Universiteit Utrecht

Faculteit Bètawetenschappen

Verdieping in het studiepad statistiek en kansrekening

Bachelor Scriptie

Lisa Broman

6403190

Wiskunde en Toepassingen

Begeleider:

Dr. Matthijs Ruijgrok
Mathematisch Instituut

24 juni 2022

1 Inleiding

Stel je voor: je hebt net je middelbare school afgerond en hebt eindelijk de knoop doorgehakt. Je gaat wiskunde studeren. Tijdens het eerste jaar van je studie wiskunde kom je er achter dat je studie niet de enige keuze is die je moet maken. Er zijn verschillende vakken waar je uit kan kiezen die allemaal weer horen bij andere studiepaden. Maar welke vakken kies je dan? Veel studenten lopen tegen dit probleem aan. Ze gaan zomaar vakken kiezen zonder rekening te houden met welke vakken er nodig zijn bij de richting die zij willen doen. Dit komt omdat het niet duidelijk is welke vakken bij elkaar in het studiepad horen of welke vakken in elkaars verlengde liggen. Om eerstejaars wiskundestudenten hierin te ondersteunen pluizen we een studiepad voor ze uit. Hierin laten we zien hoe de vakken binnen een studiepad bij elkaar passen en wat je met dit studiepad kan doen. In deze scriptie gaan we kennis maken met het studiepad statistiek en kansrekening.

Veel mensen hebben een haat-liefdeverhouding met statistiek. Schrijver Mark Twain zei ooit "There are three kinds of lies: lies, damned lies and statistics". Humorist Stephen Leacock zei: "In ancient times they had no statistics so they had to fall back on lies". Ook zijn er meerdere boeken geschreven waar de gebreken van statistiek benadrukt worden. Zoals het boek "How to lie with statistics"van Darrell Huff en het boek "Standard deviations, flawed assumptions, tortured data and other ways to lie with statistics"van Gary Smith. Aan de andere kant is statistiek ook erg belangrijk. Barack Obama zei in 2010: "Statistical data drives countless decisions which impact our Nation". Ook hebben wij zelf het belang van statistiek kunnen ervaren tijdens de coronapandemie bijvoorbeeld bij het gebruik van statistiek bij het voorspellen van het verloop van de coronabesmettingen. Het is vooral belangrijk dat statistiek op de juiste manier gebruikt wordt. Dat is alleen niet zo makkelijk. Wiskundige Paul Levrie (2016) zei: "Het grote probleem met statistiek is dat het echt niet moeilijk is om de verkeerde conclusies te trekken"[3]. Deze controversiteit van statistiek is juist wat het zo uniek en spannend maakt.

We gaan de student laten proeven aan de wondere wereld van statistiek en wie weet verandert de haat wel naar liefde en smaakt het naar meer!

Inhoudsopgave

2 Het studiepad statistiek en kansrekening

Een studiepad bevat gebonden keuzes, modellering en vrije keuzevakken. Deze vakken zijn gebasseerd op de voorkennis die nodig is bij de verschillende trajecten binnen de master mathematical sciences.

Bij het studiepad statistiek horen als gebonden keuzevakken: Functies en reeksen, Differentiaal vergelijkingen en Discrete wiskunde. Deze gebonden keuzevakken komen bijna volledig overeen met de andere studiepaden, omdat dit belangrijke basiskennis is binnen de wiskunde. Specifieker voor dit studiepad zijn de modelleringsvakken: stochastische processen en financiële wiskunde. Dan zijn er nog de vrije keuzevakken: statistiek, analyse in meerdere variabelen, speltheorie, maat en integratie en functionaalanalyse [17].

Dit studiepad is ook een goede voorbereiding voor de master econometrie. Dit geeft een toepassing van deze wiskundevakken binnen de economie. Dat zorgt voor een minder abstract en meer tastbaarder gebruik van het studiepad statistiek. De vakken van het studiepad statistiek en kansrekening zijn dan ook dezelfde vakken als voor het studiepad voorbereiding master econometrie. Hier komt nog als keuzevak econometrie bij en als profileringsvakken microeconomie, macroeconomie en microeconomie in financiële markten [17].

Om de eerstejaarsstudent kennis te laten maken met de pracht van statistiek gaan we ook kijken hoe we statistiek terugzien bij de vakken financiële wiskunde en econometrie. Dit zal laten zien welke rol statistiek heeft in meerdere economische fenomenen.

3 Statistiek in de financiële wiskunde

3.1 Voorkennis statistiek

Er zijn een paar essentiële onderwerpen die de basis vormen van vele toepassingen in de statistiek. We bespreken kort hiervan de onderwerpen die van belang zijn voor de toepassingen waar wij ons in gaan verdiepen.

3.1.1 Verwachtingswaarde en variantie

De vewachtingswaarde $\mathbb{E}[X]$ is het gewogen gemiddelde van alle mogelijke uitkomsten van stochast X met als weegfactor de kans dat een uitkomst zich voordoet [24]. We kunnen $\mathbb{E}[X]$ ook definiëren als:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx \tag{1}$$

Hierbij is X een continue stochastische variable en f_X de kansdichtheidfunctie. Voor de verwachtingswaarde gelden de volgende rekenregels [24] voor stochasten X en Y en een constante $c \in \mathbb{R}$:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

$$\mathbb{E}[c] = c$$

Als stochasten X en Y onderling onafhankelijk zijn geldt ook:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

De variantie $Var(X)$ is een maat voor de spreiding van een reeks waarden. Dit is ook wel de mate waarin de waarden onderling verschillen. De variantie van stochast X is dan gedefinieerd als de verwachtingswaarde van de kwadratische afwijkingen van de verwachtingswaarde van X:

$$\begin{aligned} Var(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned} \tag{2}$$

Voor de variantie gelden de volgende rekenregels [7] met constante $c \in \mathbb{R}$ en stochast X:

$$Var(c) = 0$$

$$Var(X + c) = Var(X)$$

$$Var(cX) = c^2 Var(X)$$

Als stochasten X en Y onafhankelijk zijn geldt:

$$Var(X + Y) = Var(X) + Var(Y)$$

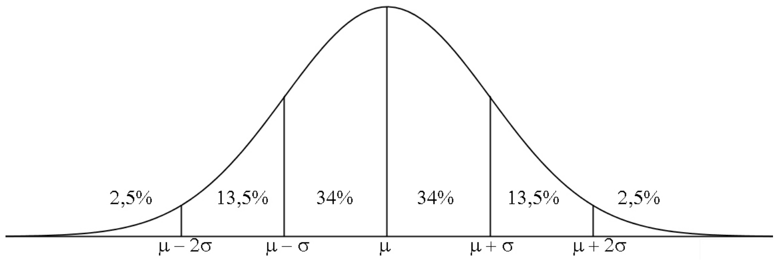
De standaardafwijking is niet hetzelfde als de variantie. Dit geeft aan hoe ver iedere waarde gemiddeld genomen van het gemiddelde verwijderd is [13]. Hierbij geldt dat de standaardafwijking $\sigma = \sqrt{Var(x)}$.

3.1.2 Normaalverdeling

Een normaalverdeling is een verdeling die bekend staat om zijn symmetrische klokvormige curve. Een normaalverdeelde stochast X heeft een verwachtingswaarde μ en een standaardafwijking σ. Het bijzondere is dat zowel de modus als de mediaan gelijk is aan het gemiddelde μ. De normaalverdeling stelt dat het waarschijnlijker is dat een waarde dichtbij de verwachtingswaarde ligt, zoals in figuur 1 te zien is. De kansdichtheidsfunctie [25] van een normaalverdeling met gemiddelde μ en standaardafwijking σ is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{3}$$

In andere woorden: Als de gemiddelde lengte van een vrouw 164 cm is met een standaard afwijking van 4 cm. Dan zal 68% van de vrouwen een lengte hebben tussen de 160 cm en 168 cm.



Figuur 1: Normaalverdeling

3.1.3 Karakteristieke functie

Een karakteristieke functie van een stochastische variable X is in de kansrekening en statistiek een functie voor een reële t:

$$\phi_X(t) = \mathbb{E}(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx \tag{4}$$

Hierbij is $\mathbb{E}$ de verwachtingswaarde en $f_X(x)$ de kansdichtheidfunctie van stochast X [27].

Het mooie aan de karakteristieke functie is dat de eigenschappen van een kansverdeling weergegeven kunnen worden. Daarnaast is deze functie altijd gedefinieerd. Er is een verband tussen het gedrag van de karakteristieke functie van een verdeling en de eigenschappen van deze verdeling. Doordat het een continue differentieerbare functie is, kan de afgeleide gebruikt worden als middel voor het beschrijven van een kansverdeling. Daarnaast is het dan ook mogelijk het limiet te bepalen en te bekijken wat er met een kansverdeling gebeurt als het aantal stochasten naar oneindig gaat.

We kunnen e^{itx} benaderen met de (oneindige) Taylorreeks. Volgens de Taylorreeks kan een reeële of complexe functie benaderd worden die differentieerbaar is in een reeël of complex getal a met de volgende machtreeks [28]:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... \tag{5}$$

Als we e^{itx} met de taylorreeks benaderen in $a = 0$ krijgen we:

$$e^{itx} = 1 + itx + \frac{(itx)^2}{2!} + ... = 1 + itx - \frac{t^2}{2}x^2 + ... \tag{6}$$

Als we deze benadering substitueren in de karateristieke functie krijgen we de volgende uitdrukking:

$$\begin{aligned}\phi_X(t) &= \int_{-\infty}^{\infty} (1 + itx - \frac{t^2}{2}x^2 + ...)f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_X(x)dx + \int_{-\infty}^{\infty} (itxf_X(x))dx - \int_{-\infty}^{\infty} (\frac{t^2}{2}x^2)f_X(x)dx + ... \\ &= 1 + it\mathbb{E}(X) - \frac{t^2}{2}\mathbb{E}(X^2) + ...\end{aligned} \tag{7}$$

Stel we nemen een stochast Y met een onbekende verdeling met $\mathbb{E}[Y] = 0$ en $Var[Y] = 1$. Voor de karakteristieke functie van Y geldt dan:

$$\phi_Y(t) = 1 + it\mathbb{E}(Y) - \frac{t^2}{2}\mathbb{E}(Y^2) + ... \tag{8}$$

Er geldt $\mathbb{E}[Y] = 0$ en $Var[Y] = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{E}(Y^2)$ dus $\mathbb{E}(Y^2) = Var[Y] = 1$. Dan krijgen we de karakteristieke functie:

$$\phi_Y(t) = 1 - \frac{t^2}{2} + ... \tag{9}$$

Deze karakteristieke functie wordt ook gebruikt voor het bewijs in 3.2. Tot nu toe hebben we het gehad over de karakteristieke functie van een onbepaalde kansverdeling. De karakteristieke functie van een normaalverdeelde stochast X met gemiddelde μ en variantie σ is:

$$\phi_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = e^{i\mu t - \frac{1}{2}\sigma t^2} \tag{10}$$

Als X normaalverdeelt is met $\mu = 0$ en $\sigma = 1$ dan geldt dat de karakteristieke functie gelijk is aan:

$$\phi_X(t) = e^{-\frac{1}{2}t^2} \tag{11}$$

Het uitrekenen van de integraal van $\phi_X(t)$ in (10) is een lastige bepaling door het complexe getal i. Om zelf een berekening te geven van de karakteristieke functie van een normaalverdeling, kijken we naar de moment genererende functie $m_X(t)$ Deze functie is bijna hetzelfde als de karakteristieke functie, behalve de complexe i. Ook deze functie wordt gebruikt als de eigenschappen van kansverdelingen weer te geven. Echter kan de moment genererende functie niet altijd gebruikt worden. Deze functie is gedefinieerd als [29]:

$$m_X(t) = \mathbb{E}(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \tag{12}$$

Als de stochast X normaal verdeeld is met $\mu = 0$ en $\sigma = 1$ krijgen we de volgende functie:

$$m_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}x^2} dx \tag{13}$$

Deze integraal kunnen wij wel makkelijk zelf bepalen door te kijken naar $e^{tx}e^{-\frac{1}{2}x^2}$. Dit kunnen we namelijk omschrijven als:

$$e^{tx}e^{-\frac{1}{2}x^2} = e^{-\frac{1}{2}(x-t)^2} e^{\frac{1}{2}t^2}$$

Dan kunnen we de term $e^{\frac{1}{2}t^2}$ uit de integraal halen. Dan krijgen we:

$$\begin{aligned} m_X(t) &= e^{\frac{1}{2}t^2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}dx \\ &= e^{\frac{1}{2}t^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2}dx \end{aligned} \tag{14}$$

In de integraal staat nu de kansdichtheidsfunctie van een normaalverdeling met gemiddelde t en variantie 1. Dan is de waarde van de integraal gelijk aan 1 [29]. Dus krijgen we dat:

$$m_X = e^{\frac{1}{2}t^2} \tag{15}$$

Hierin herkennen we de uitkomst van (11) terug, op een min naar. Dit is het gevolg van de complexe i in de karakteristieke functie.

Een karakteristieke functie heeft vele eigenschappen. Een van de eigenschappen betreft de karakteristieke functie van een lineaire combinatie van stochasten. Er geldt het volgende [27]:

Stelling 3.1. *Gegeven de onafhankelijke stochasten $X_1, ..., X_n$ en de constanten $a_1, ..., a_n$. De karakteristieke functie van de lineaire combininatie $a_1X_1 + ... + a_nX_n$ is*

$$\phi_{a_1X_1+...+a_nX_n}(t) = \phi_{X_1}(a_1t)\cdot ... \cdot \phi_{X_n}(a_nt)$$

Met al deze voorkennis kunnen we aan de slag met de centrale limietstelling.

3.2 Centrale limietstelling

Een belangrijke en vooral mooie stelling in de statstiek is de centrale limietstelling. Deze stelling wordt onder andere toegepast in de financiële wiskunde. De centrale limietstelling maakt het bijvoorbeeld makkelijker te begrijpen hoe herhaalde steekproeven zich gedragen. Hier gaan we later op in als we de statistiek gaan koppelen aan de econometrie. We beginnen met het definiëren van de stelling. Daarna verdiepen we ons in het bewijs en zullen we specifieke toepassingen van de stelling bekijken.

In 1733 was de eerste versie van de centrale limietstelling al bedacht door Abraham De Moivre. Hij gebruikte de normaal verdeling bij het werpen van een munt. Destijds was het concept niet populair. Totdat in 1812 het concept opnieuw werd geïntroduceerd door Pierre-Simon Laplace. Hij ontdekte dat het gemiddelde van onafhankelijke willekeurige variabelen de neiging heeft om een normaleverdeling te volgen. In 1901 werd de centrale limietstelling uitgebreid door Aleksandr Lyapunow [4] Hij ging het concept in algemene termen definiëren en bewijzen. Dit vormde de stelling die wij gebruiken tot op de dag van vandaag.

De centrale limietstelling is eigenlijk niet één stelling maar een collectie van stelling met allemaal andere randvoorwaarden. Zo is er de klassieke, Lyapunov, Lindeberg en multidimensionale Centrale limietstelling [21]. De bekenste is de Lindeberg–Lévy centrale limietstelling, ook wel de klassisieke centrale limietstelling genoemd. Deze stelling komt het meeste voor en zullen wij ook gebruiken.

Stelling 3.2. ***Lindeberg-Lévy centrale limietstelling*** *Gegeven n onafhankelijke, identiek verdeelde stochasten X_n met $\mathbb{E}[X_j] = \mu$ en $Var[X_j] = \sigma^2 < \infty$ en defineer $\bar{X_n} = \frac{1}{n}\sum_j^n X_j$. Dan convergeert de stochast $\sqrt{n}(\bar{X_n} - \mu)$ naar de normaal verdeling $\mathcal{N}(0, \sigma^2)$ als $n \to \infty$. [21]*

In andere woorden kunnen we ook wel zeggen dat het gemiddelde van de steekproef een normaal verdeling volgt als $n \to \infty$.

Voordat we de pracht van deze stelling gaan bekijken, gaan we eerst in op het bewijs achter de stelling om een beter inzicht te krijgen waarom de stelling geldt.

Bewijs. Gegeven is steekproefgemiddelde $\bar{X_n} = \frac{1}{n}\sum_j^n X_j$ waarbij X_j een onbekende verdeling heeft met $\mathbb{E}[X_j] = \mu$ and $Var[X_j] = \sigma^2$. Er geldt dan vanuit de rekenregels van de verwachtingswaarde en variantie dat:

$$\begin{aligned}\mathbb{E}[\bar{X_n}] &= \mathbb{E}[\frac{1}{n}\sum_{j=1}^{n} X_j] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[X_j] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mu \\ &= \frac{1}{n} \times n \times \mu = \mu\end{aligned} \tag{16}$$

$$
\begin{aligned}
Var[\bar{X_n}] &= Var[\frac{1}{n}\sum_{j=1}^{n} X_j] \\
&= (\frac{1}{n})^2 \sum_{j=1}^{n} Var[X_j] \\
&= \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n}
\end{aligned} \tag{17}
$$

We stellen de stochast Z_n op die $\bar{X_n}$ standaardiseert zodat Z_n een verdeling heeft met $\mu = 0$ en $\sigma^2 = 1$ en met $\bar{X_n}$ verdeeld met verwachtingswaarde μ en variantie $\frac{\sigma^2}{n}$. We definiëren Z_n als:

$$
\begin{aligned}
Z_n &= \frac{n\bar{X_n} - n\mu}{\sqrt{n}\sigma} \\
&= \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}\sigma} \\
&= \frac{\sum_{j=1}^{n}(X_j - \mu)}{\sqrt{n}\sigma} \\
&= \sum_{j=1}^{n}(\frac{X_j - \mu}{\sqrt{n}\sigma}) \\
&= \sum_{j=1}^{n} \frac{Y_j}{\sqrt{n}} \\
&= \frac{1}{\sqrt{n}}Y_1 + ... + \frac{1}{\sqrt{n}}Y_n
\end{aligned} \tag{18}
$$

Waarbij $Y_j = \frac{X_j - \mu}{\sigma}$ met $\mathbb{E}[Y_j] = 0$ en $Var[Y_j] = \frac{Var[X_j]}{\sigma^2} = 1$. Dus geldt er inderdaad voor Z_n dat $\mu = 0$ en $\sigma^2 = 1$ waardoor we de karakteristieke functie van Z_n kunnen definiëren vanuit stelling 3.1 als:

$$
\begin{aligned}
\phi_{Z_n}(t) &= \phi_{\frac{1}{\sqrt{n}}Y_1 + ... + \frac{1}{\sqrt{n}}Y_n} \\
&= \phi_{Y_1}(\frac{1}{\sqrt{n}}t) \cdot ... \cdot \phi_{Y_n}(\frac{1}{\sqrt{n}}t) \\
&= (\phi_Y(\frac{t}{\sqrt{n}}))^n \\
&= (1 - \frac{(\frac{t}{\sqrt{n}})^2}{2})^n \\
&= (1 - \frac{t^2}{2n})^n
\end{aligned} \tag{19}
$$

Dan geldt dat:

$$
\lim_{n \to \infty} \phi_{Zn}(t) = e^{-\frac{t^2}{2}} \tag{20}
$$

We weten vanuit functie (10) dat dit de karakteristiek functie is van een normaalverdeling met $\mu = 0$ en $\sigma^2 = 1$. Er geldt dat als de karakteristieke functie van een stochast convergeert naar een bepaalde verdeling, dan convergeert deze stochast naar dezelfde verdeling. Dus is $\lim_{n \to \infty} Z_n(t)$ convergeert ook naar een normaalverdeling met verwachtingswaarde nul en variantie een.

In (18) is Z_n gedefinieerd, deze kunnen we omschrijven zoals in 3.2.

$$\begin{aligned} Z_n &= \frac{n\bar{X}_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{n(\bar{X}_n - \mu)}{\sigma\sqrt{n}} \\ &= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \end{aligned} \tag{21}$$

Dan geldt ook dat het volgende limiet convergeert naar een normaalverdeling met $\mu = 0$ en $\sigma = 1$.

$$\lim_{n\to\infty} Z_n(t) = \lim_{n\to\infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \tag{22}$$

Na een vermenigvuldiging van σ geldt dan ook dat $\lim_{n\to\infty} \sqrt{n}(\bar{X}_n - \mu)$ convergeert naar een normaalverdeling met verwachtingswaarde nul en variantie σ^2,

Er geldt dus dat onafhankelijk van de verdeling van X_n, als $n \to \infty$ convergeert de stochast $\sqrt{n}(\bar{X}_n - \mu)$ naar de normaalverdeling $\mathcal{N}(0, \sigma^2)$. Hiermee is de Lindeberg-Lévy centrale limietstelling 3.2 bewezen. □

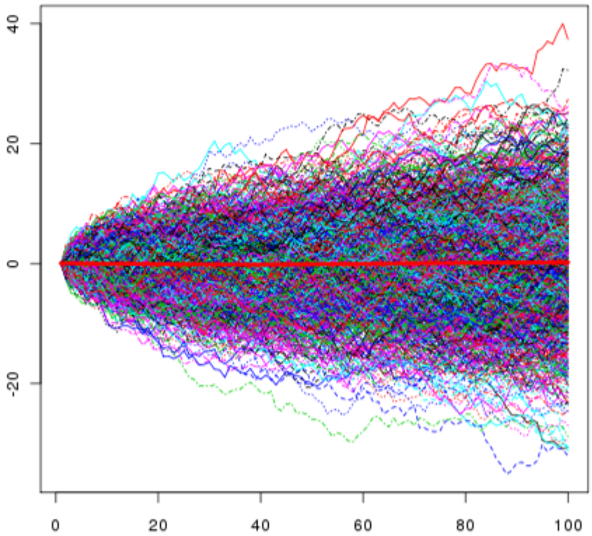
We hebben kennis gemaakt met de centrale limietstelling en ook bewezen waarom deze stelling geldt. In de volgende paragrafen gaan wij deze stelling toepassen en zien we waarom die zo mooi en nuttig is.

3.3 Random walk in de financiële wiskunde

Een fenomeen waar de centrale limietstelling gebruikt wordt, is een random walk. Een random walk is een willekeurig proces. We kunnen dit definiëren als een reeks discrete stappen dat een object zet in een willekeurige richting [6]. Deze stappen worden willekeurig bepaald door middel van een kansverdeling. De nieuwe positie van het object is onafhankelijk van de huidige positie. Een simpel voorbeeld van een random walk is "The Drunkard's walk"[20]. Een dronken persoon heeft geen voorkeur in richting waarin hij wil lopen waardoor de kans hetzelfde is voor elke richting. We definiëren een random walk als:

Definitie 3.1. *Neem een reeks $X_1, X_2, ...$ van onafhankelijk, identiek verdeelde willkeurige stochasten. Een* ***random walk*** *begint op positie $z \in \mathbb{R}$ in de reeks $(S_n)_{n \geq 0}$ met $S_0 = z$ en $S_n = S_{n-1} + X_n$ met $n \geq 1$. Waarbij X_n de stappen zijn van de random walk. [16]*

In figuur 2 is een voorbeeld van gegeven van gegenereerde random walks.



Figuur 2: Random walks

Deze random walks worden dus bepaald door een willekeurige stochast. Zoals figuur 2 al doet vermoeden, geldt voor deze stochast ook de centrale limietstelling. Hoe meer random walks we genereren hoe meer wij een normaalverdeling kunnen herkennen zoals we zagen bij figuur 1. De meeste paden liggen rond het gemiddelde $\mu = 0$. Hoe verder je van dit gemiddelde afgaat hoe minder paden er nog lopen. Dit brengt ons tot de volgende stelling [16]:

Stelling 3.3. ***Centrale limietstelling voor een random walk*** *Neem $t \geq 0$. De verdeling van een random walk $W^{(n)}(t)$ die evalueert op tijdstip t convergeert naar een normaalverdeling met gemiddelde nul en variantie t als $n \to \infty$.*

Er zijn verschillende toepassingen van de centrale limietstelling van random walks. We gaan een toepassing binnen de economie bekijken.

3.3.1 Random walk theorie

Een veelvoorkomende toepassing van de random walk is de *Random walk theorie* of ook wel de *Random walk hypothese.* Dit is een financieel model dat ervan uitgaat dat de aandelenmarkt zich op een volledig onvoorspelbare manier ontwikkelt. Dit wil zeggen dat de prijs van een effect zich willekeurig beweegt. Daarom is elke poging om toekomstige prijsbewegingen te voorspellen zinloos. Deze theorie is bekend geworden door econoom Burtoon Malkiel. Hij nam aan dat de aandelenkoers een willekeurig pad afgelegt wat dus wil zeggen dat de kans even groot is dat de koers daalt als dat die stijgt. Hij zei dat een geblinddoekte aap die een willekeurige portefeuille aandelen samenstelt even goed zou presteren als een portefeuille van een professional [9].

Er zijn twee basisveronderstellingen van de Random walk theorie [5]:

1. De prijs van een effect op de aandelenmarkt volgt een random walk.
2. De bewegingen van de prijzen van effecten zijn onafhankelijk van elkaar.

Hieruit volgt dat het dus onmogelijk is prijsveranderingen van effecten te voorspellen. Waardoor het voor handelaren onmogelijk is beter te presteren dan het algemene marktgemiddelde zonder een groot extra risico te nemen. De beste strategie is dan de buy-and-holdstrategie. Bij deze strategie wordt effecten gekocht en bewaard voor een langere periode. Hierbij wordt er dus niet gereageerd op de prijsveranderingen op korte termijn.

3.3.2 Wall Street journal dartboard contest

In 1988 werd de random walk theorie in de praktijk op de proef gesteld. De Wall Street journal organiseerde een wedstrijd geïnspireerd op het boek van Burton Malkiel: "A random walk down Wall Street". Bij deze wedstrijd waren de helft van de deelnemers handelexperts. De experts kozen hun aandelen op basis van hun kennis en ervaring. De andere helft deelnemers waren werknemers van de Wall Street journal. Zij kochten hun aandelen op basis van wat zij blind gooiden met dartpijltjes naar een bord met daarop de Wall Street journal aandelentabellen geplakt.

Na zes maanden werden de resultaten gemeten en gekeken of de aandelen van de experts of de darters het meeste winst heeft opgeleverd. Tien jaar later presenteerde de Journal de resultaten na honderd wedstrijden. De experts hebben 61 van de 100 wedstrijden gewonnen van de darters. Dat is meer dan de verwachtte 50 procent volgens de Random walk theorie. Aan de andere kant hebben de experts dus 39 keer verloren van darters terwijl die slechts willekeurig aandelen kochten. Dat is natuurlijk wel een beschamende uitkomst voor de experts en een bevestiging dat er zeker een kern van waarheid in de random walk theorie zit. Echter is dit niet een perfecte test voor de random walk theorie. De bekendheid van de Wall Street journal zorgde voor een zo genoemd mededelingseffect. Doordat er bekend werd gemaakt in welke aandelen de experts gingen investeren nam de vraag naar deze aandelen toe. Waardoor de aandelen van de experts extra in waarde stegen en zij hun kans op de winst van de wedstrijd vergrootten [19].

Uit dit experiment komt inderdaad naar voren dat het onmogelijk is altijd te voorspellen wat er met de koers gebeurt. Het is vooral een kwestie van goed geluk en risico's nemen om een hoge winst te behalen. In het algemeen zal je winst volgens de random walk theorie altijd rond het marktgemiddelde uitkomen. Hierin herkennen we ook de centrale limietstelling.

3.4 Klassenexperimenten random walk

Om zelf te ervaren hoe een random walk werkt zijn er twee klassenexperimenten om te houden met de eerstejaars wiskundestudenten. De eerste is een simpel experiment om te laten zien dat als je de steekproef vergroot dat de waarde naar het gemiddelde gaat [10]. Dit experiment kan ook werken als een introductie van de centrale limietstelling. Waarna de studenten na het experiment zelf de stelling kunnen proberen te formuleren. Het tweede experiment is een langer experiment om zelf het Wall Street journal dart toernooi na te bootsen en de random walk theorie te testen.

3.4.1 Munt werpen als een random walk

Wat heb je nodig?

- Munt
- Pen en papier
- Pion

Hoe ga je te werk?

1. Teken een getallenlijn van 20 centimeter van -10 tot 10.
2. Zet de pion op het midden bij nul.
3. Gooi met de munt. Als het kop is ga je een stap naar rechts op de getallenlijn en wanneer het munt is naar links.
4. Herhaal dit tien keer en noteer waar je op de getallenlijn geeïndigd bent.
5. Herhaal dit proces nog 9 keer.
6. Bereken het gemiddelde van de 10 getallen.

Vanuit de centrale limietstelling zou het gemiddelde nul moeten zijn. Dit is omdat de kans op kop en munt even groot is. Hoe vaker je het experiment uitvoert hoe dichter bij de nul je komt. Laat de studenten hierover discussiëren en de centale limiet stelling formuleren.

3.4.2 Je eigen Wall Street journal dart toernooi

Wat heb je nodig?

- Pen en papier
- Rekenmachine

De spelregels:

1. Verdeel de groep in tweeën. De helft zullen de experts zijn en de andere helft de "darters".
2. Noteer de verschillende bedrijven waarvan aandelen gekocht kunnen worden op een briefje voor een grabbelton.

3. De experts mogen met voorkennis beslissen in welke vier aandelen zij gaan investeren. Zij mogen hier onderzoek naar doen en andere experts vragen. De "darters"grabbelen willekeurig waarin zij gaan investeren.
4. Verdeel zogenaamd duizend euro over vier aandelen die je gaat kopen. De experts mogen zelf de verdeling maken. De darters verdelen het evenredig. Noteer de investeringen.
5. Na een week gaat iedereen berekenen wat hun aandelen op dat moment waard zijn en of zij winst of verlies hebben gemaakt.
6. Bereken de gemiddelde winst van de experts en van de darters en vergelijk dit met elkaar. Zijn er grote verschillen dussen de experts en de darters? Volgt uit dit experiment de random walk theorie?

Volgens de random walk theorie zal een willekeurig geselecteerde portefeuille een even grote kans hebben op winst dan een weloverwogen portefeuille. De winst zal altijd rond het marktgemiddelde liggen. Hierdoor zou vanuit het experiment de gemiddelde winst van de experts en de darters niet veel verschillen.

4 Statisiek in de econometrie

Econometrie is de combinatie tussen economie en wiskunde. Het is de disipline binnen de economische wetenschap die zich richt op het kwantificeren van de relaties tussen economische grootheden. Econometrie betekent letterlijk economisch meten. Bij econometrie wordt gebruik gemaakt van wiskundige technieken om verbanden in de economie te ontdekken en economische trends te voorspellen [15]. Voornamelijk kansrekening en statistiek worden gebruikt voor het economisch modelleren. Ook informatica is belangrijk binnen de econometrie voor het toetsen van econometrische modellen.

Econometrie heeft drie grote functies [15]:

1. Het beschrijven van economische activiteiten en het meten van effecten.
2. Het testen van hypothesen over economische theorieën en beleiden.
3. Het voorspellen van toekomstige ecnomische activiteiten.

Wij gaan bekijken hoe wij de kennis over statistiek kunnen gebruiken binnen de econometrie. Hiermee willen we laten zien dat dit studiepad ook een meer praktischere toepassing heeft.

4.1 Hoe wordt statistiek gebruikt?

4.1.1 Regressieanalyse

Regressieanalyse wordt gebruikt om een afhankelijke variabele te omschrijven als een functie van onafhankelijke variabelen. Je kan met regressieanalyse de samenhang tussen deze variabelen bepalen maar ook de verandering van de afhankelijke variabele voorspellen [18]. De meest voorkomende regressie is de lineaire regressie analyse. Hier zullen we ons ook op focussen. Bij een enkelvoudige lineaire regressie hangt de afhankelijke variabele slechts van één onafhanelijke variabele af [15]. Dan geldt:

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{23}$$

Hierbij geldt dat:

- Y is de afhankelijke variabele
- X is de onafhankelijk variabele
- β_0 is het startpunt van de regressielijn, de constante.
- β_1 is de gemiddelde toename van Y per eenheid van X: $\beta_1 = \frac{\Delta Y}{\Delta X}$
- ϵ is de foutterm

Er is altijd een foutterm ϵ in de regressie vergelijking, omdat de verandering van Y van meer kan afhangen dan de verandering van X. Ten eerste kunnen er andere variabelen zijn die een kleine invloed hebben op Y. Deze variabelen zijn vaak weggelaten in de regressie door de kleine impact. Ten tweede is het bijna onmogelijk geen meetfout te maken in de onafhankelijke variabele, dit heeft dan natuurlijk invloed op Y. Ten derde kan het verband tussen X en Y ook niet lineair zijn. Door het kiezen voor een lineaire regressie zal dit ook een verschil opleveren dat verwerkt zit in de foutterm. Tenslotte kunnen er ook onvoorspelbare willekeurige invloeden zijn door het menselijk gedrag. Al deze factoren zorgen ervoor dat de afhankelijke variabele Y niet alleen maar van hangt van de onafhankelijke variabele X en er daarom een foutterm aanwezig is [15]. Een onafhankelijke variabele kan ook door meerdere variabelen beïnvloed worden. Dit noemen we een meervoudige lineaire regressie. Dan geldt:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon \tag{24}$$

Er zijn twee soorten data die gebruikt kunnen worden voor de regressie: cross-sectional data en time-series data. Bij cross-sectional data bekijk je verschillende variabelen die invloed hebben op Y op hetzelfde tijdstip. Terwijl bij time-series data dezelfde variabele over een tijdperiode bekeken wordt. De soort data bepaalt hoe de variabele geïnterpreteerd kan worden [15].

4.1.2 OLS-schatter

We kunnen de coefficiënten van de lineaire regressie vergelijking (24) voorspellen met een OLS-schatter. De Ordinary Least Squares schatter bepaalt de waarden voor de onbekende coëfficienten β_k [23]. De OLS bepaalt deze waarden zodat de som van het kwadraat van de standaard afwijkingen geminimaliseerd is. Hiervoor geldt dat [15]:

$$\begin{aligned} Y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + e_i \\ \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \\ e_i &= Y_i - \hat{Y}_i \end{aligned} \tag{25}$$

Hierbij zijn $\hat{\beta}_i$ en $\hat{Y}_i$ de voorspelde waarden. Het residu e_i is de het verschil tussen de werkelijke en de verwachte waarde van Y. Het residu heeft twee eigenschappen [15]:

1. Het residu heeft een gemiddelde van nul: $\frac{1}{n}\sum_{i=1}^{n} e_i = 0$
2. Het residu heeft geen correlatie met alle onafhankelijke variabele X_k: $Corr(X_{ki}, e_i) = 0$

We gebruiken de OLS-schatter omdat volgens de stelling van Gauss-Markov dit de beste lineaire schatter is.

Stelling 4.1. ***Gauss-Markov stelling*** *Gegeven de volgende aannames:*

- *Het model is lineair in de parameters.*
- *De foutterm heeft gemiddelde nul:* $\mathbb{E}(\epsilon_i) = 0$.
- *Alle onafhankelijke variabelen zijn niet gecorreleerd met de errorterm:* $Corr(X_{ki}, e_i) = 0$.
- *Er is geen perfecte multicolineariteit. Er is geen exacte lineaire relatie tussen de variabelen.*
- *Er is homoscedasticiteit, dan is de foutterm constant en onafhankelijk van de variabelen:* $Var(\epsilon_i) = \sigma^2$
- *Er is geen autocorrelatie. De fouttermen van verschillende waarnemingen van X zijn niet gerelateerd:* $Corr(\epsilon_i, \epsilon_j) = 0$

Als deze aanamens gelden, dan is de OLS-schatter een onbevooroordeelde schatter zodat $\mathbb{E}(\hat{\beta}) = \beta$ *en de variantie van de OLS-schatter zo klein mogelijk [14].*

Deze OLS-schatter kunnen wij gebruiken om vanuit data de lineaire regressie op te stellen en hierbij de coëfficienten kunnen schatten om de invloed van de onafhankelijke variabele te bepalen op de afhankelijke variabele. De OLS-schatter kan de waarde van β_1 met een statisch computerprogramma bepalen zoals SPSS of Stata maar dit kan ook exact bepaald worden. Hierbij gebruik je de eigenschap van de OLS-schatter dat de som van het kwadraat van de som van de standaard afwijkingen SSE geminimaliseerd is.

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_i))^2 \tag{26}$$

Hieruit kan je de coëfficient β_i uit (25) bepalen met:

$$\beta_1 = \frac{\sigma_{XY}}{\sigma_{XX}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{27}$$

4.1.3 Hypothese toetsen

Door middel van hypotheses te toetsen kunnen we economische relaties kwantificeren door het analyseren van data van een steekproef. Hypothese toetsen gaat verder dan het berekenen van parameters. Aan de hand van de uitkomsten van een toetsing kunnen belangrijke vragen beantwoord worden over de echte wereld via de steekproef. Zo kan de kans bepaald worden dat het zich werkelijk voordoet of de significantie bepaald worden. Daarnaast wordt hypothese toetsen voornamelijk gebruikt om een bepaalde theorie te testen. Het is bijna onmogelijk om een theorie te bewijzen voor de gehele populatie maar met een steekproef kan er wel iets gezegd worden over de waarheid van de theorie binnen de steekproef. Zo kunnen we een hypothese afwijzen met een bepaalde zekerheid.

Om hypotheses te toetsen moeten deze hypotheses eerst worden opgesteld. Een hypothese is een vooraf geformuleerde verwachte uitkomst. Dit wordt geformuleerd als een verklaring of beschrijving. Een hypothese bestaat altijd uit twee delen: de nulhypothese en de alternatieve hypothese. De nulhypothese H_0 neemt een bepaald verband aan tussen variabelen [2]. Dit zijn vaak de waardes die je niet verwacht. Deze hypothese blijft waar tot er genoeg bewijs is om de alternatieve hypothese aan te nemen[1]. De alternatieve hypotese H_A is wat jij verwacht en geeft dus een tegengesteld verband weer. Er zijn drie soorten hypothesesets met constante $c \in \mathbb{R}$ en $n \in \mathbb{N}_{\geq 1}$ [1]:

1. Tweezijdige toetsing
$$\begin{aligned} H_0 &: \beta_n = c \\ H_A &: \beta_n \neq c \end{aligned} \tag{28}$$

2. Linkszijdige toetsing
$$\begin{aligned} H_0 &: \beta_n \geq c \\ H_A &: \beta_n < c \end{aligned} \tag{29}$$

3. Rechtszijdige toetsing
$$\begin{aligned} H_0 &: \beta_n \leq c \\ H_A &: \beta_n > c \end{aligned} \tag{30}$$

Als de hypotheses zijn opgesteld kunnen ze getoetst worden om de hypothese te kunnen accepteren of weigeren. Voor het toetsen moet er een significantieniveau α vastgesteld worden.

Het significantieniveau geeft aan met hoeveel procent zekerheid het gevonden resultaat voor de steekproef ook geldt voor de gehele populatie [8]. De meest voorkomende significantieniveau is $\alpha = 0,05$ dus er kan met 95% zekerheid gezegd worden dat de uitkomst geldt voor de gehele populatie. Echter wordt ook vaak $\alpha = 0,01$ en $\alpha = 0,1$ gebruikt. Welke waarde je neemt voor α hangt af van de situatie. Zo wordt α bepaald aan de hand van hoe groot de kans is op een Type I fout. Een Type I fout is dat je een nulhypothese verwerpt die waar is. Hoe groter α hoe groter de kans op een Type I fout. In elke situatie is de ernst van een Type I fout anders, daarom kan α verschillend zijn [8].

Er zijn verschillende manieren om een hypothese te toetsen. We zullen ons focussen op het toetsen van de invloed van één parameter β tegelijkertijd. Hieronder bespreken we de bekendste methode.

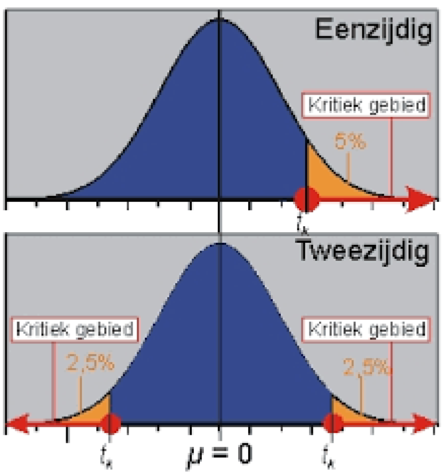
4.1.4 T-toets

De T-toets is de meest voorkomende toets binnen de econometrie. Deze toets is makkelijk te gebruiken doordat het ook werkt bij verschillende meetwijzes van variabelen en de standaard afwijkingen. De T-toets wordt voornamelijk gebruikt als de stochastische foutterm normaal verdeeld is en wanneer deze standaard afwijking geschat moet worden [26]. Voor de T-toets berekenen we de t-waarde van de geschatte coëfficienten. Bij de T-toets wordt alleen gekeken naar de coëfficienten die de gemiddelde verandering aan geven en niet naar het startpunt β_0. Deze t-statistiek voor de k-de coëfficient bereken we als volgt met $n = 1, 2, ..., N$ [15]:

$$t_n = \frac{(\hat{\beta_n} - \beta_{H_0})}{SE(\hat{\beta_n})} \tag{31}$$

- $\hat{\beta_n}$ is de geschatte regressie coëfficient van de k-de variabele
- β_{H_0} is de waarde gegeven in de nulhypothese voor β_n
- $SE(\hat{\beta_n})$ is de geschatte standaard afwijking van $\hat{\beta_n}$), ook wel $SE(\hat{\beta_n}) = \sqrt{Var(\beta_n)}$

Hoe groter de t-waarde hoe groter de kans dat de geschatte regressie coëfficient is anders de nulhypothese. Om te bepalen wanneer de nulhypothese wordt verworpen vergelijken we de berekende t-waarde met de kritieke t-waarde t_k.



Figuur 3: Kritieke waarde bij de curve

De kritieke t-waarde zijn de waardes van t als de verdeling van de waarneming perfect normaalverdeeld is. Zoals we weten vanuit de centrale limietstelling naderen de waarden van een onbekende verdeling X naar een normaalverdeling toe. Binnen een steekproef is dit ook het geval, echter is niet de een perfecte normaalverdeling. De tabel met de kritieke t-waardes in bijlage 1 is gemaakt om tijd te besparen voor onderzoekers [15]. Hierbij zijn de kritieke t-waardes berekend passend bij specifieke gedeeltes onder de curve afhankelijk voor het soort test, het significantieniveau en de vrijheidsgraden (zie figuur 3). De vrijheidsgraden = het aantal waarnemingen - het aantal variabelen - 1.

Met de kritieke waarden uit bijlage 1 kunnen we bepalen wanneer we de nulhypothses verwerpen. We verwerpen de nulhypothese als de t-waarde t_k in het kritiekgebied ligt zoals aangegeven in figuur 3. In andere woorden:

Verwerp H_0 als $|t_n| > t_k$ en als t_n hetzelfde teken heeft aangegeven in H_A.

4.2 Opdracht gemeenteraadverkiezingen

Deze opdracht is voor de eerstejaars studenten om zelf te oefenen met het hypothese testen bij een lineaire regressie. Hierbij wordt er een database gebruikt met de gegevens van 1537 politici van verschillende gemeentes in de verenigde staten. Bij deze politici gaan we kijken naar het invloed van de lengte, leeftijd en het geslacht van de politici op het aandeel van de stemmen. Figuur 4 is de tabel waarin het aantal observaties, het gemiddelde, de standaard afwijking en de minimale en maximale waarde kunnen aflezen.

Variable	Obs	Mean	Std. Dev.	Min	Max
voteshare	1537	24.09405	12.76645	0	75.97795
height	1537	172.1837	11.85932	137.5807	207.8188
age	1537	44.2674	7.940087	19	73
male	1537	.6050748	.4889937	0	1

Figuur 4: Informaties variabelen

Hiervoor geldt:

- *voteshare* is het aandeel stemmen die een politici heeft ontvangen tijdens de gemeenteraadverkiezingen in procenten
- *height* is de lengte van de politici in centimeter
- *age* is de leeftijd van de politici in jaren
- *male* is het geslacht van de politici bij een man waarde 1 en bij een vrouw waarde 0

In figuur 5 zijn de uitkomsten na het toepassen van de OLS-schatter. Door gebruik te maken van de OLS-schatter zijn de verwachte coëfficiënten β_k bepaald. Daarnaast is ook de standaard afwijking SE geschat. In de laatste kollom zijn al de t-waarden bepaald bij een nulhypothese van bijvoorbeeld $\beta_1 = 0$. In deze opdracht geldt een significantieniveau van 5%.

voteshare	Coef.	Std. Err.	t
height	.1009863	.0305956	3.30
age	-.0558723	.0392669	-1.42
male	6.119952	.7416573	8.25
_cons	5.476144	5.299002	1.03

Figuur 5: Coëfficiënten regressie

4.2.1 Vragen met uitwerking

1. Stel de lineaire regressie.

   De lineaire regressie geeft het lineaire verband tussen de onafhankelijke variabelen (lengte, leeftijd, geslacht) en de afhankelijke variabele aandeel stemmen.

$$voteshare = \beta_0 + \beta_1 heigth + \beta_2 age + \beta_3 male + \epsilon$$

$$vote\hat{s}hare = \hat{\beta_0} + \hat{\beta_1} heigth + \hat{\beta_2} age + \hat{\beta_3} male$$

2. Interpreteer de coëfficiënt van de variabele *age*.

   Vanuit figuur 5 weten we dat $\beta_2 = -0,0558723$ Dit wil zeggen dat als de politicus 1 jaar ouder wordt het aandee stemmen afneemt met 0,06%.

3. Bepaal of de variabele *height* invloed heeft op het aandeel stemmen met een T-toets.

   Stel als eerst de hypotheses op:

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

   Vervolgens bepalen we de t-waarde. Vanuit figuur 5 weten we dat $t_1 = 3.30$. Deze t-waarde kan ook bepaald worden met de formule:

$$t_1 = \frac{(\hat{\beta_1} - \beta_{H_0})}{SE(\hat{\beta_1})} = \frac{(0.1009863 - 0)}{0.0305956} = 3.30068$$

   Dan zoeken we de kritieke t waarde t_k. Dit is een tweezijdige toets met een significantie niveau van 5% dus $\alpha = 0.05$. Het aantal vrijheidsgraden= aantal observaties- aantal variabelen - 1= 1537-3-1= 1533. Dan gebruiken we tabel B-1 uit de bijlage om de kritieke t-waarde $t_{0.05,1533} = 1.960$.

   We verwerpen de nulhypothese als $|t_1| > t_k$ In dit geval geldt $3.30 > 1.960$ dan verwerpen we de nulhypothese. Dit wil zeggen dat $\beta_1 \neq 0$ dus de lengte van de politicus heeft invloed op het aandeel stemmen.

5 Samenvatting

Eerstejaars wiskundestudenten maken kennis met het studiepad statistiek en kansrekening. Dit wordt gedaan door kleine onderdelen uit de statistiek, financiële wiskunde en econometrie te introduceren. Als hoofdonderwerp is de centrale limiet stelling genomen. Deze stelling is bewezen en er zijn vervolgens toepassingen van laten zien. Zoals de random walks binnen de financiële wiskunde en hoe deze ook de centrale limietstelling volgen. Dit zagen we ook terug op de aandelenmarkt. Volgens de random walk theorie gedraagt de aandelenmarkt zich ook als een random walk waardoor je winst altijd rond het marktgemiddelde zal liggen. Tenslotte wordt het verband tussen statistiek en econometrie geïntroduceerd. Hierbij duiken we in de regressieanalyse en gebruiken we de T-toets voor het toetsen van hypotheses. In een opdracht zien we hier ook een toepassing van terug.

6 Discussie

Het doel van de scripie is eerstejaars wiskundestudenten vroegtijdig kennis te laten maken met de studiepaden binnen de bachelor wiskunde. Binnen deze scriptie is dit uitgewerkt voor het studiepad statistiek en kansrekening. Het zou mooi zijn als dit in de toekomst ook voor de andere studiepaden wordt uitgewerkt. Dan zou er een soort collegereeks gevormd kunnen worden over alle studiepaden. Ik zou dit voor me zien als een soort caleidoscoop lezingen die verspreid over het eerste jaar gegeven zullen worden. Deze scriptie kan gebruikt worden door de docent als een conceptplan voor zijn college.

Een andere optie zou zijn dat studenten zelf de scriptie kunnen doorlezen en de verschillende studiepaden gebundeld kunnen worden als een soort dictaat. Hierdoor kan de student zich zelfstandig in de studiepaden verdiepen en de studiepaden beter vergelijken. Daardoor kunnen zij een betere afweging kunnen maken voor hun keuzevakken in de rest van hun bachelor. Daarnaast zijn ze beter op de hoogte van de verschillende vakken van elk studiepad om zo ook de juiste voorbereiding voor je toekomstige master te hebben.

7 Bijlagen

Table B-1 Critical Values of the *t*-Distribution

Degrees of Freedom	Level of Significance: One-Sided: 10% Two-Sided: 20%	5% 10%	2.5% 5%	1% 2%	0.5% 1%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
(Normal)					
∞	1.282	1.645	1.960	2.326	2.576

Figuur 6: Kritieke t-waarden

8 Literatuurlijst

Referenties

[1] Bauwens, N. (2021, 27 mei). Hypothese opstellen en toetsen? Dat doe je zo! Natascha Bauwens. Geraadpleegd op 23 juni 2022, van https://www.nataschabauwens.nl/hypothese-opstellen-en-toetsen/

[2] Bevans, R. (2022, 31 januari). Een stappenplan voor hypothesetoetsing. Scribbr. Geraadpleegd op 23 juni 2022, van https://www.scribbr.nl/statistiek/hypothesetoetsing/

[3] Blogt, L. K. U. (2016, 20 oktober). De zin en onzin van statistiek? KU Leuven blogt. Geraadpleegd op 18 mei 2022, van https://kuleuvenblogt.be/2016/10/20/de-zin-en-onzin-van-statistiek/

[4] Centrale limietstelling - overzicht, geschiedenis en voorbeeld. (2019). Financiële gidsen. Geraadpleegd op 26 april 2022, van `https://nl.livingeconomyadvisors.com/1440-what-is-the-central-limit-theorem-clt`

[5] Corporate Finance Institute. (2021, 28 juli). Random Walk Theory. Geraadpleegd op 25 mei 2022, van https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/what-is-the-random-walk-theory/

[6] Datta, S. (2022, 17 februari). What is a Random Walk? Baeldung on Computer Science. Geraadpleegd op 25 mei 2022, van https://www.baeldung.com/cs/random-walk

[7] Kaspers, L. J. (2016). Statistical Review with formula of, and rules for the mean, variance, covariance, correlation coefficient. kaspercpa. Geraadpleegd op 2 juni 2022, van http://www.kaspercpa.com/statisticalreview.htm

[8] Ontdek welk alfaniveau de statistische significantie bepaalt. (2019, 31 juli). Greelane. Geraadpleegd op 12 juni 2022, van https://www.greelane.com/nl/science-tech-math/wiskunde/what-level-of-alpha-determines-significance-3126422/

[9] Random walk theorie definitie. (2016). IG. Geraadpleegd op 25 mei 2022, van https://www.ig.com/nl/woordenboek-handelstermen/random-walk-theorie-definitieinformation-banner-dismiss

[10] Random Walk. (2013). MRSEC Education Group. Geraadpleegd op 27 mei 2022, van https://education.mrsec.wisc.edu/random-walk/

[11] Snoek, H., Schut, M. (z.d.). De Centrale Limietstelling - Data Analyse en Statistiek. Data analyse en statistiek. Geraadpleegd op 26 april 2022, van `https://das.mprog.nl/module-3/de-centrale-limietstelling`

[12] Shreve, S. (2022). stochastic-calculus-for-finance-ii-continuous-time-models. Springer Finance.

[13] Scribbr. (2021, 3 november). Wat is het verschil tussen de standaarddeviatie en de variantie? Geraadpleegd op 2 juni 2022, van https://www.scribbr.nl/veel-gestelde-vragen/wat-is-het-verschil-tussen-standaarddeviatie-en-variantie

[14] Stelling van Gauss-Markov - frwiki.wiki. (2019). Wikipedia. Geraadpleegd op 23 juni 2022, van https://nl.frwiki.wiki/wiki/Th

[15] Studenmund, A. H. (2017). A practical guide to using econometrics. Pearson.

[16] UCLA. (2018). Random walks. https://www.math.ucla.edu/ biskup/PDFs/PCMI/PCMI-notes-1

[17] Universiteit Utrecht. (2019). Studiepaden. Students UU. Geraadpleegd op 18 mei 2022, van https://students.uu.nl/beta/wiskunde-en-toepassingen/onderwijs/studiepaden

[18] Van Heijst, L. (2021, 27 oktober). Regressieanalyse uitvoeren en interpreteren. Scribbr. Geraadpleegd op 23 juni 2022, van https://www.scribbr.nl/statistiek/regressieanalyse/

[19] Wall Street Journal Dartboard Contest. (2001). Invester home. Geraadpleegd op 26 mei 2022, van http://www.investorhome.com/darts.htm

[20] Wikipedia contributors. (2022a, februari 10). The Drunkard's Walk. Wikipedia. Geraadpleegd op 25 mei 2022, van $https://en.wikipedia.org/wiki/The_Drunkard\%27s_Walk$

[21] Wikipedia contributors. (2022, 7 april). Central limit theorem. Wikipedia. Geraadpleegd op 19 mei 2022, van $https://en.wikipedia.org/wiki/Central_limit_theorem$

[22] Wikipedia contributors. (2022c, mei 14). Characteristic function (probability theory). Wikipedia. Geraadpleegd op 2 juni 2022, van $https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory)$

[23] Wikipedia contributors. (2022d, juni 7). Ordinary least squares. Wikipedia. Geraadpleegd op 23 juni 2022, van $https://en.wikipedia.org/wiki/Ordinary_least_squaresProperties$

[24] Wikipedia-bijdragers. (2020, 15 juni). Verwachting (wiskunde). Wikipedia. Geraadpleegd op 2 juni 2022, van $https://nl.wikipedia.org/wiki/Verwachting_(wiskunde)Rekenregels$

[25] Wikipedia-bijdragers. (2021, 24 mei). Normale verdeling. Wikipedia. Geraadpleegd op 22 mei 2022, van $https://nl.wikipedia.org/wiki/Normale_verdeling$

[26] Wikipedia-bijdragers. (2022, 17 januari). T-toets. Wikipedia. Geraadpleegd op 23 juni 2022, van https://nl.wikipedia.org/wiki/T-toets

[27] Wikizero - Karakteristieke functie (kansrekening). (z.d.). Wikizero. Geraadpleegd op 19 mei 2022, van $https://www.wikizero.com/nl/Karakteristieke_functie_(kansrekening)$

[28] Wikizero - Taylorreeks. (z.d.). Wikizero. Geraadpleegd op 19 mei 2022, van https://www.wikizero.com/nl/Taylorreeks

[29] Zitkovic, G. (2017). Moment-generating functions. Universiteit Utrecht.