

**Practicum 4 – Full factorial design****Tips (om nog even goed naar te kijken als je tegen problemen aanloopt)**

- Stel dat de statistische toets mislukt, kijk dan eens goed of er genoeg informatie beschikbaar was voor deze toets. (Hint: Hoe groot is het aantal “error”-vrijheidsgraden?)
- Kijk ook even goed naar het aantal experimenten. Er zijn 16 mogelijke combinaties van instellingen, en die zijn allemaal 3 keer uitgevoerd. Dat zijn dus in totaal  $16 * 3 = 48$  experimenten. Heb je die ook echt als 48 experimenten opgevoerd?
- Pas op als je in Excel cellen kopieert en plakt waarvan de waarden met formules worden berekend. Wil je de getallen kopiëren (in plaats van de formules), dan kun je bij het plakken kiezen voor “waarden plakken”.
- Onderaan dit document staan nog wat R-tips.

XXX Chemicals is een bedrijf dat biologische tandpasta op de markt wil brengen. Er zijn echter twijfels over de houdbaarheid van het product. Daarom wordt een onderzoek gestart waarin verschillende conserveermiddelen worden getest. In dat onderzoek wordt de tandpasta gemaakt volgens verschillende recepten; de producten worden vervolgens twee jaar bewaard, en na die twee jaar wordt de kwaliteit onderzocht. Na die twee jaar wil men een keuze maken voor het product dat uiteindelijk op de markt wordt gebracht; er is dan geen tijd meer voor vervolgonderzoek. Het experiment wordt dan ook volledig **in drievoud** uitgevoerd.

Er zijn **vier verschillende conserveermiddelen** geselecteerd om uit te proberen, die we hier voor het gemak C1, C2, C3 en C4 zullen noemen. **Of er belangrijke interacties zijn tussen hun effecten, is niet bekend (maar wel belangrijk).** Elk conserveermiddel wordt ofwel wél ofwel níet toegevoegd; er wordt dus niet gevarieerd met de hoeveelheid.

- a) Waarom is, gezien de vraag, full factorial de meest geschikte proefopzet?
- b) Maak, volgens een full factorial design, een lijst van uit te voeren experimenten.
- c) Vul deze in, in het simulatie-Excel-sheet. De resultaten verschijnen weer “vanzelf” in de kolom achter de experimenten. De scores zeggen iets over de kwaliteit na twee jaar; een hogere score betekent een betere kwaliteit.
- d) Analyseer de resultaten. Welke hoofdeffecten en interacties zijn statistisch significant? (Zie tips hieronder.) Je hoeft alleen de 2-weg interacties, dus de interacties tussen twee factoren, mee te nemen. (Denk eraan dat je  $16*3=48$  experimenten hebt, niet 16.)
- e) Breng de statistisch significante interacties in beeld met (een) toepasselijke grafiek(en). Je kunt de functie `allEffects()` van het package `effects` gebruiken; zie weer de tips hieronder.
- f) Herhaal de analyse met regressie in Excel. Het is niet nodig de plaatjes nogmaals te tekenen. (Dat is in Excel overigens ook veel meer werk.) Je kunt je beperken tot de hoofdeffecten.
- g) Ga na of de resultaten ongeveer met elkaar overeenstemmen. Omdat je nu de interacties niet hebt meegenomen, kunnen er natuurlijk wel verschillen zijn. Maar als het goed is, lijkt het globale beeld wel op wat je in de R-plaatjes ziet.
- h) Welke combinatie geeft de beste houdbaarheid?
- i) Herhaal je statistische analyse in R nog eens, maar nu zonder de interacties. Vergelijk de uitkomst hiervan met je Excel-uitvoer (van vraag f). Als het goed is, krijg je nu (zonder interacties) uit R dezelfde resultaten als wat je uit Excel kreeg. De resultaten zullen er waarschijnlijk wel anders uitzien, maar de betekenis zou hetzelfde moeten zijn.

### Tips

Een variantieanalyse in R heb je al eerder gedaan, maar nieuw is hier het onderzoeken van interacties. Een interactie kun je in `lm()` en `Anova()` (in het package `car`) toevoegen met een `*` sterretje of `:` dubbele punt. (Dubbele punt = interactie; sterretje = hoofdeffecten én interacties.) Dus met bijvoorbeeld

```
lm( uitkomst ~ x + y + z, data = gegevens )
```

onderzoek je alleen hoofdeffecten, terwijl je met

```
lm( uitkomst ~ x + y + z + x:y, data = gegevens )
```

of

```
lm( uitkomst ~ x * y + z, data = gegevens )
```

ook de interactie tussen x en y meeneemt.

Voor het tekenen van meerdere factoren in een plaatje, om de interactie te zien, moet je deze functies wel vertellen dat je dat wilt. Dat kan bijvoorbeeld met

```
plot( allEffects( regressieresultaat ),  
      lines = list( multiline = TRUE ),  
      confint = list( style="auto" ) )
```

Hier vertel je de functie `plot()` dat je meerdere lijnen wilt (voor de verschillende instellingen) en dat je een betrouwbaarheidsinterval (`confint` betekent confidence interval) wilt, en dat je de automatische instellingen van de stijl van het betrouwbaarheidsinterval wilt.