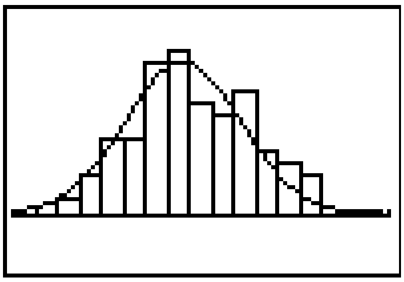
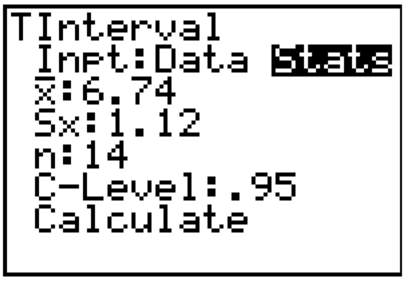
T^3 VLAANDEREN

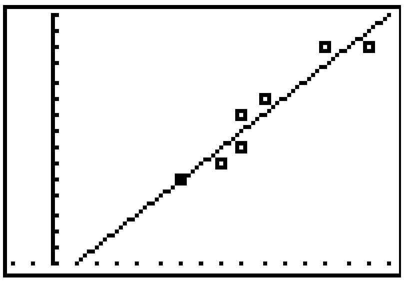
Cahiers T^3 Europe
Vlaanderen nr. 24

De TI-84 in de lessen statistiek

Praktisch gebruik

Philip Bogaert



KATHOLIEKE UNIVERSITEIT LEUVEN

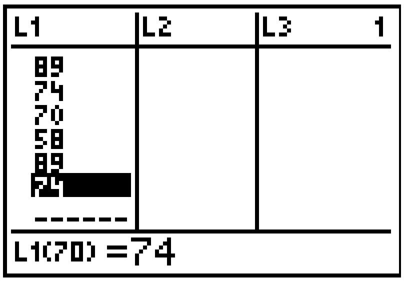
Inhoudstafel

1. Beschrijvende statistiek

1.1. Kippen op het erf

Op het erf van de oma van Robbe lopen heel wat kippen. Robbe besluit om gedurende een week alle eieren te wegen. Hier zie je de resultaten (in gram):

65	54	71	82	81	67	62	75	64	76	87	44	65	80
56	52	59	51	64	68	60	49	48	57	59	68	64	92
49	60	45	62	64	69	64	40	84	54	61	76	64	64
83	67	66	73	78	75	53	56	62	90	78	94	61	76
72	58	65	69	86	50	74	78	89	74	70	58	89	74



(a) Data ingeven

Wis al je datalijsten : 2nd [Mem] 4:ClrAllLists

Geef de gegevens in, in lijst L_1 : Stat Edit 1:Edit

(b) Karakteristieken

Stat Calc 1:1-Var Stats 2nd [L1]

grootte van de steekproef : n = ………………

(steekproef)minimum : min = ………………

(steekproef)maximum : max = ………………

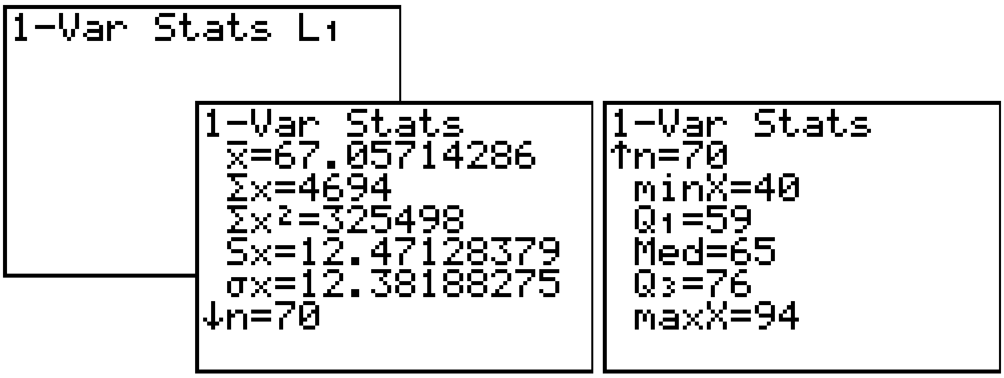
steekproefgemiddelde : $\bar{x}$ = ………………

mediaan : Me = ……………… (= tweede kwartiel)

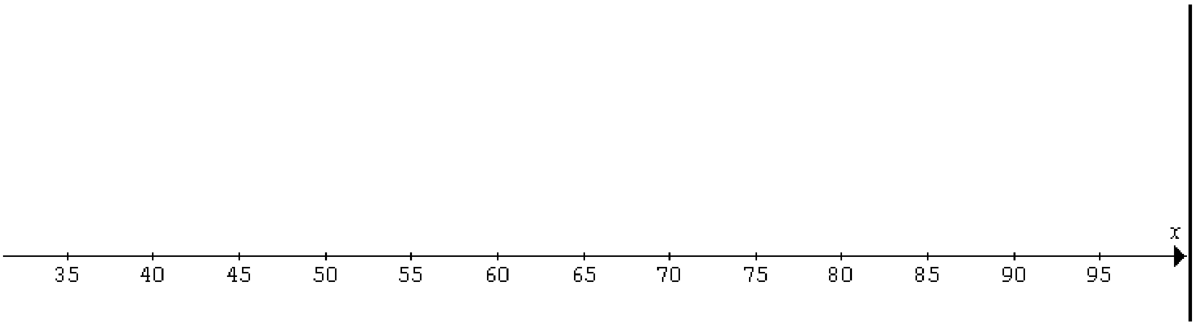
eerste kwartiel : Q_1 = ………………

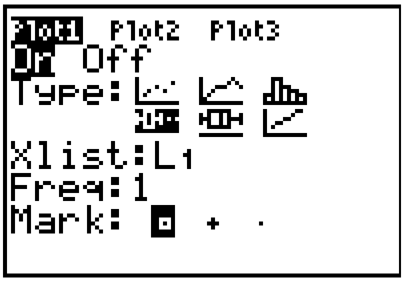
derde kwartiel : Q_3 = ………………

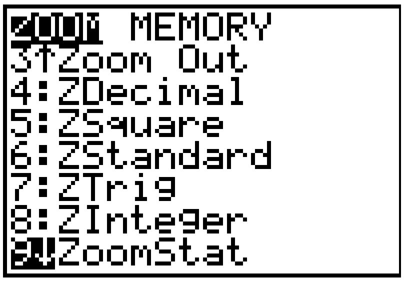
standaardafwijking (v/d steekproef) : s = ………………

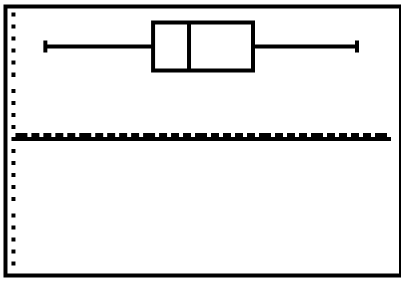


(c) Boxplot







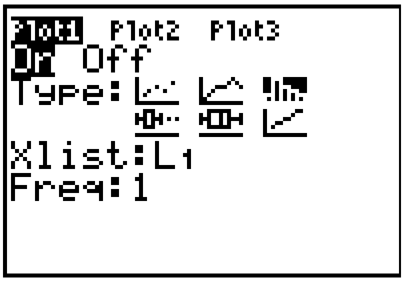


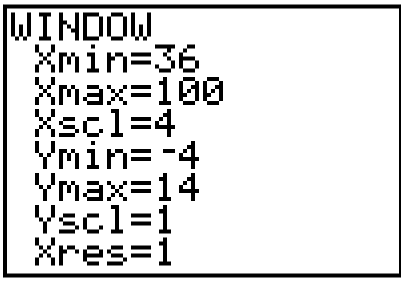
(d) Histogram

Het aantal klassen (minimaal 7, maximaal 25) en de beginklasse mag je als onderzoeker zelf kiezen.

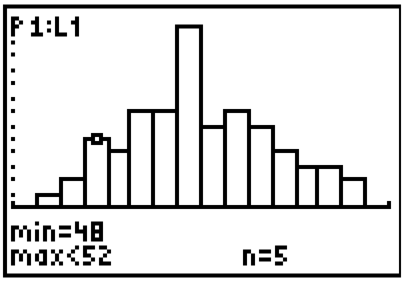
Neem als klassenbreedte 4 en als beginklasse [40,44[.

2nd [Stat Plot] 1:Plot1 Window Graph

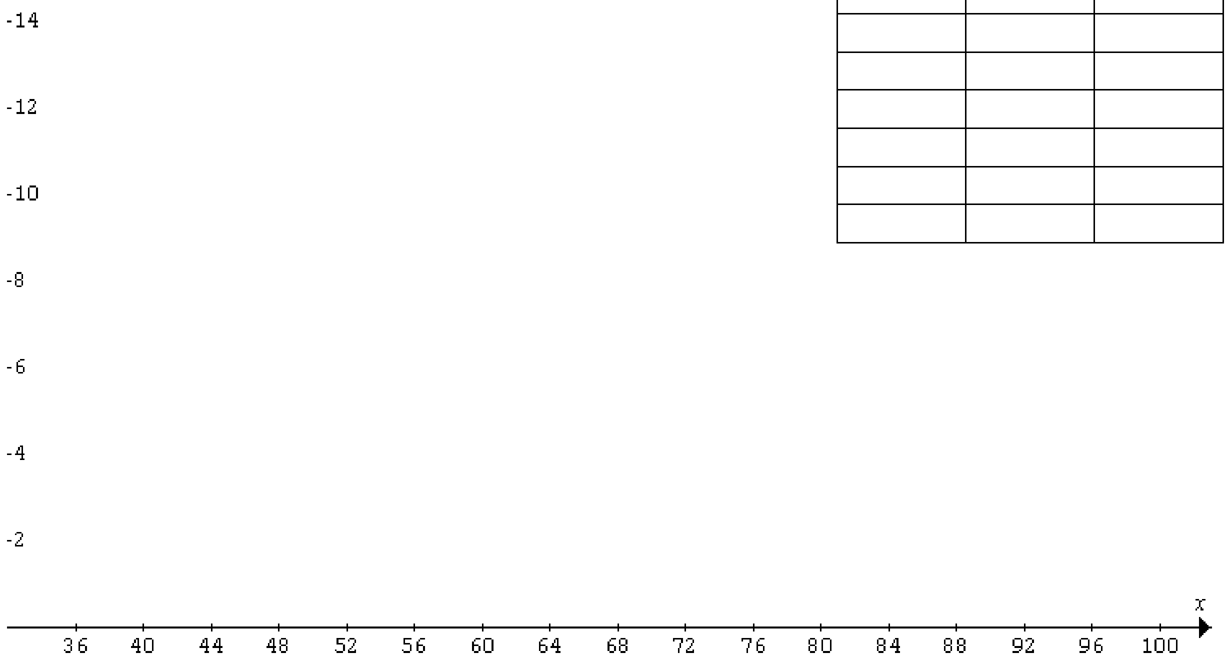




Via de optie Trace kan je op dit histogram de ***absolute frequenties*** van elke klasse aflezen zodat je de frequentietabel kan invullen:



Index	Klasse	a.f.
1	[40,44[	

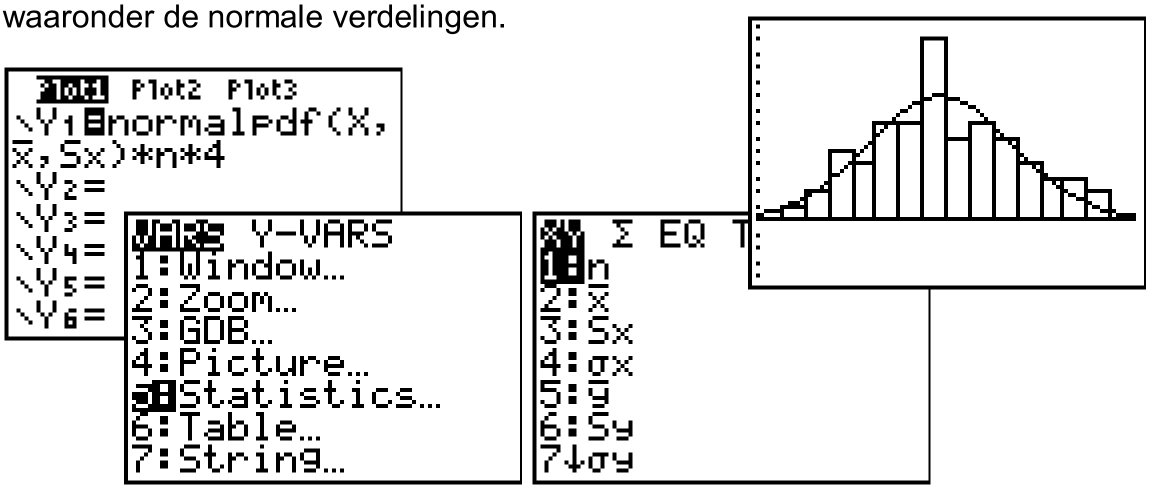


(e) Een normale verdeling (= modelleren van de waargenomen gegevens)

Plot volgende functie normalpdf(X , $\bar{x}$, s) * n * klassenbreedte.

Waarom vermenigvuldigen we met n?
Waarom vermenigvuldigen we met de klassenbreedte?

Onder de toets $\boxed{\text{2nd}}$ [Distr] zitten er een aantal statistische verdelingsfuncties, waaronder de normale verdelingen.



1.2. Reistassen voor Barcelona

Een groep leerlingen uit het laatste jaar van een scholengemeenschap trok dit jaar tijdens de paasvakantie naar Barcelona. Voor de cursus statistiek deed de leerkracht wiskunde een aselecte steekproef van 80 reistassen die hij één voor één woog.

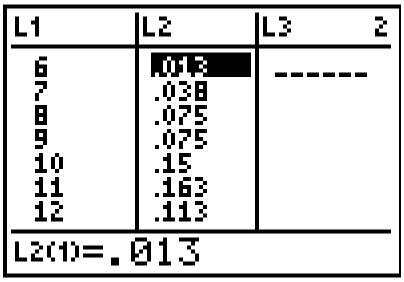
De resultaten (in kg) vind je in volgende tabel:

11,1	8,7	12,4	14,9	14,7	16,0	6,6	11,2	14,2	9,2
10,1	7,7	7,4	9,5	9,9	6,7	10,4	10,8	12,1	10,9
8,3	9,8	8,1	10,9	11,7	11,9	11,0	17,1	7,6	10,0
5,6	15,3	8,7	10,2	13,6	12,9	13,9	13,2	8,5	9,1
11,0	10,9	15,0	11,6	11,4	10,6	16,5	13,9	17,4	10,2
15,8	7,9	13,1	14,0	16,4	11,6	10,5	13,4	10,9	13,6
13,5	12,6	9,5	11,2	12,1	13,1	12,1	9,6	16,3	13,0
12,0	13,8	13,9	10,3	9,6	14,5	8,9	8,1	13,5	13,3

Verwerkt in een frequentietabel geeft dit:

massa (in kg)	klassen-midden	absolute frequentie	relatieve frequentie
[5,5 ; 6,5[	6	1	1,3 %
[6,5 ; 7,5[	7	3	3,8 %
[7,5 ; 8,5[	8	6	7,5 %
[8,5 ; 9,5[	9	6	7,5 %
[9,5 ; 10,5[	10	12	15,0 %
[10,5 ; 11,5[	11	13	16,3 %
[11,5 ; 12,5[	12	9	11,3 %
[12,5 ; 13,5[	13	8	10,0 %
[13,5 ; 14,5[	14	10	12,5 %
[14,5 ; 15,5[	15	5	6,3 %
[15,5 ; 16,5[	16	4	5,0 %
[16,5 ; 17,5[	17	3	3,8 %

(a) Frequentietabel ingeven



Stat Edit 1:Edit

L_1: klassenmiddens
L_2: relatieve frequenties

(b) Karakteristieken (op basis van de frequentietabel)

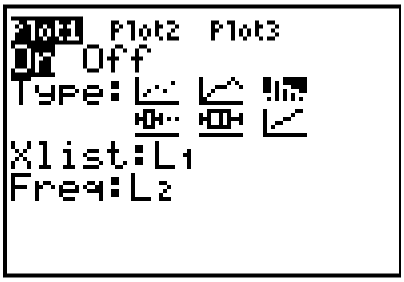
Stat Calc 1:1-Var Stats 2nd [L1] 2nd [L2]

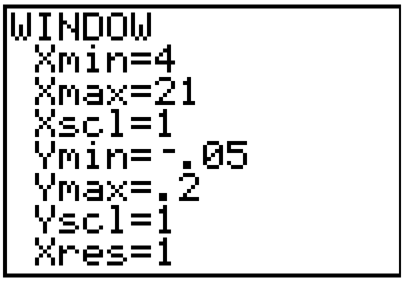
steekproefgemiddelde : $\bar{x}=$ ………………

standaardafwijking s = ………………

(c) Histogram

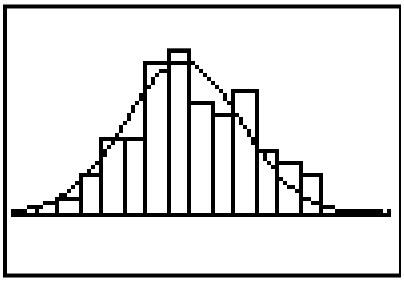
2nd [Stat Plot] 1:Plot1 Window Graph





(d) Modelleren m.b.v. een normale verdeling

\Y1= normalpdf(X , $\bar{x}$, s_x)



(e) Intervallen met centrum het gemiddelde en straal een aantal keer de standaardafwijking

$\left[\bar{x}-s,\bar{x}+s\right]=$ ……………………

hoeveel reistassen zitten er in dit interval ? ………………

hoeveel % is dat ? ………………

$\left[\overline{x}-2s, \overline{x}+2s\right] =$ ……………………

hoeveel reistassen zitten er in dit interval ? ………………

hoeveel % is dat ? ………………

$\left[\overline{x}-3s, \overline{x}+3s\right] =$ ……………………

hoeveel reistassen zitten er in dit interval ? ………………

hoeveel % is dat ? ………………

$\boxed{\text{2nd}}$ [Distr] Draw 1:ShadeNorm($\overline{x}-s, \overline{x}+s, \overline{x}, s$) = ………………

of

$\boxed{\text{2nd}}$ [Distr] 2:normalcdf($\overline{x}-s, \overline{x}+s, \overline{x}, s$) = ………………

Analoog :

$\boxed{\text{2nd}}$ [Distr] 2:normalcdf($\overline{x}-2s, \overline{x}+2s, \overline{x}, s$) = ………………

$\boxed{\text{2nd}}$ [Distr] 2:normalcdf($\overline{x}-3s, \overline{x}+3s, \overline{x}, s$) = ………………

En verder :

Hoeveel % van de reistassen weegt lichter dan 9 kg?

$\boxed{\text{2nd}}$ [Distr] 2:normalcdf($-1E99$, 9 , $\overline{x}$, s) = ………………

Hoeveel % van de reistassen weegt zwaarder dan 10 kg?

$\boxed{\text{2nd}}$ [Distr] 2:normalcdf(……………………) = ………………

2. Kansverdelingen: de normale verdeling

2.1. Pakjes suiker

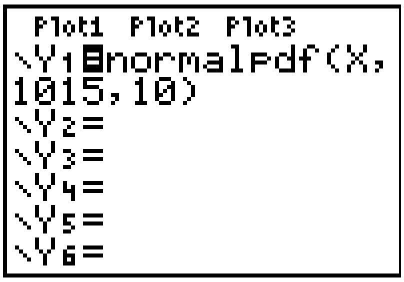
(a) probleemstelling

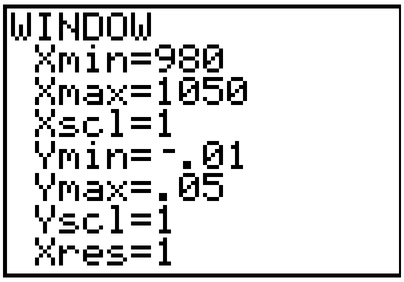
Een machine vult pakken met suiker. De massa suiker die door de machine afgeleverd wordt, is normaal verdeeld met μ = 1015 gram en σ = 10 gram.

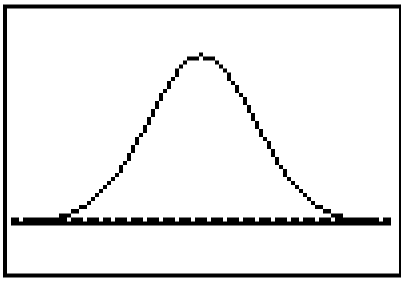
- Hoeveel % van de afgeleverde pakken bevat minder dan 1 kg?
- Boven welke gewichtsgrens ligt 10 % van de pakken suiker?
- Stel dat het mogelijk is om de afstelling van het vulapparaat (d.w.z. de gemiddelde hoeveelheid μ) te veranderen zonder dat de standaarddeviatie verandert. Hoe moet het gemiddelde gekozen worden opdat slecht 1% van de pakken suiker een massa heeft beneden de 1 kg.

(b) oplossing

Het vulgewicht van de pakjes suiker kunnen we grafisch voorstellen door de normale verdeling N(μ = 1015 ; σ = 10).



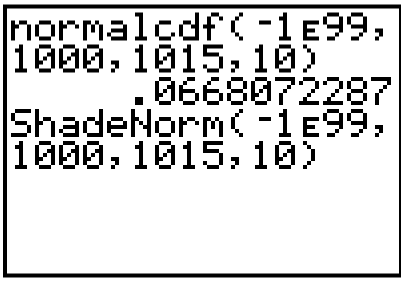


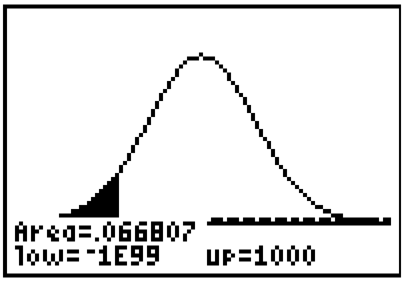


Hoeveel % van de afgeleverde pakken bevat minder dan 1 kg?

P(X < 1000) = ?

antwoord : 6,68 %

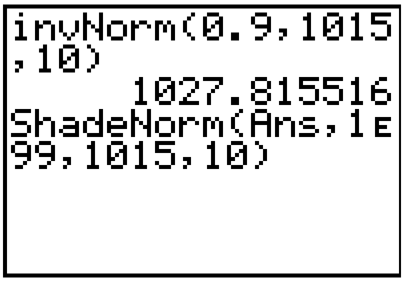


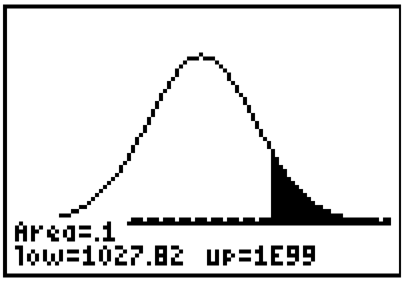


Boven welke gewichtsgrens ligt 10 % van de pakken suiker?

P(X > ?) = 10 % = 0,1

antwoord : 1027,8 gram



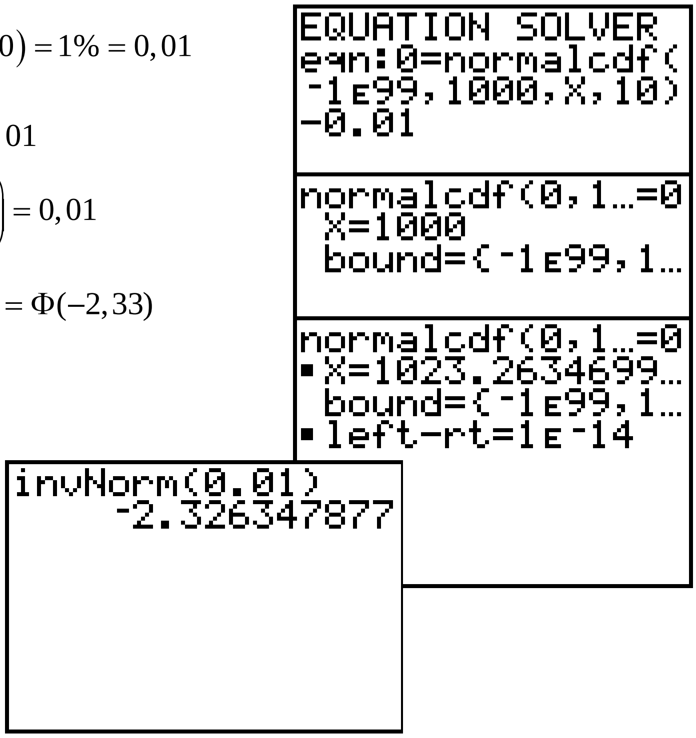


Stel dat het mogelijk is om de afstelling van het vulapparaat (d.w.z. de gemiddelde hoeveelheid μ) te veranderen zonder dat de standaarddeviatie verandert. Hoe moet het gemiddelde gekozen worden opdat slecht 1% van de pakken suiker een massa heeft beneden de 1 kg.

Bepaal μ' zodat $P\left(X \leq 1000\right) = 1\% = 0,01$

$$P\left(X \leq 1000\right) = 0,01$$
$$\Leftrightarrow P\left(Z \leq \frac{1000-\mu'}{10}\right) = 0,01$$
$$\Leftrightarrow P\left(Z \leq \frac{1000-\mu'}{10}\right) = \Phi(-2,33)$$
$$\Leftrightarrow \frac{1000-\mu'}{10} = -2,33$$
$$\Leftrightarrow \mu' = 1023,3$$

antwoord : 1023,3 gram



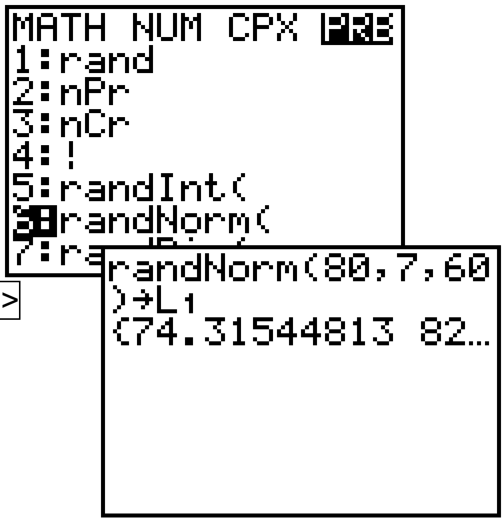
2.2. Genereren van een steekproef

- Genereer met je TI-84 een steekproef van 60 waarnemingen uit een normaal verdeelde populatie X ~N(μ = 80, σ = 7).
- Bereken vervolgens het gemiddelde en de standaardafwijking van deze steekproef.
- Ga via een normal probability plot na dat deze gegevens een steekproef zijn uit een normaal verdeelde populatie.
- Hoeveel % van de gegevens hebben theoretisch een waarde groter dan 87?
- Hoeveel % van de gegevens hebben op basis van de getrokken steekproef een waarde groter dan 87?

oplossing

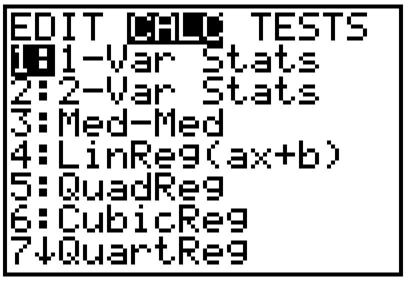
Een steekproef van 60 waarnemingen genereren met de TI-84 uit een normaal verdeelde populatie X ~N(μ = 80, σ = 7).

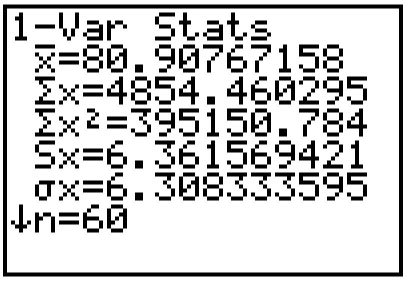
MATH PRB 6:randNorm(80, 7 , 60) STO>
2nd [L1] Enter



Bereken vervolgens het gemiddelde en de standaardafwijking van deze steekproef.

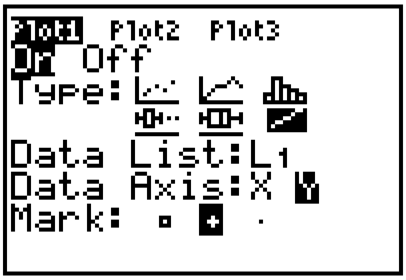
STAT CALC 1:1-Var Stats 2nd [L1] Enter

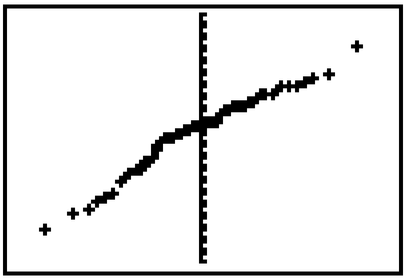




Ga via een normal probabilty plot na dat deze gegevens een steekproef zijn uit een normaal verdeelde populatie.

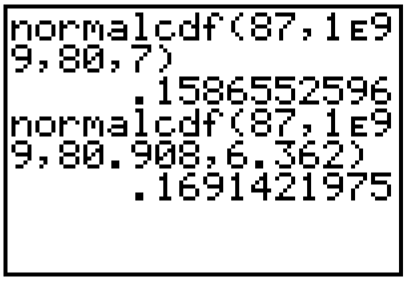
2nd [StatPlot] …
selecteer zesde grafiek type …
ZOOM 9:ZoomStat





Wanneer we de normal probabilty plot bekijken, zien we dat de punten ongeveer op een rechte liggen. We mogen dus terecht veronderstellen dat deze steekproef afkomstig is uit een normaal verdeelde populatie.

Hoeveel % van de gegevens hebben theoretisch een waarde groter dan 87?



Antwoord : 15,87 %

Hoeveel % van de gegevens hebben op basis van de getrokken steekproef een waarde groter dan 87?

Antwoord : 16,91 %

3. Kansverdelingen: de binomiale verdeling

3.1. Taartjes kopen

(a) probleemstelling

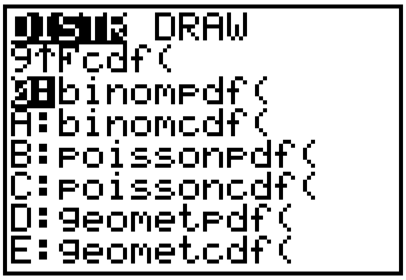
Een bakker verkoopt taartjes waarbij bij 1 op de 5 gebakjes een koffieboon in het gebakje zit.

- Als Nikolaas 24 taartjes bij deze bakker koopt, wat is de kans dat er bij 4 taartjes of meer een koffieboon inzit?
- Hoeveel taartjes moet Nikolaas kopen om meer dan 90% kans te hebben dat er bij minstens 4 taartjes een koffieboon zit?

(b) oplossing

We hebben hier duidelijk te maken met een binomiale verdeling met parameters n = 24 en p = $\frac{1}{5}$; de kans op i successen is gelijk aan:

$$P(X=i) = C_{24}^{i} \cdot \left(\frac{1}{5}\right)^{i} \cdot \left(\frac{4}{5}\right)^{24-i}$$

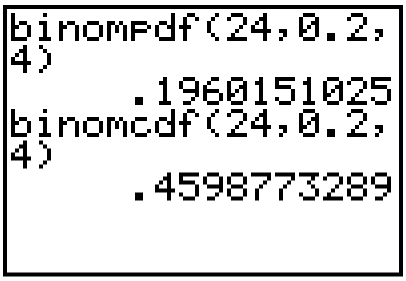


De kans op i keer succes bij een binomiale verdeling met parameters n en p bereken je met je TI-84 als volgt :

2nd [Distr] 0:binompdf(n , p , i) Enter

M.a.w. de kans op juist 4 taartjes met een koffieboon krijg je via binompdf(24, 0.2 , 4)

P(X = 4) = binompdf(24 , 0.2 , 4)



De kans op i keer succes of minder bij een binomiale verdeling met parameters n en p bereken je met je TI-84 als volgt :

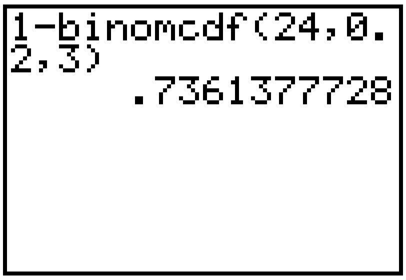
2nd [Distr] A:binomcdf(n , p , i) Enter

M.a.w. de kans op 4 taartjes of minder met een koffieboon krijg je via binomcdf(24, 0.2 ,4)

P(X $\leq$ 4) = binomcdf(24 , 0.2 , 4)

De kans op 4 taartjes op meer bereken je dan via :

$$P(X > 4) = 1 - P(X \leq 3) = 1 - \text{binomcdf}(24 , 0.2 , 3)$$



Antwoord : de kans dat er bij 4 taartjes of meer een koffieboon inzit is 73,6 %.

Als Nikolaas n taartjes koopt, is de kans dat er bij 4 taartjes of meer een koffieboon inzit gelijk aan :

$$P(X > 4) = 1 - P(X \leq 3) = 1 - \text{binomcdf}(n , 0.2 , 3)$$

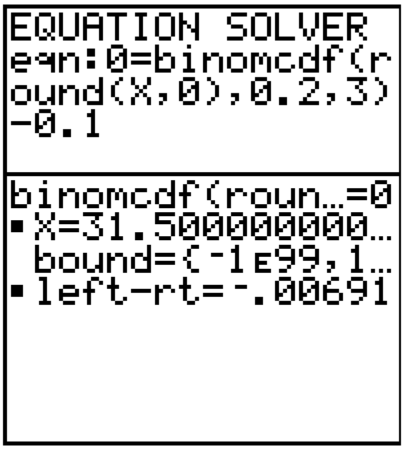
Wil deze kans groter dan of gelijk zijn aan 90%, dan moeten we n halen uit de ongelijkheid :

$$1 - \text{binomcdf}(n , 0.2 , 3) \geq 0{,}9$$

of nog :

$$\text{binomcdf}(n , 0.2 , 3) \leq 0{,}1$$

dit probleem los je met de TI-84 op via de solver.

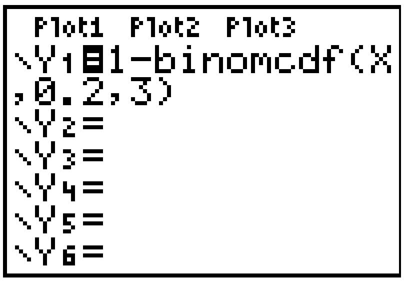


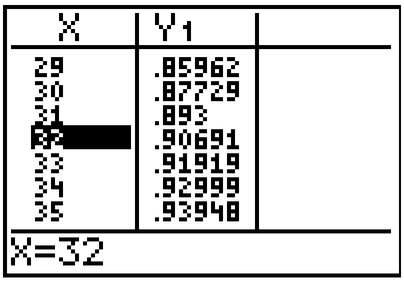
Opgelet! Omdat de binomiale verdeling werkt met positieve gehele getallen moet je bij de solver de functie round gebruiken.

Antwoord: Nikolaas moet minstens 32 taartjes kopen om meer dan 90% kans te hebben dat er bij minstens 4 taartjes een boon zit.

Alternatieve methode:

Definieer 1 – binomcdf(n , 0.2 , 3) als functie en ga via 2nd [Table] na vanaf welke (gehele) waarde van X deze functie een waarde aanneemt groter dan 0,9.





Ook hier zien we dat Nikolaas minstens 32 taartjes moet kopen om meer dan 90% kans te hebben dat er bij minstens 4 taartjes een boon zit.

3.2. genereren van een steekproef

- Genereer met je TI-84 een steekproef van 80 waarnemingen uit een binomiaal verdeelde populatie X ~Bi(n = 7, p = 0,4).
- Bereken vervolgens het gemiddelde en de standaardafwijking van deze steekproef.
- Hoeveel % van de gegevens hebben theoretisch een waarde kleiner of gelijk aan 2?
- Hoeveel % van de gegevens hebben op basis van de getrokken steekproef een waarde kleiner of gelijk aan 2?

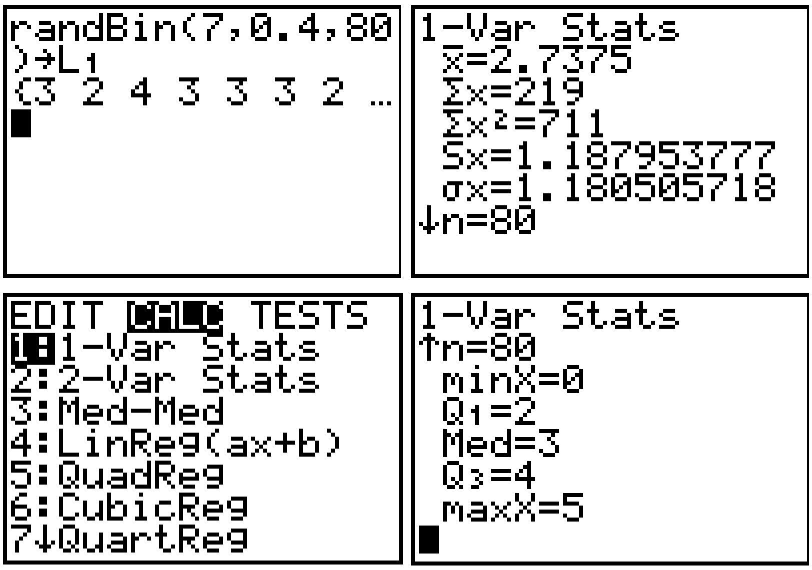
oplossing

Een steekproef van 80 waarnemingen genereren met de TI-84 uit een binomiaal verdeelde populatie X ~Bi(n = 7, p = 0.4).

MATH PRB 7:randBin(7, 0.4 , 80) STO> 2nd [L1] Enter

Bereken vervolgens het gemiddelde en de standaardafwijking van deze steekproef.

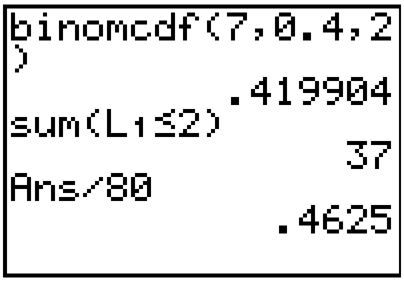
STAT CALC 1:1-Var Stats 2nd [L1] Enter



Theoretisch: $\mu = np = 7.0,4 = 2,8$; $\sigma = \sqrt{np(1-p)} = 1,296$

Hoeveel % van de gegevens hebben theoretisch een waarde kleiner of gelijk aan 2?

Antwoord : 42 %



Hoeveel % van de gegevens hebben op basis van de getrokken steekproef een waarde kleiner of gelijk aan 2?

Antwoord : 46,25 %

4. Enkele andere discrete verdelingen

4.1. De geometrische verdeling

Als we een reeks onafhankelijke Bernouilli-pogingen doen met succeskans p, kunnen we een vast aantal (n) beschouwen, zodat we maar moeten afwachten hoe vaak we succes zullen hebben. Dit leidt tot de binomiale verdeling.
Gaan we echter net zolang door tot we succes hebben, dan moeten we maar afwachten hoeveel experimenten we moeten doen. Dat aantal N is een stochastische variabele met als verdeling de geometrische verdeling.

$$P(N=k)=(1-p)^{k-1}.p$$

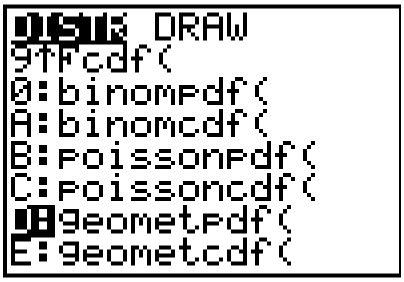
$$\mu=E[X]=\frac{1}{p} \qquad \sigma^2=\frac{1-p}{p^2}$$

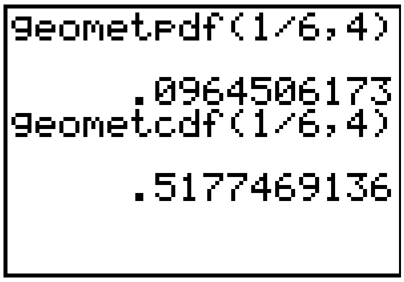
voorbeeld

Een dobbelsteen wordt net zolang geworpen tot dat er een eerste maal “1” wordt geworpen.

Hoe groot is de kans dat dit bij de 4de worp gebeurt ?

$P(N=4)=$ $\boxed{\text{2nd}}$ [Distr] D:geometpdf(p , k)   antwoord: 9,6 %





Hoe groot is de kans dat dit bij de eerste vier worpen gebeurt ?

$P(N\leq 4)=$ $\boxed{\text{2nd}}$ [Distr] E:geometcdf(p , k)   antwoord: 51,8 %

4.2. De negatief-binomiale verdeling

Als we een reeks onafhankelijke Bernouilli-pogingen doen met succeskans p, kunnen we een vast aantal (n) beschouwen, zodat we maar moeten afwachten hoe vaak we succes zullen hebben. Dit leidt tot de binomiale verdeling.
Gaan we echter net zolang door tot we voor de m-de keer succes hebben, dan moeten we maar afwachten hoeveel experimenten we moeten doen. Dat aantal N is een stochastische variabele met als verdeling de negatief-binomiale verdeling.

De geometrische verdeling is een speciaal geval van de negatief-binomiale verdeling met parameter m = 1.

$$P(N=k)=\binom{k-1}{m-1}p^m.(1-p)^{k-m}$$

$$\mu=E[X]=\frac{m}{p} \qquad \sigma^2=m.\frac{1-p}{p^2}$$

voorbeeld

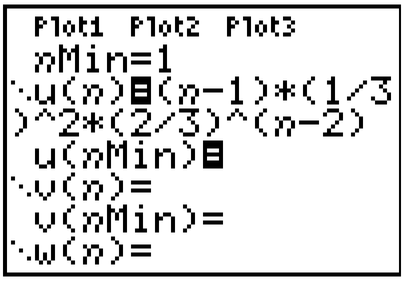
Op de Kermis heb je bij een bepaald spelletje (inleg 2 euro) één kans op drie om een plushen welpje te winnen. Je wil nu zolang blijven spelen tot dat je twee welpjes hebt gewonnen voor je tweeling Stefan en Klaasje.

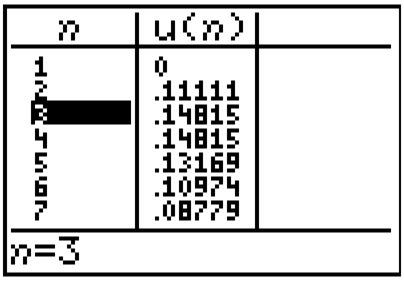
- Wat is de kans dat je bij het derde spelletje je tweede plushen welpje hebt?
- Wat is de kans dat je met een briefje van 10 euro toekomt om de twee plushen welpjes te bekomen?

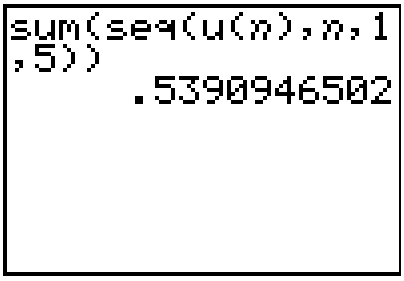
Discrete verdelingen die niet standaard voorgeprogrammeerd zijn in de TI-84 kan je altijd definiëren aan de hand van een rij (sequentie).

In ons geval is m = 2 en p = 1/3, dus:

$$P(N=k)=\binom{k-1}{1}\left(\frac{1}{3}\right)^2.\left(1-\frac{1}{3}\right)^{k-2}=(k-1)\left(\frac{1}{3}\right)^2.\left(\frac{2}{3}\right)^{k-2}$$







De kans om bij het derde spelletje de tweede plushen welp te winnen, bedraagt 14,8 %

De kans om bij het tweede, het derde, het vierde of het vijfde spelletje de tweede plushen welp te winnen, bedraagt 53,9 %

4.3. De hypergeometrische verdeling

Als we een serie trekkingen doen uit een eindige populatie dan hebben we te maken met een hypergeometrische verdeling. De populatie N is verdeeld in M elementen die een bepaald kenmerk vertonen en N-M, de rest, die dat kenmerk niet bezitten.
Uit de populatie wordt een steekproef van n elementen genomen, zonder teruglegging. Vervolgens zijn we geïnteresseerd in het aantal elementen k dat uit de deelverzameling met M elementen afkomstig is.

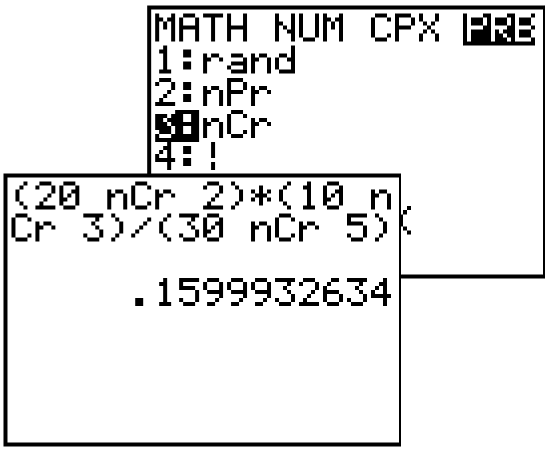
$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\mu = E[X] = n\frac{M}{N} \qquad \sigma^2 = \frac{n(N-n)(N-M)M}{N^2(N-1)}$$

voorbeeld

In een speelgoedmand liggen 30 pakjes. In 20 pakjes zit een pluchen hondje en in de andere 10 een bal. Vijf kinderen mogen nu lukraak een pakje trekken.
Wat is de kans dat er twee pluchen hondjes worden getrokken.

$$P(X = 2) = \frac{\binom{20}{2}\binom{10}{3}}{\binom{30}{5}} = 0{,}16 = 16\%$$



4.4. De Poisson verdeling

De Poisson verdeling is net zoals de binimiale verdeling een discrete kansverdeling. Het grote verschil met de binomiale verdeling is dat er bij de Poisson verdeling geen sprake is van een vast aantal pogingen. Bij de Poisson verdeling gaat het om het aantal keren succes in een bepaald tijdsinterval of specifieke plaats. Een voorbeeld hiervan is het aantal, dagelijkse ongelukken op een specifiek stuk snelweg. Het Poisson experiment heeft de volgende eigenschappen:

- het aantal keren succes dat plaatsvindt in elk willekeurig interval is onafhankelijk van het aantal keren succes in elk ander willekeurig interval
- de kans op succes in een interval is even groot in intervallen van dezelfde grootte
- de kans op succes is evenredig aan de grootte van het interval
- de kans op meer dan 1 succes in een interval ligt dichter bij 0 naarmate het interval kleiner wordt

De Poissonverdeling is genoemd naar Siméon Poisson die deze kansverdeling ontdekte en samen met zijn statistische theorie in 1838 publiceerde in zijn werk *“Recherches sur la probabilité des jugements en matières criminelles et matière civile.”*

De Poissonverdeling komt voor in relatie met zogenaamde Poissonprocessen. Hij is van toepassing op diverse fenomenen die een discrete natuur hebben (dat wil zeggen dat ze 0, 1, 2, 3... keer voorkomen gedurende een gegeven tijdsinterval of in een bepaald gebied), wanneer de kans op het evenement constant is in de tijd of in de ruimte. Voorbeelden zijn:

- het aantal binnen een bepaalde tijd vervallen radioactieve atoomkernen in een stuk radioactief materiaal
- het aantal auto's die gedurende een zekere tijd een bepaald punt van een weg passeren (maar niet als het zo druk is dat er twee of meer tegelijk voorbijkomen!)
- het aantal typefouten dat een secretaresse maakt bij het typen van een enkele pagina
- het aantal telefoontjes dat iemand op een dag krijgt.
- het aantal keren in een minuut dat een webserver wordt benaderd.
- het aantal dode dieren op een kilometer weg
- het aantal mutaties in een stuk DNA van gegeven lengte na een bepaalde stralingsdosis.
- het aantal naaldbomen op een hectare gemengd bos
- het aantal sterren in een gegeven ruimte
- het aantal op een vierkante mijl van Londen gevallen bommen gedurende een Duitse luchtaanval in het begin van de Tweede Wereldoorlog
- het aantal olietankers dat op een dag een bepaalde haven binnenvaren
- …

Als algemene regel geldt dat een Poisson stochastische variabele staat voor het aantal keren dat een relatief zeldzame gebeurtenis die willekeurig en onafhankelijk plaatsvindt. Bijvoorbeeld het aantal mensen dat bij een restaurant binnenkomen is geen Poisson verdeling, want mensen komen meestal in groepen naar een restaurant en dat is in strijd met de eerste eigenschap van het Poisson experiment, te weten het ontbreken van onafhankelijkheid.

voorbeeld 1

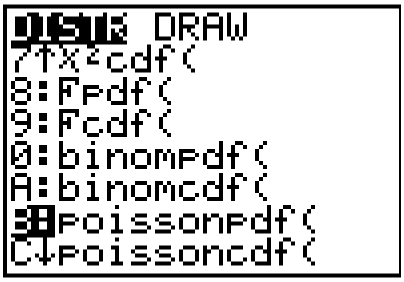
Het aantal telefoongesprekken dat binnenkomt op een centrale is Poisson-verdeeld met een gemiddelde van 2 per minuut.

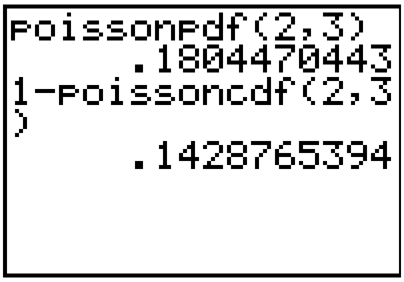
$$P(X = k) = \frac{e^{-\lambda}.\lambda^k}{k!} = \frac{e^{-2}.2^k}{k!}$$

$$\mu = E[X] = \lambda \qquad \sigma^2 = \lambda$$

Hoe groot is de kans dat er juist 3 gesprekken in een minuut binnenkomen?

$P(X = 3) =$ 2nd [Distr] B:poissonpdf(λ , k)   antwoord: 18,0 %





Hoe groot is de kans dat er meer dan 3 gesprekken in een minuut binnenkomen?

$P(X > 3) = 1 - P(X \leq 3)$

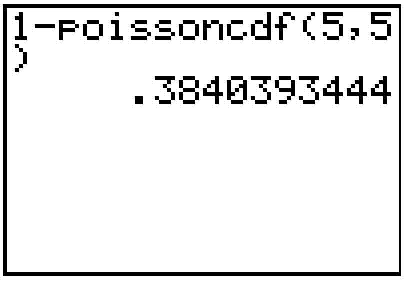
1 – 2nd [Distr] C:poissoncdf(λ , k)   antwoord: 14,3 %

voorbeeld 2

De bezoekers van een vogelreservaat kunnen aan de ingang van het park een verrekijker huren. Een bezoek duurt anderhalf uur. Er zijn gemiddeld 7,5 bezoekers per uur die een verrekijker huren. Dit aantal is Poisson verdeeld. Het park beschikt echter slechts over 5 verrekijkers.

- Het vogelreservaat opent om 10 uur ’s ochtends. Wat is de kans dat een bezoeker die om 10h40 toekomt, geen verrekijker meer kan huren.
- Over hoeveel verrekijkers moet het park minstens beschikken om ten minste 90% zekerheid te hebben dat alle bezoekers die voor 10h20 toekomen nog een verrekijker kunnen huren?

oplossing



$\lambda = 7,5/\text{uur} = 7,5/60\,\text{min.} = 5/40\,\text{min.}$

$P(X > 5) = 1 - P(X \leq 5)$

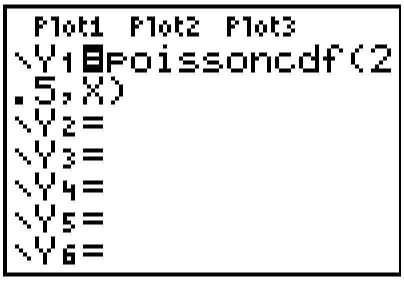
1 – $\boxed{\text{2nd}}$ [Distr] C:poissoncdf(5 , 5)

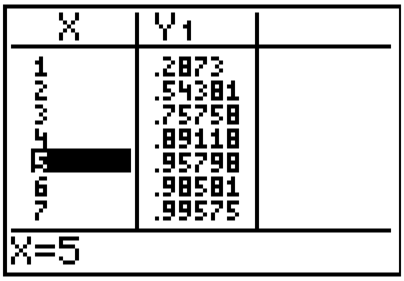
antwoord: 38,4 %

$\lambda = 7,5/\text{uur} = 7,5/60\,\text{min.} = 2,5/20\,\text{min.}$

$P(X \leq k) \geq 0,9$

definieer poissoncdf(2,5 ; X) als een functie en lees de oplossing af via $\boxed{\text{2nd}}$ [table]:





antwoord: minstens 5 verrekijkers.

5. Simuleren van kansexperimenten

5.1. Kruis of munt

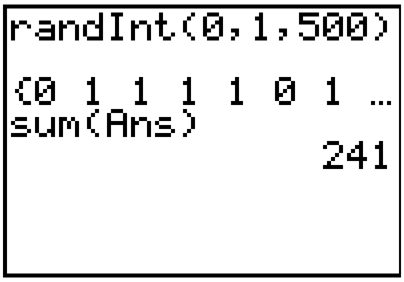
We willen een muntstuk 500 keer opgooien en tellen hoeveel keer we kruis hebben. We gebruiken de volgende code: kruis = 1 en munt = 0.

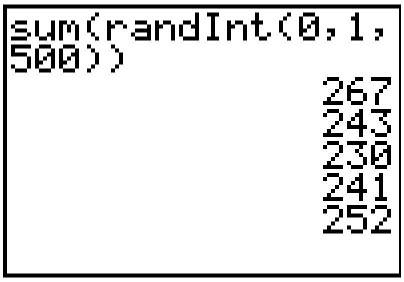
Als je nu wilt tellen hoeveel keer kruis gegooid wordt, moet je alleen maar de som van alle door de rekenmachine gegenereerde getallen maken.

Math Prb 5: randInt(0,1,500) Enter

2nd [List] Math 5: sum(2nd [Ans]) Enter

De relatieve frequentie om kruis te gooien is in dit geval $\frac{241}{500} = 0,482$. Dit is een goede benadering voor de kans om kruis te gooien, namelijk $\frac{1}{2}$.





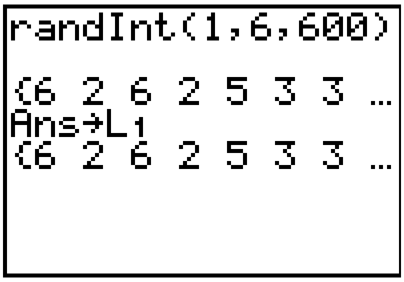
Je kan de simulatie ook in één stap uitvoeren: sum(randInt(0,1,500)). Je kan de simulatie meerdere keren uitvoeren door telkens op Enter te drukken.

5.2. Gooien met 1 dobbelsteen

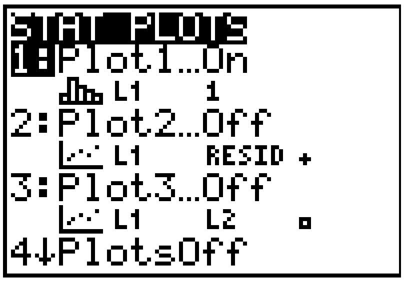
We kunnen de rekenmachine laten tellen hoeveel zessen er gegooid worden met een dobbelsteen (600 worpen). Je kan de resultaten van de simulatie opslaan in lijst 1 als volgt:

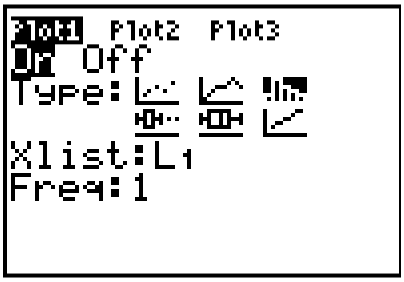
Math Prb 5: randInt(1,6,600)

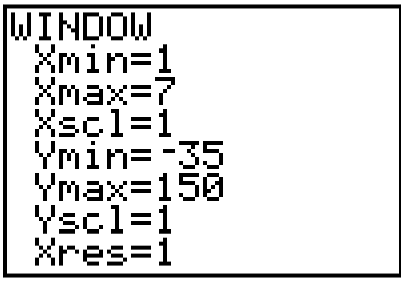
Sto> 2nd L1 Enter



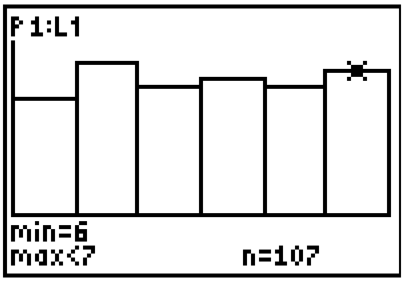
Via Stat Plot kunnen we nu een histogram maken van de gegevens uit lijst 1 (let op de Window instellingen !!).







Elk balkje van het histogram komt overeen met een bepaalde uitkomst: 1, 2, 3, 4, 5 of 6.
Door middel van de “trace”-modus kunnen de absolute frequenties van elke uitkomst afgelezen worden.

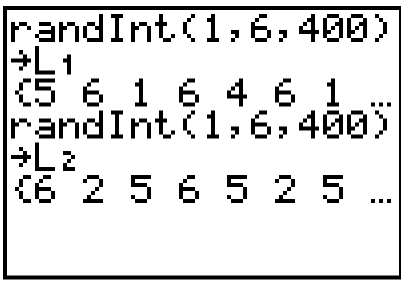


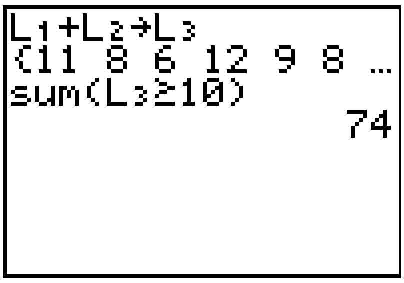
Zo blijkt in dit voorbeeld dat de uitkomst “6” 107 keer voorkomt in de rij van 600 getallen. De experimentele kans op een “6” is dus $\frac{107}{600} = 17,8\%$, wat reeds een goede benadering is van de theoretische kans van $\frac{1}{6}$ of 16,666… %.

5.3. Gooien met 2 dobbelstenen

Een gokker wil de kans kennen dat de som van de ogen van twee geworpen dobbelstenen groter dan of gelijk aan 10 is, door het experiment “gooien van twee dobbelstenen” 400 maal te herhalen.

Het aantal ogen op de eerste dobbelsteen slaan we op als lijst L1, het aantal ogen op de tweede dobbelsteen als lijst L2. Als we nu beide lijsten optellen, dan krijgen we een lijst L3 die bij elke worp de som van het aantal ogen geeft van beide dobbelstenen.





Met de gegevens hierboven vinden we:

P(som van de ogen groter of gelijk aan 10) = $\frac{74}{400} = 18,5\%$

Dit is een experimentele kans. Je kan berekenen dat de theoretische, exacte kans op deze gebeurtenis $\frac{1}{6}$ is. De berekende experimentele kans is dus een goede benadering.

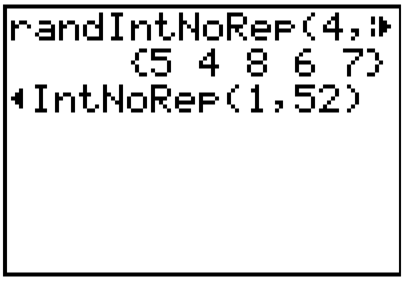
5.4. Lottogetallen genereren

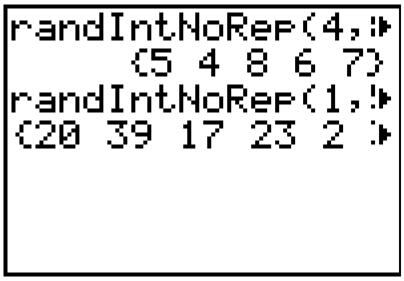
Het loont de moeite om af en toe je TI-84 te upgraden met het nieuwste OS. Als je nog ergens een TI-83 hebt, zal je merken dat in het menu 2nd [Distr] de optie invT() ontbreekt. Nochtans is deze functie onontbeerlijk bij het toetsen van hypothesen.

Vanaf het OS 2.53 MP (Math Print) zit er onder het menu Math Prb de optie randIntNoRep(). Hiermee genereer je permutaties van getallen.

Veronderstel dat je de getallen 4 tot 8 (beide getallen inclusief) wil permuteren, kan dit via randIntNoRep(4,8).

Als je de 52 kaarten van een kaartspel van een nummer zou voorzien, kan je het kaartspel dooreenschudden via randIntNoRep(1,52).



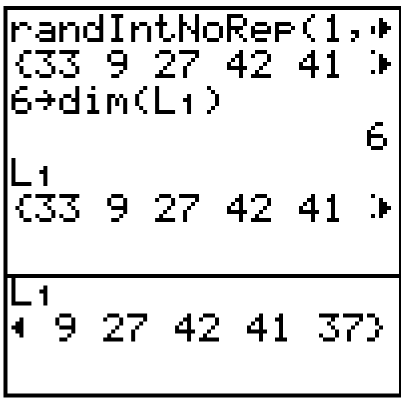


Hoe genereer je nu 6 lottogetallen?

- Schud de 42 balletjes door elkaar en bewaar deze in een lijst.
- Neem de eerste 6 balletjes uit deze lijst.
- Toon deze 6 balletjes.

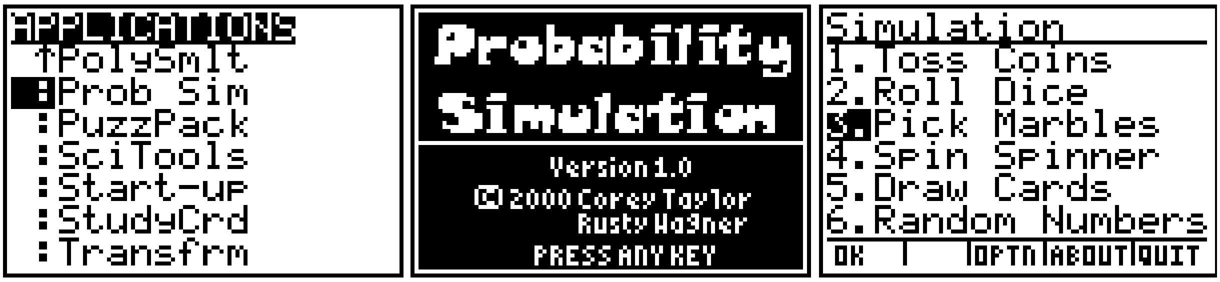
Concreet:

- randIntNoRep(1,42) → L1
- 6 → dim(L1)
- L1



5.5. Knikkers trekken uit een zak

Activeer op een TI-84 via APPS de applicatie Prob Sim waarmee je kansexperimenten kan simuleren. Kies hier de optie 3: Pick Marbles.



Via ‘SET’ (F3) kan je nu de instellingen aanpassen.

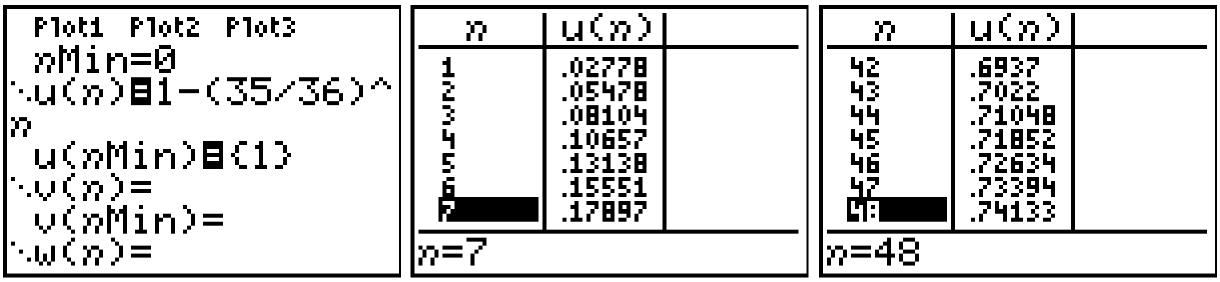
5.6. De kans op minstens éénmaal succes bij n pogingen

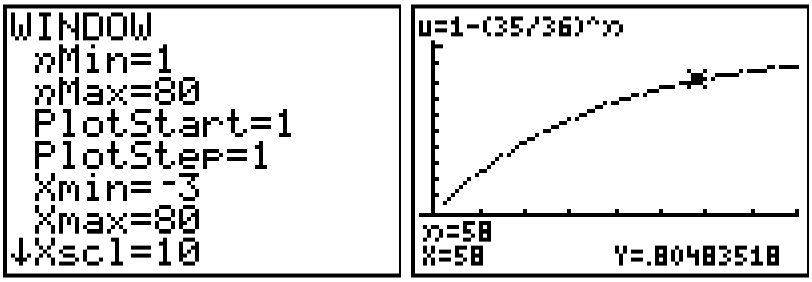
Berekenen de kans op minstens eenmaal twee zessen als je n keer met twee dobbelstenen gooit.

De kans op minstens één keer succes bij n pogingen is gelijk aan:

$1-\left(\frac{35}{36}\right)^n$ (waarom ?)

Via het gebruik van rijen (MODE SEQ) berekenen (2nd TABLE) en visualiseren (GRAPH) we de kans op succes bij 1,2, 3, … pogingen.





6. Betrouwbaarheidsintervallen

6.1. Betrouwbaarheidsintervallen voor μ met bekende σ

probleemstelling 1

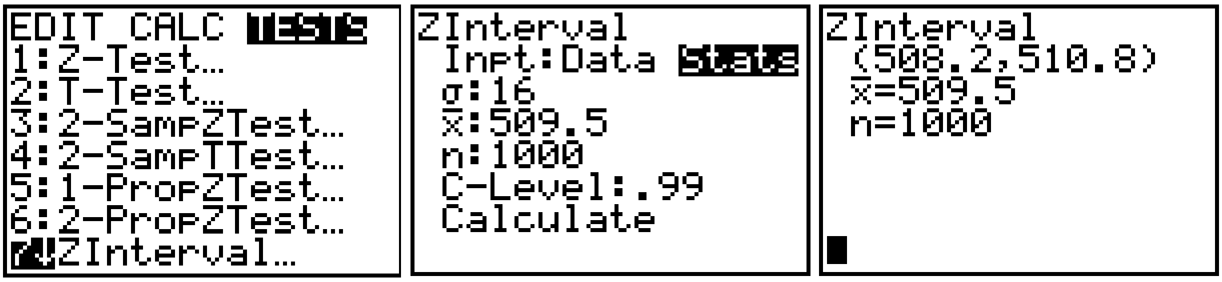
Een machine vult pakjes koffie. De inhoud van de pakjes zijn normaal verdeeld met standaardafwijking σ = 16 gram. De kwaliteitscontroleur neemt een steekproef van 1000 pakjes en zet deze op een weegschaal. Deze hebben een totale massa van 509,5 kg. Geef een 99% betrouwbaarheidsinterval voor de gemiddelde massa van één pakje koffie.

oplossing

formule:

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Stat – Test – 7: ZInterval



antwoord: [508,2 gram ; 510,8 gram]

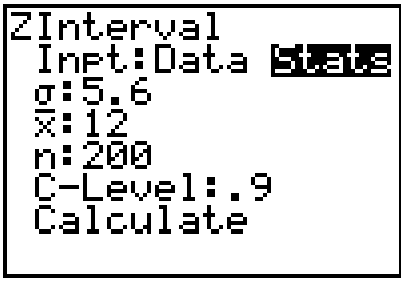
probleemstelling 2

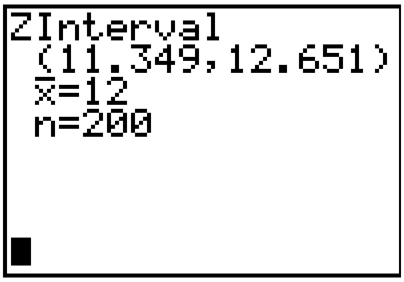
De vuilnisdienst van een stad wil weten hoeveel huisvuil een gezin uit deze stad wekelijks buitenzet. Er wordt een aselecte steekproef genomen van 200 gezinnen. Het gemiddelde gewicht huisvuil dat door deze gezinnen verbruikt werd was 12 kg, met een standaardafwijking van 5.6 kg. Omdat de steekproefomvang groot is, zal de standaardafwijking van deze steekproef bij benadering gelijk zijn aan de standaardafwijking van de populatie.

- Stel een 90%-betrouwbaarheidsinterval op voor de gemiddelde wekelijkse hoeveelheid afval dat een gezin buitenzet.
- Als de gemiddelde wekelijkse hoeveelheid afval groter is dan 13 kg, kan de vuilnisdienst de afvalverwerking niet meer aan. Heeft de vuilnisdienst reden om zich zorgen te maken?

oplossing

Omdat n = 200 relatief groot is, gaan we uit van een normale verdeling.
X = hoeveel kg huisvuil, X ~ N(μ ; σ = 5,6).





90%-betrouwbaarheidsinterval voor μ: [11,35 kg ; 12,65 kg]

Met 95% zekerheid weten we dat de gemiddelde wekelijkse hoeveelheid afval niet groter is dan 12,65 kg. De vuilnisdienst hoeft zich dus geen zorgen te maken.

6.2. Betrouwbaarheidsintervallen voor μ met onbekende σ

probleemstelling 1

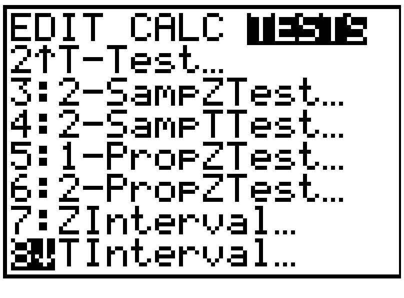
De breukspanning van katoendraden (d.w.z. het gewicht waarbij de draad breekt) is normaal verdeeld. Een reeks van 14 metingen geeft een gemiddelde breukspanning van 6,74 kg en een standaardafwijking van 1,12 kg. Bepaal een 95% betrouwbaarheidsinterval voor de gemiddelde breukspanning van de hele populatie katoendraden.

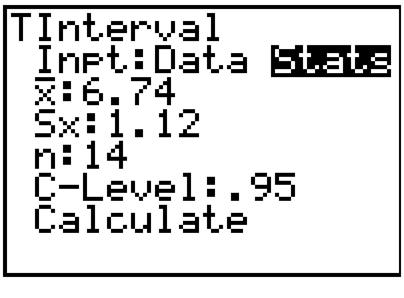
oplossing

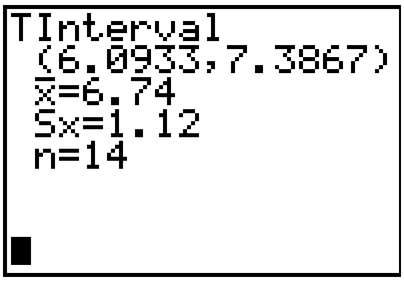
formule:

$$\left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Stat – Test – 8: TInterval







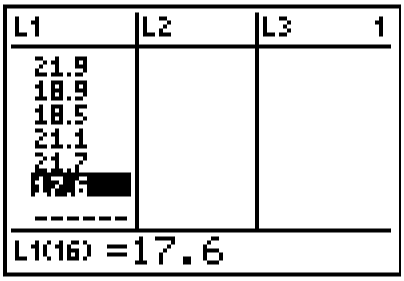
antwoord: [6,09 kg ; 7,39 kg]

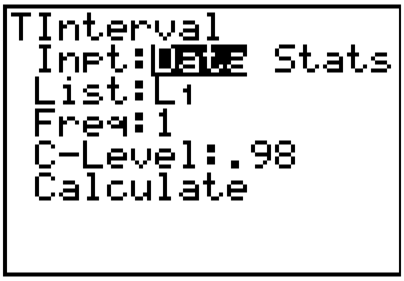
probleemstelling 2

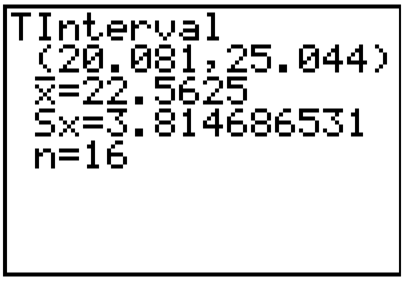
De dikte van een reeks houten platen is normaal verdeeld. Hieronder staat een reeks van 16 metingen (in mm). Bepaal een 98% betrouwbaarheidsinterval voor de gemiddelde dikte van deze reeks houten platen.

23,1	20,2	24,7	27,8	27,6	29,2	17,4	23,3
27,3	20,7	21,9	18,9	18,5	21,1	21,7	17,6

oplossing







antwoord: [20,08 mm ; 25,04 mm]

6.3. Betrouwbaarheidsintervallen voor de populatieproportie p

probleemstelling 1

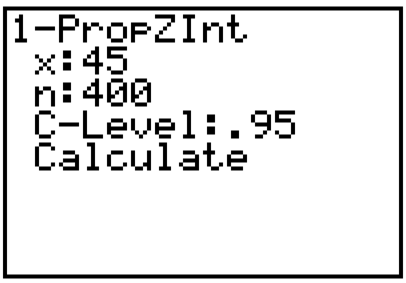
Bij een controle van een steekproef van 400 lampen vond men er 45 slechte. Vind op grond hiervan een 95% betrouwbaarheidsinterval voor het percentage (= de proportie) slechte lampen in de hele populatie.

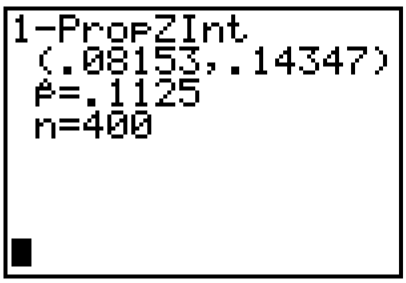
oplossing

formule:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}\ ,\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}\right]$$

Stat – Test – A: 1-PropZInt



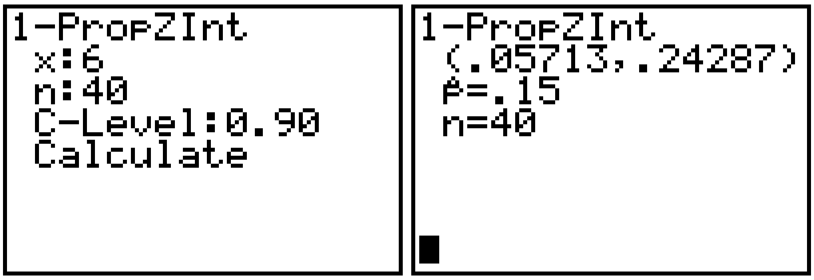


antwoord: [8,2 % ; 14,3 %]

probleemstelling 2

In de spaarpot van Kasper zitten enkel Belgische euromunten van 2 euro. Jonas wil nu weten hoeveel geld er in de spaarpot van Kasper zit zonder het geld effectief te tellen. Hij haalt 40 munten uit de spaarpot en vervangt ze door Franse euromunten van 2 euro. Nadien schudt hij de spaarpot zodat de Franse munten voldoende gemengd zijn met de Belgische en haalt hij opnieuw 40 munten (met teruglegging) uit de spaarpot. Van de 40 munten blijken er 6 Franse bij te zijn. Geef een 90% betrouwbaarheidsinterval voor het bedrag dat in Kasper zijn spaarpot zit.

oplossing



waaruit:

$$0,057 \leq p \leq 0,243$$

$$\Leftrightarrow 0,057 \leq \frac{40}{N} \leq 0,243$$

$$\Leftrightarrow 17,544 \geq \frac{N}{40} \geq 4,115$$

$$\Leftrightarrow 702 \geq N \geq 165$$

antwoord : [€ 330 ; € 1404]

7. Toetsen van hypothesen

7.1. Toetsen van hypothesen voor μ met bekende σ

probleemstelling 1

In een fabriek worden assen vervaardigd waarbij de gemiddelde diameter ingesteld wordt op 7,6 mm. De diameters van de geproduceerde assen zijn normaal verdeeld met standaardafwijking 0,4 mm. Ter controle neemt men een steekproef van 50 assen en men vindt als gemiddelde een waarde van 7,4 mm. Indien de diameters te veel afwijken, wordt het productieproces stopgezet.
Ga na, met α = 1%, of het productieproces wordt stopgezet.

oplossing

- formuleren van de hypothesen:

  H_0 : $\mu = 7{,}6$

  H_1 : $\mu \neq 7{,}6$

  Dit is een tweezijdige toets van het gemiddelde

- toetsingsgrootheid:

  $\overline{X}$ = gemiddelde diameter van 50 assen

  $\overline{X}$ ~ N(μ = 7,6 , σ = $\frac{0,4}{\sqrt{50}} = 0,0566$)

methode 1: kritieke grens

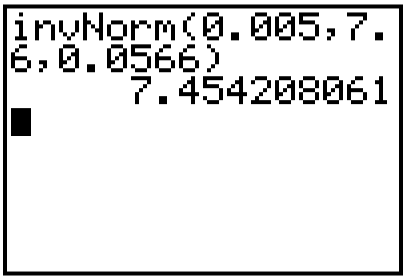
Verwerp H_0 indien $\overline{x} < k_1$ *of* $\overline{x} > k_2$

met $k_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ *en* $k_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

bepalen van de grenswaarde van het aanvaardingsgebied bij α = 1%

$P(\overline{X} \leq g_L) = 0,005$

$\Leftrightarrow g_L = 7,454$

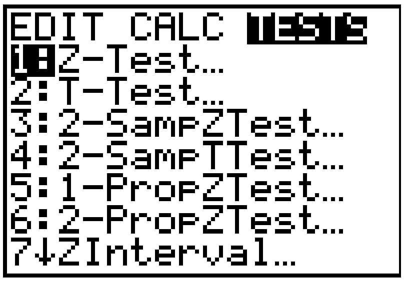


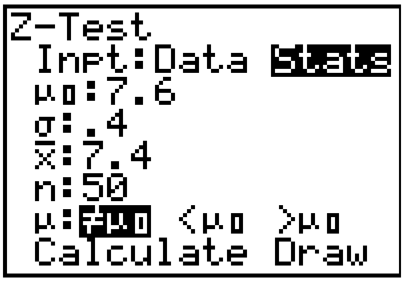
antwoord: 7,4 < 7,454 ; H_0 wordt verworpen, het productieproces wordt stopgezet.

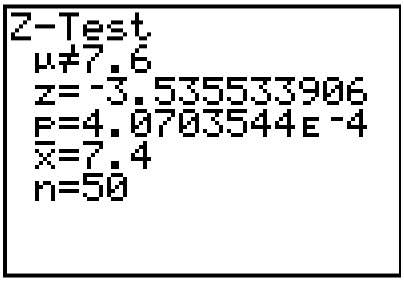
methode 2: p-waarde

Verwerp H_0 indien p-waarde $\leq \alpha$

Stat – Test – 1: Z-Test





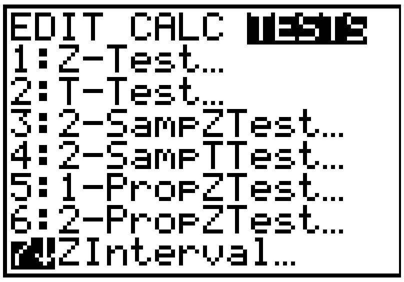


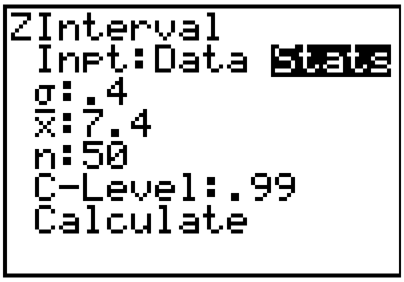
antwoord: p-waarde = 0,0004 < 0,01, H_0 wordt verworpen.

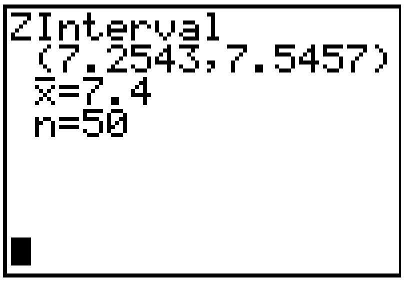
methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $\mu_0 \in \left[\bar{x} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$

Stat – Test – 7: ZInterval







antwoord: 7,6 $\notin$[7,25 ; 7,55] dus H_0 verwerpen.

probleemstelling 2

Een fabrikant van light producten beweert dat zijn producten slechts 140 calorieën (met een standaardafwijking van 20 calorieën) bevatten per pakje van 200 gram. Bij een serie controleproeven heeft de consumentenbond 20 pakjes onderzocht. Deze 20 pakjes bleken gemiddeld een voedingswaarde van 155 calorieën te bevatten.
Toets of de fabrikant gelijk kan hebben met zijn uitspraak. (α = 1%)

oplossing

- formuleren van de hypothesen:

  $H_0 : \mu = 140$

  $H_1 : \mu > 140$

  Dit is een rechts eenzijdige toets van het gemiddelde

- toetsingsgrootheid:

  $\overline{X}$ = gemiddeld aantal calorieën in 20 pakjes

  $\overline{X} \sim N(\mu = 140 , \sigma = \frac{20}{\sqrt{20}} = \sqrt{20})$

methode 1: kritieke grens

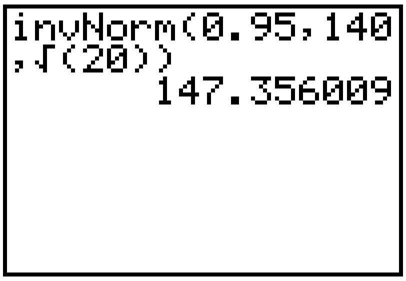
Verwerp H_0 indien $\overline{x} > k$ met $k = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$

bepalen van de kritieke grens van het aanvaardingsgebied bij α = 5%

$P(\overline{X} > g_R) = 0,05$

$\Leftrightarrow P(\overline{X} \leq g_R) = 0,95$

$\Leftrightarrow g_R = 147,4$

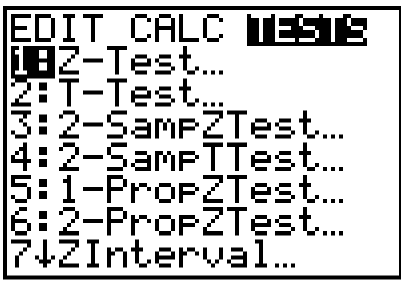


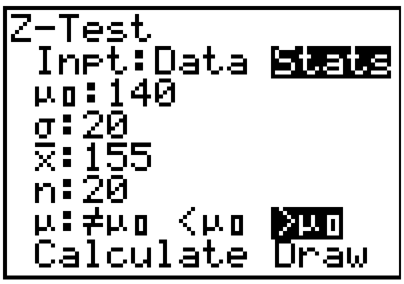
antwoord: 155 > 147, 4 de nulhypothese wordt verworpen.

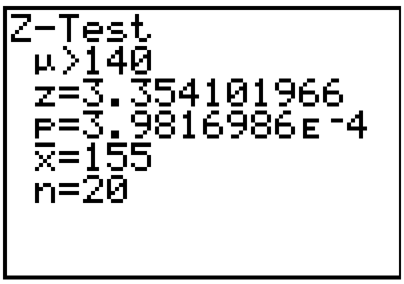
methode 2: p-waarde

Verwerp H_0 indien p-waarde $\leq \alpha$

Stat – Test – 1: Z-Test

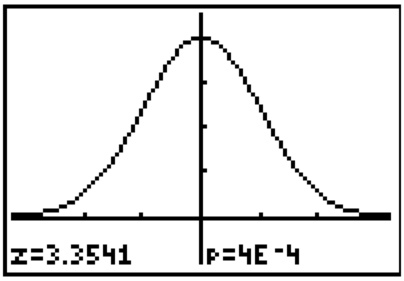






antwoord: p-waarde = 0,0004 < 0,05 ; H_0 wordt verworpen.

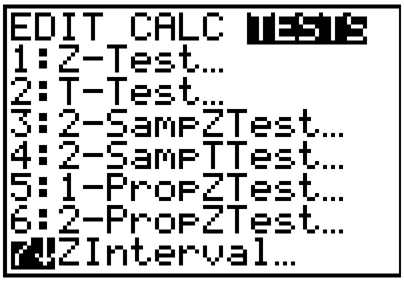
Wanneer je i.p.v. op Calculate op Draw klikt, krijg je een visuele voorstelling van de p-waarde.

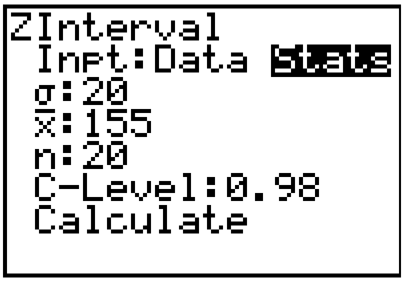


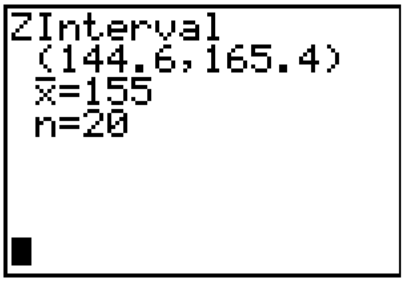
methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $\mu_0 \geq \overline{x} - z_\alpha \dfrac{\sigma}{\sqrt{n}}$

Stat – Test – 7: Zinterval (eenzijdig α = 1% → C-Level 0,98)







antwoord: 140 $\notin$[145 ; 165] dus H_0 verwerpen.

7.2. Toetsen van hypothesen voor μ met onbekende σ

probleemstelling 1

De breukbelasting van kabels is normaal verdeeld. Een industrieel beweert kabels te vervaardigen met een gemiddelde breukbelasting van 8000 kg. Een dokwerker heeft de indruk dat de gemiddelde breukbelasting minder dan 8000 kg bedraagt. Een steekproef van 6 kabels geeft een gemiddelde breukbelasting van 7750 kg en een standaardafwijking van 135 kg. Is de afwijking beduidend of niet op 5% significantieniveau?

oplossing

- formuleren van de hypothesen:

H_0 : μ = 8000

H_1 : μ < 8000

Dit is een links eenzijdige toets van het gemiddelde

- toetsingsgrootheid:

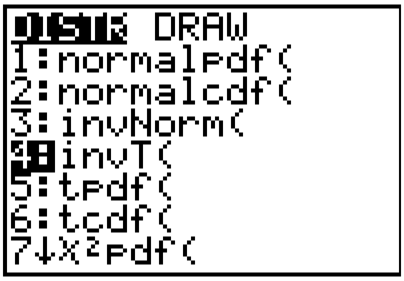
$\overline{X}$ = gemiddelde breukbelasting van 6 kabels

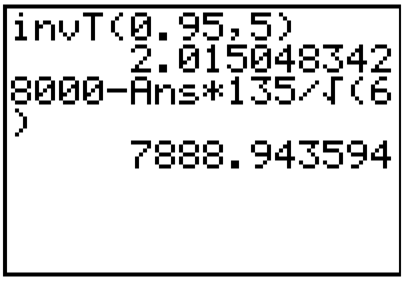
$\overline{X}$ ~ N(μ = 8000 , σ = $\frac{\sigma}{\sqrt{6}}$)

methode 1: kritieke grens

Verwerp H_0 indien $\overline{x} < k$ met $k = \mu_0 - t_{n-1,\alpha}\,\frac{s}{\sqrt{n}}$

bepalen van de kritieke grens van het aanvaardingsgebied bij α = 5%





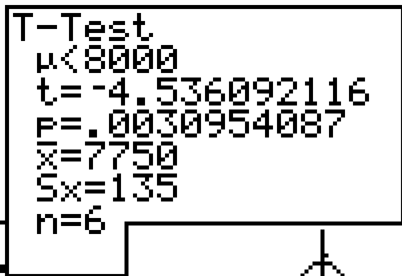
antwoord: 7750 < 7888,9 de nulhypothese wordt verworpen.

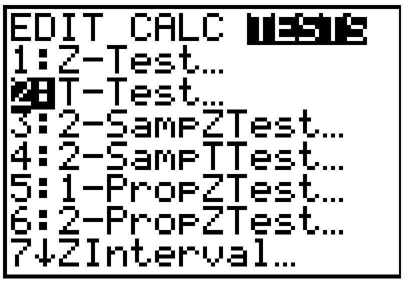
(opmerking: op de ietwat oudere versies van de TI-84 was de functie invT standaard niet aanwezig, je verkrijgt deze door het ‘besturings-systeem’ van je GRM te updaten).

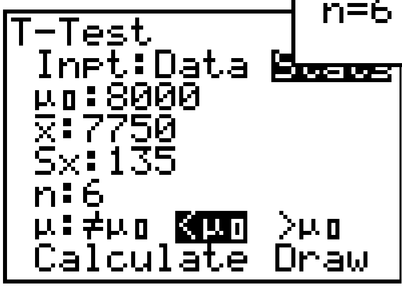
methode 2: p-waarde

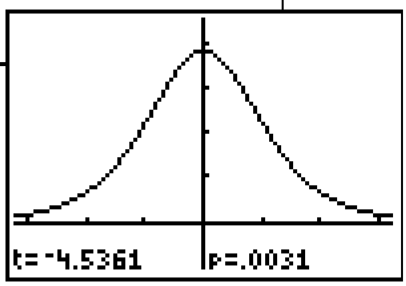
Verwerp H_0 indien p-waarde $\leq \alpha$

Stat – Test – 2: T-Test







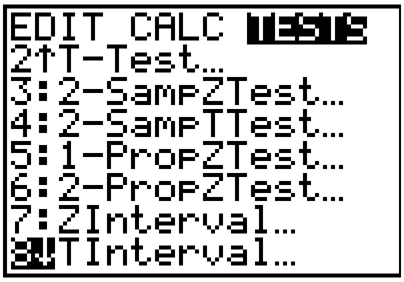


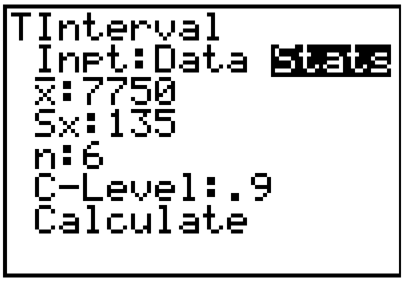
antwoord: p-waarde = 0,0031 < 0,05, H_0 verwerpen.

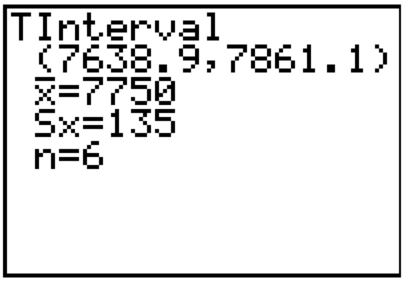
methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $\mu_0 \leq \overline{x} + t_{n-1,\alpha} \frac{s}{\sqrt{n}}$

Stat – Test – 8: Tinterval (eenzijdig α = 5% → C-Level 0,90)







antwoord: 8000 $\notin$[7639 ; 7861] dus H_0 verwerpen.

probleemstelling 2

De inhoud van potjes speculaaspasta is normaal verdeeld. De machine die deze potjes vult is zodanig ingesteld dat de gemiddelde inhoud 376 gram zou moeten bedragen. De kwaliteitsmanager haalde deze namiddag 18 potjes van de lopende band en woog deze na. De resultaten (in gram) staan in volgende tabel:

375	382	370	376	369	379	377	381	380
368	373	367	371	381	371	384	372	373

Is de bewering dat deze potjes speculaaspasta 376 gram bevatten correct op het 10% significantieniveau? (toets tweezijdig)

oplossing

- formuleren van de hypothesen:

  $H_0 : \mu = 376$

  $H_1 : \mu \neq 376$

  Dit is een tweezijdige toets van het gemiddelde

- toetsingsgrootheid:

  $\overline{X}$ = gemiddelde inhoud van een potje speculaaspasta

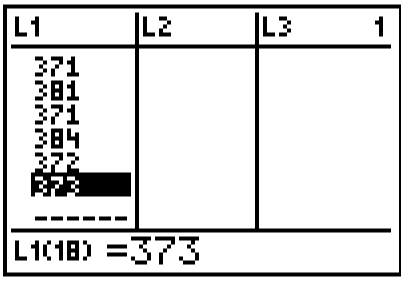
$\overline{X} \sim N(\mu = 376 , \sigma = \dfrac{\sigma}{\sqrt{18}})$

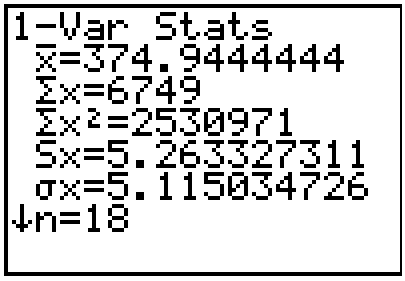
methode 1: kritieke grens

Verwerp H_0 indien $\overline{x} < k_1$ of $\overline{x} > k_2$ met $k_1 = \mu_0 - t_{n-1,\alpha/2} \dfrac{s}{\sqrt{n}}$ en

$k_2 = \mu_0 + t_{n-1,\alpha/2} \dfrac{s}{\sqrt{n}}$

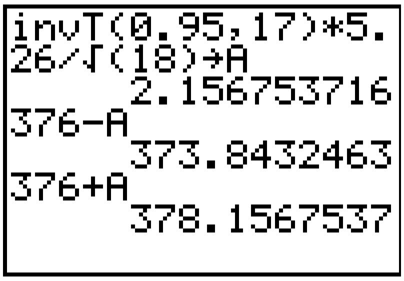
bepalen van het gemiddelde en de standaardafwijking van de steekproef





$\overline{x} = 374,9 \quad s = 5,26$

bepalen van de kritieke grenzen bij α = 10%

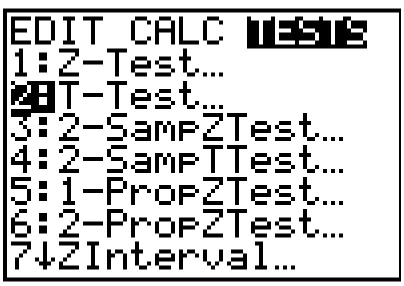


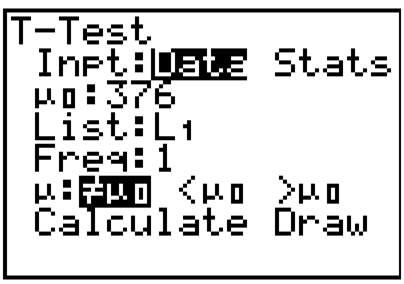
antwoord: 373,84 < 374,9 < 378,16 de nulhypothese wordt niet verworpen.

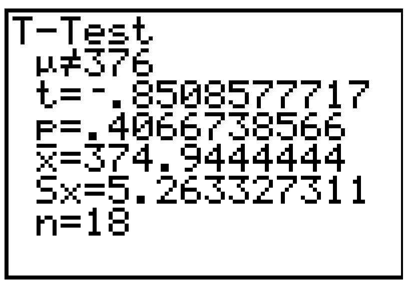
methode 2: p-waarde

Verwerp H_0 indien p-waarde $\leq \alpha$

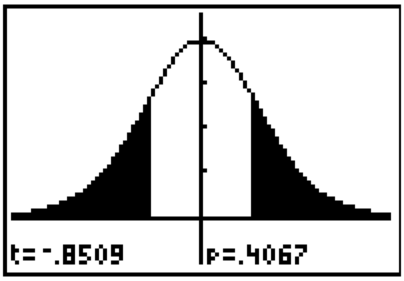
Stat – Test – 2: T-Test







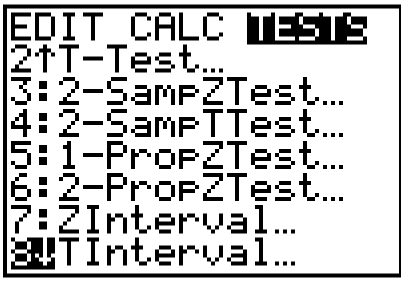
antwoord: p-waarde = 0,4067 > 0,10,
H_0 niet verwerpen.

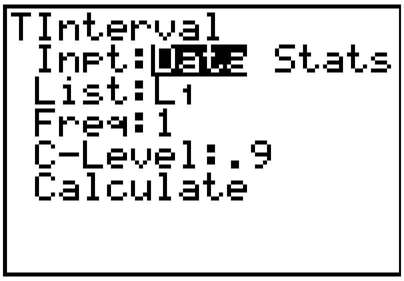


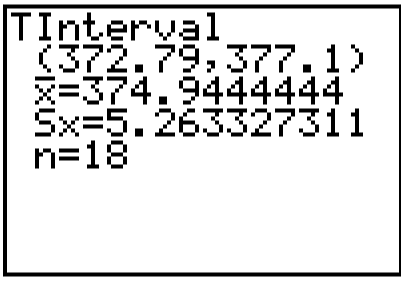
methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $\mu_0 \in \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$

Stat – Test – 8: Tinterval







antwoord: 376 $\in$[372,8 ; 377,1] dus H_0 niet verwerpen.

7.3. Toetsen van hypothesen voor de populatieproportie p

probleemstelling 1

Een fotostudio wenst een partij flitslampjes te kopen. In de partij mogen niet meer dan 6% slechte lampjes voorkomen. Om de kwaliteit van de lampjes te controleren neemt hij een steekproef van 100 stuks. Hierin bevinden zich 8 slechte lampjes. Wordt de partij goedgekeurd of niet? (α = 5%)

oplossing

- formuleren van de hypothesen:

  H_0 : p = 0,06

  H_1 : p > 0,06

  Dit is een rechts eenzijdige toets van fracties

- toetsingsgrootheid:

  X = aantal slechte lampjes in de steekproef

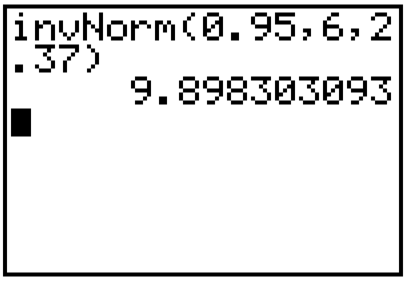
  X ~ B(n = 100 , p = 0,06)

methode 1: kritieke grens

Verwerp H_0 indien $\hat{p} > k$ met $k = p_0 + z_\alpha \sqrt{\dfrac{p_0(1-p_0)}{n}}$

X benaderen door de normale met μ = np = 6, σ = $\sqrt{np(1-p)} = 2{,}37$

bepalen van de rechtergrenswaarde (kritieke grens)van het aanvaardingsgebied bij α = 5%



$P(X > g_R) = 0{,}05$
$\Leftrightarrow P(X \le g_R) = 0{,}95$
$\Leftrightarrow g_R = 9{,}9$

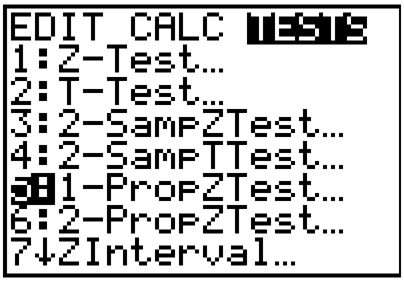
antwoord: 8 < 9,9 de nulhypothese wordt niet verworpen.

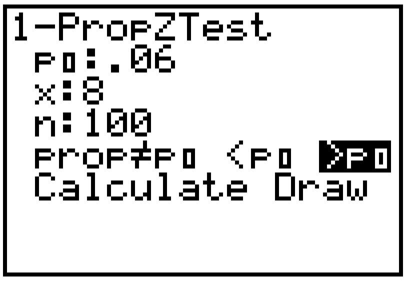
methode 2: p-waarde

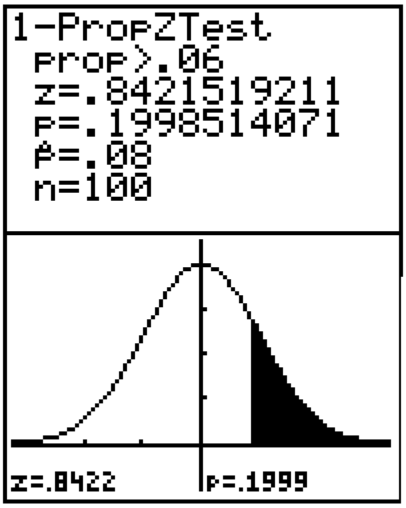
Verwerp H_0 indien p-waarde $\le \alpha$

p-waarde = $P\left(\hat{P} \ge \hat{p} \mid p = p_0\right) = P\left(Z \ge \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\right)$

Stat – Test – 5: 1-PropZTest





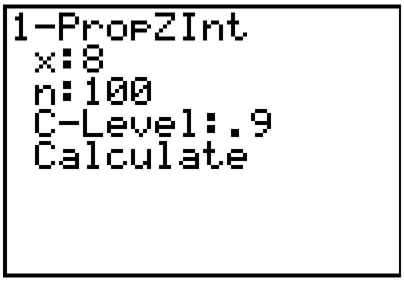


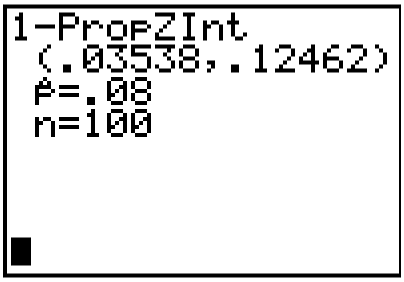
antwoord: p-waarde = 0,2 > α = 0,05 ;
H_0 wordt niet verworpen.

methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $p_0 \ge \hat{p} - z_\alpha \sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}}$

Stat – Test – A: 1-PropZInt (eenzijdig α = 5% → C-Level 0,90)





antwoord: p = 0,06 > 0,035 dus H_0 niet verwerpen.

probleemstelling 2

Een muntstuk wordt 160 keer geworpen en we verkrijgen 101 keer munt. Is dit normaal? Toets tweezijdig. (α = 5%).

oplossing

- formuleren van de hypothesen:

  H_0 : p = 0,5

  H_1 : p $\neq$ 0,5

  Dit is een tweezijdige toets van fracties

- toetsingsgrootheid:

  X = aantal keren munt in de steekproef

  X ~ B(n = 160 , p = 0,5)

methode 1: kritieke grens

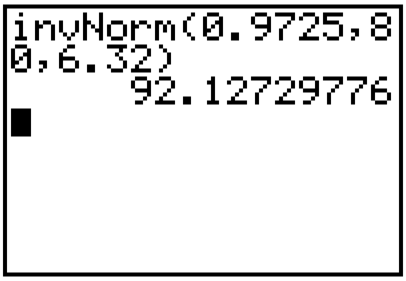
Verwerp H_0 indien $\hat{p} < k_1$ *of* $\hat{p} > k_2$

met $k_1 = p_0 - z_{\alpha/2}\sqrt{\dfrac{p_0(1-p_0)}{n}}$ *en* $k_2 = p_0 + z_{\alpha/2}\sqrt{\dfrac{p_0(1-p_0)}{n}}$

X benaderen door de normale met μ = np = 80, σ = $\sqrt{np(1-p)} = 6,32$

bepalen van de grenswaarden van het aanvaardingsgebied bij α = 5%

$P(X > g_R) = 0,025$
$\Leftrightarrow P(X \le g_R) = 0,9725$
$\Leftrightarrow g_R = 92,1$

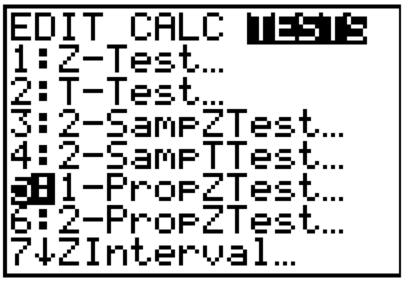


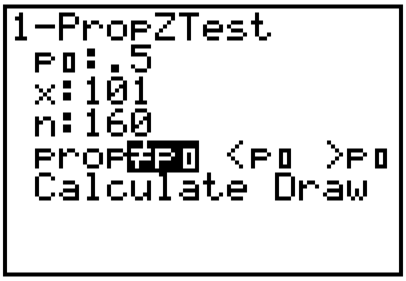
antwoord: 101 > 92,1 de nulhypothese wordt verworpen.

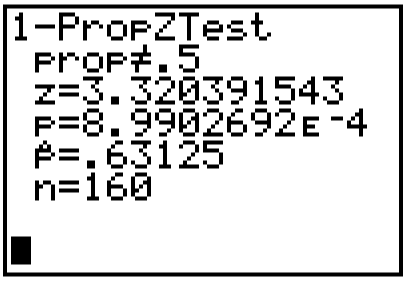
methode 2: p-waarde

Verwerp H_0 indien p-waarde $\le \alpha$

Stat – Test – 5: 1-PropZTest





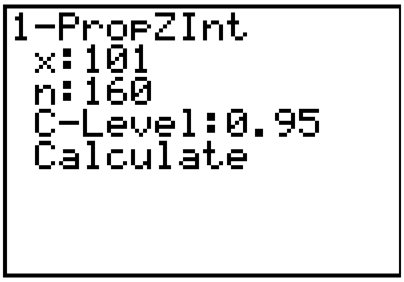


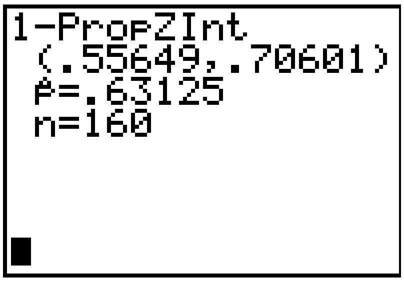
antwoord: p-waarde = 0,0009 < α = 0,05 ; H_0 wordt verworpen

methode 3: betrouwbaarheidsinterval

Verwerp H_0 **niet** indien $p_0 \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \; , \; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \right]$

Stat – Test – A: 1-PropZInt





antwoord: 0,5 $\notin$ [0,56 ; 0,71] dus H_0 verwerpen.

7.4. Het toetsen van hypothesen voor twee μ's met bekende σ's

probleemstelling 1

Dozen speculaasijs worden automatisch gevuld door twee machines. Om de gemiddelde inhoud te meten wordt van beide een steekproef genomen, dit leverde volgende resultaten.

machine A : n = 72 $\quad \overline{x} = 2,48l \quad \sigma_x = 0,08l$

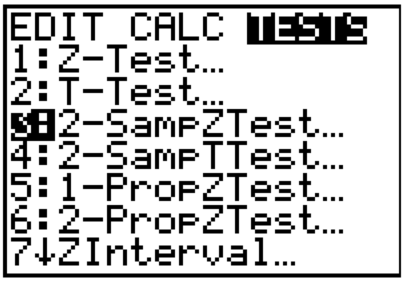
machine B : m = 54 $\quad \overline{y} = 2,55l \quad \sigma_y = 0,05l$

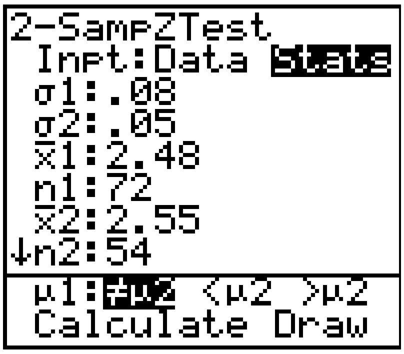
Mogen we op basis van deze gegevens besluiten dat beide machines niet dezelfde gemiddelde inhoud leveren? (b.i. betrouwbaarheidsniveau 95%).

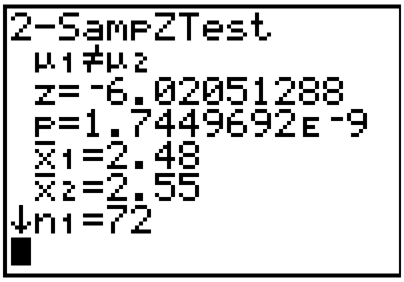
oplossing

- formuleren van de hypothesen:

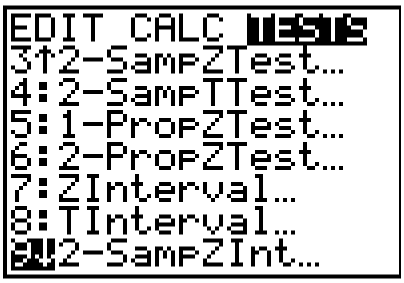
H_0 : $\mu_A = \mu_B$
H_1 : $\mu_A \neq \mu_B$

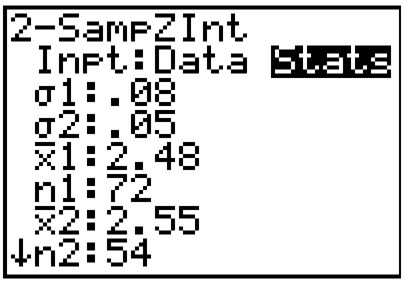


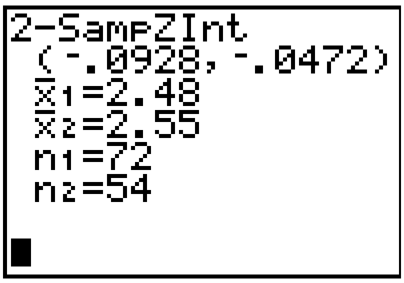




- antwoord: p-waarde zeer klein dus H_0 verwerpen of beide machines leveren inderdaad een andere gemiddelde inhoud.







- antwoord: b.i. [-0,093 ; -0,047] , omdat 0 niet tot dit interval behoort mag je besluiten dat $\mu_A \neq \mu_B$.

probleemstelling 2

Zakjes pindanootjes worden automatisch gevuld door twee machines. De kwaliteitsmanager vermoedt dat het gemiddelde bij machine A iets hoger ingesteld staat dan bij machine B. Om dit te controleren neemt hij van beide machines een steekproef. De resultaten (in gram) vind je terug in volgende tabellen:

machine A ($\sigma_x = 4\ gram$)

267	263	269	273	273	275	259	267	272	264	265	261	261
264	265	260	266	266	269	267	267	267	273	268	267	266

machine B ($\sigma_x = 5\ gram$)

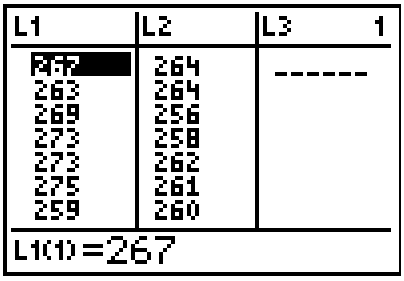
264	264	256	258	262	261	260	257	260	253
259	263	254	255	262	260	257	261	265	262
265	267	259	263	265	257	259	256	263	258

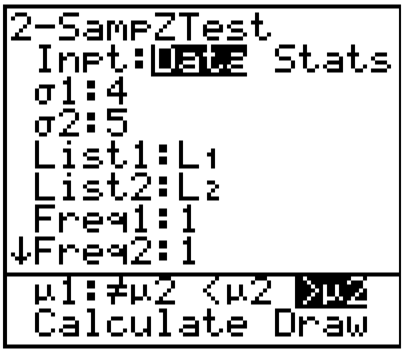
Mag de kwaliteitsmanager op basis van deze gegevens besluiten dat de gemiddelde inhoud van een zakje nootjes gevuld door machine A hoger is dan de gemiddelde inhoud van een zakje nootjes gevuld door machine B? (b.i. significantie 90%).

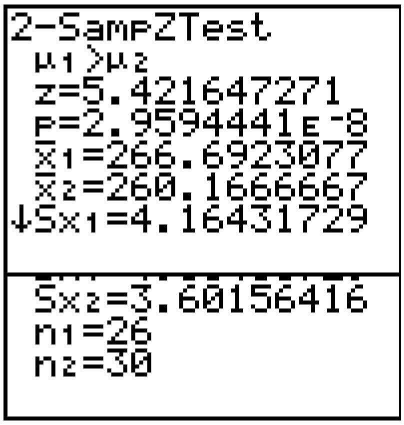
oplossing

- formuleren van de hypothesen:

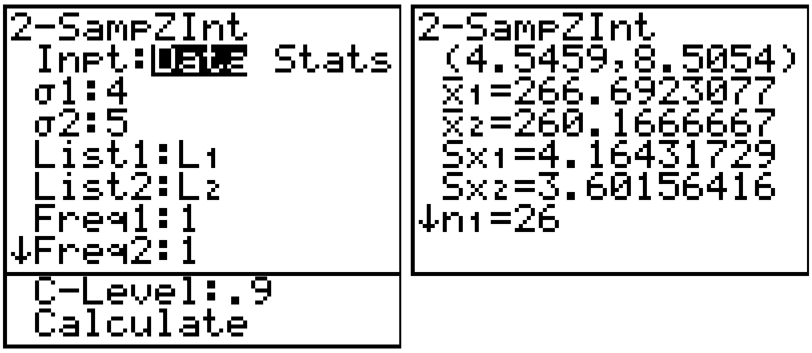
  H_0 : $\mu_A = \mu_B$
  H_1 : $\mu_A > \mu_B$







- antwoord: p-waarde zeer klein dus H_0 verwerpen of machines A levert inderdaad een groter gemiddelde inhoud dan machine B.



- antwoord: b.i. [4,546 ; 8,505] , omdat 0 niet tot dit interval behoort mag je besluiten dat $\mu_A > \mu_B$.

7.5. Het toetsen van twee σ's

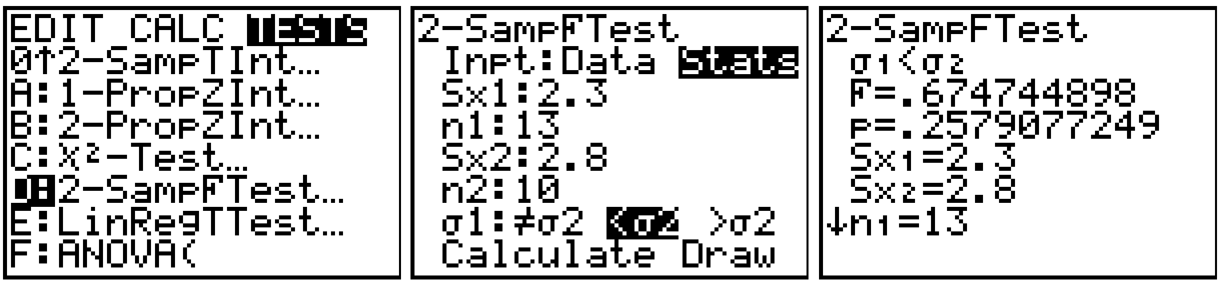
probleemstelling

In een bedrijf staan twee machines en men heeft de indruk dat machine 2 minder precies werkt dan machine 1. Om dit te onderzoeken neemt men uit de productie van machine 1 een staal van omvang 13 en men vindt $s_1^2 = 5,29$, uit een staal van 10 exemplaren van machine 2 vond men $s_2^2 = 7,84$. Vormt dit op een significantieniveau van 5% voldoende bewijs dat machine 2 minder precies werkt dan machine1? Je mag normaliteit veronderstellen.

oplossing

- formuleren van de hypothesen:

  H_0 : $\sigma_1 = \sigma_2$
  H_1 : $\sigma_1 < \sigma_2$



- antwoord: p-waarde = 0,258 > 0,05 we verwerpen H_0 niet.

7.6. Toetsen van hypothesen voor twee μ's met onbekende σ's

probleemstelling 1

Op een aantal proefterreinen werden twee soorten meststof gebruikt en werd de productie (per oppervlakte-eenheid) voor elk terrein gemeten.

meststof A : n = 180 $\bar{x} = 58$ $s_x = 11$

meststof B : m = 120 $\bar{y} = 66$ $s_y = 16$

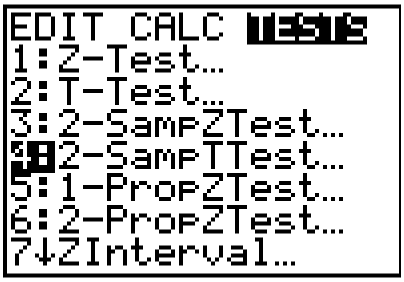
Mogen we op basis van deze gegevens besluiten dat meststof B beter is dan meststof A? (b.i. betrouwbaarheidsniveau 95%).

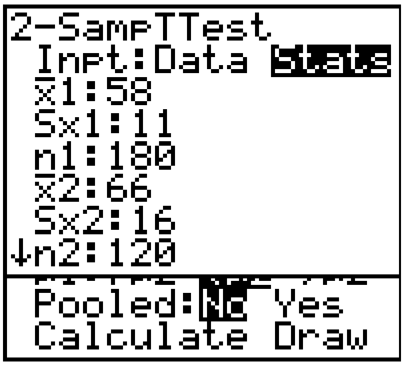
oplossing

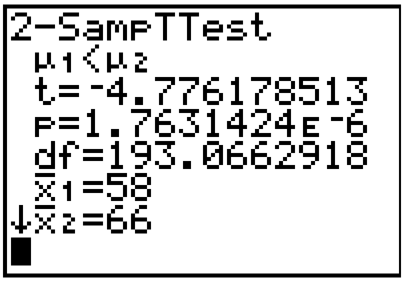
- formuleren van de hypothesen:

H$_0$: $\mu_A = \mu_B$

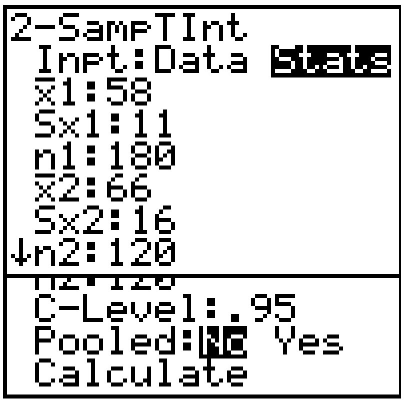
H$_1$: $\mu_A < \mu_B$

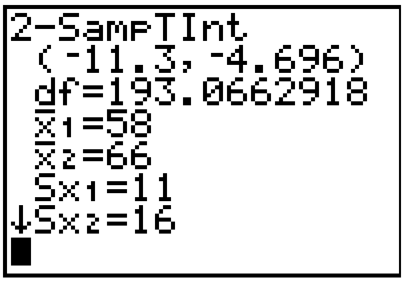






- antwoord: p-waarde zeer klein (we veronderstellen ongelijke varianties en dus pooled "no") dus H$_0$ verwerpen of meststof B is inderdaad beter dan meststof A.





- antwoord: b.i. [-11,3 ; -4,7] , omdat 0 niet tot dit interval behoort mag je besluiten dat $\mu_A < \mu_B$.

probleemstelling 2

Een bepaalde stadsschool beweert dat de leerlingen die zij aantrekt gemiddeld over een hoger IQ beschikken dan de leerlingen uit de andere stadsscholen in de buurt.
Een onderzoeker wenst deze bewering na te gaan en meet via een onafhankelijk test het IQ van alle laatstejaarsstudenten uit deze school (A) en van een andere nabij gelegen school (B).
De resultaten zijn als volgt:

school A: n = 58 $\overline{x} = 104{,}6$ $s_x = 13{,}4$

school B: m = 66 $\overline{y} = 102{,}3$ $s_y = 14{,}1$

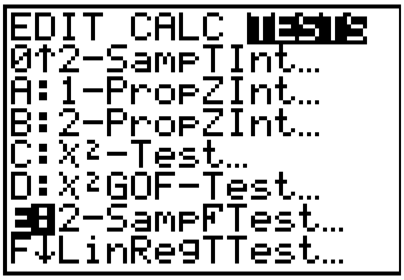
Mogen we op basis van deze gegevens besluiten dat de uitspraak van school A waar is? (Het IQ is een variabele die normaal verdeeld is).

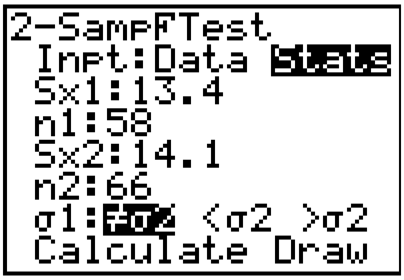
oplossing

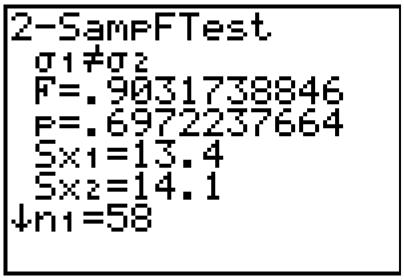
Alhoewel we vermoeden en gerust mogen veronderstellen dat de populatievarianties gelijk zijn, gaan we dit nog even na via een F-toets.

- formuleren van de hypothesen:

  H0 : $\sigma_1 = \sigma_2$
  H1 : $\sigma_1 \neq \sigma_2$



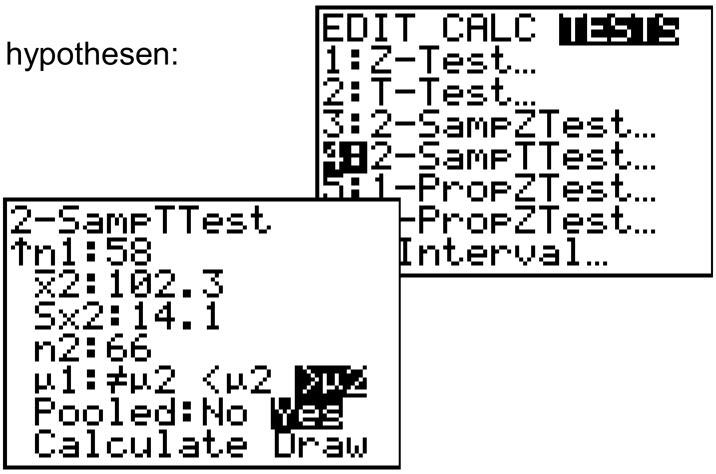


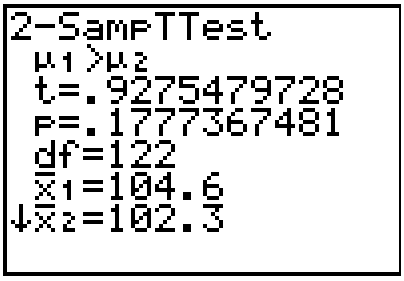


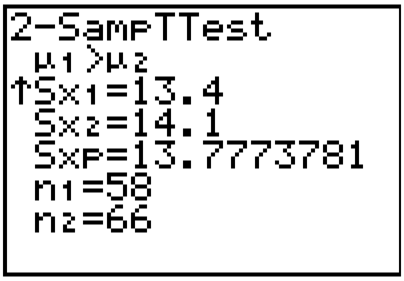
- antwoord: p-waarde = 0,697 en dus vrij hoog 0,05 we verwerpen H_0 niet.
  We veronderstellen gelijke varianties en dus pooled “yes”

- formuleren van de hypothesen:

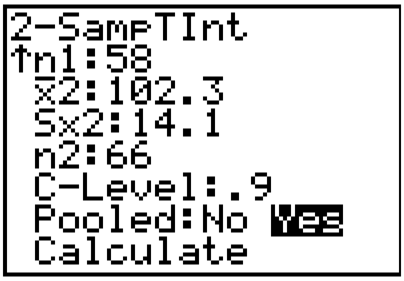
  H0 : $\mu_A = \mu_B$
  H1 : $\mu_A > \mu_B$

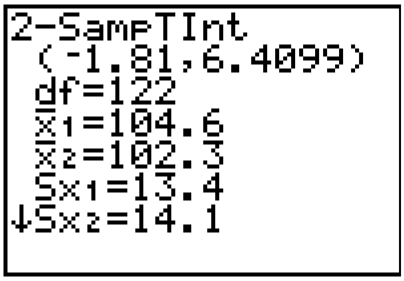






- antwoord: p-waarde = 0,178 dus H_0 wordt niet verworpen of de leerlingen van de ene school hebben geen significant hoger IQ dan de leerlingen van de andere school.





- antwoord: b.i. (significantie 90%) [-1,8 ; 6,4] , omdat 0 tot dit interval behoort mag je besluiten dat er in principe geen significant verschil is tussen de gemiddelden van beide scholen.

7.7. Toets voor twee populatieproporties

probleemstelling

Bij enquêtes in Vlaanderen (1284 ondervraagden) en Nederland (923 ondervraagden) stelden we vast dat 948 Vlamingen en 607 Nederlanders regelmatig naar het Tv-nieuws kijken.

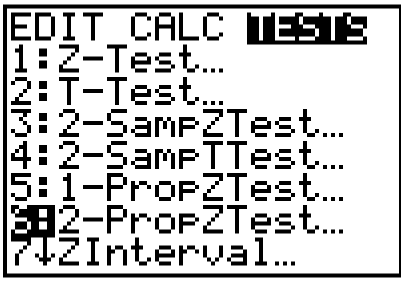
- Mogen we hieruit besluiten dat Vlamingen meer naar het journaal kijken dan Nederlanders?
- Construeer een 95% b.i. voor p_v - p_n.

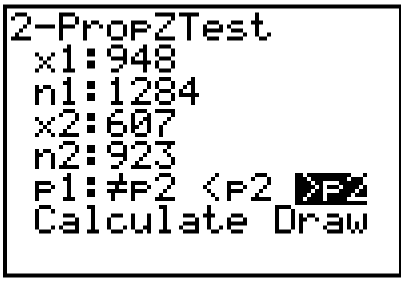
oplossing

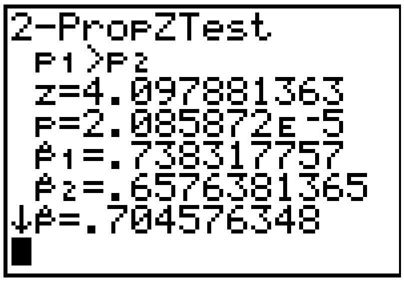
- formuleren van de hypothesen:

H_0 : $p_V = p_N$

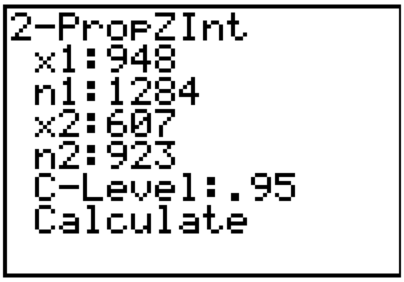
H_1 : $p_V > p_N$

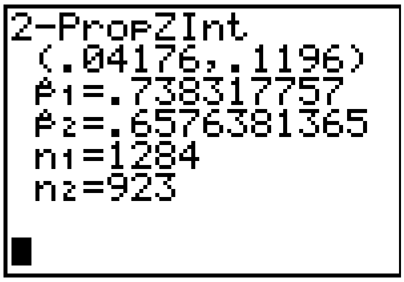






- antwoord: p-waarde = 0,00002 < 0,05 dus H_0 verwerpen, m.a.w. je mag besluiten dat Vlamingen meer naar het journaal kijken dan Nederlanders.





- antwoord: b.i. : [0,04 ; 0,12]. Omdat 0 niet tot dit interval behoort mag je besluiten dat $p_V > p_N$.

7.8. De χ^2-toets

Een chi-kwadraattoets wordt in de statistiek gebruikt om te zien of waargenomen aantallen systematisch afwijken van verwachte aantallen. Een chi-kwadraattoets wordt veel gebruikt om kruistabellen te analyseren. Omdat er geen aannamen over gemiddelden of over de populatie worden gedaan is dit een parametervrije toets.

probleemstelling

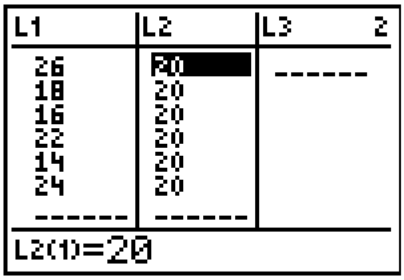
Iemand krijgt een dobbelsteen in handen die er niet erg symmetrisch uitziet. Zou de dobbelsteen wel zuiver zijn? Hij gooit er 120 keer mee en verwacht elk van de ogenaantallen ongeveer 20 keer te gooien. De resultaten van de test vind je in volgende tabel:

Aantal ogen	1	2	3	4	5	6
Aantal worpen	26	18	16	22	14	24

oplossing

- formuleren van de hypothesen:

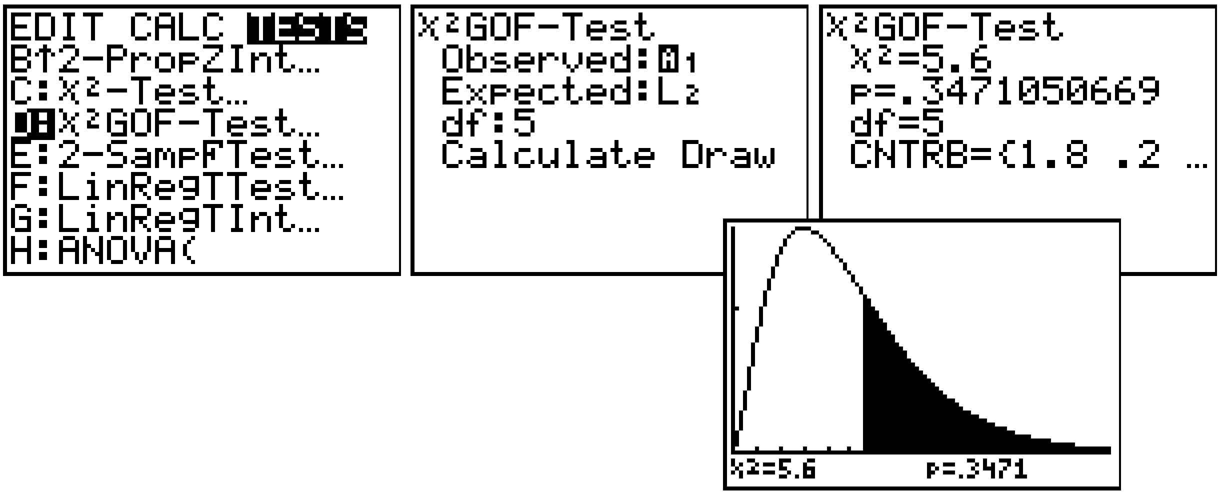
H_0 : “de dobbelsteen is zuiver”
H_1 : “de dobbelsteen is niet zuiver”



- uitvoeren van de test

  Lijst 1 : de waargenomen aantallen
  Lijst 2 : de te verwachte aantallen

  Stat Tests χ^2 GOF-Test (GOF = Goodness-of-Fit)



- antwoord: p-waarde = 0,347 en dus voldoende groot. H_0 wordt niet verworpen, we mogen veronderstellen dat de dobbelsteen zuiver is.

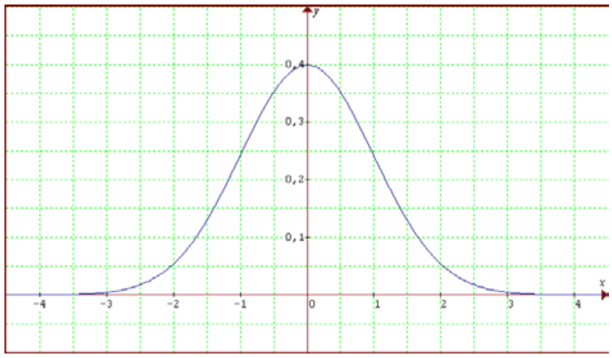
8. Enkele continue verdelingen

8.1. De standaardnormale verdeling

Een normale verdeling met verwachtingswaarde (gemiddelde) 0 en standaardafwijking 1 wordt een standaardnormale verdeling genoemd.

De kansdichtheidsfunctie wordt gegeven door:

$$f(x) = \frac{1}{\sqrt{2\pi}}.\exp\left(-\frac{x^2}{2}\right)$$



$$P\left(X \le a\right) = b$$

a gegeven ➔ b = normalcdf(-1 E 99 , a) of ShadeNorm(- 1 E 99 , a)

b gegeven ➔ a = invNorm(b)

8.2. De normale verdeling

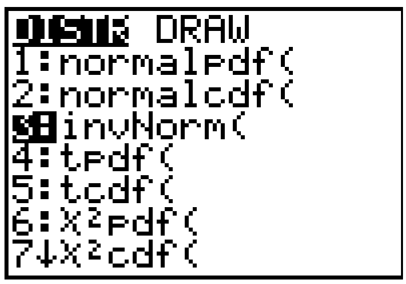
De normale verdeling of Gauss-verdeling heeft als kansdichtheidsfunctie de klokkromme of Gausscurve die gegeven wordt door:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$P\left(X \le a\right) = b$$

a gegeven ➔ b = normalcdf(-1 E 99 , a , μ , σ)

b gegeven ➔ a = invNorm(b , μ , σ)

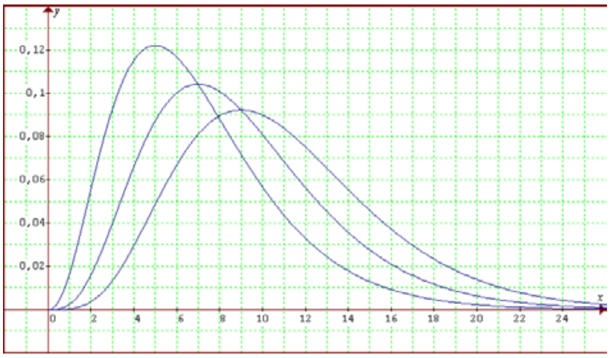


8.3. De chi-kwadraatverdeling

De chi-kwadraatverdeling of χ^2-verdeling is een verdeling van de som van de kwadraten van n onderling onafhankelijke standaardnormale variabelen.
De parameter n wordt het aantal vrijheidsgraden genoemd. De chi-kwadraatverdeling is een specifiek geval van de gamma verdeling.

De kansdichtheidsfunctie wordt gegeven door:

$$f_n(x) = \frac{1}{2^{\frac{n}{2}}.\Gamma\left(\frac{n}{2}\right)}.x^{\frac{n}{2}-1}\exp\left(-\frac{x}{2}\right)$$



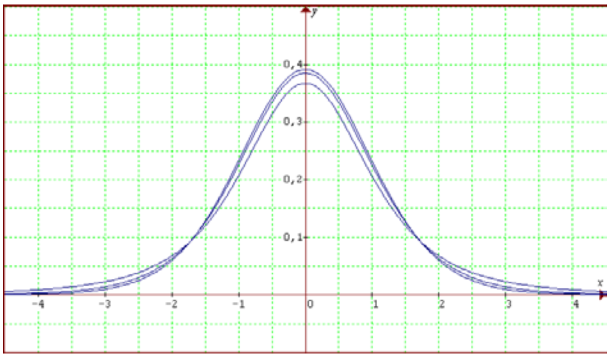
$P(X \le a) = b$ b = X^{2}cdf(-1 E 99 , a , n)

(In de praktijk is dikwijls b gegeven en a gevraagd. Dit probleem los je op via de Solver).

8.4. De t-verdeling of studentverdeling

De dichtheidsfunctie van de t-verdeling of studentverdeling wordt gegeven door:

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right).\sqrt{n\pi}}.\frac{1}{\left(1+\frac{x^2}{n}\right)^{(n+1)/2}}$$



De grafiek van deze functie lijkt wat vorm betreft sterk op standaardnormale verdeling, maar is wat ‘breder’. Hoe kleiner het aantal vrijheidsgraden (n) is, hoe ‘breder’ de grafiek van de kansdichtheid.

$P(X \le a) = b$

a gegeven ➔ b = tcdf(-1 E 99 , a , n)

b gegeven ➔ a = invT(b , n)

8.5. De F-verdeling of verdeling van Fisher

De F-verdeling, genoemd naar Sir R.A. Fisher, is de verdeling van het quotiënt van twee chi-kwadraat verdeelde grootheden. Hij vindt vooral toepassing in de variantieanalyse als verdeling van de toetsingsgrootheid van de F-toets.

De kansdichtheid van de F-verdeling met m vrijheidsgraden in de teller en n vrijheidsgraden in de noemer wordt gegeven door:

$$f_{m,n}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} . \frac{\left(\frac{m}{n}\right)^{m/2} x^{m/2-1}}{\left(1+\frac{m}{n}x\right)^{(m+n)/2}}$$

$P(X \leq a) = b$ b = Fcdf(-1 E 99 , a , m , n)

(In de praktijk is dikwijls b gegeven en a gevraagd. Dit probleem los je op via de Solver).

8.6. De exponentiële verdeling

In de kansrekening en de statistiek worden de exponentiële verdelingen vaak gebruikt voor het modelleren van de tijd tussen twee gebeurtenissen die met een constante gemiddelde snelheid voorkomen.

De kansdichtheid wordt gegeven door:

$$f(x,\lambda) = \lambda.\exp(-\lambda x) \qquad x \geq 0\ ;\ \lambda > 0$$

$$P(X \leq a) = b = 1 - \exp(-\lambda\, a)$$

8.7. De gamma verdeling

In de kansrekening en statistiek is de gammaverdeling een continue kansverdeling met twee parameters.

De kansdichtheid wordt gegeven door:

$$f(x,k,\theta) = x^{k-1}\frac{\exp(-x/\theta)}{\theta^k.\Gamma(k)} \qquad x > 0$$

Als k een natuurlijk getal is, geldt $\Gamma(k) = (k-1)!$

9. Lineaire regressie

probleemstelling

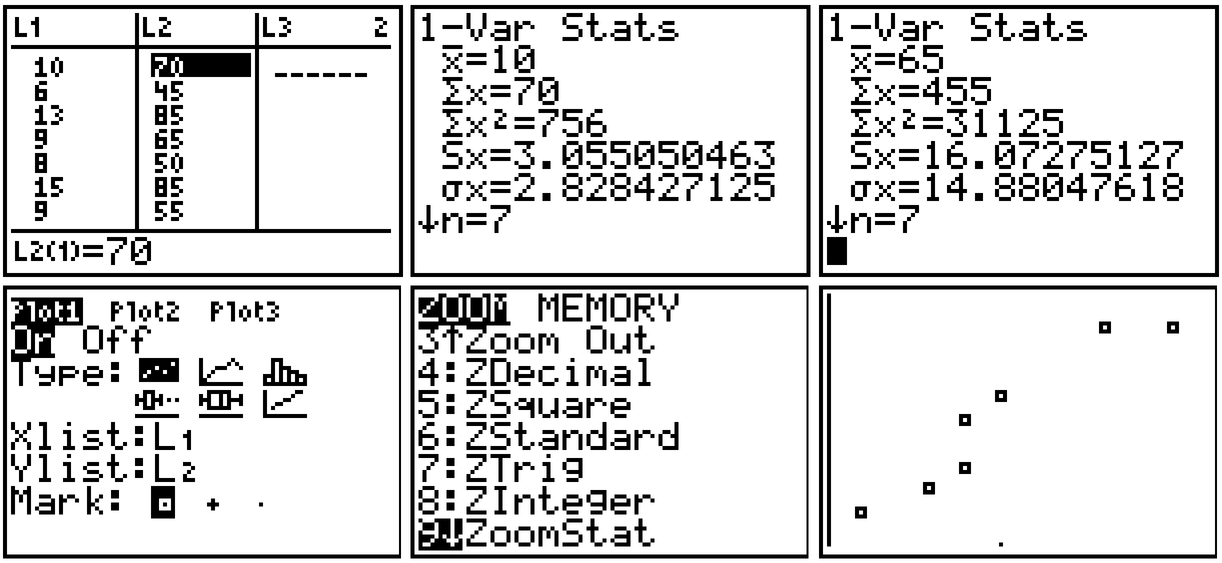
Aan een aantal leerlingen is gevraagd hoeveel tijd (in uren) ze besteed hebben aan een groepswerk geschiedenis. Verder is voor deze leerlingen het aantal punten (op 100) vastgesteld dat ze voor dit groepswerk hebben gekregen. De resultaten waren als volgt:

Groep		I	II	III	IV	V	VI	VII
Aantal gepresteerde uren	X	10	6	13	9	8	15	9
Aantal behaalde punten	Y	70	45	85	65	50	85	55

- Wat is het gemiddeld aantal uren dat men presteerde aan het groepswerk? Wat was de gemiddelde score?
- Bereken de lineaire regressie van Y op X.
- Wat zijn de te verwachten punten voor een groep die 12 uur gespendeerd heeft aan het groepswerk?
- Bereken de covariantie en de correlatiecoëfficiënt.

oplossing 1

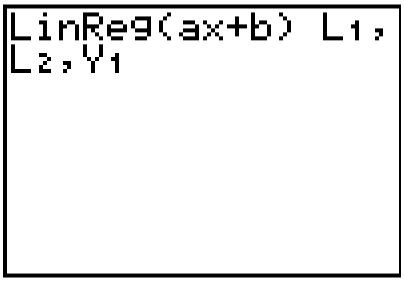
Plaats de onafhankelijke variabele X, in dit geval het aantal gepresteerde uren, in lijst L_1 en de afhankelijke variabele Y, in dit geval het aantal behaalde punten, in lijst L_2.

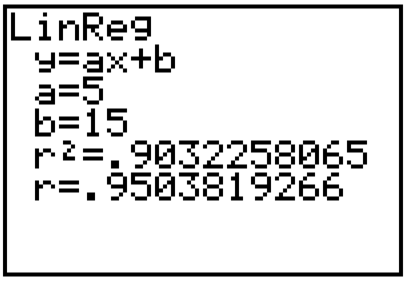


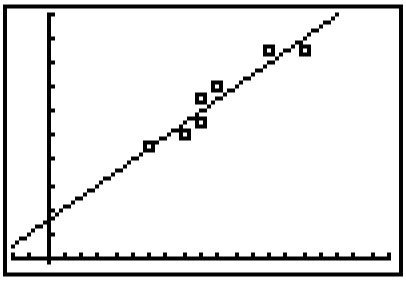
Als je de gegevens r en r² niet verkrijgt met je TI-84 moet je eerst 2nd [CATALOG] DiagnosticsOn uitvoeren vanuit het basisscherm. (Vanaf operating system 2.53 staat diagnostic standaard op on).

(De determinatiecoëfficiënt r² is een maat voor de kwaliteit van een regressiemodel dat niet noodzakelijk lineair is.)

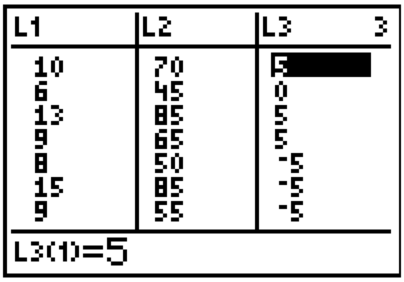
Druk nu op STAT CALC 4:LinReg(ax + b) 2nd [L1], 2nd [L2], VARS Y-VARS 1:Function 1:Y1 ENTER







Ben je geïnteresseerd in de residu's, tik je 2nd [List] :Resid STO> 2nd [L3] ENTER



Antwoorden:

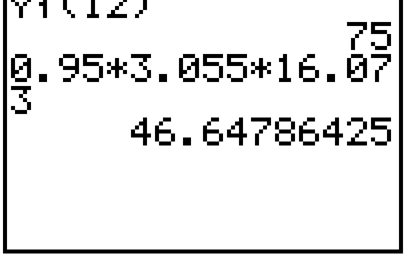
gemiddeld aantal gepresteerde uren : 10 uur
gemiddelde score : 65 op 100

regressielijn : y = 5x + 15

y(12) = 75

r = 0,95

cov(x,y) = r.s_x.s_y = 46,65



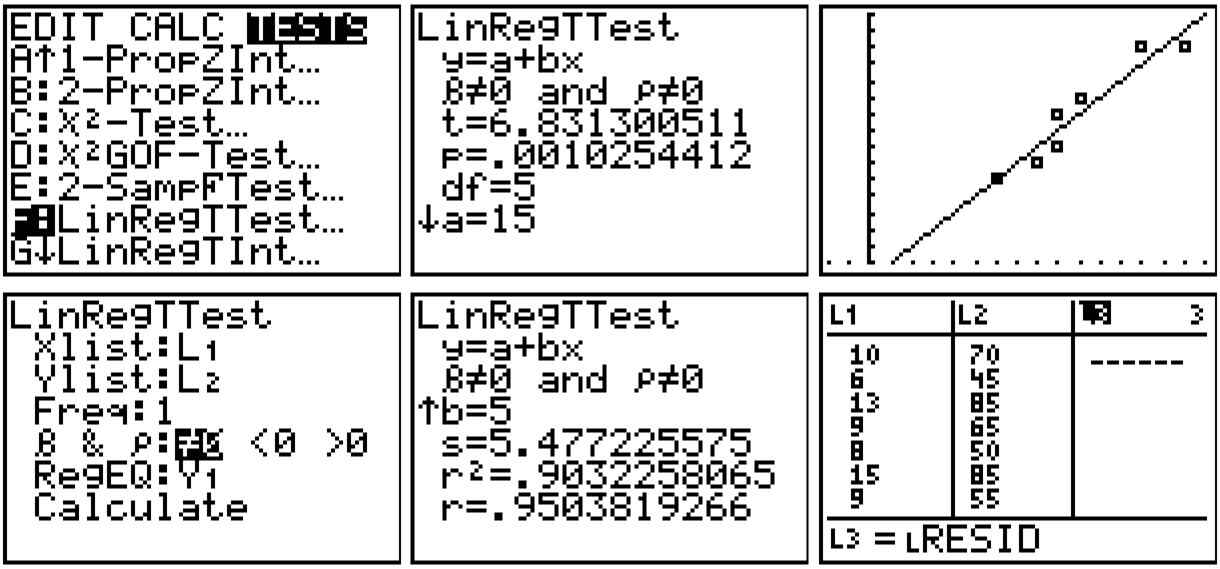
oplossing 2

De regressierechte kan je ook bekomen via een test, namelijk LinRegTTest. Deze berekent een lineaire regressie voor de ingevoerde gegevens en een t-test voor de waarde van de richtingscoëfficiënt en de correlatiecoëfficiënt.

De nulhypothese H_o : rico = 0 (en dus correlatie = 0) wordt getest ten opzichte de alternatieven rico $\neq$ 0 , rico > 0 en/of rico < 0.

Wanneer LinRegTTest wordt uitgevoerd, wordt de lijst van de resten automatisch gecreëerd en opgeslagen in de lijstnaam RESID (2nd [List] Names).

STAT Tests LinRegTTest



Bibliografie

Bij het samenstellen van dit cahier heb ik bij het hoofdstuk toetsen van hypothesen enkele opgaven gebruikt uit de verzameling oefeningen die ik van mijn vroegere leerlingen heb gekregen toen ze een woordje uitleg kwamen vragen i.v.m. hun cursus statistiek in het hoger onderwijs.

Zo komen er o.a. enkele opgaven uit de cursus "*Statistiek II-B van J.Hendrickx, A.Maes en J.Walrave uitgave 2004*". Een zeer verzorgde cursus die ik iedereen aanbeveel die zich verder wil verdiepen in de ‘hogere’ statistiek.

Honden bestaan maar dé hond bestaat niet. Hond is een verzamelnaam voor heel wat rassen elk met hun eigen uiterlijke kenmerken en karaktereigenschappen. Wiskunde bestaat, maar dé wiskunde bestaat niet. Wiskunde is een verzamelnaam voor heel wat disciplines zoals algebra, meetkunde, analyse, discrete wiskunde, … elk met hun eigen structuren en wetmatigheden die vanaf een zeker niveau met elkaar interfereren.
Zo ook kan men statistiek opvatten als een verzamelnaam en verder opdelen in verschillende subdomeinen.

Een onderzoeker of enquêteur verzamelt gegevens die hij via de beschrijvende statistiek tracht de modelleren met behulp van frequentietabellen, histogrammen en steekproefkarakteristieken.

Een wiskundige filosoof gooit niet 6000 maal met een dobbelsteen maar zal vanuit de kansrekening wiskundige kansmodellen (normale verdeling, binomiale verdeling, poissonverdeling, …) opstellen. Dit noemt men de verklarende statistiek.

Wanneer een onderzoeker zijn proefondervindelijk model wil toetsen aan een model uit de verklarende statistiek betreedt hij het terrein van de inferentiële of inductieve statistiek. We vinden dit terug in hoofdstukken zoals betrouwbaarheidsintervallen en toetsen van hypothesen.

Naast statistiek met betrekking tot slechts één toevalsvariabele, is er ook een tak van de statistiek die het verband tussen meerdere variabelen onderzoekt. Het bekendste voorbeeld is hier de lineaire regressie.

De bedoeling van dit cahier is een brug te bouwen tussen de statistiek van de derde graad secundair onderwijs en de statistiek waarmee heel wat leerlingen geconfronteerd worden in hun verdere opleiding. Het is een uitbreiding van de statistiek die leerlingen in het secundair tegenkomen. Dankzij de TI-84 vervaagt het rekenwerk en zijn formules vlotter toegankelijk, zodat leerkrachten die via projectwerk verder willen gaan, zich kunnen concentreren op de begripsvorming en de diepere achtergrond achter het verhaal van de statistiek.

PHILIP BOGAERT is leerkracht aan het Sint-Maarteninstituut te Aalst, medewerker van T³ Vlaanderen en lid van de stuurgroep Wiskunde Oost-Vlaanderen. Hij is ook medeauteur van een reeks wiskundeboeken voor de 2de en 3de graad.

juni 2010

TEXAS INSTRUMENTS