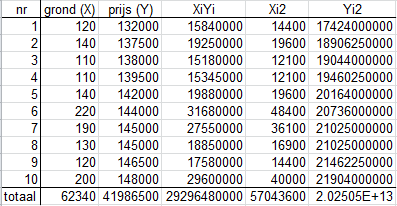
|  |
| --- |
| Hanzehogeschool |
| Statistiek |
| Week 4 |

|  |
| --- |
| Door: Martine Dambrink(326694) en Caroline Lu(316178)  5/28/2015 |

# OPDRACHT 4 (regressie en correlatie)

## Opdracht 4.1

Bestand WONINGEN. Net als in de syllabus zoeken we naar een (lineair) verband tussen de verklarende variabele “grond” (X) en de te verklaren variabele “prijs” (Y). Maak in Excel een opzet vergelijkbaar met tabel 2.11 op blz. 90 van de syllabus, maar nu natuurlijk wel voor de gehele dataset van 120 woningen. Laat in je verslag daarvan echter alleen de titelrij, de eerste 10 dataregels en de totaalrij ter illustratie zien.



## Opdracht 4.2

Gebruik voor deze opdracht de resultaten uit die totaalrij bij 4.1. Bereken de parameters b (richtingscoëfficiënt) en a (snijpunt) van de beste rechte lijn volgens de “methode van de kleinste kwadraten”, m.b.v. de formules [2.10a, b] op blz. 89 van de syllabus. Bereken vervolgens de correlatiecoëfficiënt r van X en Y , m.b.v. de formule [2.11] op blz. 90 van de syllabus. 4.2

## Opdracht 4.3

Zoek in Excel naar formules voor a, b en r, vermeld ze in je verslag. Pas ze toe in het bestand WONINGEN en vergelijk hiermee je resultaten in 4.2.

|  |  |  |
| --- | --- | --- |
|  | **Excel-formule** | **Antwoord** |
| R (correlatie) | =CORREL(E2:E121,I2:I121) | 0,639216815854823 |
| A (snijpunt) | =INTERCEPT(I2:I121,E2:E121) | 192202,408997983 |
| B (richtingscoeficient) | =SLOPE(I2:I121,E2:E121) | 303,532417713218 |

Tabel 1 Correlatie, snijpunt en richtingscoeficient berekend met excel-formules



Tabel 2 Correlatie, snijpunt en richtingscoeficient volgens excel-formules en eigen berekeningen

In tabel2 kun je zien dat de zelf berekende antwoorden voor de correlatie, de richtingscoeficient en het snijpunt hetzelfde zijn als die met excel-formules zijn berekend.

## Opdracht 4.4

Bestand WONINGEN. Doe hetzelfde nogmaals m.b.v. de optie “regressie” onder de tab gegevensanalyse. Zorg dat je de “labels” meeneemt en het bijbehorende spreidingsdiagram (vergelijk afbeelding 2.7, blz. 89 syllabus) afdrukt. Zoek in de syllabus hoe (de waarden van) b en r hier geïnterpreteerd kunnen worden en zet dat in eigen woorden in je verslag.

## Opdracht 4.5

Bestand CURSUS. Dat bevat van 50 studenten het cijfer dat zij voor het vak gehaald hebben en het aantal uren dat zij aan dat vak besteed hebben. Ga nu, op de manier van 4.4, zoeken naar een (lineair) verband tussen het aantal “uren” en het behaalde “cijfer”. Maak een verslagje van je onderzoekingen en geef zinvolle interpretaties voor de gevonden waarden van de regressiecoëfficiënt en de correlatiecoëfficiënt in de context van de probleemstelling.

## Opdracht 4.6

Ga zelf op zoek naar een zinvol (lineair) verband in de gegeven datasets, voer een regressieanalyse uit en rapporteer daarover als hierboven in 4.5.