

Academiejaar 2021–2022 – 2 ^e examenperiode		Individuele opgave
Departement: IT en Digitale Innovatie Opleiding, afstudeerrichting en jaar: Toegepaste informatica, 2TI Naam van het opleidingsonderdeel: Data Science & AI dOLOD/Deelexamen: Theorie+oefeningen Campus: Aalst, Schoonmeersen, TIAO, VC Lesgever(s): Sabine De Vreese, Stijn Lievens, Lieven Smits, Bert Van Vreckem		Examendatum: 2022-05-31 Aanvangsuur examen: 8:30
Voornaam en naam student: BENOOT Robin		
Studentennummer: 201100875		
Lector bij wie de student de onderwijsactiviteit volgde:		Lesgroep v/d onderwijsactiviteit: PBA-TIN-TIAO/2A
Het behaalde resultaat op dit examen wordt herleid tot een cijfer op 20.		

Het examen is open boek. Tijdens het examen mogen volgende hulpmiddelen gebruikt worden:

- Eigen laptop met alle daarop aanwezige software zoals Python, Excel (of andere rekenbladsoftware), rekenmachine, ...
- Internetverbinding voor gebruik van Google Colab, raadplegen elektronisch cursusmateriaal via Chamilo of Github, opzoeken van informatie.
- Afdrukt cursusmateriaal, bv. slides, eigen uitgewerkte oefeningen, Python-code, nota's, ...
- **Je mag geen enkele vorm van communicatie gebruiken tijdens de examens (chatten, mailen, Discord, Messenger, GSM, personen bij je in de buurt, ...).**

Algemene richtlijnen:

- Gebruik het bijgevoegde .ipynb-bestand (Jupyter Notebook) om je antwoorden op de vragen in te vullen, met eventueel de Python-code die je gebruikt hebt om het resultaat te bekomen.
- | |
|---|
| Je hernoemt het bestand naar dsai-benoot-robin.ipynb |
|---|
- Het moet mogelijk zijn om alle code in het notebook opnieuw uit te voeren zonder fouten.
- Wanneer de uitkomst een reëel getal is, rond dan af tot **exact drie cijfers na de komma**, behalve als dit **expliciet anders gevraagd wordt**. Er worden geen marges gerekend op de antwoorden.
- **Let op!** Geef altijd een duidelijk en expliciet antwoord op de (onderzoeks-)vraag in een Markdown-cel. Enkel de Python-code om het antwoord te bekomen of de uitvoer van de Python-code is onvoldoende. De Python-code (en de uitvoer ervan) is enkel ter info, en geldt niet als antwoord.
- Dien het ipynb-bestand in op exam.hogent.be.
- **Wie indient na de deadline krijgt Afwezig.** Wacht dus niet tot de laatste minuut om in te dienen. Dien eventueel tijdens het examen af en toe een onafgewerkte versie in.

Veel succes!

Vragen

1. (2 pt)

Wat is het meetniveau van deze variabelen?

- (a) Het melkras van een koe (roodbont, Holstein, enz)
- (b) Het gewicht van een koe (in kg)
- (c) De jaarlijkse melkproductie van een koe (in l)
- (d) Levensfase van een koe (kalf, pink, vaars, koe)

2. (2 pt)

Een marketing bedrijf werkt voor een klein aantal grote bedrijven en voor een groot aantal kleine bedrijven. Het bedrijf wenst een aselechte steekproef van 10% te nemen van zijn 320 cliënten. De cliënten worden gesorteerd op basis van jaarlijkse omzet en vervolgens neemt men client 3, 13, 23, 33, ...

- (a) Is dit een aselechte steekproef? Motiveer je antwoord.
- (b) Welk soort fout wordt hier gemaakt? Leg uit.

3. (4 pt)

Beschouw volgende kansfunctie voor een toevalsveranderlijke X :

x	-7	-4	4	7
$f_X(x)$	2/8	z	1/8	1/8

- (a) Bepaal z zodat we te maken hebben met een geldige kansfunctie.
- (b) Geef μ_X (ook genoteerd als $E(X)$)
- (c) Bereken σ_X
- (d) Bepaal $P(X \leq 4)$

4. (2 pt)

Patiënten die herstellen van een beroerte laten hun grijpkracht in elke hand laten meten om hun vooruitgang te volgen. Een bepaalde populatie van meer dan 100 mannelijke patiënten hebben grijpkracht in hun dominante handen met een gemiddelde en standaardafwijking van respectievelijk ongeveer 41 kg en 9 kg.

Stel dat we aselechte steekproeven nemen van 30 mannelijke patiënten uit deze populatie en berekenen het steekproefgemiddelde \bar{x} van elke groep patiënten. Wat zal de vorm zijn van de kansverdeling van \bar{x} ?

- (a) Benaderend normaal verdeeld.
- (b) Scheef naar rechts.
- (c) Onbekend. We hebben niet genoeg informatie om de vorm te bepalen.
- (d) Scheef naar links.

5. (2 pt)

Uit een normaal verdeelde populatie werd de hieronder gegeven steekproef getrokken. Bereken een 99%-betrouwbaarheidsinterval voor het populatiegemiddelde.

(a) Ondergrens
(b) Bovengrens

Gebruik een geschikte statistische toets (met significantieniveau 5%) om dit uit te zoeken.

- Welke toets moet je gebruiken om deze onderzoeksvraag te beantwoorden? Wees zo specifiek mogelijk!
- Formuleer de nulhypothese en de alternatieve hypothese
- Bereken het gemiddelde van beide groepen
- Bereken de p-waarde
- Trek een besluit op basis van de vorige stap en beantwoord de onderzoeksvraag.
- Bereken de effectgrootte (Cohen's d)

Gebruik een geschikte statistische toets (met significantieniveau 5%) om dit te onderzoeken.

Page 3

```

'Soda':
  ['BuzzCola', 'Nozz-A-La', 'Nozz-A-La', 'BuzzCola', 'PittCola',
   'BuzzCola', 'BuzzCola', 'BuzzCola', 'Nozz-A-La', 'PittCola',
   'Nozz-A-La', 'BuzzCola', 'BuzzCola', 'Nozz-A-La', 'BuzzCola',
   'PittCola', 'Nozz-A-La', 'Nozz-A-La', 'Nozz-A-La', 'BuzzCola',
   'BuzzCola', 'Nuka-Cola', 'BuzzCola', 'Nozz-A-La', 'BuzzCola',
   'BuzzCola', 'Nozz-A-La', 'Nozz-A-La', 'BuzzCola', 'Nozz-A-La',
   'Nozz-A-La', 'BuzzCola', 'Nozz-A-La', 'BuzzCola', 'Nuka-Cola',
   'Nozz-A-La', 'Nozz-A-La', 'PittCola', 'PittCola', 'Nuka-Cola',
   'Nuka-Cola', 'Nuka-Cola', 'BuzzCola', 'Nozz-A-La', 'Nuka-Cola',
   'PittCola', 'Nuka-Cola', 'PittCola', 'Nozz-A-La', 'Nuka-Cola',
   'Nozz-A-La', 'Nozz-A-La', 'PittCola', 'BuzzCola', 'Nozz-A-La',
   'PittCola', 'Nuka-Cola', 'PittCola', 'BuzzCola', 'Nuka-Cola',
   'Nuka-Cola', 'PittCola', 'Nuka-Cola', 'Nuka-Cola', 'BuzzCola',
   'PittCola', 'Nuka-Cola', 'Nuka-Cola', 'BuzzCola', 'BuzzCola']
})

```

- Welke toets moet je gebruiken om deze onderzoeksvraag te beantwoorden? Wees zo specifiek mogelijk!
- Bereken de teststatistiek (geef symbool en waarde)
- Bereken de p-waarde
- Trek een besluit op basis van de vorige stap en beantwoord de onderzoeksvraag.
- Is de regel van Cochran voldaan? Motiveer je antwoord!
- Bereken Cramér's V en interpreteer deze waarde.

8. (6 pt)

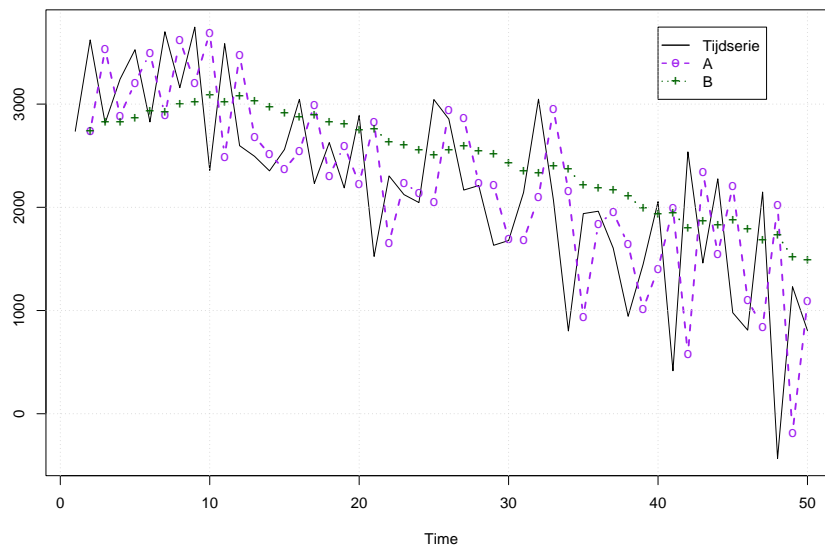
Beschouw de steekproef die hieronder gegeven is

```

sample_data = pd.DataFrame(data = {
  'x': [963, 966, 940, 1073, 1017, 1039, 978, 731, 988, 979, 1047, 918,
        1102, 1034, 1049, 1191, 1041, 907, 972, 1010, 1193, 957,
        925, 897, 963, 806, 1008, 1181, 976, 1015, 993, 1112,
        1049, 922, 1153, 980, 1054, 1038, 805, 1149, 899, 846,
        941, 1246, 975, 913, 1104, 1112, 925, 980, 1122, 869,
        1124, 831, 889, 944, 932, 939, 943, 915],
  'y': [1458, 1380, 1552, 1518, 1520, 1449, 1450, 1398, 1497, 1490,
        1382, 1496, 1481, 1463, 1622, 1497, 1418, 1418, 1467,
        1502, 1494, 1526, 1388, 1419, 1459, 1510, 1438, 1504,
        1486, 1365, 1403, 1506, 1502, 1513, 1488, 1470, 1420,
        1483, 1416, 1569, 1395, 1377, 1471, 1461, 1463, 1468,
        1502, 1499, 1398, 1445, 1520, 1423, 1474, 1358, 1490,
        1417, 1395, 1458, 1419, 1382]})

```

- Bereken de richtingscoëfficiënt van de regressierechte
- Bereken het snijpunt van de regressierechte met de y-as
- Bereken de correlatiecoëfficiënt (symbool + waarde)
- Geef een interpretatie voor de waarde van de correlatiecoëfficiënt
- Bereken de determinatiecoëfficiënt (symbool + waarde)
- Geef een interpretatie voor de waarde van de determinatiecoëfficiënt



Figuur 1: Tijdseriedata (in het zwart) met twee vormen van voortschrijdend gemiddelde (in kleur).

9. (2 pt)

In Figuur 1 vind je een grafiek van een tijdserie (in het zwart) met twee vormen van voortschrijdend gemiddelde in kleur getoond.

Wat zijn de lijnen A en B precies? Kies uit de mogelijkheden hieronder:

- Eenvoudig voortschrijdend gemiddelde met periode 10
- Eenvoudig voortschrijdend gemiddelde met periode 15
- Eenvoudig voortschrijdend gemiddelde met periode 20
- Enkelvoudige exponentiële afvlakking met $\alpha = 0.1$
- Enkelvoudige exponentiële afvlakking met $\alpha = 0.9$

Ter info is de definitie van de tijdserie hieronder gegeven, maar je hebt deze in principe niet nodig om de vraag te kunnen beantwoorden:

```
df = pd.DataFrame(data={'observations':
    [2737, 3622, 2816, 3242, 3527, 2827, 3701, 3159, 3746, 2358, 3586,
    2597, 2494, 2353, 2560, 3046, 2231, 2627, 2188, 2889, 1524,
    2304, 2123, 2046, 3044, 2855, 2167, 2211, 1632, 1678, 2143,
    3048, 2070, 802, 1939, 1962, 1605, 944, 1447, 2055, 416, 2537,
    1460, 2276, 979, 810, 2148, -433, 1231, 806]
})
```

Vermeld expliciet welke lijn welk voortschrijdend gemiddelde voorstelt.

- (a) Lijn A: ...
- (b) Lijn B: ...