



Vlaanderen
is erfgoed

Onderzoeksrapport

Forfaitaire kostprijsbepaling van archeologisch vooronderzoek met ingreep in de bodem

Agentschap
Onroerend
Erfgoed

COLOFON

TITEL

Forfaitaire kostprijsbepaling van archeologisch vooronderzoek met ingreep in de bodem

REEKS

Onderzoeksrapporten agentschap Onroerend Erfgoed nr. 119

AUTEURS

Korneel De Groot, Raf Ribbens

JAAR VAN UITGAVE

2019

Een uitgave van agentschap Onroerend Erfgoed Wetenschappelijke instelling van de Vlaamse Overheid, Beleidsdomein Omgeving
Published by the Flanders Heritage Agency Scientific Institution of the Flemish Government, policy area Environment

VERANTWOORDELIJKE UITGEVER

Sonja Vanblaere

OMSLAGILLUSTRATIE

Proefsleuf in het Sigma gebied 'Wijmeers 1' (Schellebelle, Oost-Vlaanderen), 2012
Copyright Onroerend Erfgoed, foto: Erwin Meylemans

agentschap Onroerend Erfgoed

Havenlaan 88 bus 5

1000 Brussel

T +32 2 553 16 50

info@onroerenderfgoed.be

www.onroerenderfgoed.be

Dit werk is beschikbaar onder de Modellicentie Gratis Hergebruik v1.0.

This work is licensed under the Free Open Data Licence v.1.0.

Dit werk is beschikbaar onder een Creative Commons Naamsvermelding

4.0 Internationaal-licentie. Bezoek

<http://creativecommons.org/licenses/by/4.0/> om een kopie te zien van de licentie.

This work is licensed under a Creative Commons Attribution

4.0 International License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by/4.0/>.

ISSN 1371-4678

D/2019/3241/160



////////////////////////////////////

FORFAITAIRE
KOSTPRIJSBEPALING
VAN
ARCHEOLOGISCH
VOORONDERZOEK
MET INGREEP IN DE
BODEM

////////////////////////////////////

KORNEEL DE GROOTE, RAF RIBBENS

INHOUD

| | | |
|-------|--|----|
| 1 | INLEIDING..... | 6 |
| 2 | VOORSTELLING VAN DE DATASET | 8 |
| 2.1 | INLEIDING OP DE DATASET | 8 |
| 2.2 | KOSTPRIJS ARCHEOLOGISCHE DOSSIERS | 9 |
| 2.2.1 | ONDERZOEKSMETHODES | 9 |
| 2.2.2 | BOXPLOT VAN DE KOSTPRIJS VAN ONDERZOEKSMETHODES | 10 |
| 2.2.3 | VASTSTELLINGEN UIT DE BOXPLOT | 12 |
| 2.3 | KOSTPRIJS TEGENOVER OPPERVLAKTE..... | 14 |
| 2.4 | KOSTPRIJS PER WERKDAG..... | 15 |
| 3 | BEPALEN VAN SIGNIFICANTIE | 15 |
| 3.1 | INLEIDING BIJ SIGNIFICANTIE..... | 15 |
| 3.2 | BEPALEN VAN EN OMGAAN MET EXTREME WAARDEN | 16 |
| 3.3 | CORRELATIECOËFFICIËNT VAN NUMERIEKE VARIABELEN | 16 |
| 3.4 | VARIANTIEANALYSE VAN CATEGORISCHE VARIABELEN | 17 |
| 3.5 | TEST OP VERSCHILLEN BINNEN VARIABELEN..... | 19 |
| 4 | ARCHEOLOGISCH VOORONDERZOEK MET PROEFSLEUVEN EN PROEFPUTTEN..... | 20 |
| 4.1 | INLEIDING BIJ PROEFSLEUVEN EN PROEFPUTTEN | 20 |
| 4.2 | CORRELATIEMATRIX VAN DATA PROEFSLEUVEN | 20 |
| 4.3 | ANOVA VAN DATA PROEFSLEUVEN | 22 |
| 5 | ARCHEOLOGISCH VOORONDERZOEK MET ARCHEOLOGISCHE BORINGEN..... | 23 |
| 5.1 | INLEIDING BIJ ARCHEOLOGISCHE BORINGEN | 23 |
| 5.2 | CORRELATIEMATRIX VAN DATA ARCHEOLOGISCHE BORINGEN..... | 23 |
| 5.3 | ANOVA VAN DATA ARCHEOLOGISCHE BORINGEN | 24 |
| 6 | MODELLEN OP BASIS VAN SIGNIFICANTE VARIABELEN | 25 |
| 6.1 | INLEIDING BIJ DE MODELLEN | 25 |
| 6.2 | MODELLEN VOOR DE ONDERZOEKSMETHODE PROEFSLEUVEN | 27 |
| 6.2.1 | INLEIDING..... | 27 |
| 6.2.2 | MODEL: LINEAIRE REGRESSIE VAN PROEFSLEUVEN MET 2 VERKLARENDE VARIABELEN..... | 27 |
| 6.2.3 | APART MODEL: KOSTPRIJS VAN HET UITGESTELD TRAJECT | 28 |
| 6.2.4 | MODEL: "REGRESSION THROUGH THE ORIGIN" | 29 |
| 6.2.5 | MODEL: LINEAIRE REGRESSIE OP LOGARITMISCHE SCHAAL EN NA ANALYSE RESIDUEN | 30 |
| 6.3 | MODELLEN VOOR DE ONDERZOEKSMETHODE ARCHEOLOGISCHE BORINGEN | 34 |
| 6.3.1 | INLEIDING..... | 34 |
| 6.3.2 | MODEL: REGRESSIE MET 2 VERKLARENDE VARIABELEN | 35 |



| | | |
|-------|--|----|
| 6.3.3 | MODEL: LINEAIRE REGRESSIE MET AANPASSING PRIJS PROEFSLEUVEN..... | 38 |
| 6.3.4 | MODEL: LINEAIRE REGRESSIE MET LOGARITMISCHE SCHAAL EN NA ANALYSE RESIDUEN..... | 39 |
| 7 | BESLUIT EN VOORSTEL VAN FORMULE | 42 |
| 8 | BIJLAGEN | 43 |
| 8.1 | MODALE KOSTPRIJS PER ONDERZOEKSMETHODE | 43 |
| 8.2 | HONEST SIGNIFICANT DIFFERENCES ONDERZOEKSMETHODES..... | 44 |
| 8.3 | R PACKAGES EN HUN FUNCTIES | 47 |
| 9 | BIBLIOGRAFIE | 49 |



1 INLEIDING

Het Onroerenderfgoeddecreet verplicht in bepaalde gevallen een archeologisch vooronderzoek bij vergunningsplichtige ingrepen in de bodem. Artikel 5.4.1 en 5.4.2 van het Onroerenderfgoeddecreet bepalen in welke gevallen de aanvragers van een omgevingsvergunning verplicht zijn om voorafgaandelijk een dergelijk archeologisch vooronderzoek te laten uitvoeren. Dit gebeurt door erkende archeologen. Daarbij voeren ze eerst een archeologisch vooronderzoek uit en later, indien nodig, een archeologische opgraving.

Het resultaat van een archeologisch vooronderzoek is een **archeologienota**. In dit verslag schrijft de erkend archeoloog of er op het terrein al dan niet archeologisch erfgoed aanwezig is en welke maatregelen nodig zijn om bij de werkzaamheden uit de omgevingsvergunning op een gepaste manier met dat aanwezige erfgoed om te gaan.

Om het vooronderzoek uit te voeren kan de erkend archeoloog gebruikmaken van diverse **onderzoeksmethodes**. Deze staan uitvoerig gedocumenteerd in de Code van goede praktijk voor de uitvoering van en rapportering over archeologisch vooronderzoek en archeologische opgravingen en het gebruik van metaaldetectoren (de code van goede praktijk). Grosso modo onderscheiden we 2 grote groepen: archeologisch vooronderzoek zonder ingreep in de bodem en archeologisch vooronderzoek met ingreep in de bodem.

Archeologisch onderzoek heeft een kostprijs. Die wordt betaald door de aanvrager van de omgevingsvergunning en dus de persoon die een ingreep in de bodem wenst uit te voeren en hierdoor mogelijk archeologisch erfgoed kan verstoren. Het behoud van kennis van het archeologisch verleden is van algemeen belang en het is aan de initiatiefnemer van een ontwikkelingsproject om de eventuele schade aan dat algemeen belang te vermijden of te remediëren. De wijziging van het Onroerenderfgoeddecreet en -besluit organiseren een **nieuwe premie voor verplicht uit te voeren archeologisch vooronderzoek met ingreep in de bodem bij vergunningsplichtige ingrepen in de bodem**. Daarom introduceert de minister bevoegd voor onroerend erfgoed in zijn beleidsbrief 2019 dit nieuwe instrument. Deze nieuwe premie moet, naar analogie van de bestaande premie voor buitensporige opgravingskosten, de financiële impact beperken voor initiatiefnemers die deze niet kunnen vermarkten of spreiden.

De premie komt tegemoet in de kosten voor archeologisch vooronderzoek met ingreep in de bodem. Ze is bestemd voor occasionele initiatiefnemers van bepaalde types projecten. De criteria daarvoor zijn dezelfde als bij de bestaande premie voor buitensporige opgravingskosten. Zo krijgen enkel particulieren en kleinschalige ondernemingen en verenigingen toegang, en zijn verkavelingen en professionele vastgoedontwikkelingen uitgesloten. De premie geldt enkel voor de onderzoeksmethodes met ingreep in de bodem: vooronderzoek zonder ingreep in de bodem is uitgesloten. De berekening van de premie gebeurt **op forfaitaire basis**, naar analogie met de premie voor buitensporige opgravingskosten. Het achterliggende idee is dat dergelijk vooronderzoek een vaste basiskost heeft die in positieve en negatieve zin beïnvloed wordt door bepaalde variabelen. Idealiter komt de forfaitair bepaalde kostprijs daarbij overeen met de reëel gefactureerde kostprijs. De premie bedraagt 80% van de forfaitair bepaalde kost van het vooronderzoek met ingreep in de bodem, en de kost van het vooronderzoek zonder ingreep in de bodem wordt aanzien als een franchise die volledig voor rekening van de initiatiefnemer is.



Dit onderzoek heeft als doel de **berekeningswijze van deze forfaitaire kostprijs te bepalen**. Het globale opzet is een statistisch onderzoek van reeds uitgevoerde archeologische vooronderzoeken waarbij in eerste instantie wordt vastgelegd wat de meerkost is van vooronderzoek met ingreep in de bodem ten opzichte van vooronderzoek zonder ingreep in de bodem. Anders gesteld: welk deel van de totale kostprijs van een traject van vooronderzoek (dat altijd minstens één vorm van vooronderzoek zonder ingreep in de bodem omvat) komt op het conto van het vooronderzoek zonder ingreep in de bodem en welk deel komt op rekening van het vooronderzoek met ingreep in de bodem? Het onderzoek moet in tweede instantie uitwijzen welke de bepalende factoren zijn in de variantie van de kostprijs van vooronderzoek met ingreep in de bodem en in welke mate die de kostprijs dan beïnvloeden.

Het onderzoek begint hiervoor met een **verkenning van de beschikbare data**. Sinds de invoering van het archeologisch traject van het Onroerenderfgoeddecreet gebruiken de erkende archeologen het online loket van het agentschap Onroerend Erfgoed, en dan specifiek het **Archeologieportaal**, voor het indienen van hun verslag van archeologisch (voor)onderzoek. Dit archeologieportaal is een rijke bron die informatie bevat over elk individueel archeologisch vooronderzoek.

Na de algemene verkenning van de data volgt het onderzoek naar de variabelen die mogelijk een **significant effect** hebben op de kostprijs van het archeologisch vooronderzoek. We doen dit met behulp van **correlatie en variantieanalyse**.

De hoofdstukken 4 en 5 verlaten de algemene dataset en onderzoeken in detail een aantal vooronderzoeken met ingreep in de bodem. In deze hoofdstukken werken we met **steekproeven** uit de volledige dataset en voegen we een aantal meer gedetailleerde variabelen toe die we verzamelden uit de individuele verslagen van vooronderzoeken met ingreep in de bodem. We onderzoeken vervolgens deze nieuwe variabelen op hun significantie op de kostprijs van het vooronderzoek.

In hoofdstuk 6 testen we verschillende **regressiemodellen** om te komen tot een betrouwbare inschatting van de verschillende kosten die samen de totale kostprijs maken van een archeologisch vooronderzoek met ingreep in de bodem. Stap voor stap onderzoeken we de afzonderlijke kostprijs van verschillende variabelen om te komen tot een betrouwbaar model waaruit we een formule kunnen afleiden.

Het onderzoek eindigt met een **formule** die aan de hand van objectieve criteria voorspelt wat de kostprijs van een bepaald vooronderzoek met ingreep in de bodem in normale omstandigheden zou zijn.



2 VOORSTELLING VAN DE DATASET

2.1 INLEIDING OP DE DATASET

We beginnen dit onderzoek met een verkenning van de dataset, een “**exploratory data analysis**” (**EDA**). Gedurende een EDA formuleren we verschillende **hypotheses** over de beschikbare data die we verzamelden. Behalve het formuleren van hypothesen over bepaalde fenomenen die we waarnemen in de data, gebruiken we EDA ook voor het **beoordelen van de kwaliteit van de data**. In sommige gevallen is het noodzakelijk de **data te transformeren of te vereenvoudigen**, om op die manier de relevante fenomenen te kunnen waarnemen¹.

Uit het Archeologieportaal van het agentschap genereerden we voor dit onderzoek een omvangrijke dataset met gegevens over uitgevoerde archeologische vooronderzoeken. We beslisten hierbij te werken op het niveau van het **volledige archeologische vooronderzoek**, en niet op het niveau van de individuele verslagen van dat vooronderzoek. Daarvoor koppelden we de eventuele fases van vooronderzoeken, in praktijk archeologienota’s en de daarop volgende nota’s tot één “dossier”.

De resulterende dataset uit het Archeologieportaal telt 41 verschillende kolommen voor in totaal 3.957 observaties. De dataset bevat **alle bekrachtigde archeologienota’s en alle bekrachtigde archeologienota’s waar al een bekrachtigde nota op is gevolgd** sinds het begin van de regelgeving archeologie (1 april 2016) tot 31 oktober 2018.

Erkende archeologen voeren een archeologisch vooronderzoek uit en schrijven de resultaten hiervan neer in een **archeologienota**. In bepaalde gevallen is het voor de archeoloog niet mogelijk om het vooronderzoek in één keer uit te voeren en is de archeoloog gedwongen om het vooronderzoek (of een deel ervan) uit te stellen. Dit kan bijvoorbeeld voorkomen wanneer de omgevingsvergunning handelt over een sloop en het bouwen van een nieuw gebouw, en de sloop nog moet gebeuren. In dergelijk geval kan de archeoloog zijn vooronderzoek met ingreep in de bodem niet uitvoeren alvorens het gebouw is gesloopt. De regelgeving noemt dit fenomeen “*uitgesteld vooronderzoek*” en het verslag van dit uitgesteld traject is een **nota**.

Voor dit onderzoek vertrekken we van archeologische vooronderzoeken waarvan het resultaat een bekrachtigde archeologienota is en in voorkomend geval waaraan ook één of meerdere bekrachtigde nota’s zijn gekoppeld. Vooronderzoeken waarvan de archeologienota in een uitgesteld vooronderzoek voorziet, maar dit nog niet was uitgevoerd, sloten we uit.

De **variabelen** in de dataset zijn onder meer:

- de erkende archeologen verantwoordelijk voor het vooronderzoek,
- de gehanteerde onderzoeksmethodes,
- de kostprijs,
- de totale duur in werkdagen, inclusief de rapportering,
- de behandelende organisatie,
- de ligging,
- het programma van maatregelen,

¹ Wickham & Grolemund, 2017, p. 81.



- de overlapping met vastgestelde archeologische zones of beschermde archeologische sites,
- het aantal indieningen.

In het volgende deel verrichten we een aantal verkenningen van de dataset. We doen dit om de algemene kenmerken en categorieën binnen de verschillende variabelen te verduidelijken. De focus ligt op de kostprijs van archeologische “dossiers”.

In de verkenning bekijken we volgende elementen:

- De modale kostprijs van archeologische dossiers en naargelang de gebruikte onderzoeksmethode(s);
- De kostprijs tegenover de oppervlakte van het onderzoeksgebied en per vierkante meter;
- De kostprijs per werkdag.

2.2 KOSTPRIJS ARCHEOLOGISCHE DOSSIERS

Bij deze analyse vertrekken we van de kostprijs van de archeologienota, en in voorkomend geval, de eraan gekoppelde nota. De bedragen van beide worden opgeteld om op deze manier te komen tot een totale kostprijs per dossier. Archeologienota’s waarbij de kostprijs minder bedraagt dan 500 euro, filteren we eerst uit de analyse. Bedragen lager dan 500 euro beschouwen we als onbetrouwbaar. Een honderdtal dossiers laten we hierdoor weg uit de analyse.

De **globale mediaanprijs** van de dataset bedraagt 2.900 euro terwijl het gemiddelde 4.738 euro bedraagt. Het verschil tussen beide is groot. De oorzaak hiervan zijn de extremen, de “outliers” in de dataset die beduidend hoger liggen dan de meerderheid van de dossiers. 75% van alle dossiers hebben een kostprijs lager dan 5.000 euro terwijl het dossier met de hoogste prijs een maximum heeft van 340.869 euro. Dit wil zeggen dat de kostprijs van 1/4 van de dossiers ligt tussen 5.000 euro en 340.869 euro. Dit onevenwicht heeft een sterke invloed op het gemiddelde.

2.2.1 Onderzoeksmethodes

Na de totale kostprijs onderzoeken we de modale kostprijs naargelang de gebruikte onderzoeksmethodes zoals de erkende archeologen dit aangeven in het Archeologieportaal. De erkende archeologen kunnen 9 verschillende onderzoeksmethodes gebruiken bij het uitvoeren van archeologisch onderzoek.

- Bureauonderzoek (1);
- Landschappelijk bodemonderzoek (2);
- Geofysisch onderzoek (3),
- Veldkartering (4);
- Verkennend archeologisch booronderzoek (5);
- Waarderend archeologisch booronderzoek (6);
- Proefsleuven en proefputten (7);
- Proefputten i.f.v. prehistorische artefactensites (8);
- Opgraving (9).

De erkende archeologen kunnen (combinaties van) deze onderzoeksmethodes toepassen bij het uitvoeren van het archeologisch onderzoek, zowel rechtstreeks archeologisch vooronderzoek als



vooronderzoek in uitgesteld traject. De uitzondering is de opgraving. Deze onderzoeksmethode kan een archeoloog niet uitvoeren in een vooronderzoek en laten we weg.

De onderzoeksmethode proefputten i.f.v. prehistorische artefactensites komt geen enkele keer voor in de dataset, doordat deze methode nog niet werd toegepast in het kader van het Onroerenderfgoeddecreet. We kunnen deze methode dus ook niet meenemen in dit onderzoek en de uiteindelijke prijsbepaling.

In totaal komen we tot 16 verschillende combinaties van onderzoeksmethoden die erkende archeologen geregeld hanteren. Dit gaat van archeologienota's bestaande uit enkel bureauonderzoek tot bureauonderzoek met in uitgesteld traject verkennende en/of waarderende archeologische boringen en proefsleuven en/of proefputten.

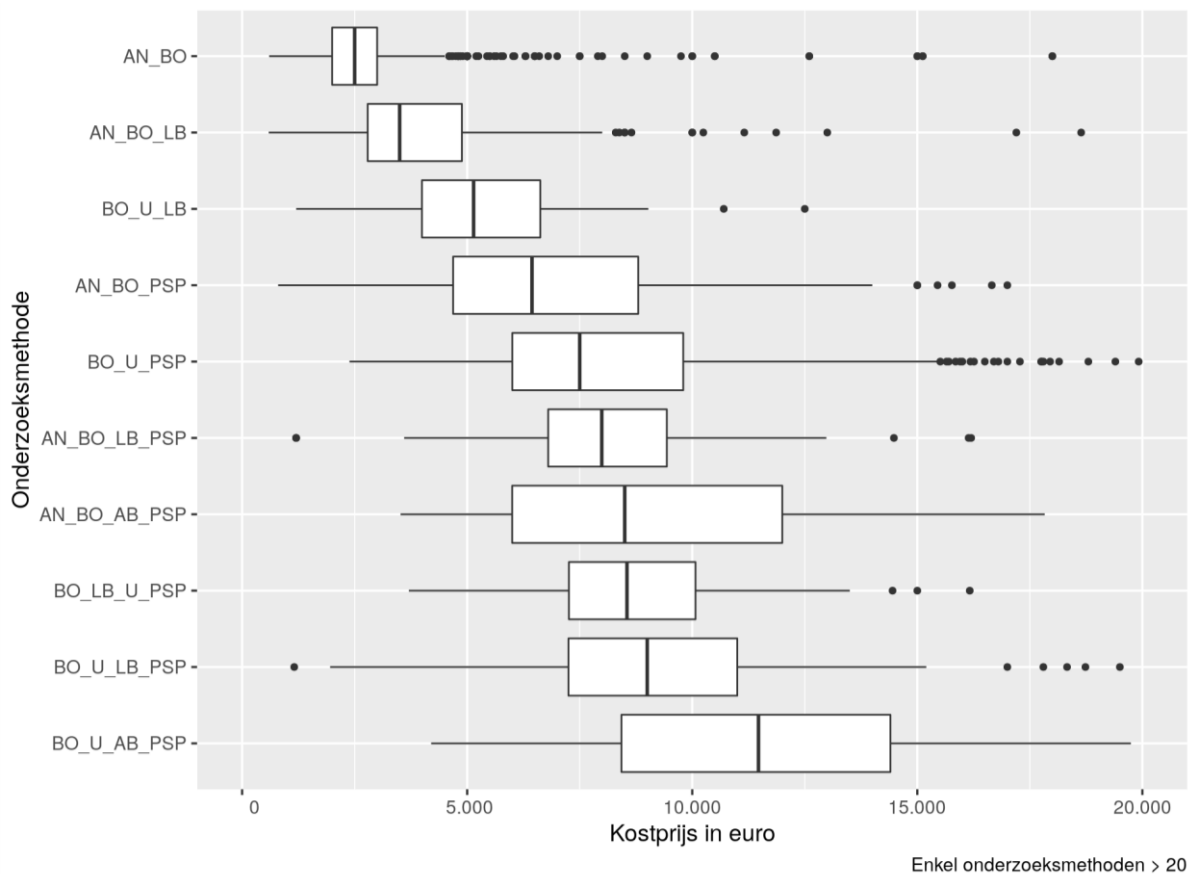
2.2.2 Boxplot van de kostprijs van onderzoeksmethodes

Onderstaande boxplot visualiseert de kostprijs van de dossiers per onderzoeksmethode (of combinatie van onderzoeksmethodes). Een **boxplot** is een compacte samenvatting van de verdeling van observaties in een dataset. De boxplot visualiseert de minimumwaarde, het eerste kwartiel, de mediaan, het derde kwartiel² en de maximumwaarde van de kostprijzen in onze dataset. De boxplot toont extreme waarden (waarden kleiner dan het eerste kwartiel of groter dan het derde kwartiel) met behulp van stippen.

Voor de verduidelijking beperkt deze boxplot zich tot onderzoeksmethodes met meer dan 20 dossiers en bevat ze geen dossiers met een hoger totaalbedrag dan 20.000 euro. We visualiseren op deze manier de grote meerderheid van de dossiers. Bijlage 1 bevat een tabel met de exacte cijfers en alle onderzoeksmethodes.

² Wanneer je een reeks opdeelt in 4 gelijke delen is het eerste kwartiel het eerste kwart (25%) van de dataset, het tweede kwartiel of de mediaan de helft (50%) van de dataset en het derde kwartiel driekwart of 75% van de dataset. Zie ook Wickham en Stryjewski, 2011.





Figuur 1: Boxplot van totale kostprijs archeologisch vooronderzoek. Onderverdeling naar onderzoeksmethode.

- Enkel bureauonderzoek = AN_BO
- Bureauonderzoek met landschappelijk bodemonderzoek = AN_BO_LB
- Bureauonderzoek met in uitgesteld traject landschappelijk bodemonderzoek = BO_U_LB
- Bureauonderzoek met proefsleuvenonderzoek = AN_BO_PSP
- Bureauonderzoek met in uitgesteld traject proefsleuvenonderzoek = BO_U_PSP
- Bureauonderzoek met landschappelijk bodemonderzoek en proefsleuvenonderzoek = AN_BO_LB_PSP
- Bureauonderzoek met archeologische boringen en proefsleuvenonderzoek = AN_BO_AB_PSP
- Bureauonderzoek met landschappelijk bodemonderzoek met in uitgesteld traject proefsleuvenonderzoek = BO_LB_U_PSP
- Bureauonderzoek met in uitgesteld traject landschappelijk bodemonderzoek en proefsleuvenonderzoek = BO_U_LB_PSP
- Bureauonderzoek met in uitgesteld traject archeologische boringen en proefsleuvenonderzoek = BO_U_AB_PSP

De methoden bureauonderzoek, landschappelijk bodemonderzoek, geofysisch onderzoek en veldkartering rekent de Code van Goede Praktijk tot onderzoeksmethoden **zonder ingreep in de bodem**. De overige, namelijk verkennend en waarderend archeologisch booronderzoek, proefsleuven en proefputten en proefputten in functie van prehistorische artefactensites zijn onderzoeksmethoden waarbij de erkende archeoloog wel een **ingreep in de bodem** uitvoert.



2.2.3 Vaststellingen uit de boxplot

Globaal kunnen we stellen dat naarmate de erkende archeoloog meer combinaties van onderzoeksmethodes samen gebruikt, **de mediaanprijs stijgt en het verschil tussen het eerste en derde kwartiel (de modale helft van de dossiers per onderzoekscategorie) toeneemt.**

Een archeologisch vooronderzoek dat enkel bestaat uit **bureauonderzoek** kost globaal genomen tussen 2.000 en 3.000 euro. De mediaan is exact **2.500 euro**. 75% van de vooronderzoeken met enkel bureauonderzoek kost minder dan 3.000 euro. De meerderheid van de archeologische vooronderzoeken in de dataset (63%) bestaat uitsluitend uit bureauonderzoek.

Archeologisch vooronderzoek bestaande uit **bureauonderzoek en landschappelijk bodemonderzoek** varieert in kostprijs naargelang de erkende archeoloog het vooronderzoek gedeeltelijk uitvoert in uitgesteld traject of niet. De mediaanprijs voor bureauonderzoek - landschappelijk bodemonderzoek ligt **op 3.500 euro**, het derde kwartiel op 4.883 euro. Wanneer de archeoloog (een deel van) het landschappelijk bodemonderzoek uitvoert in uitgesteld traject, stijgt de mediaanprijs naar 5.142 euro en het derde kwartiel naar 6.624 euro. Het verschil bij de mediaanprijs tussen een direct vooronderzoek en een vooronderzoek met uitgesteld traject bedraagt 1.642 euro.

Ook bij vooronderzoeken bestaande uit een **bureauonderzoek en proefsleuven/proefputten** zien we een gelijkaardig verschil. Hier is het verschil in mediaanprijs tussen uitvoeren in uitgesteld traject of niet kleiner, namelijk 900 euro. De erkende archeologen voeren de meerderheid van de onderzoeken met proefsleuven uit in uitgesteld traject.

Bij vooronderzoeken bestaande uit bureauonderzoek en proefsleuven/proefputten zien we het verschil tussen de mediaan en het gemiddelde fors toenemen, in tegenstelling tot vooronderzoeken bestaande uit enkel bureauonderzoek of landschappelijke boringen (waar het gemiddelde vrij dicht bij de mediaan ligt en er minder extreme waarden zijn). Bij vooronderzoek met bureauonderzoek en proefsleuven/proefputten ligt de **mediaan op 6.782 euro**, het gemiddelde al op 8.983 euro. Wanneer de erkende archeoloog de proefsleuven/proefputten deels of volledig in uitgesteld traject uitvoert bedraagt de mediaan 7.680 euro, het gemiddelde 9.134 euro. De gemiddelde kostprijs leunt bij deze onderzoeksmethode zeer dicht bij de prijs van het derde kwartiel. In totaal telt de dataset 622 vooronderzoeken met enkel bureauonderzoek en proefsleuven/proefputten als onderzoeksmethodes: 168 (27%) directe vooronderzoeken, 454 (73%) in uitgesteld traject.

Erkende archeologen combineren in hun vooronderzoek geregeld **bureauonderzoek met landschappelijk bodemonderzoek en proefsleuven/proefputten**. Dit komt voor in 130 dossiers (waarvan 60 in uitgesteld traject). De mediaanprijs van deze dossiers bedraagt **8.350 euro** bij direct vooronderzoek en 9.080 euro wanneer het onderzoek (gedeeltelijk) in uitgesteld traject gebeurt.

Opvallend is dat deze mediaanprijs niet veel verschilt van de mediaanprijzen van de afzonderlijke delen van de verschillende onderzoeksmethodes. Wanneer we de mediaanprijs nemen van dossiers met enkel bureauonderzoek en deze optellen met de mediaanprijs van dossiers met enkel bureauonderzoek en landschappelijk bodemonderzoek (met vermindering van de mediaanprijs van bureauonderzoek) en de mediaanprijs van dossiers met enkel bureauonderzoek en proefsleuven/proefputten (met vermindering van de mediaanprijs van bureauonderzoek) krijgen we het volgende: $2.500 + (3.500 - 2.500) + (6.800 - 2.500) = 7.800$ euro. Een verschil van 500 euro dat te verklaren kan zijn omwille van de relatief lage mediaan bij landschappelijk bodemonderzoek.



Wanneer we een inschatting maken van de kostprijs van landschappelijk bodemonderzoek van 1.500 euro (de gemiddelde prijs van landschappelijk bodemonderzoek) komen we op 8.300 euro. Een verwaarloosbaar verschil van 50 euro.

In het rapport evaluatie archeologie 2017 wordt de meerkost van het uitgestelde traject op 1.000 euro geraamd per archeologisch vooronderzoek.³ Dit komt overeen met de bevindingen in deze analyse. We vertrekken opnieuw van de mediaanprijs van 2.500 euro voor bureauonderzoek. Een archeologisch vooronderzoek met bureauonderzoek en landschappelijke boringen (gedeeltelijk) in uitgesteld traject heeft een mediaanprijs van 5.100 euro. We nemen aan dat het bureauonderzoek 2.500 euro is + 1.500 voor landschappelijk bodemonderzoek wat een prijs maakt van 4.000 euro, een verschil van 1.100 met de mediaanprijs wegens het uitgestelde traject.

Eenzelfde beeld valt op bij de vooronderzoeken met proefsleuven/proefputten in uitgesteld traject. $2.500 + 4.200 + 1.000 = 7.700$, een verschil van 20 euro met de mediaanprijs van bureauonderzoek met (gedeeltelijk) in uitgesteld traject uitgevoerde proefsleuven/proefputten (mediaan van 7.680).

Naast proefsleuven/proefputten zijn er nog 3 andere onderzoeksmethodes die voorkomen bij archeologisch vooronderzoek met ingreep in de bodem. Over de methode proefputten i.f.v. prehistorische artefactensites kunnen we helaas geen uitspraken doen. Momenteel zijn er amper archeologische vooronderzoeken geregistreerd waarbij archeologen van deze methode gebruikmaakten. Analyses over deze methode zijn momenteel niet mogelijk.

Verkennend archeologisch booronderzoek en waarderend archeologisch booronderzoek zijn 2 onderzoeksmethodes die zeldzaam zijn, maar toch al in voldoende vooronderzoeken toegepast om er voorzichtige uitspraken over te doen. Slechts in een heel laag aantal gevallen kiest een erkende archeoloog als onderzoeksmethode enkel bureauonderzoek met verkennend en/of waarderend archeologisch booronderzoek. Slechts in 16 dossiers komen deze onderzoeksmethodes apart voor, waarvan 6 in uitgesteld traject. De meeste dossiers met verkennend en/of waarderend archeologisch booronderzoek bevatten ook landschappelijk bodemonderzoek en/of proefsleuven/proefputten als onderzoeksmethodes. We stellen ook vast dat de archeologen de grote meerderheid van dossiers met deze combinaties aan onderzoeksmethodes uitvoeren in uitgesteld traject (53 dossiers tegenover 6 direct archeologisch vooronderzoek).

De **mediaanprijs** voor archeologisch vooronderzoek met bureauonderzoek, in sommige gevallen landschappelijk bodemonderzoek en in uitgesteld traject verkennend en/of waarderend archeologisch booronderzoek en proefsleuven/proefputten bedraagt **12.248 euro**. Nemen we de mediaanprijs van de andere onderzoeksmethoden en het uitgesteld traject in mindering, bekomen we een prijs van 3.050 euro. Deze prijs vergelijken met onderzoeken zonder uitgesteld traject is moeilijk aangezien het maar om 9 dossiers gaat.

Samenvattend komen we op volgende **kostprijzen (mediaan)** per onderzoeksmethode:

- Bureauonderzoek: 2.500 euro
- Landschappelijk bodemonderzoek: 1.500 euro
- Proefsleuven/proefputten: 4.200 euro

³ Agentschap Onroerend Erfgoed, evaluatie archeologie 2017, 2018.

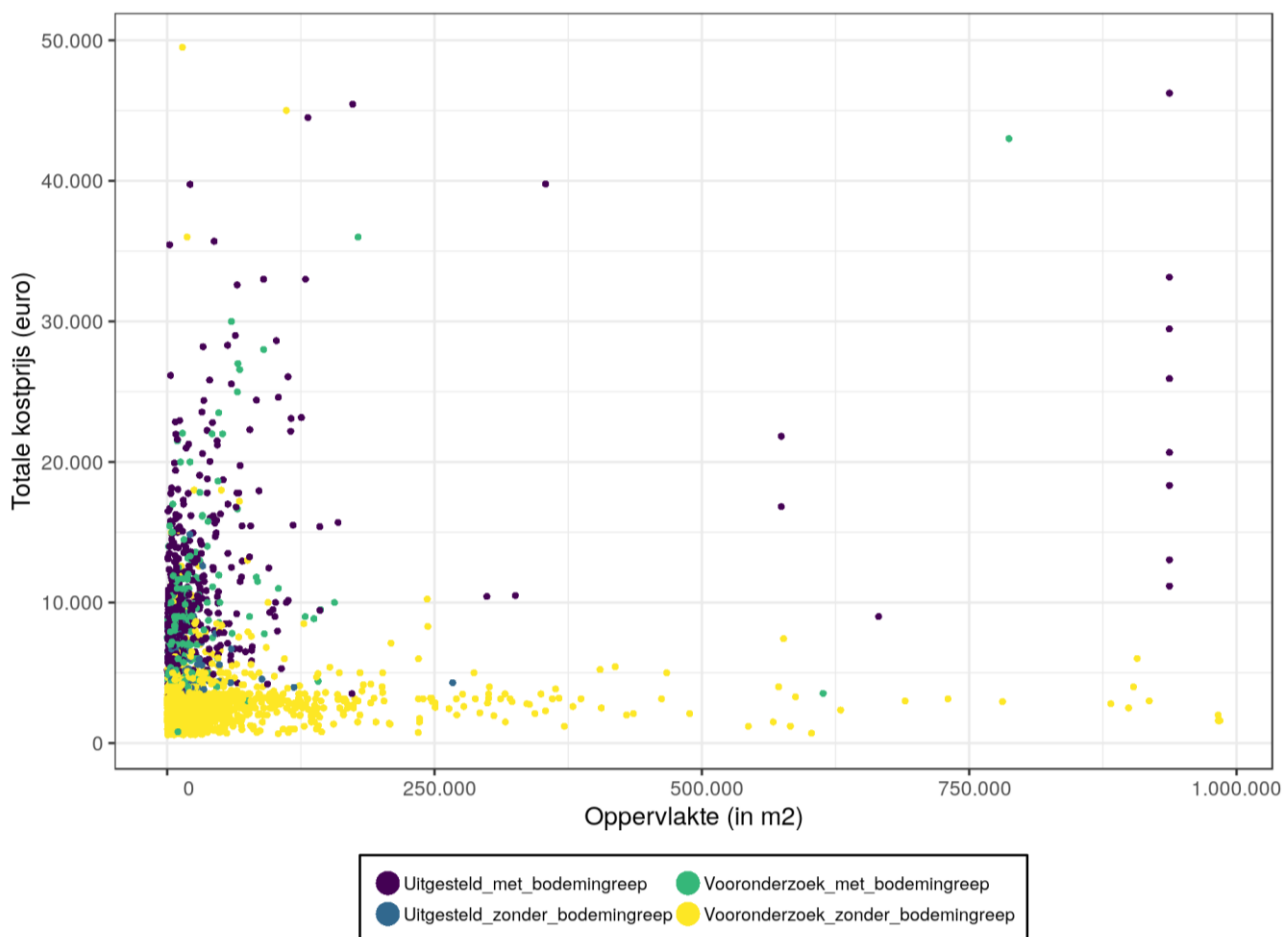


- Archeologisch booronderzoek: 3.050 euro
- Uitgesteld traject: 1.000 euro

2.3 KOSTPRIJS TEGENOVER OPPERVLAKTE

We kunnen de kostprijs afzetten tegenover de totale oppervlakte van elk projectgebied en visualiseren in een dotplot. Daaruit blijkt een grote variatie te zitten tussen de kostprijs van dossiers en hun oppervlakte. Voegen we kleur toe aan de spreiding op basis van de aanwezigheid van een ingreep in de bodem in het dossier en of het dossier is uitgesteld of niet, krijgen we een meer verhelderend beeld.

Er lijkt **geen verband tussen de kostprijs van het vooronderzoek en de oppervlakte van het terrein** bij vooronderzoek zonder ingreep in de bodem. Het merendeel van de vooronderzoeken zonder ingreep in de bodem kost minder dan 5.000 euro (bij uitsluitend bureauonderzoek is de mediaan 2.500 euro) ongeacht de oppervlakte van het terrein. De meerderheid van de vooronderzoeken met ingreep in de bodem kosten tussen 5.000 en 10.000 euro. Tegelijk heeft de overgrote meerderheid (85%) van de dossiers een onderzoeksgebied kleiner dan 50.000m² (5ha).



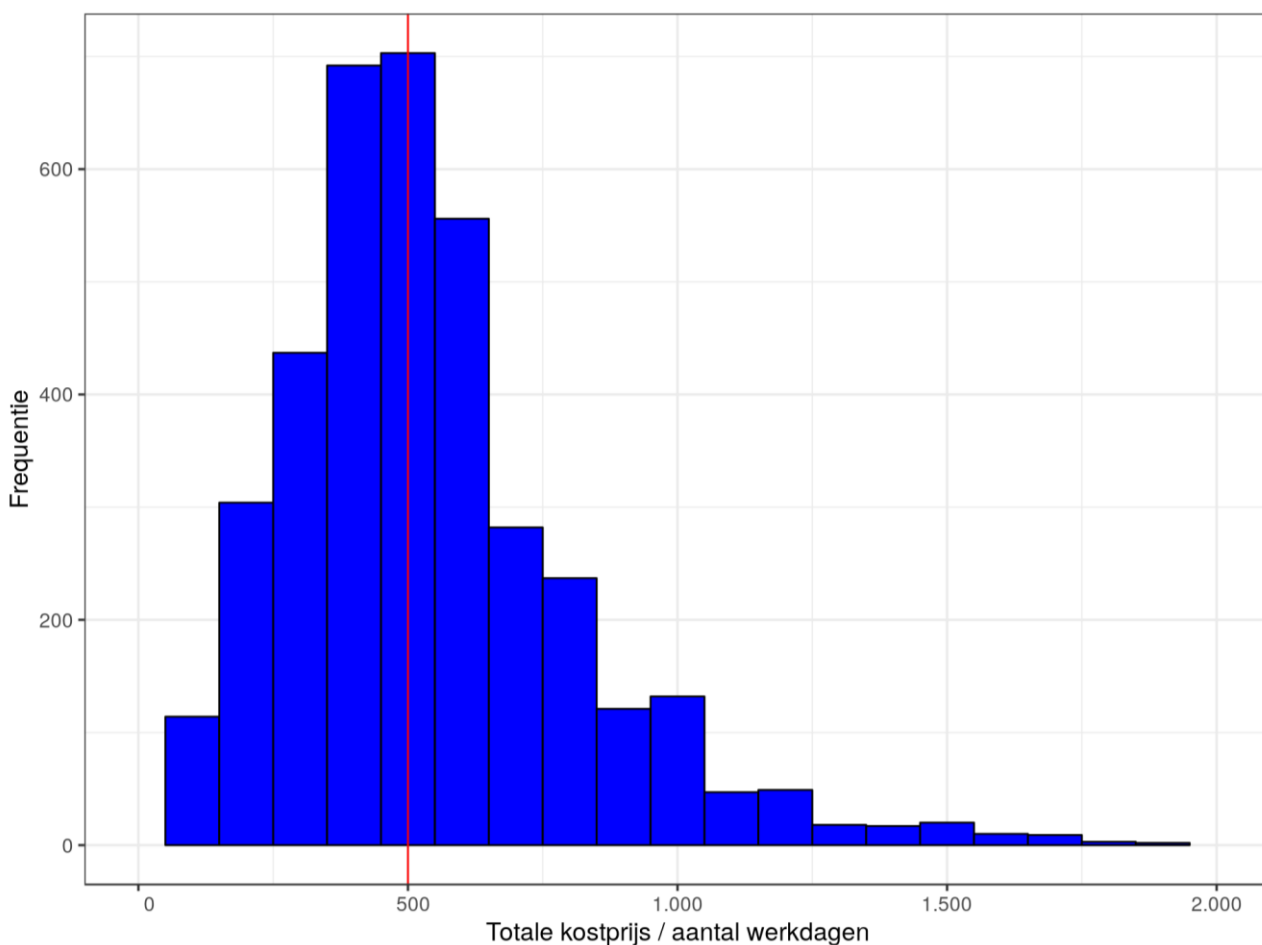
Figuur 2: Totale kostprijs vooronderzoek versus oppervlakte onderzoeksgebied. Meerderheid bureauonderzoek kost minder dan 5.000 euro, ongeacht oppervlakte.



2.4 KOSTPRIJS PER WERKDAG

De kostprijs per werkdag berekenen we door totale kostprijs per dossier te delen door het totale aantal werkdagen per dossier (de som van het aantal werkdagen per archeologienota en het aantal werkdagen per nota). Het modale archeologisch vooronderzoek duurt **6 werkdagen (mediaan)**. Het gemiddelde ligt opnieuw beduidend hoger (9 dagen) wat wijst op een beduidend aantal dossiers die veel meer werkdagen in beslag nemen (meer dan 25 tot uitzonderlijk zelfs 100).

Wanneer we de kostprijs afzetten tegenover het aantal werkdagen per dossier krijgen we een **mediaan van 500 euro per werkdag**. De verdeling is vrij normaal: naarmate het aantal werkdagen toeneemt, stijgt ook de prijs. De grote meerderheid van de dossiers kost tussen 400 en 600 euro per werkdag.



Figuur 3: Kostprijs per werkdag. Mediaanprijs per werkdag = 500 euro.

3 BEPALEN VAN SIGNIFICANTIE

3.1 INLEIDING BIJ SIGNIFICANTIE

De EDA leerde ons dat de kostprijs wordt bepaald door een aantal mogelijke variabelen, die elk een invloed kunnen spelen op de prijszetting van archeologisch vooronderzoek. Er blijkt een grote



variatie in kostprijs tussen en binnen verschillende onderzoeksmethodes, los van de oppervlakte van het onderzoeksgebied.

We toetsen in dit deel de invloed van verschillende onafhankelijke variabelen op de **responsvariabele (de totale kostprijs van het vooronderzoek)** en of deze invloed op toeval berust of niet. We willen met andere woorden weten welke variabelen een significante rol spelen in het bepalen van de kostprijs van archeologisch vooronderzoek.

We voeren de statistische analyses uit met behulp van R en een aantal R packages (zie bijlage). R heeft, net als andere statistische software zoals SPSS, SAS,... eigen algoritmes voor het berekenen van statistische significantie. De resultaten kunnen soms verschillen afhankelijk van de gehanteerde software. Indien nodig vermelden we welke functies we gebruikten voor deze analyses. Op die manier kunnen de analyses gemakkelijk worden overgedaan.

3.2 BEPALEN VAN EN OMGAAN MET EXTREME WAARDEN

De EDA liet zien dat de dataset een aantal “outliers” bevat. Deze extreme waarden verschillen sterk van de overige waarden van een bepaalde variabele. Dit verschil, de afstand tegenover de overige waarden, is subjectief. Maar zoals we konden vaststellen, hebben de “outliers” in sommige gevallen een significant effect op de verschillende centrummaten en dus op het bepalen of een bepaalde variabele al dan niet een effect heeft op de totale kostprijs van het archeologisch vooronderzoek.

De bedoeling van dit onderzoek is om het aandeel van verschillende variabelen te achterhalen op de totale kostprijs van archeologisch vooronderzoek. Daarom is het van belang om de “outliers” van de verschillende variabelen te detecteren en te onderzoeken in welke mate ze een invloed hebben op de regressiecoëfficiënt.

We beslissen om 2 soorten extreme waarden te verwijderen:

- Extreme waarden die onmogelijk zijn, bijvoorbeeld kostprijzen van 0, 1, 5, ... euro, onderzoeken waarbij het aantal werkdagen extreem hoog ligt (meer dan 500 werkdagen), vooronderzoeken met archeologische opgravingen, ...
- Extreme waarden met een uitzonderlijk hoge Cook’s Distance. Dit is een meeteenheid die de invloed uitdrukt van elke observatie op de regressiecoëfficiënt⁴. Een hoge Cook’s Distance wijst erop dat de observatie een significant effect heeft op de regressiecoëfficiënt. We kunnen deze eenheid gebruiken om de invloed van bepaalde outliers te bepalen en te beslissen welke observaties we beter uit onze analyse schrappen.

3.3 CORRELATIECOËFFICIËNT VAN NUMERIEKE VARIABELEN

De **correlatiecoëfficiënt** is een maatstaf voor verbanden tussen 2 variabelen. De coëfficiënt ligt altijd tussen -1 tot +1 (waarbij -1 een sterk negatief verband is en +1 een sterk positief verband). Wanneer de coëfficiënt dicht bij 0 ligt, wijst dit op het ontbreken van een verband tussen de variabelen. De coëfficiënt is negatief wanneer hoge waarden van de ene variabele gelinkt zijn aan lage waarden van

⁴ Dalgaard, 2008



de andere variabele en is positief wanneer beide variabelen ofwel tegelijk hoog zijn of simultaan laag zijn⁵.

We gebruiken de correlatiecoëfficiënt om verbanden te zoeken tussen de totale kostprijs van archeologisch vooronderzoek enerzijds, en andere numerieke variabelen in onze dataset anderzijds. Variabelen die niet numeriek zijn, onderzoeken we in het volgende deel met behulp van variantieanalyse.

De **totale kostprijs** geldt als onze responsvariabele als de focus van dit onderzoek. Verklarende variabelen zijn variabelen die mogelijk een invloed hebben op de responsvariabele. We onderzoeken dit voor volgende numerieke variabelen:

- De totale duur van het onderzoek (in werkdagen)
- De oppervlakte van het onderzochte gebied (we nemen altijd de oppervlakte van de bekrachtigde archeologienota hiervoor)
- Het totaal aantal keer dat het dossier is ingediend (waarbij 0 gelijk is aan rechtstreekse bekrachtiging)

In deze analyse filteren we enkel (extreme) waarden waarvan we veronderstellen dat ze onmogelijk zijn (zie eerste punt outliers). We gebruiken de functies “cor” en “cor.test” om de mogelijke relatie tussen de vermelde variabelen te detecteren. De functie cor.test genereert behalve de correlatiecoëfficiënt ook de betrouwbaarheidsinterval⁶.

Een **correlatiematrix** geeft snel een beeld over mogelijke verbanden tussen de verschillende variabelen. Zo blijkt er geen verband te zijn met de oppervlakte van het onderzoeksgebied of het aantal indieningen. De duur uitgedrukt in werkdagen blijkt wel een significant effect te hebben op de kostprijs.

| | Totale kostprijs | Totale duur vooronderzoek | Oppervlakte archeologienota | Aantal indieningen |
|-----------------------------|------------------|---------------------------|-----------------------------|--------------------|
| Totale kostprijs | 1.00 | 0.69 | 0.09 | 0.11 |
| Totale duur vooronderzoek | 0.69 | 1.00 | 0.06 | 0.13 |
| Oppervlakte archeologienota | 0.09 | 0.06 | 1.00 | 0.02 |
| Aantal indieningen | 0.11 | 0.13 | 0.02 | 1.00 |

Tabel 1: Correlatiematrix van de numerieke variabelen.

3.4 VARIANTEANALYSE VAN CATEGORISCHE VARIABLEN

In dit deel onderzoeken we of er een statistisch verband aanwezig is tussen de totale kostprijs van archeologisch vooronderzoek en een aantal **categorische variabelen**. Dit zijn niet-numerieke variabelen waarvan de waarden in te delen zijn in een aantal groepen. We testen in totaal 6 variabelen:

- De uitvoerder van de archeologienota
- De gemeente waar het vooronderzoek werd uitgevoerd
- De ligging in een beschermde archeologische site (ja/nee)

⁵ Dalgaard 2008, p. 121 - 123

⁶ R Core Team 2019



- De ligging in een vastgestelde archeologische zone (ja/nee)
- De gehanteerde onderzoeksmethodes
- Het voorgestelde programma van maatregelen

We voeren dit uit met behulp van **ANOVA (Analysis of Variance of variantieanalyse)**. We gebruiken dit om te onderzoeken of en welke verklarende variabelen een invloed hebben op de responsvariabele (de totale kostprijs van archeologisch vooronderzoek). We beginnen met eenvoudige lineaire regressie waarbij we onderzoeken of de bovenvermelde categorische variabelen een invloed hebben. Variabelen die een invloed hebben, bekijken we van naderbij met een post hoc test.

Bij een ANOVA ordenen we de data van een categorische variabele in verschillende groepen (factoren) en onderwerpen we deze groepen aan een **testhypothese**. De nulhypothese is dat de gemiddelden van de verschillende groepen dezelfde zijn (en er dus geen verschil is tussen de groepen). De alternatieve hypothese is dat het gemiddelde van minstens één groep niet gelijk is aan de andere groepen. In dat geval is er statistisch wel een verschil waar te nemen tussen de verschillende groepen.

We werpen om te beginnen een blik op de **F-waarde en p-waarde** van elke variabele. Een F-waarde van 1 geeft aan dat de gemiddelden van alle groepen aan elkaar gelijk zijn. De nulhypothese houdt dan stand. De F-waarde wijkt echter snel van één af. De vraag is dus wanneer de F-waarde groot genoeg is om de nulhypothese te verwerpen. **Aan de hand van de bijbehorende p-waarde kunnen we besluiten om de nulhypothese wel of niet te verwerpen.**

Wanneer de p-waarde valt buiten de aanvaardbare statistische interval, verwerpen we de nulhypothese. Als de **p-waarde kleiner is dan 0,05**, dan is de kans dat F toevallig is kleiner dan 5%. Het verschil tussen de gemiddelden is dan **significant**, ofwel: de categorische variabele heeft een significante invloed op de responsvariabele. *“Alternatively, you can calculate the p-value, which is the probability of obtaining a value as numerically large as or larger than the observed t and reject the hypothesis if the p-value is less than the significance level”⁷.*

Een belangrijke waarschuwing is wel dat statistische invloed niet automatisch causaliteit inhoud. Het kan een vertrekpunt vormen voor verder onderzoek naar gevolgen⁸.

We analyseerden 6 variabelen. Hiervan kunnen we 3 variabelen opmerken met een p-waarde kleiner dan 0.05).

⁷ Dalgaard, 2008, p. 96

⁸ Hoaglin, Mosteller & Tukey 1991



| Variabele | p-waarde |
|--|-----------|
| Uitvoerder | 0.0008531 |
| Gemeente (ligging) | 0.6854192 |
| Overlap met beschermde archeologische sites | 0.4300892 |
| Overlap met vastgestelde archeologische zones | 0.7746752 |
| Het voorgestelde programma van maatregelen | 0.000001 |
| De gehanteerde onderzoeksmethodes in het vooronderzoek | 0.000001 |

Tabel 2: significantie van categorische variabelen.

3.5 TEST OP VERSCHILLEN BINNEN VARIABLEN

Uit de verschillende anovatesten blijkt dat bij 3 variabelen de gemiddelde variantie van de verschillende groepen significant van elkaar verschillen. We berekenen nu welke groepen er binnen elke variabele significant van elkaar verschillen. We doen dit met behulp van een meervoudige vergelijking: de post hoc test. In dit onderzoek gebruiken we hiervoor **“Tukey Honest Significance Differences”**: **“Tukeyhsd”** (HSD)⁹. Deze functie test **alle paarsgewijze verschillen tussen de gemiddelde variantie van groepen**. De HSD functie genereert per paar een p-waarde die wijst op verschillen (p-waarde < 0.05 geeft aan dat het verschil bijna niet door toeval kan komen)¹⁰. *“Create a set of confidence intervals on the differences between the means of the levels of a factor with the specified family-wise probability of coverage. The intervals are based on the Studentized range statistic, Tukey’s ‘Honest Significant Difference’ method.”*¹¹

De ANOVA geeft een significant verschil aan bij de kostprijs van archeologisch vooronderzoek afhankelijk van de gehanteerde onderzoeksmethodes. De **“Honest Significant Differences”** versterkt dit beeld nog en vertelt ons welke groepen van onderzoeksmethodes een significant verschil van kostprijs hebben en welke niet.

54 van de 120 combinaties verschillen significant van elkaar. 47 van deze 54 combinaties bevatten **proefsleuven en proefputten als één van de onderzoeksmethodes, de overige 7 bevatten archeologische boringen** (zonder proefsleuven en proefputten). In bijlage is het resultaat te vinden van de HSD test.

De kostprijzen verschillen ook significant naargelang het bekrachtigde programma van maatregelen dat de archeoloog vooropstelde. Van de 66 combinaties zijn er 33 met een significant verschil in kostprijs. 32 hiervan bevatten in het programma van maatregelen **“uitstellen”**. De andere significante maatregel is **“opgraving”** tegenover **“geen maatregelen”**. Vooronderzoeken waarbij de archeoloog (een deel van) het onderzoek moest uitstellen of het gegeven dat na een vooronderzoek nog een opgraving moet volgen, verschillen dus sterk in prijs in vergelijking met vooronderzoeken zonder uitstel, of waar een behoud in situ volstaat als maatregel.

⁹ R Core Team 2019

¹⁰ Kabacoff 2015

¹¹ R Core Team 2019



4 ARCHEOLOGISCH VOORONDERZOEK MET PROEFSLEUVEN EN PROEFPUTTEN

4.1 INLEIDING BIJ PROEFSLEUVEN EN PROEFPUTTEN

De eerste correlatiematrix en ANOVA geven, versterkt door de “Honest Significant Differences”, aan dat de kostprijs van archeologisch vooronderzoek sterk wordt bepaald door de **gehanteerde onderzoeksmethodes en de verdere maatregelen** die nodig zijn voor het garanderen van optimaal behoud van kennis over het archeologische verleden. We bouwen verder op deze eerste resultaten om de archeologische onderzoeken waar de erkende archeologen, behalve bureauonderzoek, enkel proefsleuven of proefputten als onderzoeksmethoden inzetten, van naderbij te bekijken. De “Honest Significant Differences” gaf aan dat deze onderzoeken significant verschilden van de andere.

We nemen daarom een **steekproef** van deze archeologische vooronderzoeken om te onderzoeken welke bijkomende variabelen binnen deze groep bepalend zijn voor de kostprijs. We vullen de beschikbare informatie uit het Archeologieportaal aan met bijkomende variabelen. Op basis van de resultaten hiervan, bepalen we de kostprijs van de onderzoeksmethode “proefsleuven en proefputten”.

Uit de volledige dataset van archeologisch vooronderzoek filterden we alle vooronderzoeken met ingreep in de bodem waar de erkende archeoloog op het bureauonderzoek proefsleuven en/of proefputten liet volgen. Het gaat enkel om de **“exclusieve” proefsleuven**: landschappelijk bodemonderzoek, veldkartering, ... valt hier niet onder (ook niet in combinatie met proefsleuven).

Met behulp van een steekproef behielden we **250 archeologische vooronderzoeken** met als onderzoeksmethode enerzijds bureauonderzoek en anderzijds proefsleuven/proefputten. De steekproef had een redelijke variatie op het gebied van kostprijzen, oppervlaktes, archeologische regio’s¹² (enkel polders, duinen en de regio rond de Maas kwamen weinig of niet voor), archeologische zones, stadscontext,... 50 van de 250 vooronderzoeken waren vooronderzoeken met behulp van *proefputten in een complexe verticale stratigrafie (“stadscontext”)*.

De nieuwe dataset telt in totaal 86 verschillende variabelen. De meerderheid hiervan zijn TRUE/FALSE variabelen over bijvoorbeeld de aanwezigheid van een archeologische site, verschillende dateringen (middeleeuwen: TRUE/FALSE), ... Daarnaast zijn er numerieke en een aantal categorische variabelen.

4.2 CORRELATIEMATRIX VAN DATA PROEFSLEUVEN

Van de numerieke variabelen bouwen we een correlatiematrix. In totaal berekenen we de correlatie tussen 11 numerieke variabelen. Voor dit onderzoek gaat de aandacht uit naar de kostprijs van het vooronderzoek afgezet tegenover een aantal andere variabelen.

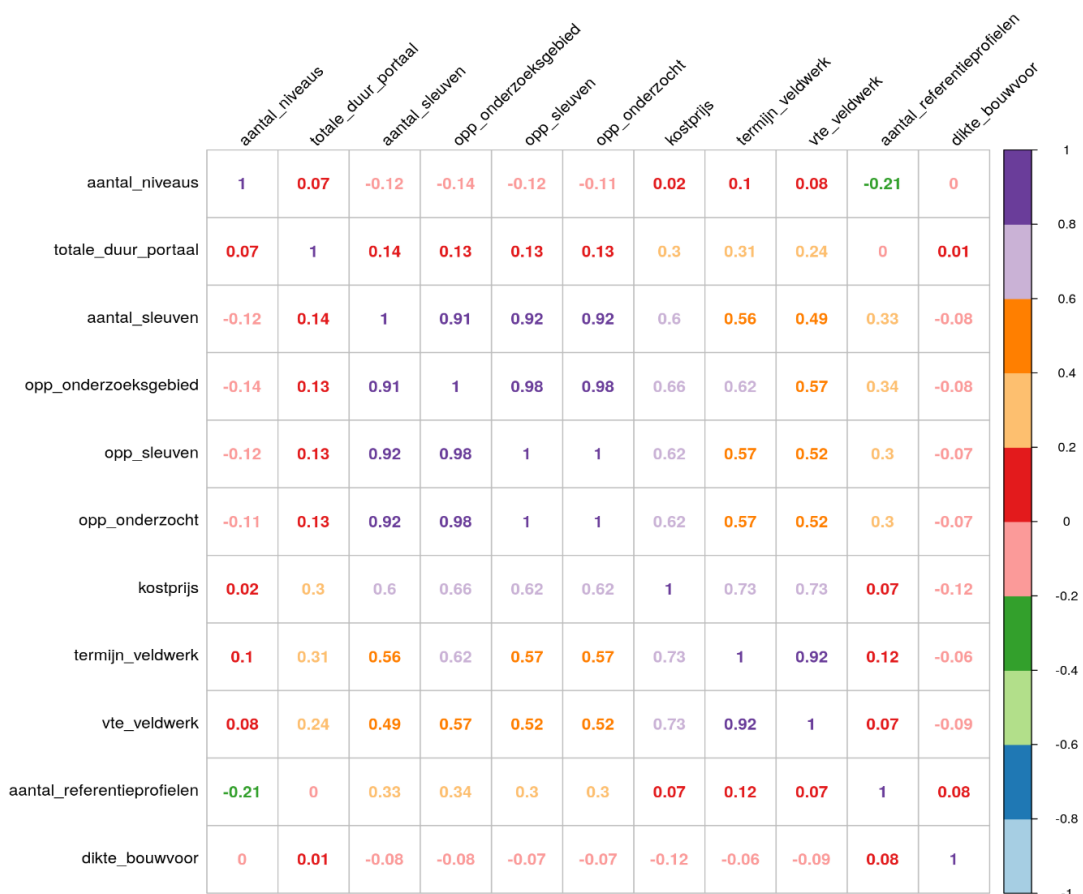
¹² Een archeoregio is “een gebied waarbinnen zowel sprake is van een globaal verband tussen landschap en bewoningsgeschiedenis als tussen landschapsvormende processen en het ontstaan van archeologische vindplaatsen, en het bodemarchief in het algemeen”. Elk van de Vlaamse archeoregio’s, die vooral afgebakend zijn op basis van de bodemassociatiekaart en de kaart van de lithologie van de oppervlakkige lagen heeft inderdaad een eigen landschapsvorming en dus ook bewoningsgeschiedenis gekend, en is gekenmerkt door specifieke condities voor de bewaring en bereikbaarheid van archeologisch erfgoed. Als het om inhoudelijke informatie gaat, is een opdeling in archeoregio’s dan ook relevanter dan een analyse per provincie of gemeente.



De correlatiematrix toont een positieve correlatie tussen de kostprijs van een vooronderzoek met proefsleuven en enerzijds variabelen die te maken hebben met de **oppervlakte van het gebied**¹³ (de oppervlakte van de bodemingreep, de effectief onderzochte oppervlakte en de totale oppervlakte van het onderzoeksgebied) en anderzijds met de **duur van het onderzoek**, namelijk de termijn van het veldwerk (aantal werkdagen) en het aantal ingezette VTE voor het uitvoeren van het veldwerk (aantal mensdagen).

De variabelen over oppervlakte hebben een correlatie van 0.6 tot 0.66, deze over de duur van het onderzoek zelfs 0.73. Beide wijzen op een relatief sterk positief verband met de kostprijs van een vooronderzoek.

Zowel de verschillende oppervlaktevariabelen als de variabelen over de duur van het onderzoek correleren ook sterk met elkaar. Uit de correlatiematrix kunnen we behalve een aantal significante variabelen ook **collineariteit** vaststellen tussen een aantal variabelen. Collineariteit wijst op de situatie waarbij twee of meer verklarende variabelen nauw verband houden met elkaar¹⁴. In het uiteindelijke model kunnen we maar 1 variabele van elk van deze overhouden om collineariteit te vermijden.



Figuur 4: Correlatiematrix data proefsleuven

¹³ De EDA wees op geen verband in de volledige dataset. Bij archeologisch onderzoek met ingreep in de bodem blijkt er wel een correlatie te zijn met de oppervlakte.

¹⁴ James et al. 2017, p. 99

4.3 ANOVA VAN DATA PROEFSLEUVEN

We analyseren de categorische variabelen met behulp van ANOVA om te bepalen welke een significant verband houden met de kostprijs.

| Variabele | p-waarde |
|--|-----------|
| Archeoregio | 0.0002010 |
| Aanwezigheid van een archeologische site | 0.0044360 |
| Gebouwde sporen gevonden | 0.0046967 |
| Programma van maatregelen | 0.0098695 |
| Vondsten verzameld | 0.0322075 |
| Gevonden sporen geregistreerd | 0.0343392 |

Tabel 3: Significantie van variabelen proefsleuven

De ANOVA levert 6 kwalitatieve variabelen die een significant effect hebben op de kostprijs. We sluiten alle andere variabelen uit omdat we geen (statistische) indicatie hebben dat ze een effect hebben op de kostprijs. De kolom van de uitvoerder van het vooronderzoek nemen we niet mee. In een normale marktwerking is het logisch dat elke uitvoerder zijn eigen prijzen bepaalt. Maar we kunnen deze variabele niet meenemen in de uiteindelijke formule voor de premie, aangezien dit een “rangorde” van archeologische uitvoerders naar kostprijs zou inhouden. Dit soort informatie zou de marktwerking te sterk verstoren en is bedrijfsgevoelige informatie.

Net als bij de correlatiematrix wijst de significantie van een aantal categorische variabelen op eenzelfde significant fenomeen dat zich uit in verschillende variabelen. Daarom moeten we vermijden om gelijkaardige variabelen mee te nemen om overfitting tegen te gaan.

De variabelen over het registreren van vondsten en sporen houden verband met **de aanwezigheid van een site** binnen het onderzoeksterrein. De aanwezigheid van een site uit zich vervolgens in het programma van maatregelen van de archeologienota. De “Honest Significant Differences” wees binnen het programma van maatregelen op 2 significante waarden:

- het al dan niet uitstellen van het vooronderzoek,
- de noodzaak om een opgraving uit te voeren na het vooronderzoek, of niet.

De significantie van de archeologische regio houdt verband met de locatie van het onderzoeksgebied. We onderscheiden in Vlaanderen 7 verschillende archeoregio’s¹⁵. Daaruit komt naar voor dat een vooronderzoek binnen een **stadscontext** significant verschilt van kostprijs met de andere vooronderzoeken. De andere archeoregio’s spelen geen significante rol, behalve de zandleem- en leemstreek. De kostprijs van deze verschilt significant van de kostprijs binnen een stadscontext. Dit kan wijzen op collineariteit bij deze variabelen. Bijna alle vooronderzoeken binnen een stadscontext zijn gelegen binnen een vastgestelde archeologische zone. De onderzoeksmethode binnen een stadscontext is anders dan erbuiten (proefputten in de plaats van proefsleuven), aangezien er binnen een stadscontext meestal sprake is van een complexe verticale stratigrafie en binnen een stadscontext vaker een opgraving volgt. De verklarende variabele is hier dus de onderzoeksmethode (proefputten) enerzijds en de maatregel (opgraving) anderzijds.

¹⁵ Het gaat om volgende regio’s: duinen, polders, zandleem- en leemstreek, zandstreek, kempen en maasland. Daar voegden we stadscontext nog aan toe voor gebieden binnen een historische stadskern.



5 ARCHEOLOGISCH VOORONDERZOEK MET ARCHEOLOGISCHE BORINGEN

5.1 INLEIDING BIJ ARCHEOLOGISCHE BORINGEN

De “*Honest Significant Differences*” wees ook archeologische boringen aan als methode die significant verschilt van de andere onderzoeksmethodes. Deze onderzoeksmethode komt veel minder vaak voor dan proefsleuven. Sinds de invoering van het Onroerenderfgoeddecreet voerden archeologen 71 vooronderzoeken uit waarbij ze **archeologisch booronderzoek** toepasten in het kader van het vooronderzoek. De code van goede praktijk maakt een onderscheid tussen verkennend archeologisch booronderzoek en waarderend archeologisch booronderzoek. Het verschil tussen beide ligt erin dat bij verkennend archeologisch booronderzoek de archeoloog een gerichte staalname van de bodem neemt door middel van grondboringen met het oog op het **opsporen** van archeologische sites, terwijl dit bij waarderend archeologisch booronderzoek gebeurt met het oog op het **evalueren** van reeds opgespoorde sites.¹⁶

We maken van alle vooronderzoeken met archeologische boringen één dataset en maken voor dit onderzoek **geen verder onderscheid tussen verkennend en waarderend archeologisch booronderzoek**. Dit is noodgedwongen, omdat het opsplitsen van de data in verkennend en waarderend archeologisch booronderzoek en een combinatie van beide, een te gefragmenteerde dataset oplevert. Bovendien is de onderzoekstechniek van beide methoden erg gelijkaardig: enkel de finaliteit verschilt.

In de volgende paragrafen bekijken we welke variabelen een significant effect hebben op de kostprijs van archeologisch booronderzoek. We doen dit op dezelfde manier als bij de proefsleuven, met correlatie en ANOVA.

5.2 CORRELATIEMATRIX VAN DATA ARCHEOLOGISCHE BORINGEN

Uit de dataset van archeologische boringen nemen we 8 numerieke variabelen mee in een correlatiematrix. Omwille van een aantal ontbrekende waarden behouden we 56 observaties waarop we de correlatie berekenen.

De correlatiematrix geeft een gelijkaardig beeld als de matrix van de proefsleuven. Opnieuw tonen de **termijn van het veldwerk**, het **aantal vte** ingezet tijdens het veldwerk en de **werkelijk onderzochte oppervlakte** van het terrein een sterke positieve correlatie op de kostprijs. Daar komt nog het **aantal uitgevoerde boringen** bij. Deze variabelen hebben een correlatie tussen de 0,63 en 0,76, wat wijst op een sterke positieve correlatie op de kostprijs van het vooronderzoek.

Tegelijkertijd stellen we vast dat een aantal van deze variabelen ook **onderling** sterk met elkaar correleren. Zoals bijvoorbeeld het aantal boringen en het aantal vte van het veldwerk (0,94). Deze correlatie is bijna dubbel zo sterk in vergelijking met de matrix van de proefsleuven (correlatie van 0,49 tussen het aantal proefsleuven en vte veldwerk). We vermijden daarom best om deze

¹⁶ Agentschap Onroerend Erfgoed, Code van Goede praktijk. 2019.

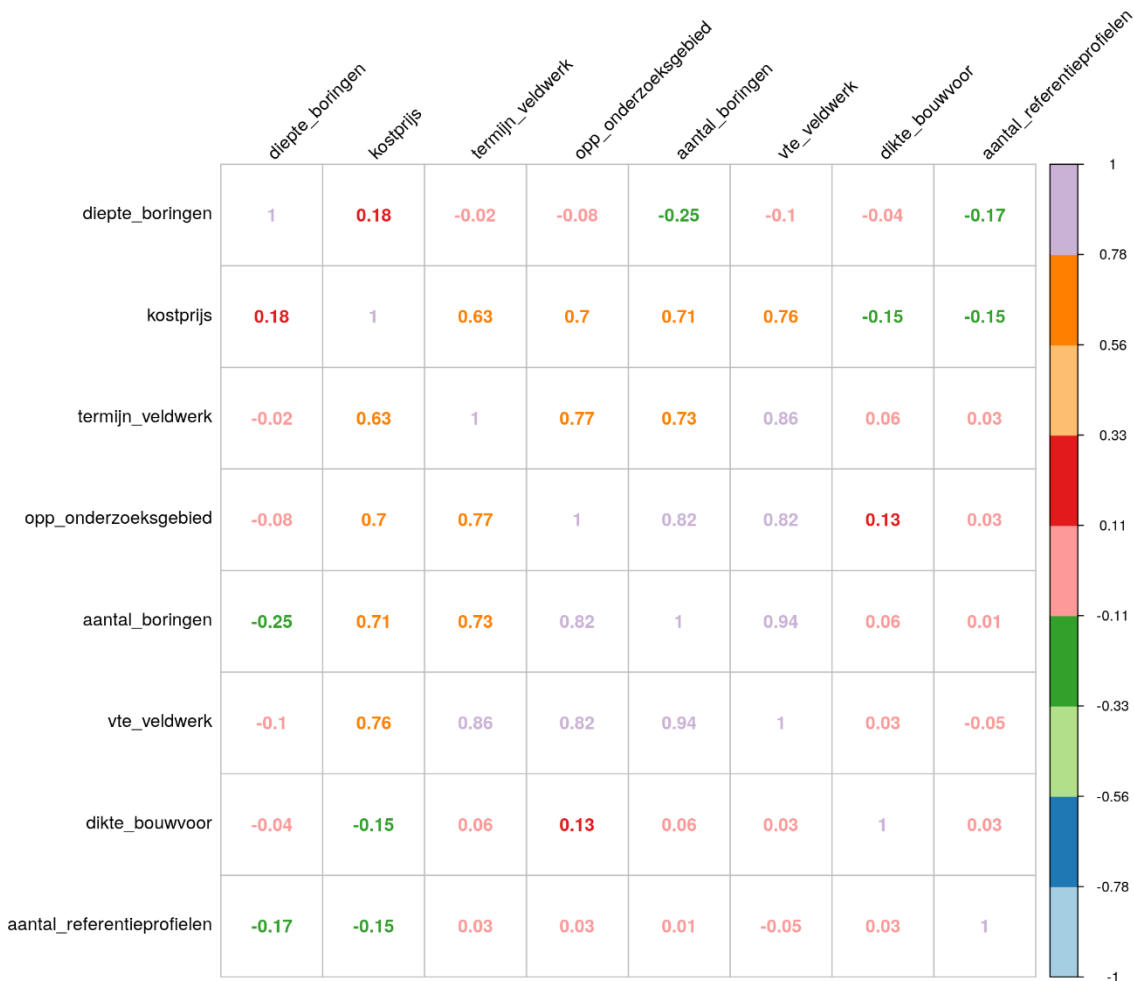


variabelen allebei mee te nemen in een regressiemodel. Merk tot slot de (zeer lichte) negatieve correlatie tussen de diepte van de boringen en het aantal boringen.

| Aantal boringen | VTE veldwerk |
|-----------------|--------------|
| 13.41968 | 13.41968 |

Tabel 4: Variance Inflation Factor van aantal boringen tegenover aantal vte van het veldwerk

De variance inflation factor (zie verder) van deze variabelen is meer dan 13, wat wijst op een zeer sterke collineariteit tussen deze variabelen.



Figuur 5: Correlatiematrix data archeologische boringen

5.3 ANOVA VAN DATA ARCHEOLOGISCHE BORINGEN

We voerden een ANOVA uit op 16 categorische variabelen uit de dataset van de archeologische boringen. Hieruit kwamen uiteindelijk **4 variabelen met een p-waarde kleiner dan 0.05** die we verder kunnen gebruiken voor het testen van regressiemodellen. De uitvoerder kwam, net als bij de proefsleuven, ook hier naar voren als significante variabele, maar nemen we verder niet mee aangezien we deze niet kunnen gebruiken in de uiteindelijke formule van de premie. De motivering hiervoor is dezelfde als bij de proefsleuven.



| Variabele | p-waarde |
|-------------------------------|-----------|
| Landgebruik | 0.0088566 |
| Archeoregio | 0.0291987 |
| Landschappelijk booronderzoek | 0.0355031 |
| bouwproject | 0.0364205 |

Tabel 5: Significantie van variabelen archeologisch booronderzoek

De meeste variabelen die ANOVA aanduidt als significant hebben te maken met de **omgeving**. Het gaat dan om het landgebruik¹⁷, het toekomstig bouwproject¹⁸ en de archeoregio. In tegenstelling tot bij de proefsleuven is het bij de archeologische boringen onmogelijk dat de ligging in een stadscontext of binnen een vastgestelde archeologische zone een rol speelt. Dit komt omdat de dataset geen archeologische boringen bevat binnen deze gebieden.

We stellen vast dat archeologische vooronderzoeken met archeologische boringen waar de erkend archeoloog ook **landschappelijk bodemonderzoek** uitvoerde, wel significant verschillen in vergelijking met andere vooronderzoeken.

Slechts een beperkt aantal vooronderzoeken met archeologische boringen gebeurden zonder andere onderzoeksmethodes. Geen enkele van deze onderzoeken resulteerde in een programma van maatregelen waar geen verdere maatregelen nodig waren of een behoud in situ voldoende bleek. De grote meerderheid (85%) van de vooronderzoeken met archeologische boringen gebeurde in combinatie met proefsleuven en/of landschappelijk bodemonderzoek. Een aantal van deze onderzoeken resulteerde wel in een programma van maatregelen waar een archeologische opgraving nodig was. Maar de eerste ANOVA uitgevoerd op deze dataset geeft “programma van maatregelen” niet aan als significante variabele bij archeologisch vooronderzoek met archeologische boringen.

6 MODELLEN OP BASIS VAN SIGNIFICANTE VARIABELEN

6.1 INLEIDING BIJ DE MODELLEN

De delen 4 en 5 leveren ons de significante variabelen aan waarop we een model kunnen bouwen dat de kostprijs inschat van de afzonderlijke variabelen in archeologisch vooronderzoek met proefsleuven en archeologisch vooronderzoek met archeologische boringen.

We stellen nu een aantal verschillende regressiemodellen op. We doen dit omdat het onmogelijk is om de responsvariabele “kostprijs van vooronderzoek” te verklaren aan de hand van slechts 1 verklarende variabele. We maken een selectie van variabelen die we als verklarende variabelen opnemen in het regressiemodel. We testen hierbij een aantal verschillende mogelijkheden en combinaties van variabelen.

¹⁷ Het landgebruik is een variabele die aangeeft hoe het projectgebied wordt gebruikt op het moment van het archeologisch vooronderzoek. We onderscheiden 6 soorten: weiland, braak, bebouwing, akker, bos of andere.

¹⁸ De variabele bouwproject geeft aan wat de initiatiefnemer van het onderzoek van plan is na het uitvoeren van het onderzoek. Hier onderscheiden we 5 soorten: lijninfrastructuur, nieuwbouw, omgevingsaanleg, verkaveling of uitbreiding/aanpassing bestaande.



We hanteren een “**mixed method**” voor het opbouwen van het model. Daarin starten we met 0 variabelen en voegen we 1 per 1 variabelen toe die het best passen. Naargelang we variabelen toevoegen zullen de p-waardes van de variabelen veranderen. Vanaf het moment dat de p-waarde van 1 van de variabelen in het model niet meer significant blijkt, verwijderen we deze variabele opnieuw. Dit blijven we doen tot het moment dat alle variabelen in het model een aanvaardbaar lage p-waarde hebben en alle variabelen die we weren uit het model een te hoge p-waarde krijgen wanneer we ze zouden toevoegen aan het model¹⁹.

Om te bepalen welke modellen als beste scoren gebruiken we verschillende statistische tests:

- *Adjusted R²*: Deze geeft het aandeel van de variantie aan dat het model kan verklaren. Een R² dicht bij 1 geeft aan dat het model een groot deel van de variantie in de responsvariabele kan verklaren²⁰.
- Akaike information criterion (**AIC**): Geeft een ingeschatte score van een model. We kunnen AIC niet op zichzelf gebruiken, maar wel relatief in verhouding tot andere modellen. Het model met de laagste AIC kunnen we beschouwen als het meest geschikte model.
- Bayesian information criterion (**BIC**): is gelijkaardig aan AIC maar gebruikt een andere manier van berekenen. Ook hier is, relatief in verhouding tot andere modellen, het model met de laagste BIC, het meest geschikte model.

Behalve de genoemde statistische tests kunnen we ook een **analyse uitvoeren van de residuen** van een lineair model om te onderzoeken in welke mate het model de variabelen vat. In een lineair model kunnen we de residuen omschrijven als het verschil tussen de geobserveerde waarde van y en de voorspelde waarde van y in het model.²¹ In dit geval wil dit zeggen dat de residuen **het verschil zijn tussen de effectieve kostprijs van een vooronderzoek en de voorspelde kostprijs** volgens het lineair model.

Wanneer we de residuen visualiseren in een grafiek zit er best geen systematisch patroon in de punten. Het gemiddelde van de residuen is idealiter ongeveer 0.

De eerste modellen bouwen we op de dataset van proefsleuven. Aan deze dataset hebben we een aantal vereenvoudigingen en veranderingen doorgevoerd:

We selecteren een subset van variabelen die ofwel een significante correlatie hadden in de correlatiematrix of die een significante p-waarde hadden in de eerste ANOVA.

- Vereenvoudiging op het programma van maatregelen
 - Een eerste vereenvoudiging voerden we door op het programma van maatregelen waarbij we de data opsplitsten tussen dossiers (deels) uitgevoerd in uitgesteld traject en dossiers die “direct” werden uitgevoerd.
 - Een tweede vereenvoudiging op het programma van maatregelen was de noodzaak tot een latere archeologische opgraving (ook als het slechts om een deel van het terrein gaat) en dossiers waar dit niet nodig was (geen maatregelen nodig, volledig behoud in situ, ...).
- Ingrepen in de kostprijs van het vooronderzoek

¹⁹ James et al. 2017, p. 78–79

²⁰ James et al. 2017, p. 80

²¹ Ismay & Kim. 2019.



- Een eerste ingreep in de kostprijs van het onderzoek is het in mindering brengen van het **uitgesteld traject**. We proberen deze kostprijs beter in te schatten door ze in te voeren als variabele in een afzonderlijk regressiemodel.
- Een tweede ingreep in de kostprijs is het in mindering brengen van het **bureauonderzoek**. Daarom brengen we 2.500 euro in mindering bij elk dossier. We doen dit omdat we de kostprijs willen weten van de ingreep in de bodem van het vooronderzoek. Uit eerdere analyses bleek dat de mediaanprijs van dossiers met enkel bureauonderzoek op 2.500 euro lag.

6.2 MODELLEN VOOR DE ONDERZOEKSMETHODE PROEFSLEUVEN

6.2.1 Inleiding

De dataset van de proefsleuven was nauw afgebakend met enkel archeologische vooronderzoeken met als onderzoeksmethode bureauonderzoek en proefsleuven of proefputten. Hierdoor kunnen we modellen testen die inschatten welke variabelen de totaalprijs van dit soort onderzoek bepalen (op basis van de eerdere ANOVA) en wat de afzonderlijke prijs ongeveer is van deze variabelen.

6.2.2 Model: Lineaire regressie van proefsleuven met 2 verklarende variabelen

De eerste reeks modellen die we testten zijn lineaire regressiemodellen waarbij één per één verklarende variabelen toevoegen om te testten of ze een significante impact hebben op de responsvariabele “kostprijs”.

Een model met slechts 1 verklarende variabele schiet ruimschoots tekort. De R^2 bedraagt 0.42. Voegen we een tweede variabele toe, stijgt de R^2 naar 0.62. Ook blijken beide variabelen significant. Dit toont aan dat minstens 2 variabelen nodig zijn.

Hieronder het resultaat van het eerste model waarbij we de kostprijs van het vooronderzoek afzetten tegenover de **oppervlakte van de proefsleuven** (de code van goede praktijk schrijft voor dat de proefsleuven als uitgangspunt 12,5% van het onderzoeksgebied moeten beslaan) en de **termijn van het veldwerk** (aantal werkdagen dat de erkende archeoloog effectief gewerkt heeft aan de proefsleuven²²).

```
## Call:
## lm(formula = kostprijs ~ opp_sleuven + termijn_veldwerk, data =
psp_model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10185.3  -1823.9   -357.1   1407.4  15107.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4719.4025    339.5268   13.900 < 2e-16 ***
## opp_sleuven     0.8593      0.1488    5.774 2.46e-08 ***
## termijn_veldwerk 1882.0562    168.6858   11.157 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

²² De termijn van het veldwerk moet de archeoloog vermelden in het dagrapport dat hij opmaakt bij het terreinwerk.



```
## Residual standard error: 3503 on 233 degrees of freedom
## Multiple R-squared: 0.6222, Adjusted R-squared: 0.6189
## F-statistic: 191.8 on 2 and 233 DF, p-value: < 2.2e-16
```

Om de R² van dit model nog te verhogen, voegen we er een derde variabele aan toe. Dit bleek echter geen effect meer te hebben, ongeacht welke derde variabele we toevoegden. De R² verhoogde amper nog en de derde variabele had geen significante p-waarde meer. Twee variabelen is echter te weinig om een degelijk model te bouwen dat de kostprijs van proefsleuven kan voorspellen. Daarom breidden we het model verder uit in het volgende deel.

6.2.3 Apart model: kostprijs van het uitgesteld traject

We stelden reeds eerder vast dat het uitgestelde traject een significante variabele is in de kostprijs van het archeologisch vooronderzoek. Het uitgestelde traject is in het leven geroepen als uitzonderingsmaatregel in gevallen wanneer het onmogelijk is om voorafgaand aan het aanvragen van de omgevingsvergunning het archeologisch onderzoek af te ronden. Deze vooronderzoeken blijken statistisch significant te verschillen van de overige en het is niet de bedoeling van de nieuwe premie voor archeologisch vooronderzoek met ingreep in de bodem om het uitstellen van het archeologisch vooronderzoek te compenseren. Daarom brengen we met een afzonderlijk regressiemodel de aparte kostprijs van het uitgestelde traject in kaart om dit bedrag in mindering te brengen van de totale kostprijs (net zoals we deden voor de kostprijs van het bureauonderzoek die we vastlegden op 2.500 euro).

We gebruiken dit regressiemodel aanvullend om de meerprijs van uitgesteld vooronderzoek exacter te bepalen. Dit doen we door de nieuw gevormde variabele “uitgesteld” versus “direct” op te nemen in het regressiemodel als tweede variabele. We nemen als responsvariabele de oorspronkelijke prijs per dossier en als verklarende variabelen de oppervlakte van de proefsleuven en het uitgesteld traject.

Het resultaat van deze formule is hieronder te zien:

```
## Call:
## lm(formula = kostprijs ~ opp_sleuven + uitgesteld_traject, data =
psp_model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8376.2 -2269.1  -935.5  1264.6 31404.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5912.7534   530.4399  11.147 < 2e-16 ***
## opp_sleuven      1.8887     0.1425  13.256 < 2e-16 ***
## uitgesteld_traject 1747.0295   605.0622   2.887  0.00425 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4263 on 233 degrees of freedom
## Multiple R-squared: 0.4403, Adjusted R-squared: 0.4355
## F-statistic: 91.66 on 2 and 233 DF, p-value: < 2.2e-16
```



In dit regressiemodel zijn alle variabelen significant. Dit model schat de meerprijs van het uitstellen in op **1.747,03 euro**. Om tegemoet te komen aan de inschatting van dit regressiemodel veranderen we de kostprijs voor het uitstellen, met een marge van 15%, tot 1.500 euro in de plaats van de eerder ingeschatte 1.000 uit de EDA. Hierdoor neutraliseren we grotendeels de kostprijs van het uitstellen van het vooronderzoek en **werken we in de nieuwe modellen verder met een aangepaste kostprijs**.

Het aanpassen van de kostprijs (mindering van bureauonderzoek en mindering van uitgesteld traject, indien van toepassing) leidt ertoe dat een aantal vooronderzoeken een negatieve prijs krijgen. We verwijderen daarom de “nieuwe” extreme waarden: alle vooronderzoeken waarbij de aangepaste kostprijs minder is dan 1.050 euro. Onze dataset behoudt hierdoor **nog 220 observaties**.

6.2.4 Model: “Regression through the origin”

Het model dat de kostprijs genereerde van het uitgesteld traject zorgt voor een **vernieuwde dataset**. In dit model werken we dus verder met kostprijzen waarbij zowel het bureauonderzoek als de prijs van het uitgesteld traject in mindering is gebracht.

Om het model te versterken voeren we nu een “*regression through the origin*” uit op het bestaande lineaire model, en met de vernieuwde kostprijsvariabele. Een “*regression through the origin*” kan een meerwaarde zijn wanneer er redenen zijn om aan te nemen dat **$y = 0$ wanneer $x = 0$** .²³ Dit is in dit onderzoek het geval. Wanneer er geen proefsleuven hoeven aangelegd te worden, legt de archeoloog ook geen oppervlak open en hoeft de archeoloog geen werkdagen uit te voeren op het terrein. In dat geval kan er geen kostprijs zijn voor de ingreep in de bodem. Er is dan geen “basiskost”.

Zoals we eerder zagen bleek in het oorspronkelijke model een derde variabele niet meer significant te zijn en hadden we een adjusted R^2 van ongeveer 0.62.

In dit model nemen we als derde variabele een vereenvoudigde vorm van het programma van maatregelen: is er een opgraving nodig na het vooronderzoek of niet?

Het resultaat staat hieronder:

```
## Call:
## lm(formula = kostprijs_bureau ~ 0 + opp_sleuven + termijn_veldwerk +
##     verder_traject, data = psp_regressiemodel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10274.1  -1758.7   -510.4   1070.8  13749.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## opp_sleuven          0.8889    0.1411   6.301 1.63e-09 ***
## termijn_veldwerk    1769.7528   165.9684  10.663 < 2e-16 ***
## verder_trajectGeen opgraving 1427.7584   333.1828   4.285 2.75e-05 ***
## verder_trajectOpgraving    2316.4947   572.7738   4.044 7.30e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

²³ Eisenhauer 2003



```
## Residual standard error: 3290 on 216 degrees of freedom
## Multiple R-squared: 0.8434, Adjusted R-squared: 0.8405
## F-statistic: 290.9 on 4 and 216 DF, p-value: < 2.2e-16
```

Een “*regression through the origin*” heeft een adjusted R² van **0.82** en bevat drie significante variabelen. Het is pas het toevoegen van een vierde verklarende variabele dat zorgt voor een verlies van significantie. De R² blijft ook ongeveer gelijk waardoor we concluderen dat 3 verklarende variabelen volstaan. Na het uitproberen van verschillende combinaties bleek de combinatie uit bovenstaande formule de beste. De variabelen zijn **de oppervlakte van de proefsleuven, de termijn van het veldwerk** aangevuld met de **noodzaak tot een latere opgraving of niet**.

Op dit model testen we de mogelijke aanwezigheid van **collineariteit** met behulp van de Variance Inflation Factor. De **Variance Inflation Factor (VIF)** is een statistische maatstaf voor het bepalen of er collineariteit aanwezig is binnen een model. De kleinste waarde is 1 en wijst op een volledige afwezigheid van collineariteit. Een VIF groter dan 5 wijst op een problematisch aandeel van collineariteit²⁴. We berekenen de VIF met behulp van `car::vif()`²⁵.

De vifcores in dit model schommelen tussen 1.2 en 2, wat wijst op de afwezigheid van collineariteit.

| Variabele | VIF |
|-----------------------------|----------|
| Oppervlakte proefsleuven | 1.456017 |
| Termijn veldwerk | 1.945857 |
| Opgraving of geen opgraving | 1.247084 |

Tabel 6: Variance Inflation Factor verklarende variabelen proefsleuven

6.2.5 Model: lineaire regressie op logaritmische schaal en na analyse residuen

De verdeling van de kostprijzen van proefsleuven is **ongelijk**. Dit bemoeilijkt het opstellen van een lineair regressiemodel. Om een accuratere inschatting te geven van de kostprijs van een vooronderzoek transformeren we nu de variabelen naar een **logaritmische schaal** en voeren we op het bestaande lineair model een **residuenanalyse** uit.

Zoals we zagen in de EDA is er een groot prijsverschil tussen de verschillende archeologische vooronderzoeken. Een klein aantal vooronderzoeken hebben een zeer hoge kostprijs en trekken de verdeling van de kostprijzen scheef. Om tot een evenrediger verdeling te komen transformeren we de prijzen naar een logaritmische schaal log₁₀. Een voorbeeld: 1.000 = 10³. Log₁₀ van 1.000 is dus gelijk aan 3.

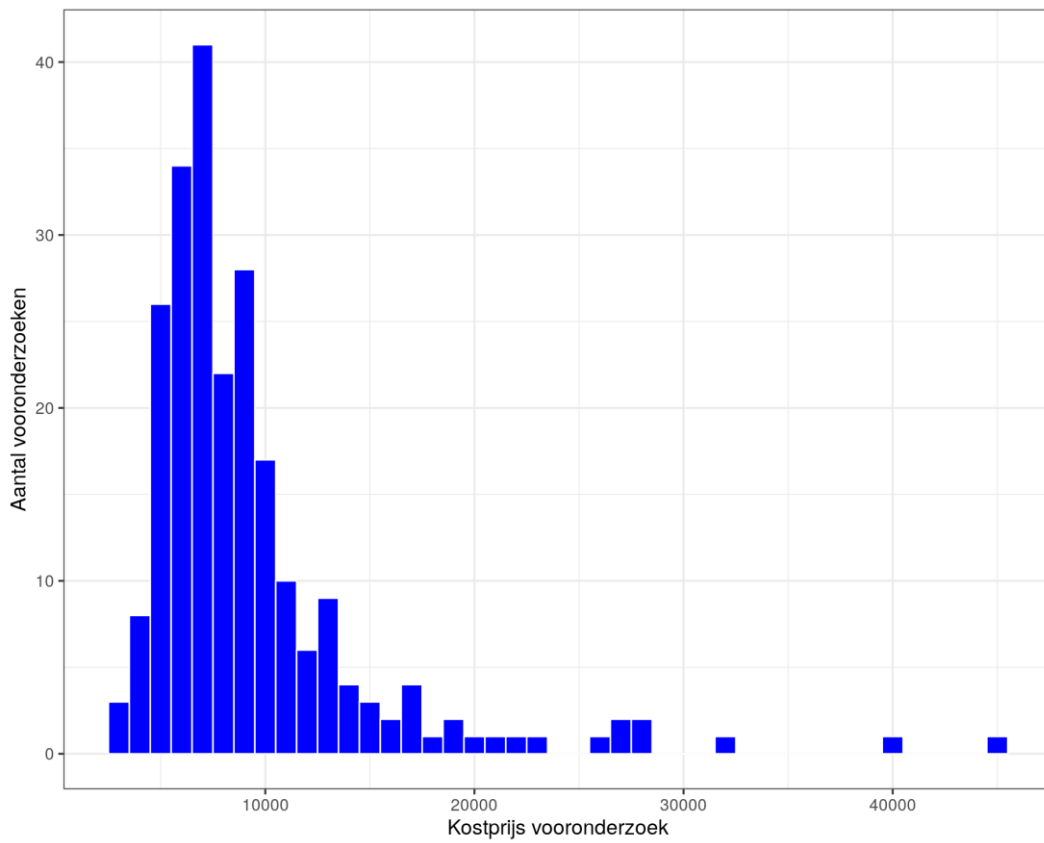
Deze transformatie zorgt voor een betere verdeling van de kostprijzen. Dit geeft een betrouwbaardere basis om een lineair model op te bouwen.

Onderstaande grafieken tonen de verdeling van de kostprijs van archeologisch vooronderzoek met proefsleuven, en kostprijzen van dezelfde dataset getransformeerd naar een logaritmische schaal.

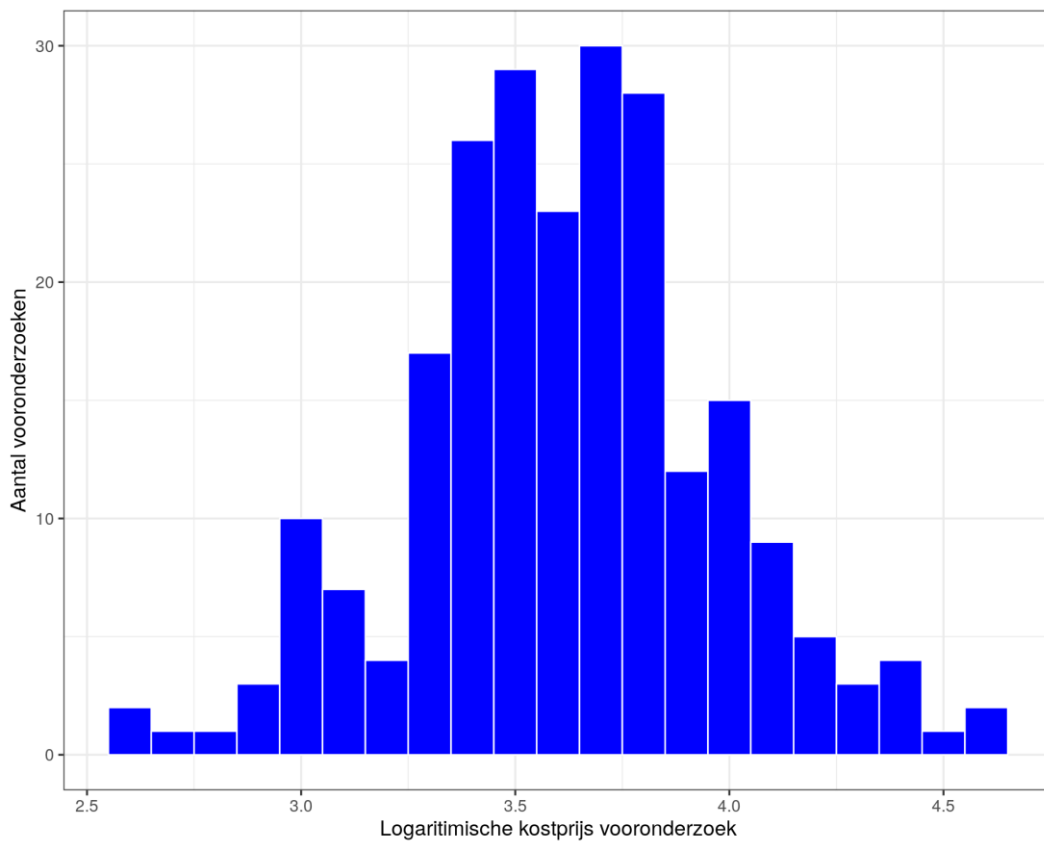
²⁴ James et al. 2017, p. 101–102

²⁵ Fox, Weisberg, and Price 2018





Figuur 6: Verdeling van de kostprijzen van archeologisch vooronderzoek met proefsleuven



Figuur 7: Verdeling van kostprijzen op logaritmische schaal van archeologisch vooronderzoek met proefsleuven



| Variabele | Kostprijs |
|---------------------|------------|
| Oppervlakte sleuven | 0,7146 |
| Termijn veldwerk | 2.008,8271 |
| Geen opgraving | 928,2984 |
| Opgraving | 2.176,2362 |

Tabel 7: Kostprijs lineair model met proefsleuven

Archeologisch vooronderzoek met proefsleuven of proefputten schatten we in als volgt:

- Aantal proefsleuven of proefputten * 0,7146
- Aantal werkdagen veldwerk * 2.008,8271
- Geen opgraving nodig + 928,2984
- Wel een opgraving nodig + 2.176,2362

We eindigen met een **quantile-quantile plot (QQ-plot)** van het model. Deze grafische vorm is een goed hulpmiddel om te testen of een statistisch model uitgaat van een normale verdeling en of de residuen in het model niet te ver afwijken van het theoretisch model. Het gemiddelde van de residuen in het finale model bedraagt 0.02. De grafiek geeft de kwantielen weer van het lineair model en zet deze af tegenover kwantielen berekend op basis van een theoretische verdeling.

De x-as van de QQ-plot bevat de theoretische waarden van het lineair model, terwijl op de y-as de waarden van de residuen staan.

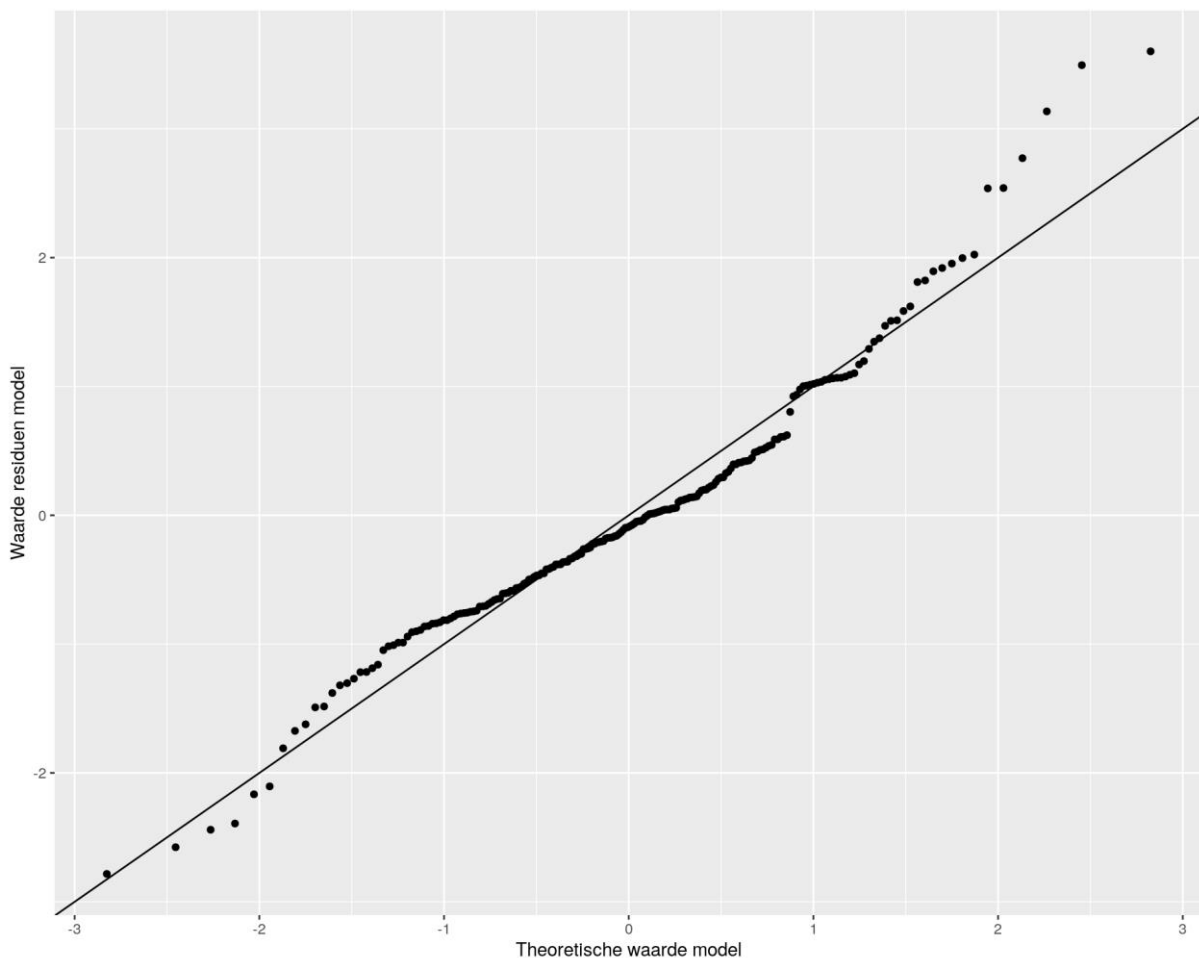
Deze werkwijze levert een accurater lineair model op, hoewel de variabelen dezelfde zijn als in het voorgaande model. De reden dat de inschattingen anders zijn, komt door het **verwijderen van de outliers** met behulp van de analyse van de residuen uit het eerdere model.

Het finale model bevat nog steeds een aantal outliers. Deze zijn ook te zien in de QQ-plot. Dit komt omdat het model uitgaat van 3 variabelen voor het bepalen van de prijs van een archeologisch vooronderzoek.

Zoals we zagen zijn deze variabelen doorslaggevend in alle vooronderzoeken met proefsleuven, maar er zijn uitzonderingen van vooronderzoeken waar bijkomende variabelen een rol kunnen spelen in de prijszetting. We opteerden ervoor om deze niet allemaal te verwijderen omdat dit de beperkte dataset te veel zou verkleinen. Enkel de variabelen met een te grote Cook's distance verwijderden we. We kunnen deze verklaren door de verscheidenheid in de archeologische markt (verschillen tussen bedrijven), verschillen in context (stadscontext met proefputten, polderlandschappen, onderzoek naar de Eerste Wereldoorlog,...), de aanwezigheid van menselijke resten in de site,...

Deze factoren kunnen allen meespelen en kunnen in sommige gevallen een effect hebben op de prijs. Ze kwamen niet als doorslaggevend naar voren uit de ANOVA, maar zorgen er wel voor dat we niet elk vooronderzoek perfect binnen dit lineair model kunnen vatten. Tegelijkertijd vinden we het niet opportuun binnen de doelstelling van het bepalen van de omvang van een premie voor archeologisch vooronderzoek met ingreep in de bodem om elk van deze vooronderzoeken uit de dataset te verwijderen. We behouden daarom de vooronderzoeken die binnen een aanvaardbare marge vallen ten opzichte van de anderen zonder te veel vooronderzoeken te verwijderen.





Figuur 8: QQ-plot lineair model proefsleuven

6.3 MODELLEN VOOR DE ONDERZOEKSMETHODE ARCHEOLOGISCHE BORINGEN

6.3.1 Inleiding

De dataset van de archeologische boringen is veel kleiner dan deze van de proefsleuven. Een tweede beperking is dat geen enkel booronderzoek zich situeerde in een vastgestelde archeologische zone en slechts 1 in een stadscontext (“site met complexe verticale stratigrafie”).

In veruit de meeste gevallen gaan archeologische boringen gepaard met een proefsleuvenonderzoek. We stelden ook vast dat van de weinige vooronderzoeken waar enkel archeologische boringen werden uitgevoerd, dus zonder proefsleuven, geen enkele leidde tot een archeologische opgraving en slechts in 1 geval een site aanwezig was (het programma van maatregelen stelde in dat geval een behoud in situ voor).

De kostprijzen van vooronderzoek met archeologische boringen zijn heel divers. Het bleek bijzonder moeilijk om patronen waar te nemen. Van de initiële kostprijs brachten we opnieuw, zoals bij de vooronderzoeken met proefsleuven, 2.500 euro in mindering als kostprijs voor het bureauonderzoek. In het geval van vooronderzoek uitgevoerd in uitgesteld traject deden we er nog eens 1.500 euro af. Dit leidde ertoe dat een aantal vooronderzoeken negatieve prijzen kregen. We beschouwen deze als



onrealistisch en verwijderen ze uit de dataset voor het bouwen van het regressiemodel. We deden dit ook voor een uitzonderlijk grootschalig project dat meer dan 300.000 euro kostte. Dit ging om een uitzonderlijk project voor de aanleg van lijninfrastructuur over een bijzonder groot oppervlak (dit project heeft een hoge Cook's distance en daarenboven komen dergelijke projecten niet in aanmerking voor de premie). Hierdoor bestaat de dataset nog uit 58 vooronderzoeken.

We bouwen het regressiemodel van de archeologische boringen op dezelfde manier als we deden bij de proefsleuven. Dit wil zeggen een *“mixed method”* waarbij we geleidelijk aan variabelen toevoegen tot we een aanvaardbare R^2 hebben en de VIF zo laag mogelijk is.

Zoals eerder vermeld bleek uit de correlatiematrix dat een aantal variabelen onderling sterk correleerden hoewel ze ook een hoge correlatie hebben met de kostprijs. We kunnen in het model geen variabelen meenemen die onderling sterk correleren, anders overschatten we het model en is er sprake van collineariteit. Dit bleek het geval tussen het aantal archeologische boringen en variabelen die handelen over personeelsinzet (termijn van het veldwerk, aantal vte).

De Variance Inflation Factor tussen archeologische boringen en termijn veldwerk bedraagt ongeveer 8, bij het aantal vte is dit zelfs 10. Een VIF hoger dan 5 wordt over het algemeen als problematisch beschouwd²⁶.

| Variabele | VIF |
|------------------|-----|
| Aantal boringen | 8 |
| Termijn veldwerk | 8 |
| VTE veldwerk | 13 |

Tabel 8: Variance Inflation Factor verklarende variabelen archeologische boringen

6.3.2 Model: Regressie met 2 verklarende variabelen

Net als bij de proefsleuven kiezen we voor een *“regression through the origin”* voor het bouwen van het model.

Als eerste variabele voor het model kunnen we kiezen tussen enerzijds het aantal archeologische boringen of een variabele over personeelsinzet. Deze hebben de hoogste correlatiewaarden. **Aantal archeologische boringen** heeft een correlatie van 0,69 op de (verwerkte) kostprijs, **vte veldwerk** zelfs 0,73.

Het zoeken naar een tweede variabele bleek veel moeilijker. In veruit de meeste gevallen bleek de significantie van de tweede variabele sterk te schommelen rond een p-waarde van 0,05. Kleiner dan 0,05 is significant, maar in de meeste modellen bleken de variabelen sterk te flirten met deze grens waarbij ze er in de meeste gevallen deels onder, deels boven zaten. We kunnen dus niet aantonen dat deze variabelen relevant zijn.

Variabelen die hier het best scoren zijn variabelen over het verder traject (het programma van maatregelen, of een vereenvoudigde versie opgraving/geen opgraving) of de aanwezigheid van landschappelijke boringen. Het nadeel van beide variabelen is de **ongelijke verdeling** ervan in de dataset. Slechts 8 vooronderzoeken met archeologische boringen leidden tot een opgraving en 8

²⁶ James et al. 2017, 101–2



omdat in dit model ook elk van de afzonderlijke variabelen een significante p-waarde heeft, wat bij de andere modellen niet het geval is.

Ook de VIF van dit model is gunstig. Er blijkt geen sprake van collineariteit.

| Variabele | VIF |
|-----------------|----------|
| Aantal boringen | 1,489154 |
| Verder traject | 1,220309 |

Tabel 9: Variance Inflation Factor verklarende variabelen archeologische boringen, regressie met 2 verklarende variabelen

| Variabele | Kostprijs lineair model |
|-----------------|-------------------------|
| Aantal boringen | 102,23 |
| Geen opgraving | 4.121,51 |
| Wel opgraving | 6.502,23 |

Tabel 10: kostprijs lineair model met archeologische boringen en verder traject opgraving / geen opgraving

6.3.3 Model: Lineaire regressie met aanpassing prijs proefsleuven

De voorgaande modellen gingen uit van de hypothese dat archeologisch vooronderzoek met archeologische boringen los staat van andere onderzoeksmethodes. Dit is in het grootste deel van de vooronderzoeken niet het geval. We gebruiken daarom het regressiemodel van de proefsleuven als inschatting van de kostprijs en brengen deze prijs in mindering van die vooronderzoeken met archeologische boringen waar de archeologen ook proefsleuven als onderzoeksmethode gebruikten. Bij de overige archeologische boringen brengen we enkel het bureauonderzoek en het eventuele uitgestelde traject in mindering.

Deze oefening resulteerde in een dataset van 49 observaties. Dit komt omdat bij een aantal vooronderzoeken na het in mindering brengen van de inschatting van de proefsleuven een negatieve prijs als resultaat naar boven kwam. Deze verwijderden we uit de dataset.

Op de nieuwe dataset testten we verschillende modellen uit, vertrekkende van de resultaten van het vorige model. Ook hier is het toevoegen van een tweede variabele al moeilijk en een derde variabele toevoegen blijkt zelfs helemaal onnodig.

Geen van de modellen uit het voorgaande deel bleken nog significant te zijn na het aanpassen van de kostprijsvariabele. Het model dat het beste scoort staat hieronder.

```
## lm(formula = verwerkte_kostprijs ~ 0 + aantal_boringen + proefsleuven,
##       data = model_prijs_psp)
##
## Residuals:
##    Min       1Q   Median       3Q      Max
## -13884   -6467   -1560    2332   41995
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## aantal_boringen         86.80     24.46   3.549 0.000905 ***
## proefsleuvenGeen proefsleuven 10997.84     3228.11   3.407 0.001374 **
## proefsleuvenProefsleuven     1002.34     2315.22   0.433 0.667086
## ---
```



```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10050 on 46 degrees of freedom
## Multiple R-squared:  0.5651, Adjusted R-squared:  0.5368
## F-statistic: 19.93 on 3 and 46 DF,  p-value: 2.019e-08
```

Het nieuwe lineaire model vertrekt van een andere kostprijs dan het vorige. Verder blijven we uitgaan van een regressie door de origine en het aantal boringen als variabele. We nemen de termijn van het veldwerk niet mee. Dit heeft een te grote collineariteit met het aantal boringen.

De variabele verder traject (opgraving of niet) bleek niet meer significant in dit model, net als landschappelijke boringen, de archeoregio,... Een variabele die degelijk scoort is “**proefsleuven**”, met andere woorden de aanwezigheid of afwezigheid van proefsleuven in het vooronderzoek. Dit geeft een redelijke R² van 0,54.

Het interessante van dit model is dat het uitgaat van een **stabiele prijs van het aantal boringen, en vervolgens een onderscheid maakt tussen vooronderzoeken met proefsleuven en vooronderzoeken zonder**. Dit maakt dit model bruikbaar als aanvulling in de formule van de proefsleuven. In het geval er proefsleuven zijn, gebruiken we het model van de proefsleuven aangevuld met de prijszetting van het aantal archeologische boringen. In het andere geval gebruiken we dezelfde inschatting van het aantal archeologische boringen en vullen we dit aan met het vast bedrag zoals het model weergeeft van de kostprijs van vooronderzoek zonder proefsleuven maar met archeologische boringen.

6.3.4 Model: Lineaire regressie met logaritmische schaal en na analyse residuen

Net als bij de proefsleuven bevat de dataset met archeologische boringen een zeer grote variatie van kostprijzen. We passen voor deze analyse dezelfde methode toe waarbij we de variabelen omzetten naar een logaritmische schaal, daarop een lineair model bouwen en vervolgens op basis van een residuenanalyse een nieuw model opbouwen.

De omzetting naar een logaritmische schaal geeft een meer normale verdeling van de kostprijzen als resultaat. Toch is het resultaat minder gunstig dan bij de proefsleuven, maar de dataset is ook veel kleiner en gevarieerder.

In dit model verwijderen we alle observaties waarvan de residuen kleiner zijn dan -0.5 of groter dan 0.5. De uitgevoerde residuenanalyse levert een dataset op van 34 observaties. Deze zouden nauw moeten aanleunen bij een ‘normaal’ archeologisch vooronderzoek met archeologische boringen.

```
## Call:
## lm(formula = unlogprijs ~ 0 + unlogboring + proefsleuven, data =
lm_ab_residuen)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4067.2 -1898.1 -297.2  1980.0  3194.5
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## unlogboring                   42.144     8.382    5.028 1.98e-05 ***
## proefsleuvenGeen proefsleuven 4661.538   920.508    5.064 1.79e-05 ***
```

//

```
## proefsleuvenProefsleuven      2310.514      666.646      3.466      0.00157 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2302 on 31 degrees of freedom
## Multiple R-squared:  0.8621, Adjusted R-squared:  0.8487
## F-statistic: 64.58 on 3 and 31 DF,  p-value: 1.964e-13
```

Dit model levert een R^2 op van 0,85. Per bijkomende archeologische boring komt er volgens dit model 42 euro bij in de totaalprijs van een vooronderzoek, rekening houdende met het al dan niet aanleggen van proefsleuven in het vooronderzoek. Bij de residuenanalyse hielden we maximaal rekening met de sterke ondervertegenwoordiging van vooronderzoeken met archeologische boringen, maar zonder proefsleuven. Toch blijft het een feit dat, gezien het zeer lage aantal van deze vooronderzoeken, het model is gebouwd op slechts een zeer klein aantal observaties met een grote variatie aan prijzen.

Dit lineair model levert volgende inschattingen op:

| Variabele | Kostprijs lineair model |
|-------------------|-------------------------|
| Aantal boringen | 42,144 |
| Geen proefsleuven | 4.661,538 |
| Wel proefsleuven | 2.310,514 |

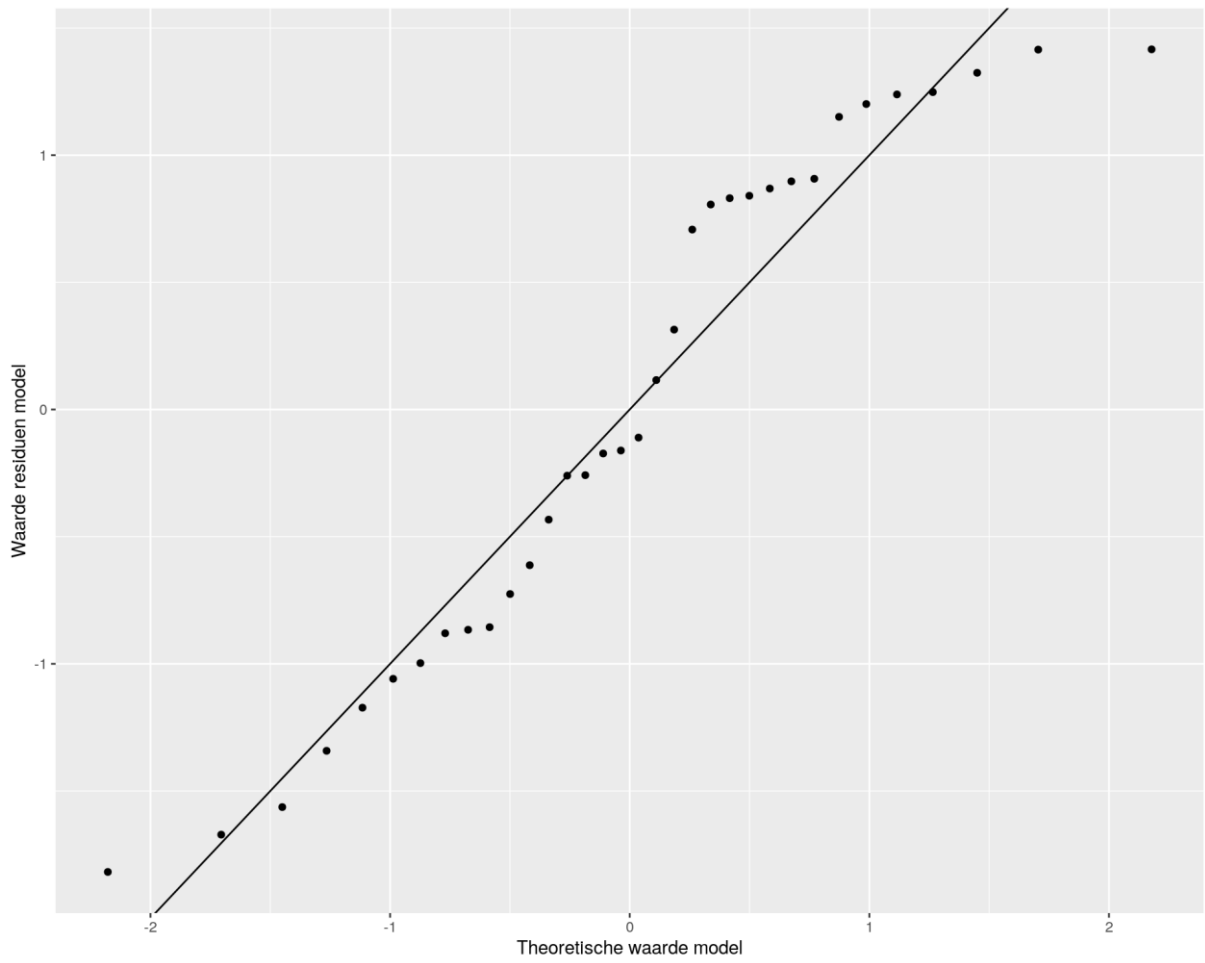
Tabel 11: Kostprijs lineair model met archeologische boringen, met prijsaanpassing model proefsleuven en na analyse van residuen

Archeologisch vooronderzoek met archeologische boringen voorspellen we op volgende manier:

- **Aantal boringen** * 42,144
- **Onderzoek bevat geen proefsleuven:** + 4.661,538
- Het model geeft ook een “rest” inschatting van de kostprijs van proefsleuven. Aangezien deze reeds vervat zit in de formule van de proefsleuven, nemen we deze niet mee in de formule van de archeologische boringen.

We eindigen opnieuw met een QQ-plot van het model. Deze grafische vorm is een goed hulpmiddel om te testen of een statistisch model uitgaat van een normale verdeling en of de residuen in het model niet te ver afwijken van het theoretisch model. Het gemiddelde van de residuen in dit model bedraagt 0,00006.





Figuur 9: QQ-plot lineair model archeologische boringen



7 BESLUIT EN VOORSTEL VAN FORMULE

Het onderzoek testte de significante variabelen op de kostprijs van een archeologisch vooronderzoek. Op basis hiervan bouwden we verschillende modellen die een inschatting maken van de kostprijs van elke afzonderlijke variabele.

Het meest geschikte model dat we genereerden uit de analyse van een dataset met als onderzoeksmethode **proefsleuven** bevatte 3 verschillende significante variabelen met een inschatting van hun prijzen. Het model is een **lineair model door de origine** en heeft volgende formule: Kostprijs vooronderzoek $\sim 0 +$ oppervlakte proefsleuven + termijn veldwerk + programma van maatregelen.²⁷

Hieruit leiden we volgende prijzen af:

| Variabele | Kostprijs model |
|---------------------|-----------------|
| Oppervlakte sleuven | 0,7146 |
| Termijn veldwerk | 2.008,8271 |
| Geen opgraving | 928,2984 |
| Opgraving | 2.176,2362 |

Tabel 12: Variabelen lineair model proefsleuven

Archeologisch vooronderzoek met archeologische boringen gaat in de meeste gevallen gepaard met een proefsleuvenonderzoek. We gebruiken daarom de kostprijzen uit bovenstaande analyse van de proefsleuven om de kostprijs van het aandeel “proefsleuven” in te schatten en in mindering te brengen in de kostprijs van de archeologische boringen. Hier voegen we de significante variabele “aantal boringen” aan toe om te komen tot een **lineair model** dat de kostprijs van archeologisch vooronderzoek met **archeologische boringen** inschat: kostprijs $\sim 0 +$ aantal boringen + proefsleuven.

Hieruit leiden we volgende prijzen af:

| Variabele | Kostprijs model |
|-------------------|-----------------|
| Aantal boringen | 42,144 |
| Geen proefsleuven | 4.661,538 |

Tabel 13: Variabelen lineair model archeologische boringen

Op basis hiervan komen we tot volgende **formule** voor het berekenen van de kostprijs van archeologisch vooronderzoek met ingreep in de bodem:

$$[(a * 42,14) + b + (c * 0,71) + (d * 2.008,83) + e]$$

- Aantal archeologische boringen (verkennend + waarderend) = **a**
- Geen proefsleuven volgend op de archeologische boringen = **b**
- Oppervlakte van aangelegde proefsleuven/proefputten in m² = **c**
- Termijn van het veldwerk van de proefsleuven/proefputten in werkdagen = **d**
- Maatregel: opgraving of geen opgraving = **e**

²⁷ Waarbij in de kostprijs het bureauonderzoek en het eventuele uitgestelde traject in mindering is gebracht. Het programma van maatregelen is vereenvoudigd naar de noodzaak om een opgraving uit te voeren of niet.



9 BIBLIOGRAFIE

Agentschap Onroerend Erfgoed. 2019. Code van Goede Praktijk Voor de Uitvoering van en Rapportering over Archeologisch Vooronderzoek en Archeologische Opgravingen en Het Gebruik van Metaaldetectoren. 4.0 ed. <https://www.onroerenderfgoed.be/de-code-van-goede-praktijk>.

Agentschap Onroerend Erfgoed. 2018. Evaluatie archeologie 2017. Evaluatie van het Onroerenderfgoeddecreet - hoofdstuk Archeologie voor het werkjaar 2017. Onderzoeksrapporten agentschap Onroerend Erfgoed 98 ISSN 1371-4678. <https://oar.onroerenderfgoed.be/item/742>.

Athanasopoulos, George & Hyndman Rob J. 2018. Forecasting: Principles and Practice. A Comprehensive Introduction to the Latest Forecasting Methods Using R. Learn to Improve Your Forecast Accuracy Using Dozens of Real Data Examples. OTexts. <https://otexts.com/fpp2>.

Cook, R. Dennis. 1977. "Detection of Influential Observation in Linear Regression." In: American Statistical Association & American Society for Quality, 19, No. 1, pp. 15-18. <http://www.jstor.org/stable/1268249>.

Dalgaard, Peter. 2008. Introductory Statistics with R. Spring Street, New York, Verenigde Staten: Springer Science.

Eisenhauer, Joseph G. 2003. "Regression Through the Origin." In: Teaching Statistics 25, nummer 3. Wiley Online Library.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. 2009. The elements of statistical learning. Data Mining, Inference, and Prediction, Second Edition. New York: Springer-Verlag.

Hoaglin, David, Mosteller, Frederick & Tukey, John. 1991. Fundamentals of Exploratory Analysis of Variance. New York, Verenigde Staten: Wiley-Interscience, John Wiley & Sons, Inc.

Huber, Peter J. 1964. "Robust Estimation of a Location Parameter." In: The Annals of Mathematical Statistics 35, Nummer 1. Institute of Mathematical Statistics. <https://www.jstor.org/stable/2238020>.

Ismay, Chester & Kim Albert Y. 2019. Statistical Inference via Data Science. A Modern Dive into R and the Tidyverse. CRC Press. <https://moderndive.com>.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2017. An Introduction to Statistical Learning with Applications in R. Baldwin Road, Shelter Island, New York, Verenigde Staten: Springer Science.

Kabacoff, Robert I. 2015. R in Action. Data Analysis and Graphics with R. Baldwin Road, Shelter Island, New York, Verenigde Staten: Manning Publications.

Wasserstein, Ronald L., Schirm Allen L. & Lazar Nicole A. 2019. Moving to a World Beyond " $p < 0.05$ ". In: "The American Statistician", 73:sup1, 1-19, <https://doi.org/10.1080/00031305.2019.1583913>.

Weisberg, Sanford & Fox, John. 2012. "Robust Regression in R: An Appendix to an R Companion to Applied Regression, Second Edition." Researchgate.



Wickham, Hadley, Cook, Dianne & Hofmann, Heike. 2015. "Visualizing Statistical Models: Removing the Blindfold." In: Statistical Analysis and Data Mining 8, issue 4, pp. 203-225. John Wiley & Sons, Inc. <http://had.co.nz/stat645/model-vis.pdf>.

Wickham, Hadley & Grolemund, Garret. 2017. R for Data Science. Import, Tidy, Transform, Visualize and Model Data. Verenigde Staten: O'Reilly Media, Inc. <https://r4ds.had.co.nz/>.

Wickham, Hadley & Stryjewski, Lisa. 2012. 40 years of boxplots. <http://vita.had.co.nz/papers/boxplots.html>.

Xie, Yihui. 2016. Bookdown: Authoring Books and Technical Documents with R Markdown. Chapman and Hall/ CRC. <https://bookdown.org/yihui/bookdown/>.

