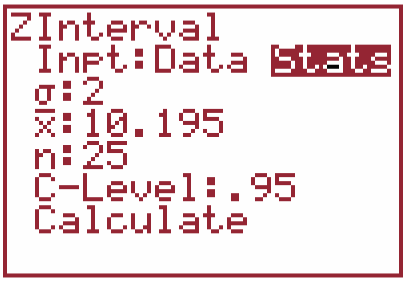
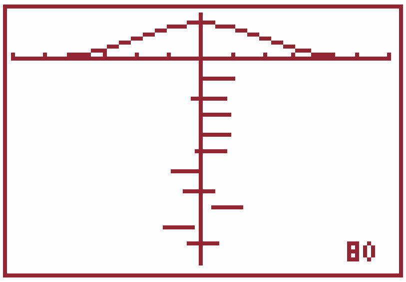
Cahiers T³ Europe
Vlaanderen nr. 8

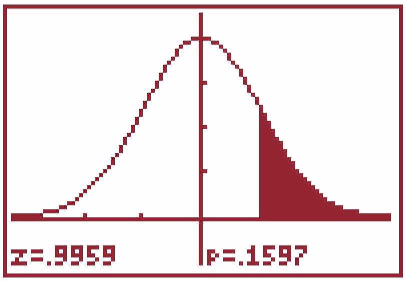
Betrouwbaarheidsintervallen en het testen van hypothesen

Van steekproef naar populatie

Guido Herweyers







KATHOLIEKE UNIVERSITEIT LEUVEN

Betrouwbaarheidsintervallen en het testen van hypothesen

Van steekproef naar populatie

Guido Herweyers

Inhoud

Betrouwbaarheidsintervallen

1 Inleiding

Statistische inferentie of Statistische besluitvorming trekt aan de hand van steekproefdata conclusies over de populatie waarvan de data afkomstig zijn.

Deze conclusies zijn enkel betrouwbaar indien de steekproefdata op een correcte wijze werden verkregen, zodat we kunnen aannemen dat de populatie goed wordt vertegenwoordigd door de steekproefdata. Eerst moet er dus veel aandacht gaan naar het *verwerven van gegevens*; uit slordig verkregen data kan men geen zinvol besluit formuleren. In wat volgt nemen we aan dat we werken met een *enkelvoudige aselecte steekproef* uit een populatie, waarbij elke steekproef van dezelfde omvang dezelfde kans op selectie heeft.

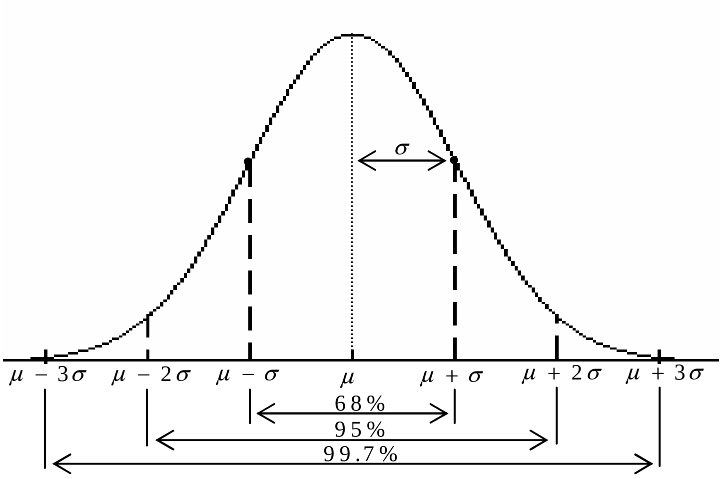
In de onderstaande tekst bespreken we twee belangrijke begrippen uit de statistische besluitvorming: betrouwbaarheidsintervallen en hypothesetesten. We beperken ons hierbij tot toepassingen op de normale en de binomiale verdeling, verdelingen die worden bestudeerd in het secundair onderwijs.

Het is de bedoeling om vertrouwd te geraken met de begrippen en de toepassingsgebieden aan de hand van concrete voorbeelden, hierbij worden het grafisch rekentoestel TI-84 Plus en Java-applets van de K.U.Leuven ingezet.

2 De normale verdeling

In deze paragraaf vatten we eerst enkele eigenschappen van de normale verdeling en van toevalsvariabelen samen. Voor een overzicht van de normale verdeling met de TI-84 Plus verwijzen we naar paragraaf 5.

1) Over de grafiek van de normale dichtheidsfunctie:



- Populatiedata worden o.a. gekenmerkt door het *populatiegemiddelde* μ en de *populatiestandaardafwijking* σ, als maat voor de spreiding van de data t.o.v. μ. Vaak neemt het dichtheidshistogram van de populatiedata, getekend met een kleine klassenbreedte, de klokvorm aan van de normale dichtheidsfunctie (zie voorgaande grafiek). We zeggen dan dat de populatiedata normaal verdeeld zijn. De normale dichtheidsfunctie is dus een *wiskundig model* voor de verdeling van dergelijke populatiedata.
  Relatieve frequenties of *proporties* van *populatiedata* van een variabele *X*, gelegen in een interval, worden genoteerd met $P(a < X < b)$, $P(X < a)$, $P(X > a),\ldots$ en berekend als oppervlakte onder de dichtheidskromme boven het beschouwde interval. De aard van de beschouwde ongelijkheden (al dan niet strikt) speelt hierbij geen rol.

- De oppervlakte van het gebied begrensd door de dichtheidskromme en de *x*-as is gelijk aan 1.

- De kromme is klokvormig en symmetrisch t.o.v. de rechte $x = \mu$, het gemiddelde valt dus samen met de mediaan: $\mu = \text{Med}$.

- Ook de standaardafwijking σ heeft een meetkundige betekenis: de punten op de kromme gelegen op een afstand σ van de symmetrieas zijn de *buigpunten* van de kromme. Dit zijn de punten waar de vorm van de kromme overgaat van bol naar hol (opwaarts kijkend naar de kromme) of omgekeerd. Een buigpunt wordt ook gekenmerkt als een punt waar de raaklijn het steilst is. De hoogte van het buigpunt is ongeveer 60% van de tophoogte.

- De 68 - 95 - 99.7 regel:
  voor een normale verdeling met gemiddelde μ en standaardafwijking σ geldt:
  - Ongeveer 68% van de waarnemingen ligt binnen een afstand σ van μ
  - Ongeveer 95% van de waarnemingen ligt binnen een afstand 2σ van μ
  - Ongeveer 99.7% van de waarnemingen ligt binnen een afstand 3σ van μ

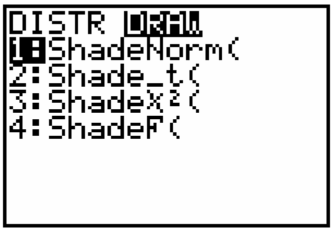
  Deze regel illustreert de rol van σ als maat van de spreiding van de populatiedata omheen het populatiegemiddelde μ.

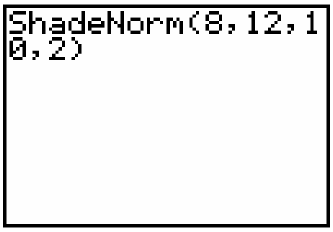
We verifiëren de 68 - 95 - 99.7 regel voor een normale verdeling met gemiddelde $\mu = 10$ en standaardafwijking $\sigma = 2$. Na keuze van een geschikt venster kan de oppervlakte ook worden gearceerd via 2nd[DISTR] <DRAW> 1:ShadeNorm(.

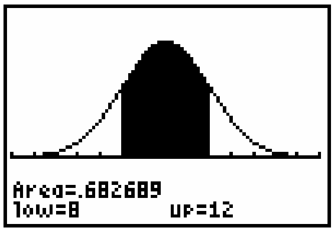
```
normalcdf(8,12,1
0,2)
       .6826894809
normalcdf(6,14,1
0,2)
        .954499876
```

```
normalcdf(4,16,1
0,2)
       .9973000656
```

```
WINDOW
 Xmin=3
 Xmax=17
 Xscl=1
 Ymin=⁻.1
 Ymax=.25
 Yscl=1
 Xres=1
```







2) Elke *toevalsvariabele X* (ook *stochastische veranderlijke* of *stochast* genoemd) wordt gekenmerkt door de volgende getallen:

- $E(X)$, ook genoteerd met μ , d.i. de *verwachtingswaarde* of het *gemiddelde* van X .
- $Var(X) = E\left((X-\mu)^2\right)$, ook genoteerd met σ^2 , d.i. de *variantie* van X .
- i.p.v. de variantie vermeldt men vaak $\sigma = \sqrt{Var(X)}$, d.i. de *standaardafwijking* van *X*.

De notatie μ voor $E(X)$ en σ^2 voor $Var(X)$ is verantwoord: als *X* een lukrake trekking is uit een populatie, dan is $E(X)$ het *populatiegemiddelde* μ en $Var(X)$ de *populatievariantie* σ^2 . We spreken dan kort over de *populatie met variabele X*.

Voor een normaal verdeelde populatie geeft de dichtheidskromme nu ook de *kansverdeling* aan van de toevalsvariabele *X*, met deze verdeling wordt het toevalsmechanisme aangestuurd. In deze context is $P(a < X < b)$ nu de *kans* dat de variabele X , lukraak gekozen uit de populatie, een waarde aanneemt tussen *a* en *b*. Na een lukrake keuze uit de populatie verkrijgen we een concreet getal *x*, d.i. een waarde die de variabele *X* door het toevalsmechanisme heeft aangenomen.

Indien we vaak een lukrake trekking doen uit een normaal verdeelde populatie (en het verkregen getal telkens terugplaatsen in de populatie), dan zal het dichtheidshistogram van deze lukraak verkregen data *op de lange duur* de vorm van de normale dichtheidsfunctie aannemen (als we een kleine klassenbreedte kiezen). Via *simulatie* kan men op die manier de kansverdeling laten ontstaan van een toevalsvariabele.

Je kan het ook als volgt bekijken: met simulatie van trekkingen uit een populatie A produceer je *op de lange duur* een nieuwe populatie B met dezelfde verdeling (en dus dezelfde μ en σ^2) als populatie A.

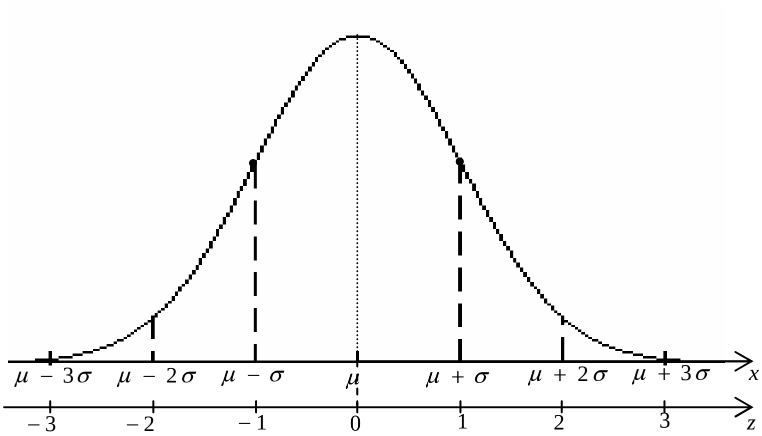
3) Als de toevalsvariabele *X* normaal verdeeld is met gemiddelde μ en standaardafwijking σ , dan noteren we dit als volgt: $X \sim N(\mu,\sigma)$.

4) Als $X \sim N(\mu,\sigma)$ dan is $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$ (*Z* is *standaard normaal* verdeeld).

Door deze transformatie wordt een concrete waarde x van *X* getransformeerd in de waarde $z = \dfrac{x-\mu}{\sigma}$ van Z . We noemen het getal *z* de standaardscore of *z-score* van *x*.

Uit de formule $x = \mu + z \cdot \sigma$ leren we dat de *z*-score aangeeft hoeveel standaardafwijkingen σ de waarde *x* verwijderd is van het gemiddelde μ.

Een *z*-score tussen -1 en 1 is "normaal" en een *z*-score met een absolute waarde groter dan 2 is zelden (zie ook volgende figuur). De *z*-scores worden o.a. gebruikt om data van verschillende normale verdelingen te vergelijken.



5) Gegeven een populatie met variabele X, met gemiddelde μ en standaardafwijking σ.

Beschouw een steekproef $X_1, X_2, X_3, \ldots X_n$ uit de populatie (dit zijn *onafhankelijke* toevalsvariabelen met dezelfde verdeling als X), dan heeft het steekproefgemiddelde $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ ook gemiddelde μ maar standaardafwijking $\dfrac{\sigma}{\sqrt{n}}$.

a) Als *X* normaal verdeeld is, dan is ook $\overline{X}$ normaal verdeeld:

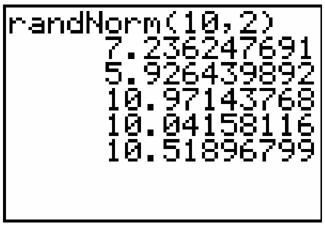
Als $X \sim N(\mu, \sigma)$ dan is $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

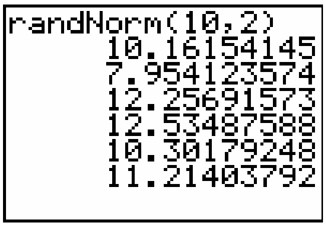
b) Als de populatie (met eindige variantie) niet normaal verdeeld is dan geldt toch bij benadering dat $\overline{X}$ normaal verdeeld is met gemiddelde μ en standaardafwijking $\dfrac{\sigma}{\sqrt{n}}$ *als n voldoende groot is* (centrale limietstelling).

3 Betrouwbaarheidsinterval voor het gemiddelde van een normale verdeling met gekende standaardafwijking

Met het TI-84 Plus commando randNorm(μ,σ) kun je simulaties uitvoeren van lukrake trekkingen uit een normaal verdeelde populatie met gemiddelde μ en standaardafwijking σ.

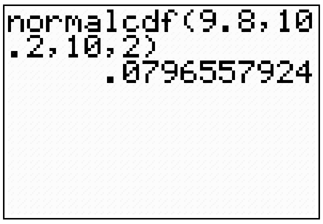
Je vindt het commando onder MATH <PRB> 6: randNorm(



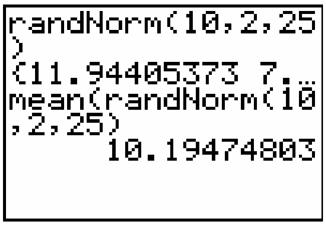


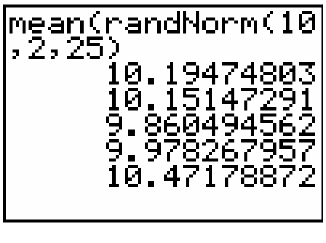
Enkele simulaties tonen de variabiliteit van de verkregen data bij trekkingen uit een normale verdeling met gemiddelde $\mu = 10$ en standaardafwijking $\sigma = 2$. De data schommelen rond het gemiddelde.

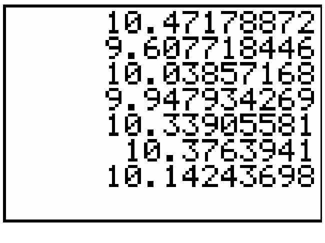
Als $X \sim N(10,2)$, dan is $P(9.8 < X < 10.2) = 8.0\%$:



We simuleren nu enkele steekproeven met grootte $n = 25$ uit de populatie, telkens berekenen we hierbij het steekproefgemiddelde $\overline{x}$, d.i. telkens een concreet verkregen waarde van de toevalsvariabele $\overline{X}$.



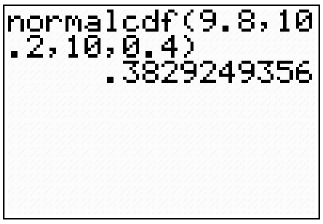




We stellen vast dat er *minder variabiliteit* is bij de steekproefgemiddelden, die tevens schommelen rond het populatiegemiddelde $\mu = 10$. Elk van die steekproefgemiddelden is een (*punt*)*schatting* voor het populatiegemiddelde.

Als $X \sim N(10,2)$ dan is $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_{25}}{25} \sim N\left(10, \dfrac{2}{\sqrt{25}}\right)$ of $\overline{X} \sim N(10,\ 0.4)$.

Nu is $P\left(9.8 < \overline{X} < 10.2\right) = 38.3\%$:



Wat is eigenlijk de betekenis van $\overline{X} \sim N\left(10,\, 0.4\right)$?

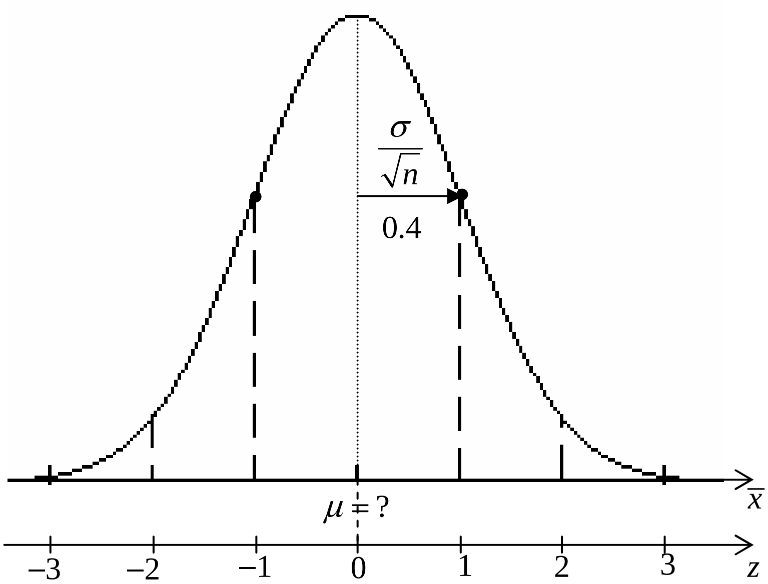
Beschouw alle mogelijke geordende steekproeven (met terugleggen) $\left(x_1, x_2, x_{3,} \ldots, x_{25}\right)$ die je kunt vormen uit de populatiedata en bepaal telkens het steekproefgemiddelde $\overline{x}$. De populatie van al deze steekproefgemiddelden is normaal verdeeld met gemiddelde 10 en standaardafwijking 0.4. Het gemiddelde van al die steekproefgemiddelden is dus gelijk aan het gemiddelde van de oorspronkelijke populatie, maar de spreiding is verkleind met een factor $\sqrt{25}$.

De toevalsvariabele $\overline{X}$ wordt gestuurd door een $N\left(10,\, 0.4\right)$-verdeling. Bij simulatie van lukrake steekproeftrekkingen zal de vorm van het dichtheidshistogram van de lukraak verkregen steekproefgemiddelden op de lange duur de vorm van $N\left(10,\, 0.4\right)$-dichtheidsfunctie aannemen.

Meestal trekken we echter maar één steekproef uit een populatie.

Beschouw het resultaat 10.195 (afgerond) van het eerste steekproefgemiddelde en stel dat je enkel weet dat de steekproef afkomstig is uit een normaal verdeelde populatie met standaardafwijking $\sigma = 2$, *hoe goed is deze schatting* $\overline{x} = 10.195$ voor het ongekende gemiddelde μ van de populatie?

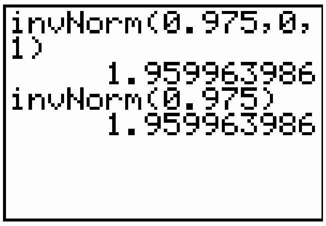
Hiertoe redeneren we grafisch:

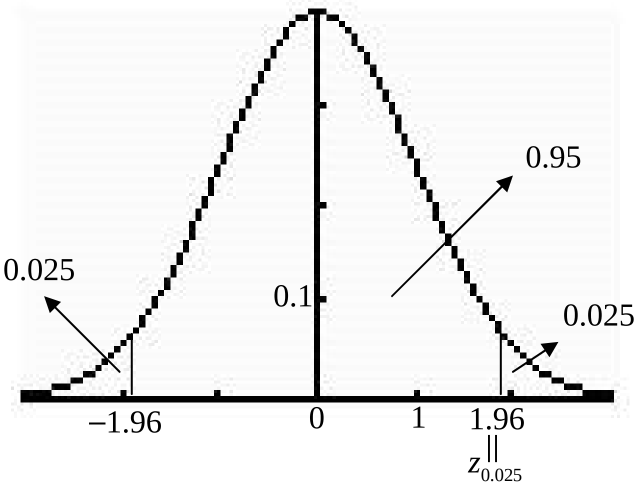


In de voorgaande figuur zien we de verdeling van de steekproefgemiddelden van alle mogelijke steekproeven met grootte *n*, het gemiddelde van al die steekproefgemiddelden valt samen met het populatiegemiddelde μ en de standaardafwijking is $\frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{25}} = 0.4$. Op basis van één lukraak verkregen steekproefgemiddelde $\overline{x} = 10.195$ wensen we een uitspraak te doen over het ongekende populatiegemiddelde μ .

We weten dat *ongeveer* 95% van de steekproefgemiddelden gelegen zijn binnen een afstand van 2 standaardafwijkingen $\frac{\sigma}{\sqrt{n}}$ van μ . Meer precies kunnen we zeggen dat 95% van de steekproefgemiddelden gelegen zijn in het interval $\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$.

Met de functie invNorm vinden we immers de *standaardscore* $z_{0.025} = 1.96$; het argument van deze functie is de oppervlakte 0.975 *links* van $z_{0.025}$ onder de standaard normale verdeling:





De notatie $z_{0.025}$ wordt gebruikt om aan te geven dat de oppervlakte *rechts* van $z_{0.025}$ gelijk is aan 0.025 .

95% van de data $\overline{x}$ liggen dus binnen een afstand $1.96\frac{\sigma}{\sqrt{n}}$ van μ, maar dit betekent ook dat voor 95% van de steekproefgemiddelden $\overline{x}$ het populatiegemiddelde μ zal gelegen zijn binnen een afstand $1.96\frac{\sigma}{\sqrt{n}}$ van $\overline{x}$!

Derhalve noemen we voor een concrete observatie $\overline{x}$ het interval $\left[\overline{x}-1.96\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}+1.96\frac{\sigma}{\sqrt{n}}\right]$ een 95% *betrouwbaarheidsinterval* voor μ, of een betrouwbaarheidsinterval met *betrouwbaarheidsniveau* 95%.

We noteren dit interval ook met $\overline{x}\pm 1.96\frac{\sigma}{\sqrt{n}}$.

Formeel loopt de bovenstaande redenering als volgt:

Aangezien $\overline{X}=\frac{X_1+X_2+\ldots+X_n}{n}\sim N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$ is $Z=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$

Uit $P(-1.96\le Z\le 1.96)=0.95$

volgt dat $P(-1.96\le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le 1.96)=0.95$

of $P(-1.96\frac{\sigma}{\sqrt{n}}\le \overline{X}-\mu\le 1.96\frac{\sigma}{\sqrt{n}})=0.95 \qquad (1)$

uit (1) volgt nu $\boxed{P(\mu-1.96\frac{\sigma}{\sqrt{n}}\le \overline{X}\le \mu+1.96\frac{\sigma}{\sqrt{n}})=0.95}$

uit (1) volgt echter ook $P(-\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\le -\mu\le -\overline{X}+1.96\frac{\sigma}{\sqrt{n}})=0.95$

waaruit $P(\overline{X}+1.96\frac{\sigma}{\sqrt{n}}\ge \mu\ge \overline{X}-1.96\frac{\sigma}{\sqrt{n}})=0.95$

of $\boxed{P(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\le \mu\le \overline{X}+1.96\frac{\sigma}{\sqrt{n}})=0.95}$

bijgevolg noemen we $\left[\overline{x}-1.96\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}+1.96\frac{\sigma}{\sqrt{n}}\right]$ een 95% betrouwbaarheidsinterval voor μ.

Voor onze eerste observatie $\bar{x} = 10.195$ verkrijgen $\mu = 10.195 \pm 1.96 \cdot 0.4 = 10.195 \pm 0.784$ als betrouwbaarheidsinterval met 95% betrouwbaarheid. Dit interval $[9.411, 10.979]$ bevat inderdaad het populatiegemiddelde $\mu = 10$.

Bij elke steekproef verkrijgen we meestal een ander gemiddelde $\bar{x}$ en dus een ander 95% betrouwbaarheidsinterval $\bar{x} \pm 0.784$.

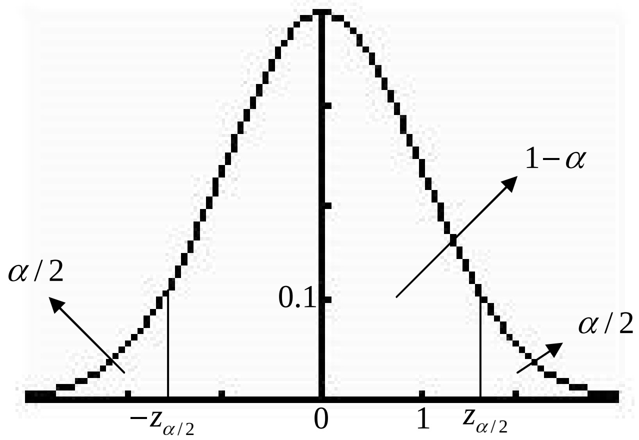
Bij honderd steekproeven verkrijgen we honderd betrouwbaarheidsintervallen, waarvan we mogen verwachten dat er *ongeveer* 95 intervallen het populatiegemiddelde μ zullen bevatten.

Simuleer zelf enkele steekproeven en bepaal de bijbehorende betrouwbaarheidsintervallen voor μ, ga na welke intervallen inderdaad $\mu = 10$ bevatten.

Merk op dat de *foutmarge* $1.96\dfrac{\sigma}{\sqrt{n}} = 0.784$ voor elk interval ongewijzigd blijft (bij vaste σ en *n* en 95% betrouwbaarheid), terwijl het centrum $\bar{x}$ van het interval door het toeval varieert van steekproef tot steekproef.

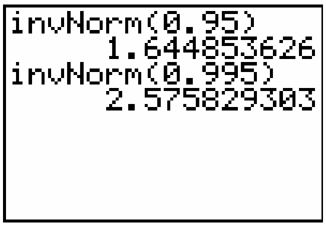
Men kan *het betrouwbaarheidsniveau* $95\% = 0.95$ wijzigen tot een bepaalde waarde $1-\alpha$ (met $0 < \alpha < 1$). Het betrouwbaarheidsniveau wordt genoteerd met $1-\alpha$ omdat α kan worden geïnterpreteerd als het significantieniveau van een tweezijdige hypothesetest (zie volgend hoofdstuk).

Een betrouwbaarheidsinterval met betrouwbaarheidsniveau $1-\alpha$, voor het gemiddelde μ van een normaal verdeelde populatie met gekende standaardafwijking σ, is $\bar{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$, waarbij de oppervlakte *rechts* van $z_{\alpha/2}$ onder de standaardnormale dichtheidskromme gelijk is aan $\alpha/2$, zodat de oppervlakte tussen $-z_{\alpha/2}$ en $z_{\alpha/2}$ gelijk is aan het gewenste betrouwbaarheidsniveau $1-\alpha$.
Met de TI-84 Plus vind je $z_{\alpha/2} = \text{invNorm}\left(1-\dfrac{\alpha}{2}\right)$.



Merk op dat de foutmarge $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ afhankelijk is van het gekozen betrouwbaarheidsniveau $1-\alpha$, de populatiestandaardafwijking σ en de steekproefgrootte *n*.

- Hoe groter het betrouwbaarheidsniveau, hoe groter de foutmarge.
  Bij een betrouwbaarheid van 90% is $z_{0.05} = 1.645$, voor een betrouwbaarheid van 99% is $z_{0.005} = 2.576$.



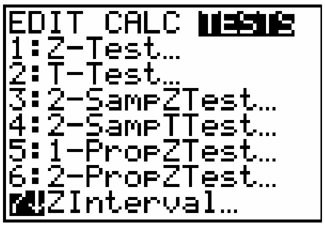
  Halveert de foutmarge als we het betrouwbaarheidsniveau halveren (bvb. van 90% naar 45%)?

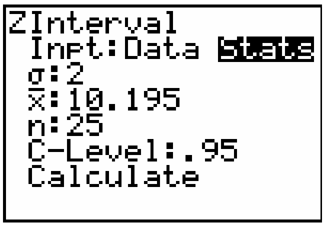
- Hoe groter de populatiestandaardafwijking σ, hoe groter de foutmarge.
  Als σ verdubbelt, verdubbelt dan ook de foutmarge?

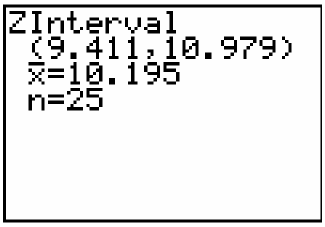
- Hoe groter de steekproefgrootte *n*, hoe kleiner de foutmarge.
  Als *n* verdubbelt, halveert dan de foutmarge?

Betrouwbaarheidsintervallen kunnen ook rechtstreeks worden berekend met de TI-84 Plus via

STAT <TESTS> 7 : ZInterval : vul de gegevens in, ga met de cursor naar Calculate en druk ENTER.







Simulatie van betrouwbaarheidsintervallen

Het programma CI simuleert tien steekproeven van eenzelfde grootte uit een populatie die normaal verdeeld is met $\mu = 10$ en $\sigma = 2$.

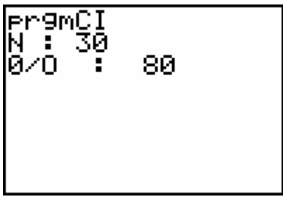
Voor iedere steekproef wordt een betrouwbaarheidsinterval berekend en deze intervallen worden grafisch voorgesteld samen met de dichtheidsfunctie van de normale verdeling.

Na het starten van het programma geef je de gegevens in :

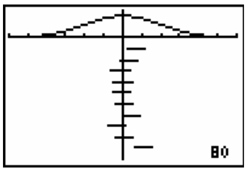
N = steekproefgrootte,

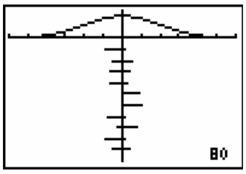
O/O = betrouwbaarheidsniveau

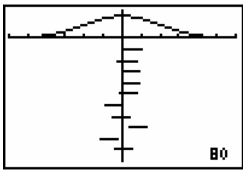
Dit doe je door de waarde in te tikken achter "**:**" en dan op ENTER te drukken.



Hier volgen enkele resultaten en het programma:

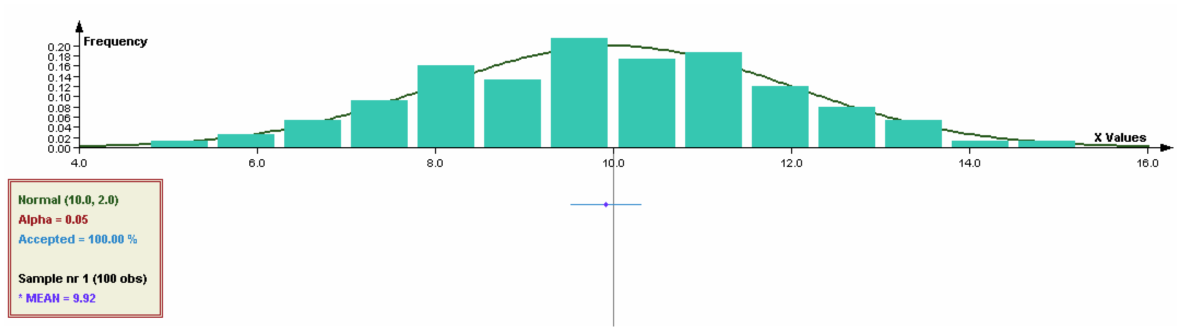






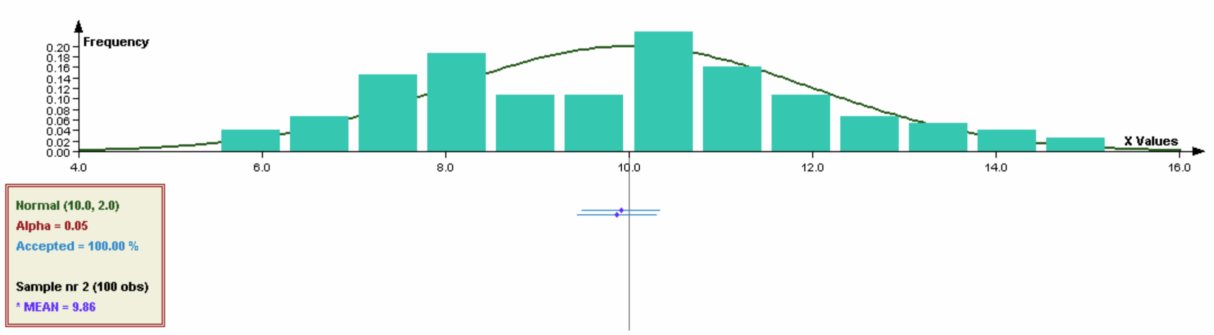

```
** Input % **
Input "N    :  ",N
Input "O/O  :  ",P
P/1ØØ→P
** Trekken van tien steekproeven **
For(I,1,1Ø)
   randNorm(1Ø,2,N)→L1
   ZInterval 2,L1,P
   lower→ L2(I)
   upper→L3(I)
End
** Definitie van het grafisch venster **
FnOff
PlotsOff
4→Xmin
16→Xmax
1→Xscl
⁻1.1→Ymin
.25→Ymax
Ø→Yscl
1→Xres
**Tekenen van dichtheidsfunctie en intervallen**
"normalpdf(X,1Ø,2)"→Y1
P*1ØØ→P
Text(56,84,P)
Vertical 1Ø
   For(J,1,1Ø)
   Line(L2(J),J*⁻.1,L3(J),J*⁻.1)
End
```

Met de Java-applets van de webpagina www.kuleuven.be/ucs/java kun je ook betrouwbaarheidsintervallen simuleren via “Tests”, vervolgens “Confidence interval for mean”.
Bij “settings” stellen we $\mu = 10$, $\sigma = 2$ en $\alpha = 5\%$ voor een 95% betrouwbaarheidsinterval. Als steekproefgrootte kiezen we 100 en voor het aantal steekproeven nemen we 1000. Nu levert “step” alvast een eerste simulatie:



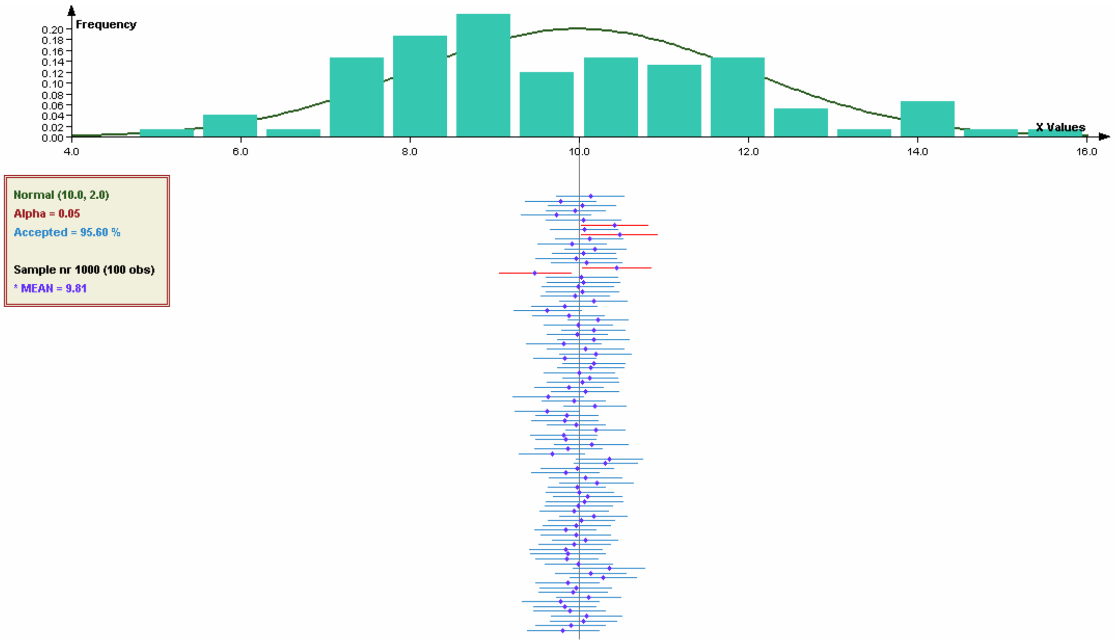
We zien de verdeling van de steekproefgegevens en het bijbehorende betrouwbaarheidsinterval, dat het populatiegemiddelde $\mu = 10$ bevat. Het kader links kun je met de muis verplaatsen.

Nog eens klikken op “step” levert een tweede steekproef met zijn betrouwbaarheidsinterval:



Ook het tweede betrouwbaarheidsinterval bevat μ.

Met “walk” zie je de volgende steekproefresultaten, met “run” gebeurt dit in versneld tempo. Uiteindelijk verkrijgen we een zicht op de laatste 100 betrouwbaarheidsintervallen en een samenvattend overzicht in het kader:



De betrouwbaarheidsintervallen die μ niet bevatten worden op de website aangegeven in het rood. In het kadertje lezen we af ("Accepted") dat 95,60 % van de duizend betrouwbaarheidsintervallen μ bevatten, dit is in overeenstemming met het gekozen betrouwbaarheidsniveau 95 %.

Opgaven

1) Een test op het kaliumgehalte in bloed is niet helemaal nauwkeurig. Bovendien varieert de feitelijke hoeveelheid kalium in iemands bloed van dag tot dag enigszins. Veronderstel dat herhaalde metingen op verschillende dagen bij dezelfde persoon variëren volgens een normale verdeling met standaardafwijking $\sigma = 0.2$ (in millimol per liter).

   (a) Het kaliumgehalte van Mark wordt één keer gemeten, het resultaat is $x = 3.4$. Bepaal een 95% betrouwbaarheidsinterval voor het gemiddelde kaliumgehalte van de normale verdeling.

   (b) Als er op vier verschillende dagen metingen werden gedaan met gemiddelde $\overline{x} = 3.4$, wat is dan het 95% betrouwbaarheidsinterval voor het gemiddelde kaliumniveau van de normale verdeling?

2) Een labo analyseert de concentratie van een actieve ingrediënt in een geneesmiddel. Drie verschillende metingen van de concentratie levert als resultaat in g/l :

   0.8403 0.8363 0.8447

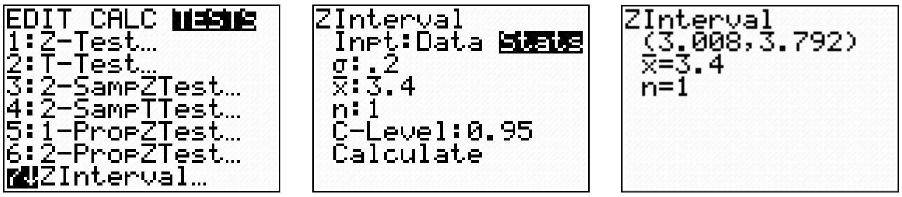
   Veronderstel hierbij dat de data afkomstig zijn uit een normale verdeling met gekende standaardafwijking $\sigma = 0.0068$ g/l en dat het gemiddelde van die verdeling de exacte concentratie is (het labo maakt geen systematische fouten).

   (a) Bepaal een 90 % betrouwbaarheidsinterval voor de exacte concentratie.

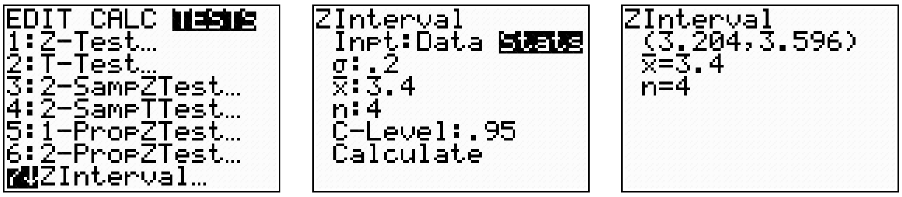
   (b) Hoeveel metingen moet men doen om de exacte concentratie met 95% zekerheid te kennen met een maximale foutmarge van 0.005 ?

Oplossingen

1) a) Met $1-\alpha = 0.95$ vinden we $\alpha = 0.05$ en $z_{\alpha/2} = z_{0.025} =$ invNorm(0.975) = 1.96
Bovendien is $\sigma = 0.2$ en $n = 1$ zodat $\overline{x} = x$.
Het 95% betrouwbaarheidsinterval $\overline{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$ wordt dan $x \pm z_{\alpha/2} \cdot \sigma$ of $3.4 \pm 1.96 \cdot 0.2$ of 3.4 ± 0.392 , dit is het interval $[3.008, 3.792]$.

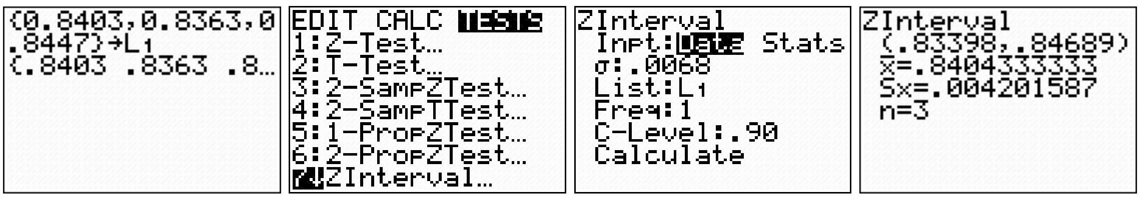


b) Het 95% betrouwbaarheidsinterval $\overline{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$ wordt $3.4 \pm 1.96 \cdot \dfrac{0.2}{\sqrt{4}}$ of 3.4 ± 0.196, dit is het interval $[3.204, 3.596]$.



2) a) We berekenen eerst het steekproefgemiddelde $\overline{x} = 0.8404$,
Met $1-\alpha = 0.90$ vinden we $\alpha = 0.10$ en $z_{\alpha/2} = z_{0.05} =$ invNorm(0.95) = 1.64
Het 90% betrouwbaarheidsinterval $\overline{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$ wordt $0.8404 \pm 1.64 \cdot \dfrac{0.0068}{\sqrt{3}}$ of 0.8404 ± 0.0064 of het interval $[0.8340, 0.8468]$

Met de TI-84 Plus kun je eerst de data ingeven in lijst L1 en het rekentoestel zelf het gemiddelde laten berekenen:



2) b) Er moet gelden dat $z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}} \le 0.005$ of $1.96 \cdot \dfrac{0.0068}{\sqrt{n}} \le 0.005$ of

$\left(1.96 \cdot \dfrac{0.0068}{0.005}\right)^2 \le n$ zodat $n \ge 7.1$; de steekproefgrootte $n = 8$ volstaat dus.

Opmerking

Het zwakke punt van onze redenering is dat we veronderstellen dat de standaardafwijking σ van de populatie gekend is, dit is echter zelden het geval waardoor het interval $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ in de praktijk zelden bruikbaar is. We veronderstellen tevens dat de populatie normaal verdeeld is.

Indien we echter een *voldoend grote steekproef* (vuistregel $n \geq 30$) nemen uit een populatie met een willekeurige verdeling en eindige standaardafwijking σ, dan zegt de *centrale limietstelling* dat het steekproefgemiddelde $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ bij benadering *normaal verdeeld* is met gemiddelde μ en standaardafwijking $\frac{\sigma}{\sqrt{n}}$.

Bovendien zal de *steekproefstandaardafwijking* s dan dicht liggen bij de ongekende σ, zodat $\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ dan een *benaderend* betrouwbaarheidsinterval geeft voor μ (indien er geen uitschieters zijn in de steekproef want deze beïnvloeden $\bar{x}$ en s sterk).

Zo wordt de foutmarge van een steekproef (uit een willekeurige verdeling) van 100 data met een steekproefstandaardafwijking $s = 3$ bij een 95% betrouwbaarheidsinterval bij benadering gegeven door $z_{0.025} \cdot \frac{s}{\sqrt{n}} = 1.96 \cdot \frac{3}{\sqrt{100}} = 0.588$.

4 Betrouwbaarheidsinterval voor een proportie

Om de proportie *p* van de Vlamingen met bloedgroep O te bepalen, worden 200 personen onderzocht; 80 onder hen hebben bloedgroep O.

Als *schatting* voor *p* nemen we de *steekproefproportie* $\hat{p} = \frac{80}{200} = 0.40$.

We willen een 95% betrouwbaarheidsinterval voor p bepalen, d.w.z. een interval $\hat{p} \pm \text{foutmarge}$ waarbinnen *p* met 95% zekerheid gelegen is.

Stel X het aantal personen met bloedgroep O bij een steekproef van 200 personen, dan is X binomiaal verdeeld met parameters $n = 200$ en p = kans op succes (bloedgroep O) :

$$X \sim B(n, p) \text{ met } E(X) = \mu_X = np \text{ en } Var(X) = {\sigma_X}^2 = np(1-p)$$

Door de *centrale limietstelling* weten we dat X bij benadering normaal verdeeld is met gemiddelde np en standaardafwijking $\sqrt{np(1-p)}$.

We kunnen X immers schrijven als $X = X_1 + X_2 + X_3 + \ldots X_n$, d.i. een som van n onafhankelijke toevalsvariabelen met een Bernouilli verdeling met parameter p. Hierbij neemt X_i de waarde 1 aan met kans p als de i-de persoon bloedgroep O heeft en de waarde 0 als dit niet het geval is. Men rekent vlug na dat $E(X_i) = p$ en $Var(X_i) = p(1-p)$,

zodat $$E(X) = E(X_1) + E(X_2) + \ldots E(X_n) = np$$

en $$Var(X) = Var(X_1) + Var(X_2) + \ldots Var(X_n) = np(1-p)$$

Hieruit volgt dat de toevalsvariabele $\hat{P} = \dfrac{X}{n}$ bij benadering normaal verdeeld is met gemiddelde p en standaardafwijking $\sqrt{\dfrac{p(1-p)}{n}}$:

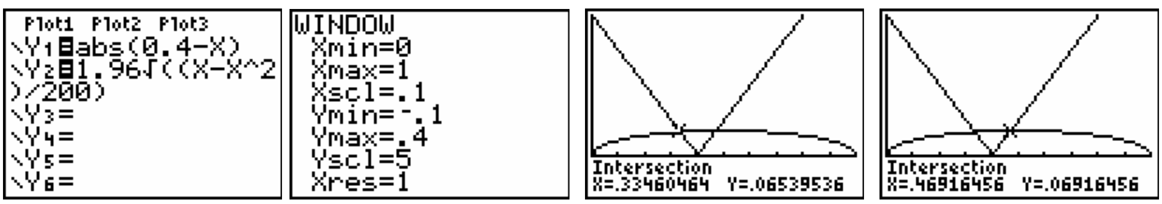
$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = p \quad \text{en} \quad Var(\hat{P}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{p(1-p)}{n}$$

---

a) Exacte oplossing:

In 95% der gevallen neemt $\hat{P}$ dus een waarde $\hat{p}$ aan binnen 1.96 standaardafwijkingen van het gemiddelde p of $|\hat{p} - p| \le 1.96\sqrt{\dfrac{p(1-p)}{n}}$ met 95% zekerheid.

Concreet vonden we voor onze steekproef $x = 80$ en $n = 200$ zodat $\hat{p} = \dfrac{80}{200} = 0.40$ en $|0.4 - p| \le 1.96\sqrt{\dfrac{p(1-p)}{200}}$ met 95% zekerheid. We lossen de ongelijkheid grafisch op:



Als interval verkrijgen we $0.3346 < p < 0.4692$ met 95 % zekerheid.

b) Benaderende oplossing:

Met 95 % zekerheid is $\hat{p}$ gelegen binnen 1.96 standaardafwijkingen van het gemiddelde p of het gemiddelde p is gelegen binnen 1.96 standaardafwijkingen van $\hat{p}$:

$$\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}$$

We benaderen $\sqrt{\frac{p(1-p)}{n}}$ door $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, dit levert het 95 % betrouwbaarheidsinterval

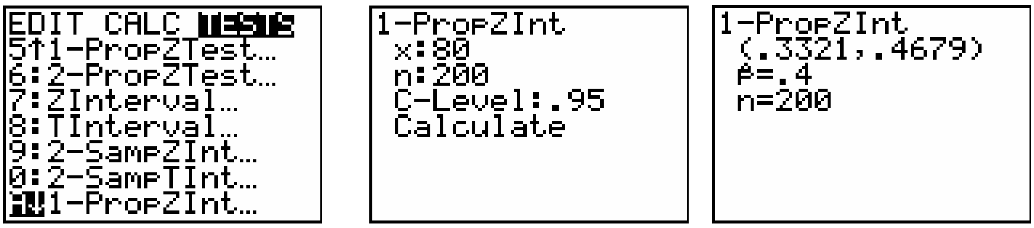
$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

of $\quad 0.3321 < p < 0.4679 \quad$ (vergelijk met het exacte interval)

Algemeen :

Een benaderend betrouwbaarheidsinterval met betrouwbaarheidsniveau $1-\alpha$,voor een populatieproportie *p*, wordt gegeven door $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, waarbij $\hat{p}$ een steekproefproportie is.

De TI-84 Plus levert het betrouwbaarheidsinterval met STAT <TESTS> A: 1-PropZInterval ; vul de gegevens in, ga met de cursor naar Calculate en druk ENTER:



Opmerking

Bovenstaande werkwijze (a) of (b) mogen we *niet* gebruiken bij een kleine steekproef, bijvoorbeeld $n = 5$, aangezien de centrale limietstelling dan niet van toepassing is; de steekproefproportie $\widehat{P} = \frac{X}{n}$ is dan niet normaal verdeeld.

Opgave

Van 20000 Belgen onderzoekt men de bloedgroep met als resultaat:

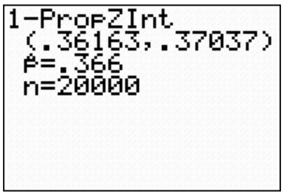
Bloedgroep	A	B	AB	O
%	43.2	14.2	6	36.6

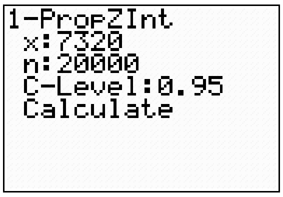
a) Bepaal betrouwbaarheidsintervallen voor de proportie *p* der Belgen met bloedgroep O, met betrouwbaarheidsniveau's 80% , 95% , 99% .

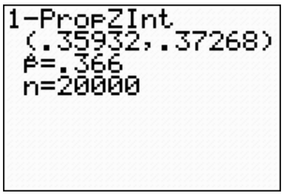
b) Hoe groot moet een (aselecte) steekproef zijn, om de proportie *p* der Belgen met bloedgroep O tot op 0.01 nauwkeurig te kennen met 95% zekerheid?

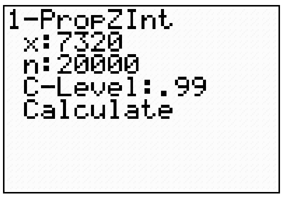
Oplossing

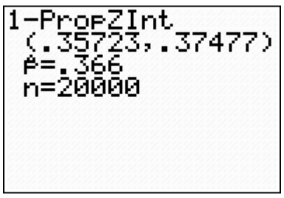
a)











Met 99% zekerheid kunnen we dus zeggen dat de ongekende proportie *p* gelegen is tussen 35.7% en 37.5% !

b) Er moet gelden dat $1.96 \cdot \sqrt{\dfrac{p(1-p)}{n}} \leq 0.01$ of $n \geq 38416 \cdot p(1-p)$ (1)

We kennen p niet maar

$$p(1-p) = -(p^2 - p) = -\left(\left(p - \frac{1}{2}\right)^2 - \frac{1}{4}\right) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2$$

zodat $p(1-p) \leq \dfrac{1}{4}$ (gelijkheid treedt op als $p = \dfrac{1}{2}$).

Bijgevolg is $38416\,p(1-p) \leq 9604$

zodat een steekproefgrootte $n = 9604$ zeker volstaat.

Schatten we echter in (1) p met $\hat{p} = 0.366$ dan vinden we

$$n \geq 38416 \cdot 0.366 \cdot 0.634$$

of $n \geq 38416 \cdot 0.232044$

of $n \geq 8914.2$

zodat $n = 8915$ volstaat (de besparing is gering aangezien $\hat{p}(1-\hat{p})$ dicht bij $1/4$ gelegen is).

5 De normale verdeling met de TI-84 Plus: een overzicht

De commando's i.v.m. normale verdelingen op een TI-84 (Plus) zijn terug te vinden onder

a) 2nd [DISTR] 1: normalpdf(
2nd [DISTR] 2: normalcdf(
2nd [DISTR] 3: invNorm(

b) 2nd [DISTR] <DRAW> 1: SchadeNorm(

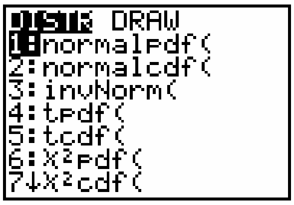
c) MATH <PRB> 6: randNorm(

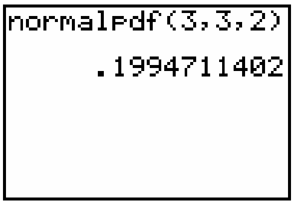
Vooraf merken we algemeen op dat de commando's onder (a) en (b) *eindigen* met het opgeven van de parameters μ en σ. Deze parameters mag je ook weglaten voor de standaard normale verdeling.

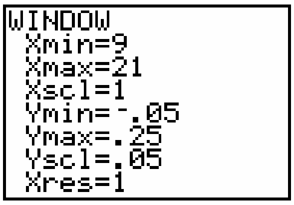
5.1 normalpdf

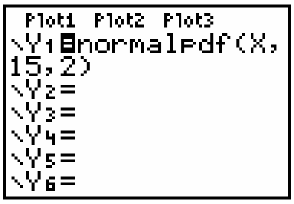
normalpdf is de dichtheidsfunctie van een normale verdeling. De syntax voor de functiewaarde in het punt x is normalpdf(x , μ , σ) .

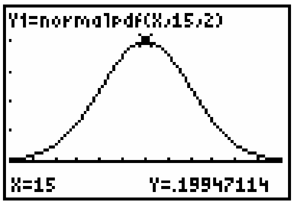
Om de grafiek te tekenen van de normale verdeling met gemiddelde 15 en standaardafwijking 2 bepaal je eerst een geschikt venster met WINDOW : nagenoeg alle data bevinden zich in het interval $[\mu-3\sigma,\mu+3\sigma]=[15-3\cdot 2,15+3\cdot 2]=[9,21]$ en de maximale functiewaarde is normalpdf (μ , μ , σ) = normalpdf $(3,3,2)$











5.2 normalcdf

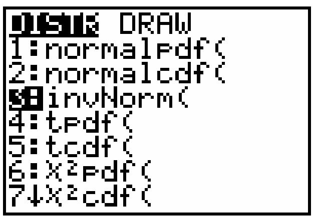
Met normalcdf bereken je de proportie van de data gelegen in een interval , de syntax is normalcdf (ondergrens , bovengrens , μ , σ) .

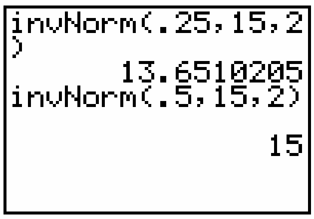
I.p.v. $+\infty$ neem je $10\wedge 99=10^{99}$, i.p.v. $-\infty$ neem je $-10\wedge 99$.

Zo levert `normalcdf(`$-10\wedge 99$, x , μ , σ) de *verdelingsfunctie* van de normale verdeling, d.i. de proportie van de data kleiner dan of gelijk aan x , of de oppervlakte onder de normale verdeling links van x.

5.3 invNorm

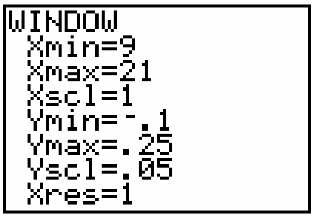
invNorm is de inverse functie van de verdelingsfunctie , de syntax is invNorm (y, μ , σ) . Voor een gegeven getal y tussen 0 en 1 bereken je hiermee de grens x waarvoor de oppervlakte onder de normale verdeling links van x gelijk is aan y . Hiermee bepaal je o.a. de kwartielen van een normale verdeling.

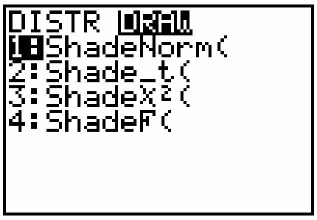


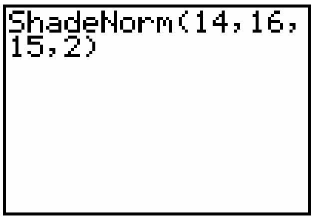


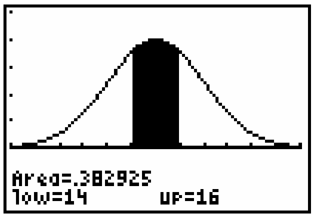
5.4 ShadeNorm

Met ShadeNorm doe je hetzelfde als normalcdf , bovendien wordt de normale verdeling getekend en de berekende oppervlakte gearceerd. Je moet dus eerst de geschikte venstergrenzen met WINDOW instellen en met Y= de eerder gedefinieerde functies uitzetten. De syntax is ShadeNorm (ondergrens , bovengrens , μ , σ) . De opdracht wordt gegeven in het basisscherm en dus pas uitgevoerd na het drukken op ENTER.







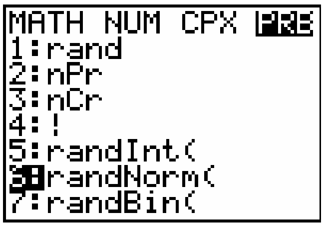


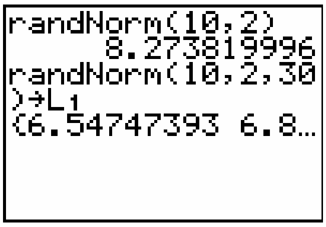
5.5 randNorm

Met randNorm(μ,σ) kun je simulaties uitvoeren van lukrake trekkingen uit een normaal verdeelde populatie met gemiddelde μ en standaardafwijking σ .

Je vindt het commando onder MATH <PRB> 6: randNorm(

Een steekproef met grootte n uit een normaal verdeelde populatie met gemiddelde μ en standaardafwijking σ wordt als een lijst van data gegenereerd met randNorm(μ,σ,n) .





Het testen van hypothesen

1 Inleiding

Een eerste kennismaking met de basisprincipes van hypothesetesten gebeurt best met een eenvoudig voorbeeld, waarbij meteen de *context* van het toepassingsgebied het uitgangspunt is: *het bestaan van een onderzoeksvraag, waarop een antwoord moet worden geformuleerd.*

Met dit voorbeeld kan de student stapsgewijs kennismaken met de vele nieuwe begrippen zoals statistische hypothesen, p-waarde, het significantieniveau α, type I en type II fout, tweezijdig of eenzijdig testen, het al dan niet significant zijn van de steekproefdata, …
De hele procedure van het testen van hypothesen komt hiermee aan bod als een natuurlijk denkproces, in plaats van een opsomming van uit te voeren stappen en te berekenen formules.

De volgende kennismaking gebruikt enkel de *binomiale verdeling* als vereiste voorkennis en wordt aangeboden als een werktekst. De antwoorden bij die tekst vind je aansluitend op pagina 28. Paragraaf 5 toont een overzicht van de TI-84 Plus commando's voor de binomiale verdeling.

2 Kennismaking: Eenden met voorkeur voor groen?

I De onderzoeksvraag

Een student biologie heeft als opdracht een antwoord te vinden op de volgende onderzoeksvraag:

Bij een bepaalde eendensoort (de slobeend , zie http://users.telenet.be/subv/) hebben de mannetjes een mooie glanzend groene kop, met als doel de vrouwtjes aan te trekken. De vraag is of de vrouwtjeseenden ook aangetrokken worden tot een groene kleur in hun voedsel, bijvoorbeeld brood?

II Statistische hypothesen

De onderzoeksvraag kan worden vertaald in de confrontatie van twee ideeën:

1) vrouwtjeseenden zijn onverschillig voor groen of gewoon brood
of
2) vrouwtjeseenden verkiezen groen brood boven gewoon brood

Als een vrouwtjeseend van die eendensoort geconfronteerd wordt met twee stukjes brood (gewoon en groen) , dan noteren we de kans op de keuze van het groene brood met p.

De ideeën worden dan kort:

1) $p =$ __________
2) $p >$ __________

Deze confronterende ideeën of uitspraken noemen we ***statistische hypothesen*** , de eerste hypothese stelt dat de eenden even graag gewoon als groen brood eten. Dit noemen we de ***nulhypothese***, voorgesteld door H_0, aangezien dit het idee vertegenwoordigt dat er *geen verschil* is (Engelse term: ‘null-change’). Het tweede idee drukt wel degelijk een verschil uit: de eenden *verkiezen* groen brood i.p.v. gewoon brood. Daarom wordt dit de ***alternatieve hypothese*** genoemd en genoteerd met H_1.

We verkrijgen dus de volgende hypothesen:

H_0: $p = 1/2$
H_1: $p > 1/2$

De reden waarom we hier de alternatieve hypothese H_1 formuleren als $p > 1/2$ (***éénzijdige test***) en niet $p \neq 1/2$ (***tweezijdige test***) is omdat de onderzoeksvraag enkel vraagt om na te gaan of de eenden groen brood *verkiezen* i.p.v. gewoon brood. Men gaat er dus a priori van uit dat de eenden zeker geen voorkeur hebben voor het gewone brood.

Uiteindelijk moeten we *beslissen* welke van de twee hypothesen voor ons de voorkeur krijgt; we moeten “kiezen” voor H_0 of H_1 , we zeggen dat we H_0 *testen versus* H_1.

Er zijn twee mogelijkheden (let op de werkwoordkeuze bij de hypothesen):

1) We *verwerpen* H_0 (we *aanvaarden* H_1)
of
2) We *verwerpen* H_0 *niet* (we *aanvaarden* H_1 *niet*)

De beslissing wordt meestal geformuleerd in termen van H_0: we verwerpen H_0 of we verwerpen H_0 niet.

III Bewijsmateriaal verzamelen om te kunnen beslissen

De student ontwerpt een experiment om tot een beslissing te komen: zal hij H_0 verwerpen of niet verwerpen?
Hij gaat naar een meer in de omgeving van de campus, waar slobeenden in grote aantallen aanwezig zijn, en kiest lukraak 10 vrouwtjeseenden. Elke eend krijgt twee stukken brood aangeboden: een gewoon stuk en een stuk met een groen laagje verf. Hij vat het resultaat samen in een rapport waarin hij het aantal eenden vermeldt die eerst naar het groene brood zwemmen.

Denk na over de toevalsvariabele X = het aantal eenden in de *steekproef* die groen brood verkiezen. De steekproefgrootte is hier gelijk aan $n = 10$.

Als de eenden onverschillig zijn voor gewoon of groen brood, wat is dan de verdeling van de variabele X ?

IV De weg naar de beslissing

Als de vrouwtjeseenden echt onverschillig zijn t.o.v. groen of gewoon brood, hoeveel van de 10 eenden verwacht je dan die eerst naar het groene brood zwemmen? ________________

Natuurlijk zal dit resultaat in werkelijkheid niet steeds hieraan gelijk zijn, ook al is de nulhypothese juist, omwille van de mogelijke variaties in de steekproeven te wijten aan het toeval.

De student biologie stelt vast dat 9 van de 10 eenden in zijn steekproef het groene brood verkiezen. Stel dat de populatie van alle slobeenden onverschillig zijn voor gewoon of groen brood (stel dat H_0 waar is), wat is dan de kans om in een steekproef van 10 eenden er 9 aan te treffen die eerst het groene brood oppikken? __________________________

Negen van de tien lijkt erop te wijzen dat de eendenpopulatie eerder het groene brood verkiest dan gewoon brood. Als er meer dan 9 eenden het groene brood zouden kiezen, dan zou dit nog meer wijzen op een situatie verschillend van de nulhypothese! Daarom zijn we geïnteresseerd in de kans dat er 9 (het geobserveerde aantal in de steekproef van de student) of meer (nog overtuigender wijzend in de richting van de alternatieve hypothese) vrouwtjeseenden eerst het groene brood gaan kiezen. We willen dus de kans kennen op een resultaat dat minstens even groot is als het steekproefresultaat, in de veronderstelling dat de nulhypothese waar is. Bereken deze kans $P\left(X \geq 9\right)$:

____

Deze kans om een resultaat te verkrijgen gelijk aan het steekproefresultaat of nog extremer (wijzend in de richting van de alternatieve hypothese) noemen we de '***p-waarde***' of *overschrijdingskans*.

Vind je deze *p*-waarde klein of groot? ______________________

Geloof je dat de nulhypothese H_0 waar is, rekening houdend met deze *p*-waarde? (omcirkel je antwoord).

Ja Neen

Welke hypothese krijgt dus je voorkeur, H_0 of H_1 ? ________________

Verwerp je H_0 of verwerp je H_0 niet? _____________________

Formuleer je antwoord op de onderzoeksvraag:

____

____

V Samenvatting van het denkproces

Herlees de stappen I tot IV , onze redenering om tot een beslissing te komen kunnen we als volgt samenvatten:

(a) Formuleer de onderzoeksvraag. (Verkiezen vrouwtjes slobeenden groen brood boven gewoon brood?)

(b) Bepaal een grootheid waar het onderzoek eigenlijk over gaat en waarvan we de waarde niet kennen. In dit voorbeeld is die grootheid de proportie van de vrouwtjeseenden die het groene brood verkiezen of de kans p dat een lukraak gekozen vrouwtjeseend het groene brood kiest. In het algemeen wordt die grootheid een ***parameter*** genoemd.

(c) Schrijf de twee confronterende statistische hypothesen in termen van de parameter waar het om gaat.
In het voorbeeld zijn de statistische hypothesen $H_0: \; p = 1/2$ en $H_1: \; p > 1/2$.

(d) Verzamel steekproefdata en bereken de verkregen waarde van de ***teststatistiek*** (d.i. een *concreet getal* verkregen m.b.v. de steekproefgegevens, in dit geval het aantal eenden (9 op 10) die groen brood verkiezen. De teststatistiek of ***testvariabele*** zelf is een *toevalsvariabele* (het aantal eenden X die het groene brood verkiezen) waarvan we de concrete waarde $x = 9$ hebben geobserveerd.

(e) Bereken de p-waarde, d.i. de *kans* dat we het resultaat van de steekproef verkrijgen of een nog extremer resultaat (wijzend in de richting van de alternatieve hypothese), *in de veronderstelling dat de nulhypothese waar is*.

(f) Beslis of de p-waarde klein of groot is. In ons voorbeeld gaf de kleine p-waarde aanleiding tot het verwerpen van de nulhypothese.

Deze redeneervorm wordt "***testen van hypothesen***" genoemd. De redeneervorm kan worden toegepast in verschillende situaties waarin een onderzoeksvraag over een populatie wordt gesteld en steekproefdata worden verzameld om een antwoord te kunnen geven op de vraag.

Het is duidelijk dat (f) een delicaat punt is dat nog meer aandacht verdient, welke p-waarde kunnen we immers als "klein" beschouwen? Dit bespreken we in de volgende paragraaf.

VI Op hoeveel verschillende manieren kunnen we een verkeerde beslissing nemen?

Bij het testen van hypothesen moeten we H_0 of H_1 selecteren. Natuurlijk nemen we graag de correcte beslissing, maar soms kunnen we ook een foute beslissing nemen. Hoe zou je in woorden de volgende situaties kunnen omschrijven? Formuleer dit in termen van wat de eenden verkiezen en wat wij *beweren* dat ze verkiezen.

(1) We verwerpen de nulhypothese H_0 , terwijl H_0 in werkelijkheid waar is.

____

____

(2) We verwerpen de nulhypothese H_0 niet, terwijl H_0 in werkelijkheid niet waar is.

____

____

(1) noemt men een ***type I fout*** (onterecht H_0 verwerpen).

(2) noemt men een ***type II fout*** (onterecht H_0 niet verwerpen).

De volgende tabel geeft een overzicht van de mogelijke situaties:

	je verwerpt H_0	je verwerpt H_0 niet
H_0 is waar	Type I fout	juiste beslissing
H_0 is niet waar	juiste beslissing	Type II fout

Natuurlijk wensen we de kans op het maken van een fout zeer klein te houden.
Onze aandacht gaat vooral uit naar het klein houden van de kans op een type I fout, omdat de nulhypothese een ‘status quo’ of neutrale situatie inhoudt, een traditionele situatie.
Als we de nulhypothese verwerpen, dan zeggen we dat iets beter is of verkozen wordt, of slechter, of verschillend, afhankelijk van hoe de alternatieve hypothese geformuleerd is. Als bijvoorbeeld twee medicamenten worden vergeleken in een onderzoek, dan zou een type I fout beweren dat een medicament beter is terwijl de medicamenten in werkelijkheid gelijkaardige effecten hebben. De nulhypothese is de ‘klassieke’ of ‘traditionele hypothese’ en een type I fout (het ten onrechte verwerpen van deze klassieke hypothese) beschouwt men dan erger dan een type II fout (het ten onrechte niet verwerpen van de klassieke hypothese).

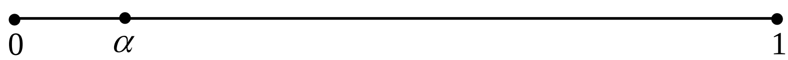
De kans op het maken van een type I fout wordt het ***significantieniveau*** genoemd en genoteerd met α. We wensen die kans α onder controle te hebben, de keuze van α wordt mee bepaald door het belang en het gevolg van onze beslissing (cfr. een rechtszaak: is de beschuldigde schuldig of niet?)

Stel dat je *op voorhand* beslist de nulhypothese te verwerpen van zodra er minstens 8 eenden het groene brood verkiezen, wat is dan de kans op het maken van een type I fout?
Of, anders geformuleerd, wat is de kans dat er minstens 8 eenden het groene brood verkiezen als H_0 waar is?

$\alpha = P\left(X \geq 8\right) =$ ______________________

Het is belangrijk dat de kans op een type I fout *op voorhand* wordt vastgelegd, bij het begin van de studie, vooraleer de data worden verzameld en de teststatistiek wordt waargenomen. Een keuze achteraf zou betekenen dat je de beslissing (H_0 verwerpen of niet) kunt aanpassen zoals je zelf wil in plaats van te proberen om de waarheid te achterhalen.

Voor welke *p*-waarde zullen we H_0 verwerpen, voor welke niet? Het significantieniveau α geeft de grens aan. Duid in onderstaand interval $[0,1]$ het gebied van *p*-waarden aan waarvoor H_0 wordt verworpen en het gebied van *p*-waarden waarvoor H_0 niet wordt verworpen.



We verwerpen H_0 als de *p*-waarde ________ α. In dit geval zeggen we dat de steekproefdata ***statistisch significant*** zijn (ze leveren het statistisch bewijs waarmee we de traditionele nulhypothese kunnen weerleggen)

We verwerpen H_0 niet als de *p*-waarde ________ α.

Denk nu eens na over het volgende: als je overdrijft in je voorzichtigheid bij het vermijden van een type I fout (door een zeer kleine α te kiezen), welk effect heeft dit dan volgens jou op het maken van een type II fout?

____

De kans op een type II fout wordt β genoemd. Om β onder controle te houden mogen we α niet overdreven klein maken, tenzij we voldoende bewijs (data) kunnen aanleveren om overtuigend te kunnen beslissen (zie volgende paragraaf).

VII De invloed van de steekproefgrootte (optie 1)

Men kiest vaak $\alpha = 5\%, 10\%$ of 1%.
Student *A* werkt met 10 eenden, student *B* met 20 eenden, student *C* met 40 eenden.
Veronderstel dat de drie studenten de opdracht krijgen om te werken met een significantieniveau $\alpha \simeq 5\%$, zo dicht als mogelijk bij 5% (bij een discrete testvariabele is $\alpha = 5\%$ exact immers meestal niet mogelijk).

Bepaal (met je rekentoestel) voor elk van de studenten het minimum aantal eenden dat het groene brood moet oppikken om de nulhypothese H_0 te kunnen verwerpen (De *p*-waarde bij dit minimum is gelijk aan het significantieniveau α). Vul de volgende tabel in, waarbij *x* het aantal eenden is dat eerst het groene brood oppikt.

Student	n	H_0 niet verwerpen als $x \leq$	H_0 verwerpen als $x \geq$
A	10		
B	20		
C	40		

Veronderstel nu dat H_0 niet waar is; stel dat 80% van de slobeendenpopulatie het groene brood verkiezen. Maak gebruik van de waarden van *x* die je in bovenstaande tabel hebt genoteerd en werk met je rekentoestel.

Student	n	β (met $\alpha \simeq 5\%$ en $p = 0.8$)
A	10	
B	20	
C	40	

Schrijf een kort verslag over de invloed van de steekproefgrootte op β.

VIII Werd de nulhypothese terecht verworpen? (Optie 2)

Het ***onderscheidingsvermogen*** $1-\beta$ van een test (Engels: power) geeft aan hoe goed de test in staat is om de nulhypothese terecht te verwerpen (de alternatieve hypothese is inderdaad waar). Bespreek, met de resultaten van voorgaande paragraaf, de invloed van de steekproefgrootte op het onderscheidingsvermogen.

	je verwerpt H_0	je verwerpt H_0 niet
H_0 is waar	Type I fout kans α = *significantieniveau*	juiste beslissing kans $1-\alpha$ = ***betrouwbaarheidsniveau***
H_0 is niet waar	juiste beslissing kans $1-\beta$ = *onderscheidingsvermogen*	Type II fout Kans β

Veronderstel nu dat H_0 niet waar is, maar dat slechts 70% van de slobeendenpopulatie het groene brood verkiezen. Bepaal opnieuw de kans β op een type II fout en het onderscheidingsvermogen $1-\beta$? Vergelijk dit met de situatie $p = 0.8$

Student	n	β (met $\alpha \simeq 5\%$ en $p = 0.7$)
A	10	
B	20	
C	40	

We hebben de basisbegrippen van hypothesetesten besproken. Herlees nu de opeenvolgende stappen (a) tot (f) in paragraaf V. Herinner je dat je vóór stap (d) een significantieniveau α moet vastleggen waarmee je zal werken. Herbekijk de nieuwe woorden (vetjes gedrukt) die werden ingevoerd, begrijp je hun betekenis? De stappen (a) tot (f), samen met de vooropgestelde α, zullen nuttig zijn in andere situaties waarbij je een antwoord moet zoeken op een onderzoeksvraag m.b.v. het uitvoeren van een hypothesetest.

Antwoorden bij de werktekst (I tot VIII)

II Statistische hypothesen

……..

De ideeën worden dan kort:

1) $p = 1/2$

2) $p > 1/2$

III Bewijsmateriaal verzamelen om te kunnen beslissen

……………..

Als de eenden onverschillig zijn voor gewoon of groen brood, wat is dan de verdeling van de variabele X ?

X is binomiaal verdeeld met parameters $n = 10$ en $p = 1/2$: $X \sim B(10, 1/2)$

IV De weg naar de beslissing

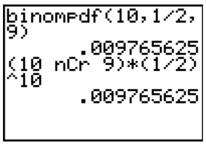
Als de vrouwtjeseenden echt onverschillig zijn t.o.v. groen of gewoon brood, hoeveel van de 10 eenden verwacht je dan die eerst naar het groene brood zwemmen?

Vier of vijf of zes (ongeveer de helft). Het is natuurlijk maar een steekproef dus het hoeft niet exact vijf te zijn. Negen of tien verwachten we zeker niet.

……..

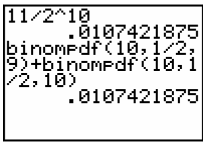
Stel dat de populatie van alle slobeenden onverschillig zijn voor gewoon of groen brood (stel dat H_0 waar is), wat is dan de kans om in een steekproef van 10 eenden er 9 aan te treffen die eerst het groene brood oppikken?

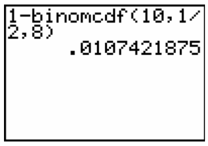
$$P(X = 9) = \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 = 10 \cdot \frac{1}{2^{10}} \simeq 0.98\%$$



………We willen dus de kans kennen op een resultaat dat minstens even groot is als het steekproefresultaat, in de veronderstelling dat de nulhypothese waar is. Bereken deze kans

$$P(X \geq 9) = P(X = 9) + P(X = 10) = \frac{10}{2^{10}} + \frac{1}{2^{10}} \simeq 1.07\%$$





……..

Vind je deze *p*-waarde klein of groot? Deze kans is zeer klein

Geloof je dat de nulhypothese H_0 waar is, rekening houdend met deze *p*-waarde? (omcirkel je antwoord). Neen

Welke hypothese krijgt dus je voorkeur, H_0 of H_1 ? H_1

Verwerp je H_0 of verwerp je H_0 niet? Ik verwerp H_0.

Formuleer je antwoord op de onderzoeksvraag:

De vrouwtjes slobeenden verkiezen groen brood boven gewoon brood.

VI Op hoeveel verschillende manieren kunnen we een verkeerde beslissing nemen?

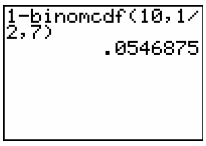
……..

(1) We verwerpen de nulhypothese H_0 , terwijl H_0 in werkelijkheid waar is.

We beweren dat de eenden het groene brood verkiezen terwijl ze in werkelijkheid geen voorkeur hebben voor groen of gewoon brood.

(2) We verwerpen de nulhypothese H_0 niet, terwijl H_0 in werkelijkheid niet waar is.

We beweren dat de eenden onverschillig zijn voor groen of gewoon brood, terwijl ze in werkelijkheid het groene brood verkiezen.



……………….

$\alpha = P(X \geq 8) = 1 - P(X \leq 7) \simeq 5.5\%$

…………..

We verwerpen H_0 als de *p*-waarde $\leq \alpha$.

We verwerpen H_0 niet als de *p*-waarde $> \alpha$.

Denk nu eens na over het volgende: als je overdrijft in je voorzichtigheid bij het vermijden van een type I fout (door een zeer kleine α te kiezen), welk effect heeft dit dan volgens jou op het maken van een type II fout?

De kans op het maken van een type II fout wordt groter.

VII De invloed van de steekproefgrootte (optie 1)

Student	n	H_0 niet verwerpen als $x \leq$	H_0 verwerpen als $x \geq$
A	10	7	8 ($\alpha \simeq 5.5\%$)
B	20	13	14 ($\alpha \simeq 5.8\%$)
C	40	25	26 ($\alpha \simeq 4.0\%$)

Student	n	β (met $\alpha \simeq 5\%$ en $p = 0.8$)
A	10	$P(X \leq 7) =$ binomcdf (10, 0.8, 7) $\simeq 32.2\%$
B	20	$P(X \leq 13) =$ binomcdf (20, 0.8, 13) $\simeq 8.7\%$
C	40	$P(X \leq 25) =$ binomcdf (40, 0.8, 25) $\simeq 0.8\%$

De kans β op een type II fout wordt gevoelig kleiner met toenemende steekproefgrootte *n*, terwijl het significantieniveau α of de kans op een type I fout (nagenoeg) constant blijft.

VIII Werd de nulhypothese terecht verworpen? (Optie 2)

…..

Bespreek, met de resultaten van voorgaande paragraaf, de invloed van de steekproefgrootte op het onderscheidingsvermogen.

Het onderscheidingsvermogen $1-\beta$ neemt toe met toenemende steekproefgrootte *n* (bij constante α).

Veronderstel nu dat H_0 niet waar is, maar dat slechts 70% van de slobeendenpopulatie het groene brood verkiezen. Bepaal opnieuw de kans β op een type II fout en het onderscheidingsvermogen $1-\beta$? Vergelijk dit met de situatie $p = 0.8$

Student	n	β (met $\alpha \simeq 5\%$ en $p = 0.7$)
A	10	$P(X \le 7) =$ binomcdf (10,0.7, 7) $\simeq 61.7\%$
B	20	$P(X \le 13) =$ binomcdf (20,0.7, 13) $\simeq 39.2\%$
C	40	$P(X \le 25) =$ binomcdf (40,0.7, 25) $\simeq 19.3\%$

Het onderscheidingsvermogen $1-\beta$ neemt af (bij vaste n en α) naarmate het werkelijke percentage van de slobeenden die kiezen voor het groene brood dichter bij 0.5 gelegen is.

3 Benaderende test voor een proportie bij een grote steekproef

We wensen na te gaan of een individu uit de populatie een bepaalde eigenschap bezit.

$p = \dfrac{\text{aantal individuen in de populatie met de eigenschap}}{\text{aantal individuen in de populatie}}$ is de *populatieproportie*.

De *steekproefproportie* is de toevalsvariabele :

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

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

Voor n voldoende groot is de variabele $Z = \dfrac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{X - np}{\sqrt{np(1-p)}}$ *bij benadering* standaard normaal verdeeld omwille van de centrale limietstelling (zie ook pagina 16).

Als vuistregel voor n voldoende groot nemen we de voorwaarden :

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

Het uitvoeren van de hypothesetest $H_0 : p = p_0$ versus b.v. $H_1 : p < p_0$ met $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ als testvariabele kan met de TI-84 Plus via het commando STAT<TESTS> 5:1-PropZTest.

Opgaven:

1) Een krantenartikel beweert dat 40% van de bevolking (leeftijdsgroep ≥ 12 jaar) rookt. Je vermoedt dat dit niet klopt; een steekproef van 100 personen (≥ 12 jaar) levert 52 rokers.

Test $H_0 : p = 0.4$ versus $H_1 : p \neq 0.4$ (tweezijdige test) op het significantieniveau $\alpha = 5\%$.

a) Exact met het aantal rokers X als teststatistiek met een binomiale verdeling.

b) Benaderend met X als teststatistiek met een bij benadering normale verdeling.

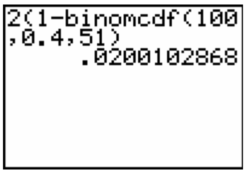
c) Benaderend met $Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ als teststatistiek met een standaard normale verdeling

d) Los de onderzoeksvraag (spreekt het artikel de waarheid?) op m.b.v. een betrouwbaarheidsinterval met betrouwbaarheidsniveau $1-\alpha = 95\%$

2) Een toets bestaat uit 10 meerkeuzevragen met elk 3 mogelijke antwoorden waarvan er één juist is. Een juist antwoord levert 1 punt op, een fout antwoord 0.
Een student haalt 8 op 10 op die toets en beweert gegokt te hebben.
Geloof je hem?

Oplossingen

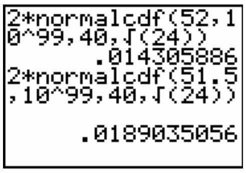
Vraag 1) a) Als H_0 waar is dan geldt $X \sim B(100,\ 0.4)$. Als *p*-waarde vinden we $2 \cdot P(X \geq 52) = 2 \cdot (1 - P(X \leq 51)) \simeq 2.0\%$. Merk op dat we hierbij $P(X \geq 52)$ verdubbelen omdat het een tweezijdige test betreft: kleine steekproefresultaten die minder dan 40% doen vermoeden (met dezelfde kans als grote resultaten ≥ 52 die meer dan 40% doen vermoeden) maken H_0 even verdacht. Aangezien de *p*-waarde $\leq 5\%$ is verwerpen we H_0.



b) Aangezien $X \sim B(100,\ 0.4)$ is $E(X) = n \cdot p = 40$ en $Var(X) = n \cdot p \cdot (1-p) = 100 \cdot 0.4 \cdot 0.6 = 24$; omwille van de centrale limietstelling geldt bij benadering $X \sim N\left(40, \sqrt{24}\right)$.

Als *p*-waarde vinden we met deze normale verdeling $2 \cdot P(X \geq 52) \simeq 1.4\%$.

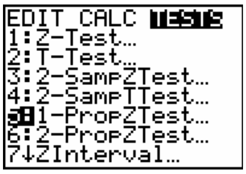
Met continuïteitscorrectie verkrijgen we een betere *p*-waarde $= 2 \cdot P(X \geq 51.5) \simeq 1.9\%$

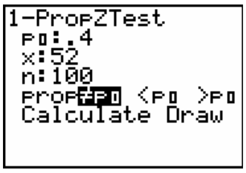


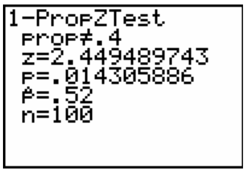
c) De teststatistiek $Z=\dfrac{\hat{P}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

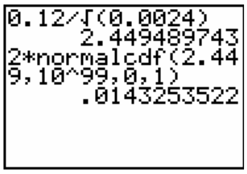
neemt de waarde $z=\dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}=\dfrac{0.52-0.4}{\sqrt{\dfrac{0.4\cdot 0.6}{100}}}\simeq 2.449$ aan voor onze steekproef .

De p-waarde is $2\cdot P\left(Z\geq 2.449\right)\simeq 1.4\%$.



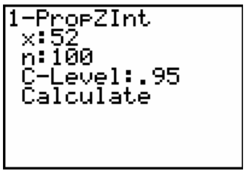


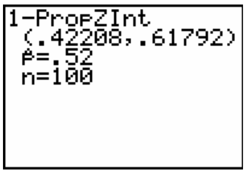




d) Een benaderend betrouwbaarheidsinterval met betrouwbaarheidsniveau $1-\alpha$, voor een populatieproportie p, wordt gegeven door $\hat{p}\pm z_{\alpha/2}\cdot\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}}$, waarbij $\hat{p}$ een steekproefproportie is (zie pagina 17).

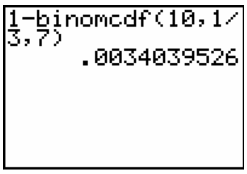
We vinden een 95% betrouwbaarheidsinterval dat het getal 0.4 *niet* bevat. Bijgevolg verwerpen we H_0.





Vraag 2)

Zij X het aantal juiste antwoorden, stel p de kans om een vraag juist te beantwoorden (we veronderstellen dat alle vragen dezelfde moeilijkheidsgraad hebben).
We testen $H_0: p=1/3$ (de student heeft gegokt)
versus $H_1: p>1/3$ (de student heeft gestudeerd en niet gegokt).
Als H_0 waar is dan geldt $X\sim B\left(10, 1/3\right)$.
De p-waarde is $P\left(X\geq 8\right)=1-P\left(X\leq 7\right)\simeq 0.3\%$, dit is zeer klein.
We hebben een overtuigend bewijs tegen H_0 en geloven de student niet…



4 Test voor het gemiddelde van een normale verdeling met gekende standaardafwijking

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 testen houdt hij gegevens bij van dertig kinderen.

de hieronder afgebeelde data (in kg) van de gynaecoloog plaatsen we in de lijst L1.

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

De gynaecoloog berekent het gemiddelde van zijn steekproef en bekomt als resultaat $\bar{x} = 3.483\text{ kg}$. 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 testen, 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 gemiddelde en standaardafwijking deze van gans Vlaanderen: $\sigma = 0.55\text{ kg}$

Vervolgens formuleert hij twee hypothesen over de parameter μ van de populatie: de *nulhypothese* 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 in zijn kliniek 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 testen $H_0 : \mu = 3.3$ versus $H_1 : \mu > 3.3$.

De hypothesetest verloopt verder als volgt. We veronderstellen in de redenering dat de nulhypothese H_0 *waar* is. Indien het gemiddelde geboortegewicht $\bar{x}$ van de steekproef van de gynaecoloog *te groot* is, verwerpen we de nulhypothese 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?

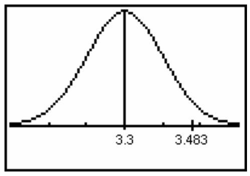
Het verkregen steekproefgemiddelde $\bar{x} = 3.483\text{ kg}$ is een door het toeval verkregen waarde van de stochast of toevalsvariabele $\bar{X} = \frac{1}{30}(X_1 + ... + X_{30})$, waarbij elke X_i verdeeld is zoals de populatievariabele X.

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 :

$$X \sim N(3.3, 0.55) \text{ en } \bar{X} \sim N(3.3, \frac{0.55}{\sqrt{30}}) \text{ of } \bar{X} \sim N(3.3, 0.1)$$

De toevalsvariabele $\bar{X}$ is de *testvariabele,* dit is een functie van de steekproef waarmee we gaan beslissen of we kiezen voor H_0 of H_1.

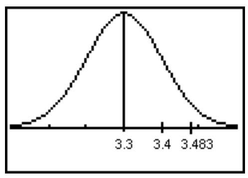
De waarde $\bar{x} = 3.483$ bevindt zich erg rechts; $\bar{x}$ ligt zelfs zo ver naar rechts dat de kans (indien H_0 waar is) om zo'n geobserveerde waarde te krijgen of groter, zeer klein is.



Dichtheidsfunctie van $\bar{X}$

Een dergelijke geobserveerde waarde onder de nulhypothese 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 $\bar{x} = 3.483\text{ kg}$. Het is niet meer zo evident om de nulhypothese te verwerpen.



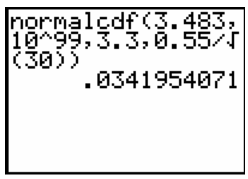
Dichtheidsfunctie $\bar{X}$

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

Een getal dat 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 :



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

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* α: verwerp H_0 bij een *p*-waarde $\leq \alpha$ (we zeggen dat de steekproefdata statistisch significant zijn op het α-niveau), verwerp H_0 niet bij een *p*-waarde $> \alpha$.

Om historische redenen 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 testgrootheid te nemen dan $\overline{X}$, nl.

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

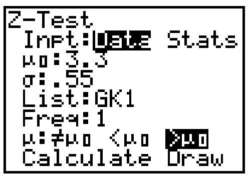
We spreken dan over de *Z-test*. We berekenen de *p*-waarde in dit geval als volgt :

$$P\left(\frac{\overline{X} - 3.3}{0.55/\sqrt{30}} \geq \frac{\overline{x} - 3.3}{0.55/\sqrt{30}}\right) = P\left(Z \geq \frac{3.483 - 3.3}{0.55/\sqrt{30}}\right)$$

$$= P\left(Z \geq 1.822\right) = 0.034$$

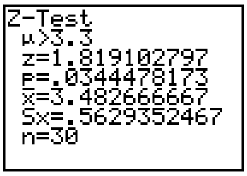
Met de TI-84 voer je deze test uit met het commando STAT<TESTS> 1: Z-Test.

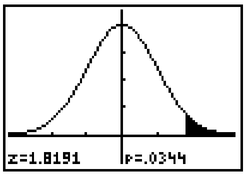
Vul dan het Z-Test-venster in zoals hiernaast is afgebeeld.
Indien de resultaten van de steekproef in een lijst gegeven zijn, selecteer je voor het item Inpt de optie Data en als je enkel statistische kengetallen kent van de steekproef, bv. $\overline{x} = 3.4$, selecteer je de optie Stats.



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

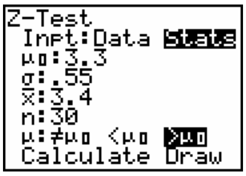
Activeer opnieuw het Z-Test-venster, selecteer Draw en druk op ENTER. Het resultaat vind je rechts op de onderstaande figuur.

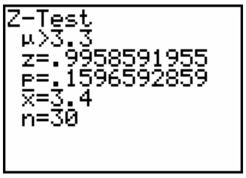


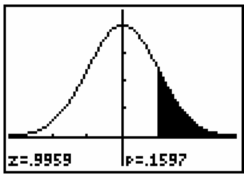


z is de geobserveerde waarde van de testgrootheid $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$, zodat $z = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$.

Voor het geval $\overline{x} = 3.4$ kg vind je hieronder de resultaten van de Z-Test.





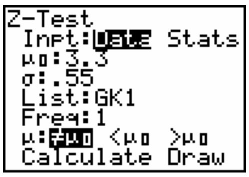


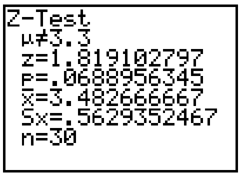
De p-waarde bij $\overline{x} = 3.4$ kg of $z = 0.9958$ is 0.1597. Dit is gevoelig groter dan $0.05 = 5\%$, zodat we de nulhypothese niet kunnen verwerpen op een 5% significantieniveau. Let op de terminologie: we verwerpen de nulhypothese niet, het is dus *mogelijk* dat $\mu = 3.3$. We spreken liever niet over het aanvaarden van de nulhypothese, aangezien we niet hebben “bewezen” dat μ exact gelijk is aan 3.3.

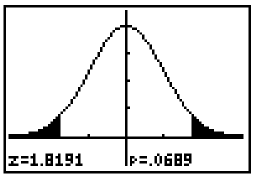
Bovenstaande test noemt men een *éénzijdige test* omdat men op voorhand vermoedt in welke richting een afwijking van H_0 wordt verwacht.

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

Het uitvoeren van deze test verloopt analoog. Het resultaat van deze test is :







Merk op dat in dit geval de p-waarde het dubbel is van de p-waarde van de éénzijdige test. Dit is zo omdat we voor een tweezijdige test 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 test is de p-waarde = 0.068 > 0.05 , we verwerpen de nulhypothese dus niet op het 5% significantieniveau.

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

Indien we op voorhand geen vermoeden hebben over het gemiddelde voeren we een tweezijdige test uit m.b.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 éénzijdige test uit te voeren met dezelfde steekproef, zou dit statistisch oneerlijk zijn t.o.v. de onwetendheid waarmee we gestart zijn.

5 De binomiale verdeling en de TI-84 Plus

Werp 20 keer een dobbelsteen. De kans dat hierbij

- juist 4 keer een zes geworpen wordt, is : $\binom{20}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{16} = 0.202$,
- juist 8 keer "zes of drie" : $\binom{20}{8}\left(\frac{1}{3}\right)^8\left(\frac{2}{3}\right)^{12} = 0.148$
- en hoogstens 4 keer een zes : $P(X \le 4) = \sum_{x=0}^{4}\binom{20}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{20-x} = 0.769$

Je kan deze kansen met de TI-84 Plus berekenen via :

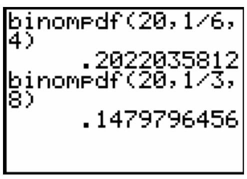
2nd[DISTR] 0:binompdf(en 2nd[DISTR] A:binomcdf(

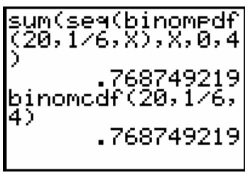
Voor $X \sim B(n,p)$ is :

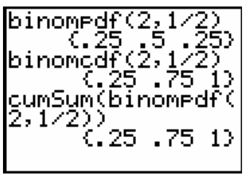
binompdf(n,p,x) = $P(X = x)$, d.i. de kansfunctie. Er geldt dat $P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$.

binomcdf(n,p,x) = $P(X \le x) = \sum_{k=0}^{x} P(X = k)$ d.i. de cumulatieve kansfunctie of de verdelings-functie.

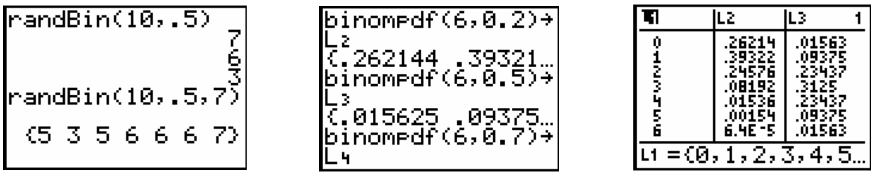
binompdf(n,p) genereert de kansverdeling van X in een lijst (d.i. de lijst van getallen $P(X = x)$ met $x = 0,1,2,\ldots,n$) en binomcdf(n,p) levert de cumulatieve kansverdeling op van X in een lijst (d.i. de lijst van getallen $P(X \le x)$ met $x = 0,1,2,\ldots,n$).



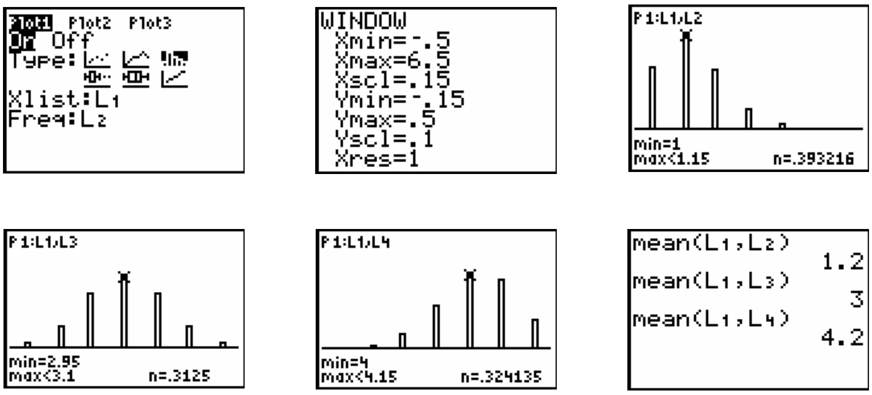




Een lukrake waarde van de binomiaal verdeelde toevalsvariabele X verkrijg je met randBin(n,p) en een lijst van k zo'n waarden met randBin(n,p,k). Zo simuleert randBin(1Ø,Ø.5) het aantal keer kop bij tien keer werpen van een muntstuk.



In wat volgt worden de binomiale kansverdelingen gegenereerd voor $n = 6$ en $p = 0.2, 0.5, 0.7$. Op de volgende wijze kun je ze grafisch voorstellen :



We verkrijgen een symmetrische kansverdeling als $p = 1/2$.
Men kan aantonen dat $P(X = x)$ maximaal is voor $x = \text{int}[(n+1)p]$, d.i. het grootste geheel getal kleiner dan of gelijk aan $(n+1)p$. Deze maximale waarde treedt op in de buurt van $E(X) = np$. Het laatste plaatje bevestigt dat $E(X) = np$.

6 Bronnen

J. Beirlant, G. Dierckx, M. Hubert, *Statistiek en wetenschap*, Acco, Leuven, 2005.

E. Seier, C. Robe, *Ducks and green - an introduction to the ideas of hypotheses testing*, in Teaching Statistics, Vol. 24, Nr. 3, p 82- 85

A. Engel, *Wahrscheinlichkeitsrechnung und Statistik, Band 1*, Ernst Klett Verlag, Stuttgart, 1976.

G. Herweyers, K. Stulens, *Statistiek met een grafisch rekentoestel*, Acco, Leuven, 2000.

D.S. Moore, G.P. McCabe, *Statistiek in de Praktijk, Theorieboek*, Academic Service, Schoonhoven, 2001.

D.S. Moore, G.P. McCabe, *Statistiek in de Praktijk, Opgavenboek*, Academic Service, Schoonhoven, 2001.

Java-applets : http://www.kuleuven.ac.be/ucs/java

Dit cahier behandelt betrouwbaarheidsintervallen en het testen van hypothesen, twee belangrijke begrippen uit de statistische besluitvorming, waarbij men aan de hand van steekproefdata conclusies trekt over de populatie waarvan de data afkomstig zijn.

De voorbeelden worden beperkt tot toepassingen op de normale en de binomiale verdeling: betrouwbaarheidsinterval en test voor het gemiddelde van een normale verdeling met gekende standaardafwijking, betrouwbaarheidsinterval en test voor een proportie.

De begrippen worden helder aangebracht. Simulaties van steekproeftrekkingen illustreren de rol van het toeval bij betrouwbaarheidsintervallen. Een werktekst introduceert de procedure bij hypothesetesten als een natuurlijk denkproces, waarbij de nieuwe terminologie stapsgewijs wordt ingevoerd.

GUIDO HERWEYERS doceert wiskunde en statistiek aan het Departement Industriële Wetenschappen en technologie van de Katholieke Hogeschool Brugge-Oostende en is wetenschappelijk medewerker aan de K.U.Leuven.

Texas Instruments