de ingeënte dieren levensgevaarlijke neveneffecten veroorzaakt. De firma zelf beweert dat de kans op neveneffecten slechts 20 % bedraagt. Een kippenhouder met zo'n 5000 ingeënte kippen, raadpleegt zijn veearts om een steekproef uit te voeren onder zijn kippen ter controle van het optreden van de neveneffecten. De veearts voert een controle uit op 20 kippen en vindt 6 kippen met neveneffecten. Wat is zijn besluit op het 5% niveau ?*

Elke hypothesetoets kunnen we onderverdelen in :

- het kiezen van het significantieniveau  $\alpha$
- het formuleren van de nulhypothese en alternatieve hypothese omtrent een parameter van een vooropgesteld populatiemodel
- het kiezen van de toetsingsgrootte met zijn verdeling
- het berekenen van de aangenomen waarde van de toetsingsgrootte via de steekproefdata en het bepalen van de  $p$ -waarde
- het nemen van de beslissing

De veearts controleert 20 kippen. Elke kip vertoont neveneffecten met kans  $p$ , met  $p$  de fractie van de 5000 kippen met nevenverschijnselen. Definieer de stochast  $X$  als het aantal kippen met nevenverschijnselen in de groep van 20 ( $X \sim B(20, p)$ ).

Een steekproef zonder terugleggen is hier toegestaan omwille van de grote populatie.

De veearts zal zolang mogelijk vasthouden aan het rapport (de norm) tot op het ogenblik dat de steekproefresultaten hem doen geloven dat de firma gelijk heeft.

We formuleren de hypothesen als volgt :

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p = 0.2 \end{cases}$$

We kiezen de stochast  $X$  als toetsingsgrootheid en berekenen de  $p$ -waarde :  $P(X \leq 6) = 0.057 = 5.7\%$  (eenzijdige test). Dit is geen significant resultaat op het 5%-niveau en de veearts aanvaardt de nulhypothese.

We kunnen ook eerst het aanvaardings- en verwerpingsgebied opstellen alvorens we een steekproef nemen. We bepalen hiervoor een kritisch punt  $k$  met de volgende beslissingsregel :

$$\begin{cases} \text{aanvaard } H_0 \text{ als de waarde van } X \geq k \text{ en} \\ \text{verwerp } H_0 \text{ als de waarde van } X < k. \end{cases}$$

Op deze manier wordt de verzameling van waarden van  $X$  opgesplitst in twee disjuncte verzamelingen  $V = \{1, \dots, k-1\}$  en  $A = \{k, \dots, 20\}$ .  $V$  noemt men het verwerpingsgebied en  $A$  het aanvaardingsgebied. De kans dat we terecht komen in  $V$  (type I fout) mag hoogstens gelijk zijn aan  $\alpha$  (gelijkheid is meestal niet mogelijk bij een discrete verdeling)

Voor de berekening van deze  $k$  vervangt de **TI-83** op een zeer handige manier de tabel met de kansverdelingen. Definieer de lijsten **L1**, **L2** en **L3** als volgt :

**L1 = binompdf(20, .5)**

**L2 = cumSum(L1)**

**L3 = seq(X, X, 0, 20)**

L1	L2	L3	Z
9.5E-7	9.5E-7	0	
1.9E-5	2E-5	1	
1.8E-4	2E-4	2	
.00109	.00129	3	
.00462	.00591	4	
.01479	.02069	5	
.03696	.05766	6	
L3 = {0, 1, 2, 3, 4, 5...			

Uit de hierboven afgebeelde tabel volgt dat we  $k$  gelijk nemen aan 6. M.a.w. indien de veearts 5 of minder kippen vindt met neveneffecten in zijn steekproef van 20 ingeënte kippen zal hij het rapport van het ministerie verwerpen op een 5% significantieniveau.

Het gevonden aantal, 6, ligt in het aanvaardingsgebied.

We kunnen ook werken met een andere toetsingsgrootheid.

In dit voorbeeld zijn we erin geïnteresseerd of een individu uit de populatie een bepaalde eigenschap (= nevenverschijnsel) bezit.

$$p = \frac{\text{aantal individuen met de eigenschap}}{\text{aantal individuen in de populatie}} \text{ noemt men de populatieproportie.}$$

De steekproefproportie is de stochast :

$$\hat{p} = \frac{X}{n} = \frac{\text{aantal individuen in de steekproef met de eigenschap}}{\text{aantal individuen in de steekproef}}$$

Merk op dat  $\hat{P}$  een onvertekende schatter is van  $p$  want  $E(\hat{P}) = p$ .

We weten dat voor  $n$  groot genoeg de stochast  $Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{X - np}{\sqrt{np(1-p)}}$  bij

*benadering* standaard normaal verdeeld is.

Als vuistregel voor  $n$  groot genoeg nemen we de voorwaarden :

$$\begin{cases} np \geq 5 \\ n(1-p) \geq 5 \end{cases}$$

Het uitvoeren van de hypothesetoets  $H_0 : p = p_0$  versus  $H_1 : p < p_0$  met  $Z$  als toetsingsgrootte kan met de **TI-83** met het commando

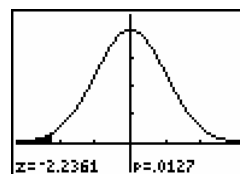
### STAT<TESTS> 5: 1-PropZTest.

Stel dat in ons voorbeeld de veearts 5 kippen met neveneffecten had gevonden in zijn steekproef en vul het **PropZTest**-venster in zoals hiernaast aangegeven.

```
1-PropZTest
P0: .5
x: 5
n: 20
PROP#P0 <P0 >P0
Calculate Draw
```

De resultaten van de **1-PropZTest** vind je hieronder. Uit de **p**-waarde kan je besluiten dat de geobserveerde waarde van de veearts significant is.

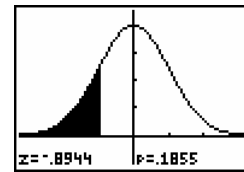
```
1-PropZTest
PROP<.5
z=-2.236067977
P=.0126736174
p̂=.25
n=20
```



Hieronder vind je nog de resultaten indien de veearts in zijn steekproef 8 kippen met neveneffecten had gevonden.

```
1-PropZTest
P0: .5
x: 8
n: 20
PROP#P0 <P0 >P0
Calculate Draw
```

```
1-PropZTest
PROP<.5
z=-.894427191
P=.1855466488
P=.4
n=20
```



In dit voorbeeld was het eerder aangewezen te werken met de discrete binomiaal verdeelde toetsingsgrootheid  $X$ , aangezien  $Z$  slechts bij benadering standaard normaal verdeeld is en er geen continuïteitscorrectie wordt toegepast door de **TI-83**.

## 7.5 Opdrachten

1. In een dobbelspel heeft een speler een vermoeden dat er met valse dobbelstenen gespeeld wordt. Om zijn tegenspelers niet onterecht te beschuldigen, voert hij een controle uit naar de correctheid van de dobbelstenen.

Hij werpt iedere dobbelsteen tien keer en telt het aantal keren dat hij zes gooit.

- a. Stel een toets op die test of iedere dobbelsteen correct is of dat er vaker zes ogen verschijnen dan met een eerlijke dobbelsteen.
- b. In de veronderstelling dat de dobbelsteen eerlijk is (de nulhypothese), wat is dan de kansverdeling van  $X$  = het aantal keren zes bij tien worpen.
- c. Daar de speler het vertrouwen van zijn tegenstanders niet wil verliezen, wil hij de kans om zijn tegenspelers onterecht te beschuldigen van vals spelen minimaliseren.

Bereken de kans op een foute beschuldiging (type I fout) indien hij de nulhypothese (eerlijke dobbelsteen) verworpt bij het bekomen van minstens

- (i) 3 keer zes op 10 worpen,
- (ii) 4 keer zes op 10 worpen,
- (iii) 5 keer zes op 10 worpen,
- (iv) 6 keer zes op 10 worpen.

- d. Welke beslissingsregel kan de speler toepassen opdat de kans op een onterechte beschuldiging ten hoogste 5 % bedraagt.

Verwerp  $H_0$  als  $X > k$  en aanvaard  $H_0$  als  $X \leq k$ . Bepaal  $k$ .

- e. Na afloop twijfelt de speler nog steeds en breidt hij zijn controle uit tot 100 worpen. Bepaal opnieuw een beslissingsregel op een 5 % significantieniveau.

- f. Stel  $X \sim B(n, p)$ . Indien  $n$  groot genoeg, geldt dat  $\frac{X - np}{\sqrt{np(1-p)}}$  standaard normaal verdeeld is. Bepaal voor  $n = 100$  het kritische punt met de normale verdeling.

2. Een onderwijzer beweert dat de gemiddelde lengte van een tienjarige 140 cm bedraagt. De directeur, voor wie statistiek een hobby is, trekt uit zijn totale populatie van leerlingen een steekproef van grootte 30 en wil hiermee zijn onderstelling dat tienjarige kinderen gemiddeld kleiner zijn dan 140 cm, toetsen.

De gegevens (in cm) die de directeur verzameld heeft, zijn :

120	157	141	132	145	142	128	121	130	128
128	142	121	133	142	144	146	128	127	124
139	138	128	137	136	133	141	136	133	136

Toets, met deze steekproef van grootte 30, of de uitspraak van de onderwijzer te verwerpen is op een 5% significantieniveau indien  $\sigma = 7.18$  en indien  $\sigma$  ongekend is.

3. Veronderstel dat de lengte van de kinderen in België bij hun geboorte normaal verdeeld is met gemiddelde lengte 52 cm.

Beschouw de onderstaande tabel als een steekproef van grootte 30 en ga na of de resultaten al dan niet significant verschillend zijn van de gemiddelde lengte bij de geboorte van de populatie.

53	53	51	53	51	54	48	43	48	53
53	51	52	53	53	53	56	53	53	53
52	51	53	53	55	53	53	52	53	55

4. Algemeen gekend is dat geneesheren aspirines voorschrijven aan hartlijders ter preventie van de vorming van verstoppingen in de aders ten gevolge van bloedklonters. De volgende studie werd uitgevoerd om na te gaan of het gebruik van aspirines een invloed heeft op klontervorming in het bloed.

Voor twaalf volwassen mannen observeerde men de tijd tussen de start van een protrombase-trombase reactie en de vorming van een bloedklonter. De stollingstijd werd gemeten vóór de mannen aspirines innamen en drie uur na het innemen van twee aspirines. De data vind je hieronder (rij 2 = voor inname en rij 3 = na inname) :

1	2	3	4	5	6	7	8	9	10	11	12
11.3	11	11	11.5	11.8	11.3	12.5	13	13	12	12	12.3
11.5	11	11.5	11.5	11.3	10.3	12.5	13	12	12.5	12.3	12

Toets a.h.v. deze steekproef de efficiëntie van het toedienen van aspirines op een 5% significantieniveau.

### Methode

Noem  $\mu_1$  het steekproefgemiddelde voor de inname en  $\mu_2$  het steekproefgemiddelde na de inname. De toets die we hier wensen uit te voeren is :

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2 \text{ of } H_0 : \mu_2 - \mu_1 = 0 \text{ versus } H_1 : \mu_2 - \mu_1 < 0.$$

Plaats hiervoor bv. de data voor de inname in lijst **L1**, de data na de inname in lijst **L2** en het verschil **L2-L1** in lijst **L3**.

5. De firma Texas Instruments meldde dat het commando **rand** van de **TI-83** toevalsgetallen genereert uit een uniforme verdeling op het interval  $[0,1)$  met gemiddelde 0.5 en standaardafwijking  $\sqrt{1/12} = 0.2887$ .

- a. Genereer zo'n 100 getallen uit het interval  $[0,1)$  en bereken het gemiddelde van deze 100 getallen.

Komt het resultaat overeen met de informatie van Texas Instruments ?

Herhaal dit experiment enkele keren.

- b. Het genereren van 100 zo'n getallen kan je beschouwen als een steekproef van 100 onafhankelijke stochasten  $X_1, X_2, \dots, X_{100}$  met dezelfde verdeling.

Wat is de verdeling van  $\bar{X} = \frac{X_1 + X_2 + \dots + X_{100}}{100}$  ?

- c. Stel een toets op met de toetsingsgrootheid  $\bar{X}$  om na te gaan of de informatie over het **rand**-commando correct is.
- d. Bereken rechtstreeks (met **normalcdf**) de **p**-waarde voor een steekproef van 100 getallen die je doen twijfelen over de informatie van Texas Instruments.

Wat besluit je op een 5% significantieniveau ?

- e. Bepaal de **p**-waarde voor een steekproef van 100 getallen zowel rechtstreeks als met de **TI-83** en de **Z-Test**.

Vergelijk de resultaten van deze twee manieren om de **p**-waarde te berekenen. Kan je dit verklaren ?

6. Een student moet een test afleggen die bestaat uit 10 meerkeuzevragen. Elke vraag bestaat uit 3 alternatieven.

De score op iedere vraag is 0 (fout) of 1 (juist).

Een student haalt een score van 8 / 10 en beweert gegokt te hebben.

Geloof je hem ?

- a. Stel een toets op om na te gaan of het kan dat de student gegokt heeft.

- b. Bepaal rechtstreeks de **p**-waarde (met **binompdf**) . Wat besluit je op een 5% significantieniveau ?
- c. Simuleer 100 keer het lukraak invullen van de test. Bepaal met deze simulatie een schatting voor de **p**-waarde.
- d. Bepaal de **p**-waarde voor dit toetsingsprobleem met de **TI-83** en het commando **1-PropZTest**.

Is dit resultaat correct ? Verklaar.