

Data in onderzoek: Experimental design.

Wat is dataverzameling voor onderzoek?

Dataverzameling is het systematisch en gestructureerd verzamelen van gegevens, met als doel betrouwbare analyses uit te voeren die het mogelijk maken om de vooropgestelde onderzoeksvragen te beantwoorden.

Steekproef vs. proefopzet

Het belangrijkste verschil tussen een steekproef (sampling) en een experimenteel design (proefopzet) is dat een steekproef bedoeld is om personen, objecten, planten etc. te selecteren die de populatie representeren, terwijl een experimenteel design gericht is op het bewust veranderen of toepassen van een behandeling om oorzaak-gevolgrelaties te onderzoeken.

Experimental design: basisbegrippen

Experiment: Een experiment is een studie waarbij onderzoekers één of meerdere onafhankelijke variabelen systematisch variëren om het effect op afhankelijke variabelen te meten.

Doel: oorzaak-gevolg relaties vaststellen.

Factoren / Onafhankelijke variabelen: Dit zijn de variabelen die de onderzoeker verandert of test in het experiment.

Bijvoorbeeld: bemestingstype, irrigatieniveau, ras, andere behandeling.

Response / Afhankelijke variabelen: Dit zijn de resultaten die gemeten worden om het effect van de factoren te beoordelen.

Bijvoorbeeld: hoogte van planten, opbrengst in kg, aantal bloemen.

Randomisatie: Toewijzing van behandelingen aan experimenteenheden op een willekeurige manier. Verwijdert bias en zorgt dat verschillen tussen groepen aan de behandelingen te wijten (of te danken) zijn.

Blokken / Blocking: Groepeer vergelijkbare eenheden in blokken (homogene groepen) om variatie te verminderen.

Bijvoorbeeld: veld verdeeld in zand- en kleiblokken; elke behandeling random binnen blokken toegewezen.

Replicatie: Herhaal behandelingen meerdere keren om de precisie van het experiment te verhogen en betrouwbaarheid van resultaten te verbeteren.

Notaties, assumpties en R-checks

Belangrijke notaties

- Y_{ij} : Waarde afhankelijke variabele, eenheid j , behandeling i
- μ : Populatiegemiddelde
- τ_i : Effect van behandeling i
- ϵ_{ij} : Random fout, $\epsilon_{ij} \sim N(0, \sigma^2)$
- F : F-waarde, p : p-waarde, α : significantieniveau

ANOVA assumpties

- Onafhankelijkheid waarnemingen
- Normaliteit residuen
- Homogeniteit van varianties
- Randomisatie binnen blokken (bij RCBD, split-plot)

R-code quick checks

- Normaliteit:
`shapiro.test(resid(model))`
- Homogeniteit:
`leveneTest(Y ~ Treatment, data=df)`
- Residuen plots: `plot(model, which=1:2)`
- Interactieplot:
`interaction.plot(Factor1, Factor2, Y, data=df)`

Post-hoc tests

- Tukey HSD: `TukeyHSD(aov model)`
- LSD / Bonferroni:
`agricolae::LSD.test(aov model, "Treatment")`
- Effect size: `effectsize::eta squared(model)`

Tips / reminders

- Controleer randomisatie en replicaties
- Bij interacties: interpreteer main effect apart van interacties

Statistische begrippen

Interpretaties

- Grote F : behandeling heeft waarschijnlijk effect
- Kleine p ($< \alpha$) : statistisch significant
- Interacties altijd interpreteren voor de hoofdeffecten
- α = Type I fout: kans dat H_0 wordt verworpen terwijl deze waar is
- β = Type II fout: kans dat H_0 niet wordt verworpen terwijl deze onwaar is

Valkuilen

Veel gemaakte fouten

- Onvoldoende herhalingen: lage power
- Blokken te heterogeen: confounding
- Verkeerde of geen randomisatie: bias
- Negeren van interacties: verkeerde conclusies

Experimental designs: overzicht

Design	Beschrijving	Voorbeeld	R-code	Illustratie
Completely Randomised Design (CRD)	Elke experimenteenheid wordt volledig random toegewezen aan een behandeling. Eenvoudig design, geschikt voor homogene populaties.	5 behandelingen in 4 herhalingen, alles random toegewezen (vb. potproef).	<pre>model1=lm(Y ~ Treatment, data=df) print(Anova(model1, type="3")) model2=aov(Y ~ Treatment, data=df) print(summary(model2))</pre>	5 treatments, 4 replicates
Randomized Complete Block Design (RCBD)	Populatie wordt verdeeld in homogene blokken; randomisatie vindt plaats binnen elk blok. Vermindert variatie door blokken.	5 behandelingen in 4 blokken, waarbij elke behandeling 1 keer in elke blok voorkomt (vb. veldproef).	<pre>model1=lm(Y ~ Treatment + Block, data=df) print(Anova(model1, type="3")) model2=aov(Y ~ Treatment + Block, data=df) print(summary(model2))</pre>	5 treatments, 4 blocks Plot within Block
2-Factor Randomized Complete Block Design (2F-RCBD)	Een RCBD waarbij twee onafhankelijke factoren worden getest binnen hetzelfde blok. Dit design laat ook toe om interacties tussen de factoren te evalueren.	Proefveld met 3 rassen × 3 bemestingsniveaus binnen 4 blokken; elke combinatie random toegewezen binnen elk blok.	<pre>model1=lm(Y ~ Factor1 * Factor2 + Block, data=df) print(Anova(model1, type="3")) model2=aov(Y ~ Factor1 * Factor2 + Block, data=df) print(summary(model2))</pre>	2 Factors (A en B), 4 blocks Plot within Block
Split-Plot Design	Twee factoren: hoofdplotfactor wordt toegewezen aan grote eenheden, subplotfactor aan kleinere eenheden binnen hoofdplot.	vb. Hoofdplot: bemesting; subplots: ras (voordeel je kan de bemesting over grotere oppervlakte toepassen).	<pre>model1=lme(Y ~ Fertilizer*Variety, random = ~1 Block, data=df) print(Anova(model1, type="3"))</pre>	3 main plots, 3 subplots, 4 blocks Main plots (A) with nested subplots (B)
Latin Square	Controleert 2 bronnen van variatie door blokken in rijen en kolommen; randomisatie per rij en kolom.	4 rijen × 4 kolommen van proefvelden; 4 behandelingen.	<pre>model1=lm(Y ~ Treatment + Row + Column, data=df) model2=aov(Y ~ Treatment + Row + Column, data=df)</pre>	Latin Square Design (4 × 4)
Augmented Design	Een design waarbij nieuwe behandelingen (bv. veel rassen of behandelingen) worden toegevoegd aan een basisset van controles, zonder volledige replicatie van alle nieuwe behandelingen. Ideaal bij veel te testen rassen met beperkte middelen.	Proefveld met 26 maïsrassen, 2 standaard-controle rassen. Nieuwe rassen worden 1 keer gezaaid, controles meerdere keren.	<pre>library(augmentedRCBD) augmentedRCBD(block=Block, treatment=Treatment, y=Yield, checks=c("Ctrl1","Ctrl2"), method.comp="lsd", alpha=0.05, group=TRUE)</pre>	26 treatments (incl. 2 references) Plot within block