Statistiek voor MBW / KW
Uitwerkingen huiswerkopgaven

Dr. ir. D.A.M.P. Blom & Dr. J.B.M. Melissen

2025

Week 5: de normale verdeling en de centrale limietstelling

Hoofdstuk 5

Opdracht 5.m1: Voor een willekeurige normale verdeling geldt dat de volgende drie grootheden aan elkaar gelijk moeten zijn:

(a) gemiddelde, standaarddeviatie en mediaan.

(b) modus, de totale kans en de variantie

(c) mediaan, gemiddelde en modus

(d) standaarddeviatie, variantie en spreidingsbreedte

Uitwerking

Het juiste antwoord is (c).

Opdracht 5.m3: Voor de populatie van mannelijke economiestudenten is bekend dat het lichaamsgewicht kan worden beschreven door middel van een normaal verdeelde kansvariabele X met $\mu = 75$ kilogram en een standaarddeviatie van 6 kilogram. Hieruit volgt dat het percentage studenten van deze populatie dat een gewicht van meer dan 85 kilogram heeft ongeveer bedraagt:

(a) $95,2$

(b) 66

(c) 34

(d) $4,8$

Uitwerking

Om dit percentage te berekenen, berekenen we eerst de z-score behorende bij $x = 85$:

$$z = \frac{x-\mu}{\sigma} = \frac{85-75}{6} \approx 1,6667$$

maken we gebruik van de functie normalcdf op de grafische rekenmachine. In het bijzonder geldt namelijk dat de kans op een gewicht van meer dan 85 kilogram gelijk is aan:

$$\begin{aligned} P(X > 85) &= P(Z > 1,6667) \\ &= \text{normalcdf}(a = 1,6667; b = 10^{10}; \mu = 0; \sigma = 1) \\ &\approx 0,0478. \end{aligned}$$

Met ongeveer $4,78\%$ kans heeft een economiestudent een gewicht van meer dan 85 kilogram. Het juiste antwoord is dus (d).

Opdracht 5.m4: Het aantal jaarlijkse reiskilometers aan woon-werkverkeer per auto is voor de werknemers van een groot concern nauwkeurig geregistreerd. Het bleek hierbij dat dit aantal goed kon worden beschreven door een normale verdeling met $\mu = 8000$ kilometer en $\sigma = 2500$. Op basis van deze verdeling kan worden vastgesteld dat de 10% werknemers die het grootste aantal woon-werkkilometers maken, per jaar per persoon meer reizen dan ongeveer ...

(a) 8000 km.

(b) 10500 km.

(c) 11200 km.

(d) 12100 km.

Uitwerking

In dit geval willen we een grenswaarde berekenen, dus moeten we gebruik maken van de functie invNorm op de grafische rekenmachine. Noteer met X de normaal verdeelde kansvariabele die het aantal woon-werkkilometers van een willekeurige werknemer beschrijft. In het bijzonder geldt dat de grenswaarde g van het aantal kilometers zodat 10% van de werknemers meer dan g km per persoon per jaar reizen gelijk is aan:

$$\begin{aligned} P(X > g) = 0,10 \rightarrow g &= \text{invNorm}(opp = 1 - 0,10 = 0,90; \mu = 8000; \sigma = 2500) \\ &\approx 11204\ km. \end{aligned}$$

Het juiste antwoord is dus (c).

Opdracht 5.m5: Van een normaal verdeelde kansvariabele X is gegeven dat deze een standaarddeviatie heeft die gelijk is aan 9. Verder is gegeven dat $P(X < 15,5) = 0,3085$. De waarde van μ van deze variabele is dan ...

(a) 16

(b) 17

(c) 20

(d) 11

Uitwerking

In dit geval willen we bepalen wat de waarde van μ is zodanig dat geldt:

$$\text{normalcdf}(a = -10^{10}; b = 15,5; \mu = ?; \sigma = 9) = 0,3085.$$

In de grafische rekenmachine kun je in het functiescherm het volgende invullen:

$$y_1 = \text{normalcdf}(-10^{10}; 15,5; x; 9)$$
$$y_2 = 0,3085$$

De optie intersect (2nd-TRACE-5) geeft in dat geval een snijpunt voor $x = 20,0010$. Dit betekent dat μ een waarde van ongeveer 20 heeft. Het juiste antwoord is dus (c).

Opdracht 5.2: Een vulmachine vult pakken rijst waarop staat dat daarin 1000 gram rijst zit. Omdat er altijd wat variatie zit in de afgeleverde hoeveelheden, staat de machine ingesteld op een gemiddeld vulgewicht van 1010 gram. We mogen ervan uitgaan dat de afgeleverde hoeveelheden beschreven worden door een normale verdeling met $\mu = 1010$ gram. De standaarddeviatie is echter onbekend.

(a) Bij nauwkeurig nawegen van een groot aantal pakken bleek dat in 20% van de gevallen een pak minder dan 1000 gram rijst bevat. Bereken op basis van deze informatie de standaarddeviatie van de vulgewichten.

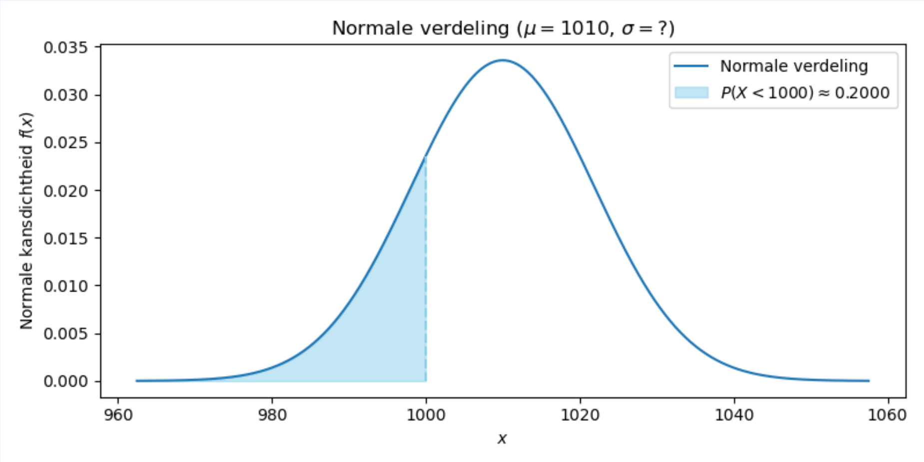
Uitwerking

Laat X de normaal verdeelde kansvariabele zijn met verwachtingswaarde $\mu = 1010$ en onbekende standaarddeviatie σ die het gewicht van een pak rijst beschrijft. In dat geval is de kansvariabele $Z = \frac{X-\mu}{\sigma}$ standaardnormaal verdeeld, oftewel $Z \sim N(0,1)$.
De z-waarde waarvoor geldt dat $P(Z < z) = 0,20$ is gelijk aan

$$z = \text{InvNorm}(opp = 0,20; \mu = 0; \sigma = 1) \approx -0,8416.$$

In andere woorden: de waarde 1000 g ligt 0,8416 standaarddeviaties onder het gemiddelde, oftewel:

$$\frac{1000 - 1010}{\sigma} = -0,8416 \rightarrow 1000 = 1010 - 0,8416 \cdot \sigma \rightarrow \sigma \approx 11,8818$$



(b) Hoe groot is de kans dat een willekeurig pak rijst meer weegt dan 1020 gram?

Uitwerking

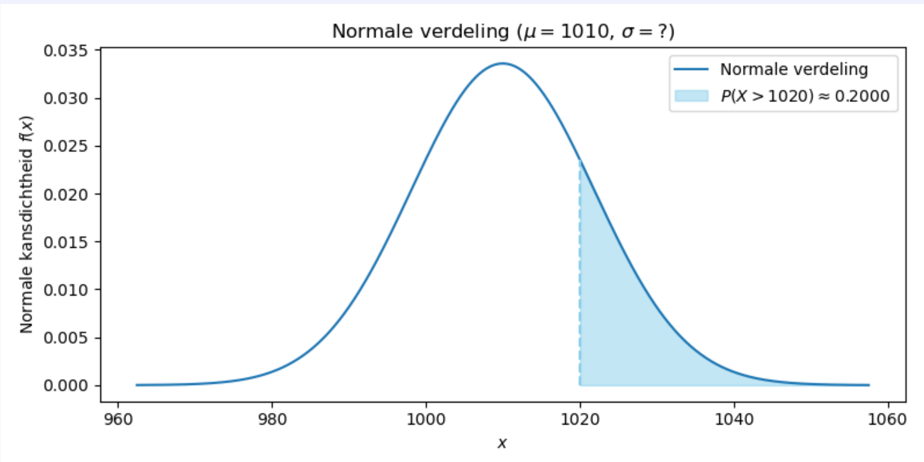
Zodra we de verwachtingswaarde μ en de standaarddeviatie σ van een normale verdeling weten, kunnen we de z-waarde bepalen van een specifieke uitkomst x. In dat geval is de z-waarde die hoort bij 1020 gram gelijk aan

$$z = \frac{x - \mu}{\sigma} = \frac{1020 - 1010}{11,8818} \approx 0,8416.$$

De kans dat een willekeurig pak rijst meer dan 1020 gram weegt is gelijk aan:

$$\begin{aligned} P(X > 1020) &= P(Z > 0,8416) \\ &= \text{normalcdf}(a = 0,8416; b = 10^{10}; \mu = 0; \sigma = 1) \\ &\approx 0,2000. \end{aligned}$$

Merk op dat deze kans hetzelfde is als de kans $P(X < 1000)$, aangezien de kansverdeling van X symmetrisch is rond $\mu = 1010$.



(c) Stel we kiezen willekeurig 16 pakken rijst. Hoe groot is de kans dat deze gemiddeld minder dan 1000 gram rijst bevatten?

Uitwerking

Laat $X_1, X_2, \ldots, X_16$ de gewichten zijn van 16 willekeurig gekozen pakken rijst. Volgens de centrale limietstelling geldt voor het gemiddelde gewicht van deze 16 pakken rijst:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_{16}}{16} \sim N(\mu_{\overline{X}} = 1010; \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{16}} \approx 2,9705)$$

.

We berekenen allereerst de z-score behorende bij een gemiddeld gewicht van 1000 gram. Die is gelijk aan:

$$z = \frac{x - \mu}{\sigma} = \frac{1000 - 1010}{2,9705} \approx -3,3665.$$

De kans dat de 16 willekeurig gekozen pakken rijst gemiddeld minder dan 1000 gram rijst bevatten is dan gelijk

$$\begin{aligned} P(\overline{X} < 1000) &= P(Z < -3,3665) \\ &\text{normalcdf}(a = -10^{10}; b = -3,3665; \mu = 0; \sigma = 1) \\ &\approx 3,8072 \cdot 10^{-4}. \end{aligned}$$

Opdracht 5.4: Bepaal voor de standaardnormaal verdeelde variabele Z:

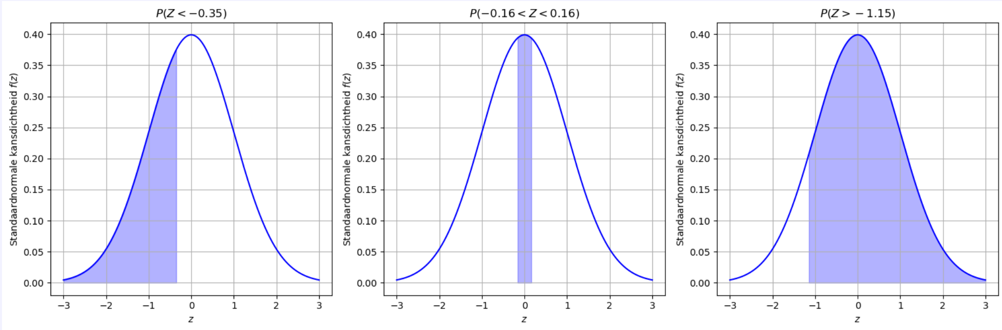
- $P(Z < -0,35)$
- $P(-0,16 < Z < 0,16)$
- $P(Z > -1,15)$

Uitwerking

We kunnen deze kansen bepalen aan de hand van de functie normalcdf op de grafische rekenmachine.

$$\begin{aligned} P(Z < -0,35) &= \text{normalcdf}(a = -10^{10}; b = -0,35; \mu = 0; \sigma = 1) \\ &\approx 0,3632 \\ P(-0,16 < Z < 0,16) &= \text{normalcdf}(a = -0,16; b = 0,16; \mu = 0; \sigma = 1) \\ &\approx 0,1271 \\ P(Z > -1,15) &= \text{normalcdf}(a = -1,15; b = 10^{1}0; \mu = 0; \sigma = 1) \\ &\approx 0,8749 \end{aligned}$$

Deze kansen kunnen ook visueel geïllustreerd worden als de oppervlaktes van de blauw gearceerde gebieden in de onderstaande figuur:



Opdracht 5.5: Een garagebedrijf bestudeert de tijdsduur waarin de jaarlijkse servicebeurt van een auto kan worden uitgevoerd. Deze kan worden beschouwd als een normaal verdeelde variabele X met $\mu = 120$ minuten en $\sigma = 20$ minuten.

(a) Hoe groot is de kans dat een servicebeurt meer dan 150 minuten vergt?

Uitwerking

De z-score die hoort bij een uitkomst van $x = 150$ minuten is gelijk aan:

$$z = \frac{x-\mu}{\sigma} = \frac{150-120}{20} = 1,5$$

De kans dat een servicebeurt meer dan 150 minuten vergt is dan gelijk aan:

$$\begin{aligned} P(X > 150) &= P(Z > 1,5) \\ &= \text{normalcdf}(a = 1,5; b = 10^{10}; \mu = 0; \sigma = 1) \\ &\approx 0,0668 \end{aligned}$$

(b) Een klant komt zijn auto brengen voor service. De werkzaamheden gaan onmiddellijk beginnen. De garagehouder zegt tegen de klant: “Komt u over x minuten maar terug”. Hoe groot moet x worden gekozen als we eisen dat de klant minder dan 5% kans wil hebben dat hij nog moet wachten als hij zich na X minuten weer meldt bij de garage? En hoe zit dat bij 1% kans?

Uitwerking

De klant heeft minder dan 5% kans om nog te moeten wachten als hij terugkomt bij de garage na x minuten als geldt:

$$P(X > x) < 0,05 \rightarrow P(X \leq x) > 1 - 0,05 = 0,95$$

We willen dus een grenswaarde bepalen, en dat kunnen we doen met de functie InvNorm:

$$x = \text{InvNorm}(opp = 0,95; \mu = 120; \sigma = 20) \approx 152,8971 \text{ minuten.}$$

In het geval van minder dan 1% is de berekening op een soortgelijke manier te maken, namelijk:

$$P(X > x) < 0,01 \rightarrow P(X \leq x) > 1 - 0,01 = 0,99$$

We willen dus een grenswaarde bepalen, en dat kunnen we doen met de functie InvNorm:

$$x = \text{InvNorm}(opp = 0,99; \mu = 120; \sigma = 20) \approx 166,5270 \text{ minuten}.$$

Opdracht 5.6: De gewichten van appels uit een grote partij blijken normaal verdeeld te zijn met $\mu = 100$ gram en $\sigma = 20$ gram. We willen deze appels in vijf gewichtsklassen verdelen die allemaal evenveel appels bevatten.

(a) Wat is de klassengrens van de 20% appels met het geringste gewicht?

Uitwerking

Laat X de normaal verdeelde kansvariabele zijn die het gewicht van een appel uit een grote partij meet. Er geldt dus $X \sim N(\mu = 100; \sigma = 20)$. Je wilt de appels opdelen in vijf gewichtsklassen die elk evenveel appels bevatten, d.w.z. dat de kans dat een appel in een bepaalde gewichtsklasse zit moet gelijk zijn voor elke gewichtsklasse. Er zijn vijf klassen dus die kans is telkens $\frac{1}{5} = 0,2$. De klasse met de 20% lichtste appels loopt van $-\infty$ tot g_1 voor een bepaald gewicht g_1 en er geldt $P(X < g_1) = 0,2$. Dat betekent dus dat

$$g_1 = \text{InvNorm}(opp = 0,2; \mu = 100; \sigma = 20) \approx 83,1676 \text{ g}.$$

(b) Bepaal ook de andere klassengrenzen?

Uitwerking

Om de overige klassengrenzen te bepalen, gebruiken we soortgelijke berekeningen. Voor de tweede klassegrens g_2 geldt

$$P(g_1 < X < g_2) = 0,2$$

Omdat voor g_1 geldt dat $P(X < g_1) = 0,2$, geldt nu dat:

$$P(X < g_2) = P(X < g_1) + P(g_1 < X < g_2) = 0,2 + 0,2 = 0,4.$$

We kunnen de tweede klassegrens dus bepalen als volgt:

$$g_2 = \text{InvNorm}(opp = 0,4; \mu = 100; \sigma = 20) \approx 94,9331 \text{ g}.$$

Op een soortgelijke manier vinden we g_3 en g_4:

$$g_3 = \text{InvNorm}(opp = 0,6; \mu = 100; \sigma = 20) \approx 105,0669 \text{ g}.$$
$$g_4 = \text{InvNorm}(opp = 0,8; \mu = 100; \sigma = 20) \approx 116,8324 \text{ g}.$$

LET OP: de eerste klasse loopt van $-\infty$ tot g_1. Aangezien appels geen negatief gewicht kunnen hebben lijkt dit raar. Dit is echter een gevolg van de aanname dat we met een normale verdeling werken, die alle reële getallen als uitkomst kan hebben. De kans op een appel met een negatief gewicht is echter bijzonder klein. De z-score is gelijk aan $z = \frac{0-100}{20} = -5$, dus

$$\begin{aligned} P(X < 0) &= P(Z < -5) \\ &= \text{normalcdf}(a = -10^10; b = -5; \mu = 0; \sigma = 1) \\ &= 2,8711 \cdot 10^{-7} \end{aligned}$$

Opdracht 5.10: In een kliniek wordt van een groep 50-plussers bloed afgenomen voor de bepaling van het cholesterolgehalte. Het is een bekend gegeven dat de hoeveelheid cholesterol per buisje bloed kan worden weergegeven door een normaal verdeelde variabele X met $\mu = 5,20$ mg en $\sigma = 0,9$ mg. Het 98-ste percentiel van een verdeling is het punt op de x-as waarboven zich 2% van de waarnemingen bevindt, dus waaronder 98% van de waarnemingen ligt.

(a) Welke x-waarde geeft hier het 98-ste percentiel aan?

Uitwerking

Gegeven is dat de hoeveelheid cholesterol (in mg) per buisje bloed wordt gegeven door $X \sim N(\mu = 5,20; \sigma = 0,9)$. Percentielen kunnen we berekenen met behulp van de functie InvNorm. Voor het 98-ste percentiel gebruiken we als oppervlakte 0,98:

$$x = \text{InvNorm}(opp = 0,98; \mu = 5,20; \sigma = 0,9) \approx 7,0484 \text{ mg}.$$

(b) Hoeveel standaarddeviaties ligt dat punt verwijderd van het gemiddelde van 5,20 mg.

Uitwerking

Om te bepalen hoeveel standaarddeviaties x verwijderd ligt van het gemiddelde 5,20 mg, moeten we de z-score berekenen:

$$z = \frac{x-\mu}{\sigma} = \frac{7,0484 - 5,20}{0,9} \approx 2,0537.$$

De waarde $x \approx 7,0484$ mg ligt dus meer dan 2 standaarddeviaties boven het gemiddelde $\mu = 5,20$ mg.

(c) In een gemeente wordt een bevolkingsonderzoek gehouden naar het cholesterolgehalte. Hierbij waren 8500 personen betrokken. Mensen met een cholesterolwaarde boven het 98-ste percentiel worden voor nader onderzoek opgeroepen. Hoeveel personen krijgt naar verwachting een oproep?

Uitwerking

Per definitie van percentielen, geldt dat 2% van de bevolking een cholesterolwaarde heeft boven het 98-ste percentiel. In een groep van 8500 personen verwacht je dus dat $0,02 \cdot 8500 = 170$ mensen een oproep zullen krijgen.

(d) Voor een willekeurig persoon uit de groep 50-plussers bepalen we het cholesterolgehalte. Noem deze uitkomst y. Wat is het twee-sigmagebied voor y? en het drie-sigmagebied?

Uitwerking

Het twee-sigmagebied voor y is het gebied tot maximaal 2 standaarddeviaties van het gemiddelde, oftewel het interval tussen $\mu - 2 \cdot \sigma$ en $\mu + 2 \cdot \sigma$. Er geldt dat:

$$\begin{aligned} \mu - 2 \cdot \sigma &= 5,20 - 2 \cdot 0,9 = 3,40 \\ \mu + 2 \cdot \sigma &= 5,20 + 2 \cdot 0,9 = 7,00 \end{aligned}$$

Het twee-sigmagebied loopt dus van $3,40$ tot $7,00$. Op een soortgelijke manier vinden we dat het drie-sigmagebied loopt van $2,50$ tot $7,90$.

Opdracht 5.12: De tijd die een vertegenwoordiger nodig heeft voor het bezoeken van een klant wordt weergegeven door de kansvariabele X. Op grond van ervaring is bekend dat X normaal verdeeld is met $\mu = 45$ minuten en $\sigma = 10$ minuten (in de tijdsduur X is ook de reistijd opgenomen). De tijdsduren van bezoeken zijn onderling onafhankelijk.

(a) Hoe groot is de kans dat een willekeurig bezoek meer dan 60 minuten vergt?

Uitwerking

Gegeven is dat de tijdsduur $X \sim N(\mu = 45; \sigma = 10)$. De kans dat een willekeurig bezoek meer dan 60 minuten vergt is dus $P(X > 60)$.
De z-score van $x = 60$ gegeven $X \sim N(\mu = 45; \sigma = 10)$ is gelijk aan:

$$z = \frac{x - \mu}{\sigma} = \frac{60 - 45}{10} = 1,5.$$

Hierdoor geldt dus

$$\begin{aligned} P(X > 60) &= P(Z > 1,5) \\ &= \text{normalcdf}(a = 1,5; b = 10^{10}; \mu = 0; \sigma = 1) \\ &\approx 0,0668. \end{aligned}$$

Met $6,68\%$ kans vergt een willekeurig bezoek meer dan 60 minuten.

(b) Hoe groot is de kans dat acht bezoeken meer dan $6\frac{1}{2}$ uur vergen?

Uitwerking

Laat nu X_i de kansvariabelen zijn die de tijd van het bezoek aan klant $i = 1, 2, \ldots, 8$ meet. Volgens de centrale limietstelling geldt dat de som $\sum X = X_1 + X_2 + \ldots + X_8$ normaal verdeeld is met gemiddelde $n \cdot \mu = 8 \cdot 45 = 360$ minuten en standaardafwijking $\sigma \cdot \sqrt{n} = 10 \cdot \sqrt{8} \approx 28,2843$.

De z-score van $x = 390$ ($6\frac{1}{2}$ uur $= 390$ minuten) gegeven $\sum X \sim N(360; 28,2843)$ is gelijk aan

$$z = \frac{x-\mu}{\sigma} = \frac{390-360}{28,2843} \approx 1,0607$$

De kans dat acht bezoeken meer dan $6\frac{1}{2}$ uur vergen is gelijk aan:

$$\begin{aligned} P(\sum X > 390) &= P(Z > 1,0607) \\ &= \text{normalcdf}(a = 1,0607; b = 10^{10}; \mu = 0; \sigma = 1) \\ &\approx 0,1444 \end{aligned}$$

Met $14,44\%$ kans zullen de acht bezoeken in totaal meer dan $6\frac{1}{2}$ uur vergen.

(c) De vertegenwoordiger moet op een zekere dag tien klanten bezoeken. Hoeveel tijd T moet hij reserveren om met 95% kans de tien bezoeken binnen de tijdsduur T te kunnen afhandelen?

Uitwerking

Laat nu X_i de kansvariabelen zijn die de tijd van het bezoek aan klant $i = 1, 2, \ldots, 10$ meet. Volgens de centrale limietstelling geldt dat de som $\sum X = X_1 + X_2 + \ldots + X_8$ normaal verdeeld is met gemiddelde $n \cdot \mu = 10 \cdot 45 = 450$ minuten en standaardafwijking $\sigma \cdot \sqrt{n} = 10 \cdot \sqrt{10} \approx 31,6228$. We willen nu de grens T berekenen waarvoor geldt dat $P(\sum X \leq T) = 0,95$. De bijbehorende z-score is

$$P(Z \leq z) = 0,95 \rightarrow z = \text{InvNorm}(opp = 0,95; \mu = 0; \sigma = 1) \approx 1,6449.$$

Terugrekenen naar de bijbehorende uitkomst T voor de kansverdeling $N(450; 31,6228)$ van $\sum X$ geeft:

$$T = \mu + z \cdot \sigma = 450 + 1,6449 \cdot 31,2668 \approx 502,0148.$$

De vertegenwoordiger moet meer dan 8 uur en 22 minuten reserveren om de tien bezoeken mete 95% kans binnen de tijd af te kunnen handelen.