Statistiek hoofdstuk 8 / werkcollege 7

Vraag 1

1. Het is handig om dan even de tabel uit te breiden met interactieeffecten:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Experiment** | **S** | **R** | **T** | **Gem.** | **SR** | **ST** | **RT** | **SRT** |
| 1 | - | - | - | 0,084 | + | + | + | - |
| 2 | + | - | - | 0,099 | - | - | + | + |
| 3 | - | + | - | 0,082 | - | + | - | + |
| 4 | + | + | - | 0,049 | + | - | - | - |
| 5 | - | - | + | 0,097 | + | - | - | + |
| 6 | + | - | + | 0,076 | - | + | - | - |
| 7 | - | + | + | 0,080 | - | - | + | - |
| 8 | + | + | + | 0,070 | + | + | + | + |

Vervolgens bereken ik het gemiddelde van alle waarden behorende bij de minnen van een bepaalde variabele en het gemiddelde van de waarden behorende bij de plussen van een bepaalde variabele, en dat voor elke variabele. Zie tabelletje hieronder.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **S** | **R** | **T** |
| **+** | 0,074 | 0,070 | 0,081 |
| **-** | 0,086 | 0,089 | 0,079 |
| ***Hoofdeffecten:*** | *-0,012* | *-0,019* | *0,002* |

Ditzelfde kan ik doen voor de interactie-effecten.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Interactie** | **SR** | **ST** | **RT** | **SRT** |
| **+** | 0,075 | 0,078 | 0,083 | 0,087 |
| **-** | 0,084 | 0,081 | 0,076 | 0,072 |
| ***Effect:*** | *-0,009* | *-0,003* | *0,007* | *0,015* |

De schuin gedrukte waardes zijn dus de gevraagde effecten.

1. Als je lineaire regressie zou gebruiken zou je uitkomen op de helft van de bij A berekende waarden.

Dit hoef je niet te berekenen maar heb ik hieronder toch gedaan:

Daartoe moeten we eerst de gegevens kwantitatief omschrijven / coderen:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Experiment** | **S** | **R** | **T** | **Gem.** | **SR** | **ST** | **RT** | **SRT** |
| 1 | -1 | 1 | 1 | 0,084 | -1 | -1 | 1 | -1 |
| 2 | 1 | 1 | 1 | 0,099 | 1 | 1 | 1 | 1 |
| 3 | -1 | 2 | 1 | 0,082 | -2 | -1 | 2 | -2 |
| 4 | 1 | 2 | 1 | 0,049 | 2 | 1 | 2 | 2 |
| 5 | -1 | 1 | 18 | 0,097 | -1 | -18 | 18 | -18 |
| 6 | 1 | 1 | 18 | 0,076 | 1 | 18 | 18 | 18 |
| 7 | -1 | 2 | 18 | 0,080 | -2 | -18 | 36 | -36 |
| 8 | 1 | 2 | 18 | 0,070 | 2 | 18 | 36 | 36 |

Ik heb ervoor gekozen om per eenheid te werken, dus daarom is alleen S gecodeerd met behulp van plussen en minnen. Bij de interactie heb ik daarom ook gewoon de factoren vermenigvuldigd. Dat lijkt me het meest logisch omdat je dan nog steeds per eenheid werkt.

Vervolgens kun je dit als vergelijking in Matlab stoppen en uitrekenen. De vergelijking die je moet oplossen is:

S, R, T en A t/m E zijn hierbij de onbekenden.

Matlab scriptje:

|  |
| --- |
| % Regressie bij statistiek hoofdstuk 8  clc;    % Vergelijkingsmatrices:  % y = a + Sx1 + Rx2 + Tx3 + Ax1x2 + Bx1x3 + Cx2x3 + Dx1x2x3;    x1 = [ -1 1 -1 1 -1 1 -1 1 ]';  x2 = [ 1 1 2 2 1 1 2 2 ]';  x3 = [ 1 1 1 1 18 18 18 18 ]';    y = [ 0.084 0.099 0.082 0.049 0.097 0.076 0.080 0.070 ]';    fbase = [x1, x2, x3, x1.\*x2, x1.\*x3, x2.\*x3, x1.\*x2.\*x3, ones(8,1)];  outcome = fbase\y |

Dan krijg je:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S | R | T | A (SR) | B (ST) | C (RT) | D (SRT) | E |
| 0.0343 | -0.0269 | -0.0011 | -0.0257 | -0.0028 | 0.0009 | 0.0017 | 0.1186 |

Als je de resultaten overigens gecodeerd berekend krijg je het volgende (maal twee om goed te kunnen vergelijken, in voorbeeld 8.6 net onder de tabel staat beschreven waarom):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S | R | T | A (SR) | B (ST) | C (RT) | D (SRT) | E |
| -0.0123 | -0.0187 | 0.0023 | -0.0093 | -0.0033 | 0.0073 | 0.0147 | 0.1593 |

Dit komt dus inderdaad exact overeen met eerder berekende waarden.

|  |  |  |  |
| --- | --- | --- | --- |
| Meting 1 | Meting 2 | di^2/2 |  |
| 0,080 | 0,088 | 3,2E-05 |  |
| 0,098 | 0,100 | 0,000002 |  |
| 0,084 | 0,080 | 8E-06 |  |
| 0,050 | 0,048 | 0,000002 |  |
| 0,095 | 0,099 | 8E-06 |  |
| 0,080 | 0,072 | 3,2E-05 |  |
| 0,085 | 0,075 | 5E-05 |  |
| 0,075 | 0,064 | 0,0000605 |  |
|  |  |  | + |
|  |  | 0,0001945 |  |
|  |  |  |  |
|  | /8 | 2,43125E-05 |  |

*In het geval van het berekenen van de variantie met gegevens in duplo geldt dat de afstand tot het gemiddelde gelijk is aan de helft van het verschil tussen de twee variabelen. Daarom kun je de standaarddeviatie zo makkelijk berekenen (zie vb. 8.4). De gepoolde variantie kun je simpelweg berekenen door de som hiervan te delen door N (totaal aantal metingen), want bij elke meting geldt dat er nog 1 vrijheidsgraad over is, dus je moet alle varianties met 1 vermenigvuldigen, sommeren, en door acht maal (n-1) delen (n is aantal herhalingen, ≠ N), ofwel delen door 8.*

Dus seffect is 0,00246538537 (denk aan kwadraat).

1. De standaardfout was 0,0025. Dus alle effecten die meer dan 2-3 groter zijn als de standaardfout (> 0,0050-0,0075) worden nader onderzocht.

Als we dan naar de tabel kijken die volgt uit de gecodeerde waarden (want je was geïnteresseerd in deze range), dan blijken alle resultaten significant te zijn:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S | R | T | A (SR) | B (ST) | C (RT) | D (SRT) | E |
| Waarde | -0.0123 | -0.0187 | 0.0023 | -0.0093 | -0.0033 | 0.0073 | 0.0147 | 0.1593 |
| Nader onderzoek | Ja | Ja | Nee\* | Ja | Nee | Ja | Ja | n.v.t. |

\*Aangezien T wel in interacties betrokken is moet deze toch verder onderzocht worden. Alleen de interactie ST is niet interessant.

Vraag 2

1. In dat geval kun je dus de standaardfout alleen bepalen uit de vier duplo metingen. Hierbij neem je dus aan dat de variantie onafhankelijk is van de meting, zodat je toch de variantie voor alle metingen kunt bepalen.

De standaardafwijking is dan als volgt:

|  |  |  |  |
| --- | --- | --- | --- |
| **Meting 1** | **Meting 2** | **di^2/2** |  |
| 0,080 | 0,088 | 3,2E-05 |  |
| 0,098 |  |  |  |
| 0,084 |  |  |  |
| 0,050 | 0,048 | 0,000002 |  |
| 0,095 |  |  |  |
| 0,080 |  |  |  |
| 0,085 |  |  |  |
| 0,075 |  |  |  |
|  |  |  | + |
|  |  | 3,4E-05 |  |
|  |  |  |  |
|  | /2 | 0,0000170 |  |

Let op: afhankelijk van of bij het betreffende interactie-effect een plus dan wel een min staat, moeten de tweede metingen worden geteld voor N- of N+. Daarom kun je dus niet de term 4/N gebruiken. Onderstaande berekening geldt dan eigenlijk ook alleen voor de interacties waarbij 0,088 minus is en 0,048 plus. Varianties voor effecten waar dit niet geldt zouden apart berekend moeten worden (dit heb ik echter nagelaten omdat ik lui ben).

Ook de waarden moeten opnieuw berekend worden (voor uitgebreidere tabellen, zie vorige opgave):

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Gem. /waarde | 0,084 | 0,098 | 0,084 | 0,049 | 0,095 | 0,08 | 0,085 | 0,075 |
|  |  |  |  |  |  |  |  |  |
|  | **S** | **R** | **T** |  |
| **+** | 0,076 | 0,073 | 0,084 |  |
| **-** | 0,087 | 0,089 | 0,079 |  |
| **Hoofdeffecten:** | -0,012 | -0,016 | 0,005 |  |
|  |  |  |  |  |
| **Interactie** | **SR** | **ST** | **RT** | **SRT** |
| **+** | 0,076 | 0,081 | 0,086 | 0,088 |
| **-** | 0,087 | 0,082 | 0,077 | 0,075 |
| **Effect:** | -0,011 | -0,001 | 0,009 | 0,014 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **S** | **R** | **T** | **A (SR)** | **B (ST)** | **C (RT)** | **D (SRT)** |
| **Waarde** | -0,012 | -0,016 | 0,005 | -0,011 | -0,001 | 0,009 | 0,014 |
| **Nader onderzoek** | Ja | Ja | Nee\* | Ja | Nee | Ja | Ja |

*Variabelen met waarden onder de 0.0052-0.0078 hoeven niet verder te worden onderzocht, behalve variabelen betrokken bij een interactie. T is in dit geval op hoofdefect dus een twijfelgeval, maar moet sowieso verder worden onderzocht vanwege het interactie-effect. In feite blijft de tabel dus hetzelfde.*

Vraag 3

*Noot van de nakijker: ‘kan eenvoudiger, maar niet zo relevant.’*

Vraag 4

(Tentamen 4a.) Met andere woorden: maak zo’n proefopzet als hierboven voor een gen, met als variabelen tijd en behandeling. Dus:

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment** | **Tijd (T)** | **Behandeling (B)** | **TB** |
| **1 (duplo)** | + | + | + |
| **2** | + | - | - |
| **3** | - | + | - |
| **4 (duplo)** | - | - | + |

|  |  |  |
| --- | --- | --- |
| **Factor** | **Tijd (T)** | **Behandeling (B)** |
| **Laag extreem** | 1 | Geen behandeling |
| **Hoog extreem** | 10 | Wel behandeling |

Door experiment 1 en 4 in duplo uit te voeren heb je alle extremen ‘geduplood’ (T+, T-, B+, B-). Verder neem ik de buitenste waarden van de tijd als extremen, en bij behandeling uiteraard geen behandeling en wel behandeling. De significantie wordt net als in de vorige opgave bekeken met behulp van de standaardfout (>2-3 maal seffect wordt verder onderzocht), die kan worden berekend analoog aan de vorige opgave.