

Opdracht 14a

Twee-factor ANOVA

In een groot research-project bestudeerde men de fysische eigenschappen van multiplex houtmaterialen, vervaardigd door kleine plakjes hout aan elkaar te hechten. Er waren verschillende boomsoorten gebruikt, en de plakjes hadden verschillende afmetingen. Tot de fysische eigenschappen die werden bestudeerd, behoorde de trek-elasticiteitsmodulus in de richting loodrecht op de richtlijn van de houtplakjes, gemeten in kilogram per vierkante centimeter. In de volgende tabel staan enkele gegevens. De boomsoorten zijn Esp, Berk en Esdoorn. De afmetingen van de houtplakjes zijn S1 (0.038 bij 5.08 cm) en S2 (0.064 cm bij 5.08 cm). (Gegevens verstrekt door Michael Hunt en Bob Lattanzi, Purdue University Forestry Department.)

 Afmeting plakjes

Soort S1 S2

Esp 21.7 19.5
 30.1 28.0
 20.0 23.3

Berk 15.0 37.5
 30.4 36.0
 16.2 22.5

Esdoorn 19.1 11.1
 26.4 35.4
 22.6 15.5

- a. ANOVA veronderstelt dat de populaties normaal verdeeld zijn met eventueel verschillende gemiddelden, maar met dezelfde standaarddeviatie. Het tekenen van een normaal-kwantiel-plot voor elk van de 6 groepen is vrij veel werk. Laat het tekenen van de plots daarom achterwege en neem gewoon aan dat de populaties normaal verdeeld zijn.
- b. Bereken de gemiddelden en standaarddeviaties voor de drie waarnemingen in elke groep (Soort/Afmeting). Wat is de verhouding van de grootste tot de kleinste standaarddeviatie? Mogen we twee-factor ANOVA toepassen?
- c. Geef de twee-factor ANOVA-tabel.
- d. Formuleer voor het hoofdeffect Afmeting H_0 en H_a . Is dit hoofdeffect significant?
- e. Formuleer voor het hoofdeffect Soort H_0 en H_a . Is dit hoofdeffect significant?
- f. Formuleer voor de interactie Afmeting*Soort H_0 en H_a . Is deze interactie significant?
- g. We doen nader onderzoek naar eventuele interactie tussen boomsoort en afmeting. Teken een grafiek waarin de gemiddelden zijn weergegeven. Twee opeenvolgende punten moeten door een rechte lijn met elkaar zijn verbonden. De x-as geeft de boomsoorten en de y-as de trek-elasticiteitsmodulus. In de grafiek moet je zowel voor S1 als voor S2 de gemiddelden weergeven. Wat laat de grafiek zien?

Opdracht 14a - S-PLUS

Voer de gegevens in in een tabel. Maak een tabel met drie kolommen: 'afmeting' (1=S1, 2=S2), 'soort' (1=esp, 2=berk, 3=esdoorn) en 'modulus'.

- a. Het tekenen van een normaal-kwantiel-plot voor elk van de 6 groepen laten we achterwege.
- b. Kies >Statistics >Data Summaries >Summary Statistics. Selecteer onder Data en achter Data Set de oorspronkelijke tabel, en achter Variables de variabele 'modulus'. Selecteer onder Summaries by Group en achter Group Variables de variabelen 'afmeting' en 'soort'. Klik daartoe eerst op de variabele 'afmeting', houd vervolgens de Ctrl-toets ingedrukt in klik op de variabele 'soort'. Klik hierna bovenaan in het window op >Statistics. Zorg dat alleen >Mean en >Std. Deviation aan staan. Klik op >OK.

Kies >Data >Change Data Type. Selecteer onder From en achter Data Set de oorspronkelijke tabel en achter Columns de variabelen 'afmeting' en 'soort'. Klik daartoe eerst op de variabele 'afmeting', houd vervolgens de Ctrl-toets ingedrukt en klik op de variabele 'soort'. Selecteer hierna onder Type en achter New Type het type factor. Klik op >OK.

Kies >Statistics >ANOVA >Fixed Effects. Selecteer onder Data en achter Data Set de oorspronkelijke tabel. Selecteer onder Variables en achter Dependent de variabele 'modulus'. Klik op >Create Formula. Selecteer onder Variable en achter Choose Variables de variabelen 'afmeting' en 'soort'. Klik daartoe eerst op de variabele 'afmeting', houd vervolgens de Ctrl-toets ingedrukt en klik op de variabele 'soort'. Klik hierna onder Add op >Main+Interact.:(*). Klik op >OK. Klik nogmaals op >OK.

- c. In het Report-Venster vormen de volgende gegevens de gevraagde ANOVA-tabel:

| | Df | Sum of Sq | Mean Sq | F Value | Pr(F) |
|----------------|----|-----------|----------|----------|-----------|
| AFMETING | 1 | 41.4050 | 41.40500 | 0.671487 | 0.4285091 |
| SOORT | 2 | 63.1944 | 31.59722 | 0.512429 | 0.6115747 |
| AFMETING:SOORT | 2 | 162.1900 | 81.09500 | 1.315161 | 0.3044839 |
| Residuals | 12 | 739.9400 | 61.66167 | | |

- d. In de ANOVA-tabel vinden we in de rij 'AFMETING' de gegevens voor het hoofdeffect Afmeting. Onder 'F Value' vinden we de F-waarde en onder 'Pr(F)' vinden we de P-waarde.
- e. In de ANOVA-tabel vinden we in de rij 'SOORT' de gegevens voor het hoofdeffect Soort. Onder 'F Value' vinden we de F-waarde en onder 'Pr(F)' vinden we de P-waarde.
- f. In de ANOVA-tabel vinden we in de rij 'AFMETING:SOORT' de gegevens voor de interactie Afmeting*Soort. Onder 'F Value' vinden we de F-waarde en onder 'Pr(F)' vinden we de P-waarde.
- g. Voor het maken van deze grafiek moeten we een tabel maken met de gemiddelde modulus per afmeting per soort. Deze gemiddelden heb je bepaald bij b. en kun je vinden in het Report-Venster. Kies >File >New en selecteer >Data Set. We kunnen nu de gegevens kolom voor kolom invoeren. Het is van belang om per kolom steeds eerst de gegevens in te voeren, en pas daarna via >Format >Selected Object(s) de kolomnaam te geven door achter Name de door S-PLUS voorgestelde naam te vervangen door de gewenste naam.

De tabel moet er als volgt uit komen te zien:

| soort | modulus.afmeting1 | modulus.afmeting2 |
|---------|-------------------|-------------------|
| esp | 23.93333 | 23.60000 |
| berk | 20.53333 | 32.00000 |
| esdoorn | 22.70000 | 20.66667 |

Controleer na het intypen van de tabel de kolom 'soort'. Selecteer de kolom 'klasse' door op de kolomnaam te klikken. Kies >Format >Selected Objects. Verwijder achter Factor Levels de klassen die je niet wilt hebben. Klik op >OK.

Kies >View >Toolbars. Zet onder Toolbars de optie Plots 2D aan. Klik op >OK. Kies in de Plots 2D-toolbar voor Line Scatter. Kies onder Data Columns en achter Data Set de tabel met de gemiddelden. Kies achter x Columns de variabele 'soort' en achter y Columns de variabele 'modulus.afmeting1'. Klik op >OK.

Klik in het Grafiek-venster buiten de grafiek. Het de Shift-toets ingedrukt en klik in de Plots 2D-toolbar opnieuw voor Line Scatter. Kies onder Data Columns en achter Data Set de tabel met de gemiddelden. Kies achter x Columns de variabele 'soort' en achter y Columns de variabele 'modulus.afmeting2'. Klik op >OK.

De onderste toolbar is de Graph Toolbar. Klik in deze toolbar op het icoontje dat precies onder Linear in de Standard Toolbar zit. Als je de cursor op dit icoontje plaatst, moet als bijschrift verschijnen: Auto Legend. Klik op dit icoontje. De legenda wordt nu in de grafiek geplaatst.

Opdracht 14a - SPSS

Voer de gegevens in in een tabel. Maak een tabel met drie kolommen: 'afmeting' (1=S1, 2=S2), 'soort' (1=esp, 2=berk, 3=esdoorn) en 'modulus'.

Kies >Statistics >General Linear Model >GLM - General Factorial. Plaats 'Modulus' onder Dependent Variable. Plaats 'Afmeting' en 'Soort' onder Fixed Factor(s).

Klik op >Plots. Plaats 'Soort' onder Horizontal Axis en plaats 'Afmeting' onder Separate Lines. Klik vervolgens op >Add. Klik op >Continue.

Klik op >Options. Zet onder Display 'Descriptive statistics' aan. Klik op >Continue. Klik op >OK. In het output-window vinden we nu drie tabellen en een grafiek.

- Het tekenen van een normaal-kwantiel-plot voor elk van de 6 groepen laten we achterwege.
- De tweede tabel in het output-window heet 'Descriptive Statistics'. Deze tabel bevat voor elk van de zes groepen het gemiddelde en de standaarddeviatie. We zien dat de kleinste standaarddeviatie gelijk is aan 3.651 en dat de grootste standaarddeviatie gelijk is aan 12.948.
- De derde tabel in het output-window heet 'Tests of Between-Subjects Effects' en is de ANOVA-tabel.
- In de ANOVA-tabel vinden we in de rij 'AFMETING' de gegevens voor het

hoofdeffect Afmeting. Onder 'F' vinden we de F-waarde en onder 'Sig.' vinden we de P-waarde.

- e. In de ANOVA-tabel vinden we in de rij 'SOORT' de gegevens voor het hoofdeffect Soort. Onder 'F' vinden we de F-waarde en onder 'Sig.' vinden we de P-waarde.
- f. In de ANOVA-tabel vinden we in de rij 'AFMETING*SOORT' de gegevens voor de interactie Afmeting*Soort. Onder 'F' vinden we de F-waarde en onder 'Sig.' vinden we de P-waarde.
- g. De grafiek in het output-window heet 'Estimated Marginal Means of MODULUS'.

Opdracht 14a - verslag

In een groot research-project bestudeerde men de fysische eigenschappen van multiplex houtmaterialen, vervaardigd door kleine plakjes hout aan elkaar te hechten. Er waren verschillende boomsoorten gebruikt, en de plakjes hadden verschillende afmetingen. Tot de fysische eigenschappen die werden bestudeerd, behoorde de trek-elasticiteitsmodulus in de richting loodrecht op de richtlijn van de houtplakjes, gemeten in kilogram per vierkante centimeter. In de volgende tabel staan enkele gegevens. De boomsoorten zijn Esp, Berk en Esdoorn. De afmetingen van de houtplakjes zijn S1 (0.038 bij 5.08 cm) en S2 (0.064 cm bij 5.08 cm).

- a. ANOVA veronderstelt dat de populaties normaal verdeeld zijn met eventueel verschillende gemiddelden, maar met dezelfde standaarddeviatie. Het tekenen van een normaal-kwantiel-plot voor elk van de 6 groepen is vrij veel werk. Laat het tekenen van de plots daarom achterwege en neem gewoon aan dat de populaties normaal verdeeld zijn.
- b. Bereken de gemiddelden en standaarddeviaties voor de drie waarnemingen in elke groep (Soort/Afmeting). Wat is de verhouding van de grootste tot de kleinste standaarddeviatie? Mogen we twee-factor ANOVA toepassen?

| | AFMETING | | | |
|---------|----------|---------------|--------|---------------|
| | S1 | | S2 | |
| | Mean | Std Deviation | Mean | Std Deviation |
| SOORT | | | | |
| Esp | 23.933 | 5.408 | 23.600 | 4.258 |
| Berk | 20.533 | 8.566 | 32.000 | 8.261 |
| Esdoorn | 22.700 | 3.651 | 20.667 | 12.948 |

De verhouding van de grootste standaarddeviatie tot de kleinste standaarddeviatie is $12.948/3.651 = 3.546$, wat groter is dan 2. We mogen er niet van uitgaan dat de drie populaties gelijk standaarddeviaties hebben. Toepassing van de ANOVA-toets zal geen betrouwbare resultaten geven.

- c. Geef de twee-factor ANOVA-tabel.

| Source | Sum of Squares | df | Mean Square | F | Sig |
|-----------------|----------------|----|-------------|-------|------|
| AFMETING | 41.405 | 1 | 41.405 | .671 | .429 |
| SOORT | 63.194 | 2 | 31.597 | .512 | .612 |
| AFMETING*SOORT | 162.190 | 2 | 81.095 | 1.315 | .304 |
| Error | 739.940 | 12 | 61.662 | | |
| Corrected Total | 1006.729 | 17 | | | |

- d. Formuleer voor het hoofdeffect Afmeting H_0 en H_a . Is dit hoofdeffect significant?

Hoofdeffect AFMETING

H_0 : $\mu_{S1} = \mu_{S2}$

H_a : $\text{not}(\mu_{S1} = \mu_{S2})$

SPSS geeft als tweezijdige P-waarde 0.429 en als F-waarde 0.671. We willen ook tweezijdig toetsen omdat we willen onderzoeken of twee of meer groepen significant verschillend zijn ten opzichte van elkaar. De kans, berekend onder de aanname dat H_0 waar is, dat F een waarde zou aannemen die even extreem is als of nog extremer is dan 0.671 is gelijk aan 0.429. Omdat 0.429 groter is dan 0.05, wordt H_0 aangenomen. Dit hoofdeffect is niet significant.

- e. Formuleer voor het hoofdeffect Soort H_0 en H_a . Is dit hoofdeffect significant?

Hoofdeffect SOORT

H_0 : $\mu_{\text{esp}} = \mu_{\text{berk}} = \mu_{\text{esdoorn}}$

H_a : $\text{not}(\mu_{\text{esp}} = \mu_{\text{berk}} = \mu_{\text{esdoorn}})$

SPSS geeft als tweezijdige P-waarde 0.612 en als F-waarde 0.512. We willen ook tweezijdig toetsen omdat we willen onderzoeken of twee of meer groepen significant verschillend zijn ten opzichte van elkaar. De kans, berekend onder de aanname dat H_0 waar is, dat F een waarde zou aannemen die even extreem is als of nog extremer is dan 0.512 is gelijk aan 0.612. Omdat 0.612 groter is dan 0.05, wordt H_0 aangenomen. Dit hoofdeffect is niet significant.

- f. Formuleer voor de interactie Afmeting*Soort H_0 en H_a . Is deze interactie significant?

Interactie AFMETING*SOORT

H_0 : $\mu_{S1_esp} = \mu_{S1_berk} = \mu_{S1_esdoorn} =$

$\mu_{S2_esp} = \mu_{S2_berk} = \mu_{S2_esdoorn}$

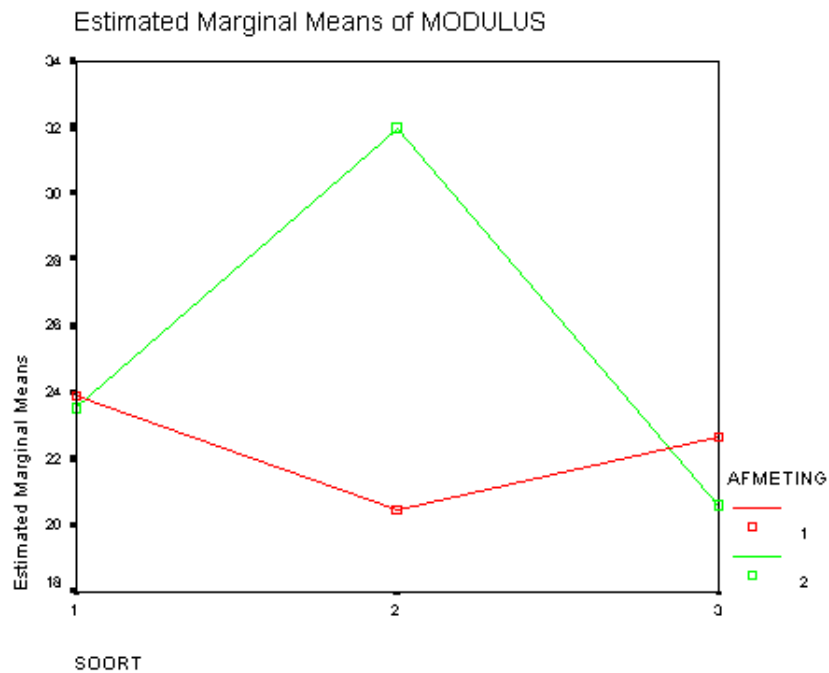
H_a : $\text{not}(\mu_{S1_esp} = \mu_{S1_berk} = \mu_{S1_esdoorn} =$

$\mu_{S2_esp} = \mu_{S2_berk} = \mu_{S2_esdoorn})$

SPSS geeft als tweezijdige P-waarde 0.304 en als F-waarde 1.315. We willen ook tweezijdig toetsen omdat we willen onderzoeken of twee of meer groepen significant verschillend zijn ten opzichte van elkaar. De kans, berekend onder de aanname dat H_0 waar is, dat F een waarde zou aannemen die even extreem is als of nog extremer is dan 1.315 is gelijk aan 0.304. Omdat 0.304 groter is dan 0.05, wordt H_0 aangenomen. Deze interactie is niet significant.

- g. We doen nader onderzoek naar eventuele interactie tussen boomsoort en afmeting. Teken een grafiek waarin de gemiddelden zijn weergegeven. Twee opeenvolgende punten moeten door een rechte lijn met elkaar zijn verbonden. De x-as geeft de boomsoorten en de y-as de trek-elasticiteitsmodu-

lus. In de grafiek moet je zowel voor S1 als voor S2 de gemiddelden weergeven. Wat laat de grafiek zien?



Tussen lijnen die geheel parallel lopen is geen interactie. Tussen lijnen die geheel of gedeeltelijk convergeren of divergeren, of die elkaar snijden is wel interactie. We zien dat het verschil in gemiddelde trek-elasticiteitsmodulus tussen de afmetingen (de gekleurde lijnen) niet voor elke boomsoort (de punten op de x-as) hetzelfde is. Dit duidt op interactie tussen afmeting en boomsoort. Met ANOVA bepaalden we reeds of deze interactie wel of niet significant is.