## Practicumopdrachten Beschrijvende Statistiek BuijsH1H2 2013-2014 kw 4

Laatste wijziging 6-5-2013

De benodigde datasets in Excel vind je op BlackBoard: P 1.4 Bi+I+Ti:

ITSM 2013-14 - kw 4 > Ondersteunend materiaal > Beschrijvende Statistiek

ARCHITECT.xls (46 KB) BANKIERS.xls (17 KB) CURSUS.xls (16 KB) MULTIFOOD.xls (14.5 KB) TOP100.xls (24.5 KB) WINKELS.xls (24.5 KB) WONINGEN.xls (22 KB) KOERSEN.xls (27 KB) template-boxplots.xls (31 KB).

Bestudeer voor het samenstellen van grafieken en tabellen nauwkeurig de voorschriften in de paragrafen 1.2.4. (blz. 33) (tabel) en 1.3.1 (blz. 35) (grafiek) van de syllabus (Buijs).

Maak de opdrachten niet als een sommetje waar iets uit moet komen, maar als een rapportage voor een opdrachtgever. Dat betekent dat het op te leveren document verzorgd, begrijpelijk, zelfstandig leesbaar, volledig en uiteraard juist moet zijn en moet voldoen aan de richtlijnen uit de syllabus (Buijs H1 en H2).

Nadat je alles goed hebt uitgewerkt, plak je telkens de Excel-resultaten van een opdracht in een (1!) word-document en je voorziet alle rapportages van voldoende toelichting en van hun opdrachtnummer. Je mag de opdrachten in vaste 2-tallen uitwerken. Zet bovenaan het document de datum, je namen en studentnummers, groepsnummer en opdrachtnummer.

Er zijn 5 practica en 4 weekopdrachten. Lever per 2-tal telkens een hardcopy(!!) van de weekopdracht in, uiterlijk (!) bij het begin van het daarop volgende practicum.

OPDRACHT 1 (grafieken en tabellen)

* 1. Zie de frequentietabel tabel 1.7 uit de syllabus(blz. 31). Maak dezelfde frequentietabel (absoluut en relatief), maar nu met vaste klasse-intervallen van 150.000. (bestand WONINGEN). Je kunt dit doen met de optie “histogram” onder de tab “gegevensanalyse”. Die tab moet echter eerst geïnstalleerd worden in Excel. Zie voor instructies blz. 53, inclusief voetnoot.
  2. Maak ook het bijbehorende histogram. (let op de waarschuwing op blz. 54 van de syllabus over de ruimte tussen de staven van een histogram). Geef een goede titel.
  3. Maak ook een cumulatieve frequentietabel, zowel absoluut als relatief.
  4. Bekijk de Excel-datasets op Bb. Geef voor elk van de 4 schalen uit afb. 1.1 (blz. 28) 2 voorbeelden van variabelen. Noem telkens schaal - dataset – variabele.
  5. Bekijk de dataset KOERSEN. Geef de variabelen in dat bestand weer in 1 lijndiagram. (Zie voor een voorbeeld van een lijndiagram afb. 1.2 op blz. 36).
  6. Bekijk de datasets. Zoek uit de datasets 2 variabelen die zinvol in een staafdiagram zijn weer te geven. Maak die twee staafdiagrammen.
  7. Bekijk de datasets. Zoek uit 1 dataset 2 variabelen die zinvol (oorzaak-gevolg-relatie) in een spreidingsdiagram zijn weer te geven. Maak dat spreidingsdiagram. (Zie voor een voorbeeld van een spreidingsdiagram afb. 1.5 op blz. 39).

OPDRACHT 2 (maatstaven voor ligging)

2.1 Bepaal in het bestand TOP100 “op eigen wijze” (d.w.z. zonder van de statistische formules bij 2.2 gebruik te maken) de gemiddelde omzet89, de modale branchcode, de mediane winst90, maximale omzet90 en minimaal aantal werkn90. Licht toe hoe je dat gedaan hebt.

2.2. Bepaal dezelfde resultaten als in 2.1, maar nu m.b.v. de statistische functies in Excel voor gemiddelde, mediaan, modus, max en min. Raadpleeg eventueel de helpfunctie om die statistische functies te vinden en om te ontdekken hoe ze werken. Bespreek en verklaar eventuele verschillen met de resultaten onder 2.1

2.3 Bepaal dezelfde gegevens m.b.v. de optie “beschrijvende statistiek” onder ”gegevensanalyse”. (zie voor “gegevensanalyse” opdracht 1.1) Bespreek en verklaar eventuele verschillen met de resultaten onder 2.2

2.4 Zie het bestand CURSUS. Maak een frequentieverdeling van de variabele “cijfer” (kolom B) op basis van klassen met een breedte van 10 . Benader (!) uit die frequentieverdeling het gemiddelde (zie blz. 71 syllabus) en de modus (zie blz. 74 syllabus). Bereken (!)vervolgens m.b.v. de Excel-formules (zie 2.2) het echte gemiddelde en de echte modus. Vergelijk de benadering en de exacte berekening en verklaar de eventuele verschillen.

2.5 Het gewogen rekenkundig gemiddelde wordt besproken op blz. 69 van de syllabus Buijs. Zie het bestand CURSUS. Stel dat de variabele “cijfer” de deelresultaten van 50 door jouw behaalde studieonderdelen bevat. De variabele “uren” bevat de voor die onderdelen vastgestelde studiebelasting. Je eindcijfer wordt bepaald door elk deelresultaat te wegen met de bijbehorende studiebelasting. Bereken dan je eindcijfer uit die 50 deelresultaten.

2.6 De kwartielen worden besproken op blz. 74 van de syllabus Buijs. Zie het bestand KOERSEN. Bereken het 1e, 2e (mediaan) en 3e kwartiel van de 3 variabelen uit dat bestand. Zoek de functies in Excel om dit te berekenen.

2.7 De BoxPlot wordt besproken op blz. 78/79/80 van de syllabus Buijs. Zoek in de help-functie van Excel op “Een BoxPlot/en Whisker grafiek maken in Excel” of maak gebruik van de “template-boxplots” bij de datasets. Verzamel de benodigde gegevens voor “simpele” boxplots (zie afb. 2.5 op blz. 79 syllabus) van de variabelen “dollar”, “olie”. Daarnaast bestaat er ook nog een “gewone” boxplot zie afb. 2.6 op blz. 80). Verzamel de benodigde gegevens van de variabele “dowjones”, door de uitbijters te bepalen. Maak deze 3 boxplots (2 “simpele”en 1 “gewone”) en voorzie ze van een toelichting.

OPDRACHT 3 (maatstaven voor spreiding)

3.1 Bestudeer in de syllabus Buijs de spreidingsbreedte. Bestand WINKELS. Bepaal de spreidingsbreedte van de “winst92” en leg uit hoe je het resultaat gevonden hebt.

3.2 Bestudeer in de syllabus Buijs de gemiddelde absolute afwijking. Bestand WINKELS. Neem de eerste 10 omzetten in 1991. Bepaal daarvan het gemiddelde. Plaats naast die eerste 10 omzetten de absolute afwijking van hun gemiddelde. Bepaal vervolgens de GAA van de eerste 10 omzetten. Zoek daarna in Excel de formule voor de GAA en vergelijk daarmee je antwoord.

3.3. Bepaal nu in de dataset BANKIERS de GAA van alle leeftijden.

3.4 Bestudeer in de syllabus Buijs de interkwartiel range (IQR). Bestand BANKIERS. Bepaal de IQR van de salarissen.

3.5 Bestudeer in de syllabus Buijs hoe in een groep gegevens uitbijters kunnen worden gedefinieerd. Welke salarissen uit het bestand BANKIERS zijn als uitbijters te beschouwen?

3.6 Bestudeer in de syllabus Buijs de variantie. Bestand BANKIERS. Neem de eerste 10 leeftijden. Bepaal daarvan het gemiddelde. Plaats naast die eerste 10 leeftijden de afwijking van hun gemiddelde. Plaats daarnaast in de volgende kolom de kwadraten van die die afwijkingen. Bereken het gemiddelde van die kwadraten. Bereken vervolgens de wortel uit dat gemiddelde. Zoek daarna in Excel de formules voor de variantie en de standaarddeviatie en vergelijk daarmee je antwoorden.

3.7 Bepaal nu in de dataset BANKIERS de variantie en de standaarddeviatie van alle leeftijden. Voor de te gebruiken formules kan het uitmaken of je de dataset als een steekproef of als een populatie beschouwt (zie blz. 82/83 syllabus). Zoek uit welke verschillende formules daarvoor in Excel gebruikt worden. Pas ze allebei toe en licht de verschillen toe.

3.8 Bestudeer in de syllabus Buijs de variatiecoëfficiënt . Bestand TOP100. Bereken en vergelijk de twee variatiecoëfficiënten van de omzetten in 1989 en 1990. Becommentarieer de verschillen.

OPDRACHT 4 (regressie en correlatie)

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.

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

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.

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.

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.

4.5 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.