TEST MATLAB (STATISTIEK EN WISKUNDIGE DATA-ANALYSE)
(1^{ste} zit '20-'21, reeks B)
Opleiding industrieel ingenieur

FACULTEIT INGENIEURSWETENSCHAPPEN EN ARCHITECTUUR

Naam : Correctiesleutel Richting: /20

Schrijf netjes. Vul in op de opengelaten plaatsen.
De MATLAB-code komt in de kadertjes.
Geen gsm, smartphone, rekentoestel .... Veel succes!

UNIVERSITEIT GENT

1. Formuleer in deze oefening telkens H_0 en H_1 en maak een schets met alle informatie indien van toepassing; vermeld de berekende waarden gegenereerd door MATLAB (geen code).
Men wil weten wat de invloed is van industriële activiteiten op de kwaliteit van de bodem. Er werden stalen genomen van verschillende bodems (met of zonder industriële activiteiten in het verleden). Analyse gaf onderstaande waardes als resultaat. /4

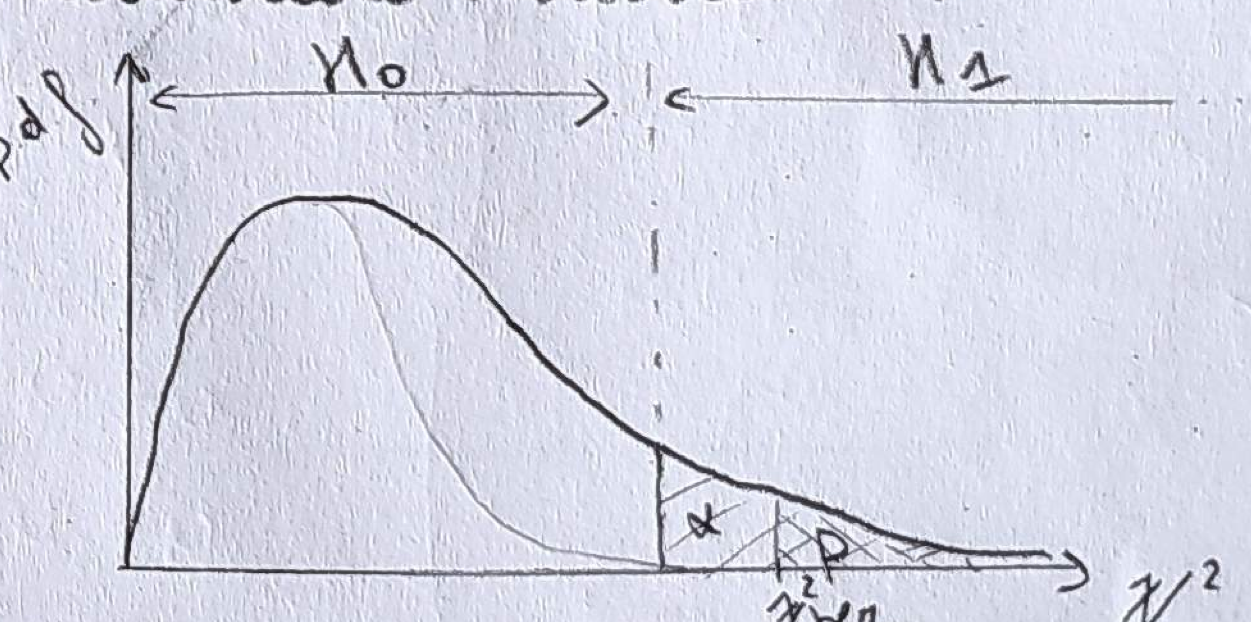
	Bodemvervuiling vastgesteld	Geen bodemvervuiling vastgesteld
Industriële activiteiten	50	34
Geen industriële activiteiten	37	47

Test met 90% betrouwbaarheid of de kwaliteit van de bodem in verband kan worden gebracht met industriële activiteiten uit het verleden. Welke test gebruik je? Geef de waarde van de testveranderlijke en leg je besluit uit.

We gebruiken een χ^2-test voor het testen van (on)afhankelijkheid.

H_0: Er is geen afhankelijkheid tussen industriële activiteiten en bodemvervuiling

H_1: Er is wel afhankelijkheid



$p = 0,0474 < 0,1$

$\Rightarrow H_0$ verwerpen

$\Rightarrow$ Met 90% betrouwbaarheid bestaat er afhankelijkheid tussen industriële activiteiten en vastgestelde bodemvervuiling

Hoeveel van de 168 stalen zouden er zijn voor grond met bodemvervuiling en zonder industriële activiteiten indien er perfecte onafhankelijkheid zou zijn tussen vervuiling van de bodem en het al dan niet hebben plaatsgevonden van industriële activiteiten?

aandeel met vervuiling $= \dfrac{50+37}{168} = 0,5179$

aandeel zonder industrie $= \dfrac{37+47}{168} = 0,5$

Aantal met vervuiling en zonder industrie $= 0,5179 \cdot 0,5 \cdot 168 = 43,5 \approx 44$

2. Los volgend stelsel op:

$$\begin{cases} 8x + y + 6z = 7.5 \\ 3x + 5y + 7z = 4 \\ 4x + 9y + 2z = 12 \end{cases}$$

$x =$ 1,2931

$y =$ 0,8972

$z =$ -0,6236

(numeriek, 4 cijfers na de komma)

```
A=[8,1,6 ; 3,5,7;4,9,2]
b=[7.5;4;12]
a = A\b
  of
D = [A,b]
rref(D)
```

/2.5

3. Bepaal het laplacebeeld van

$$f(t) = \begin{cases} 4 & 0 < t < 1 \\ 4e^{1-t} & t > 1 \end{cases}$$

$F(s) = \dfrac{4e^{-s}}{s+1} - \dfrac{4(e^{-s}-1)}{s}$

```
syms s t
laplace(4*heaviside(1-t)...
+4*exp(1-t)*heaviside(t-1), t, s)
```

/2.5

4. Is er een $x \in \mathbb{R}$ waarvoor de drie eigenwaarden van $\begin{pmatrix} 0 & 2 & 0 \\ 2 & x & 0 \\ 3 & x & x \end{pmatrix}$ dezelfde waarde hebben? Neen

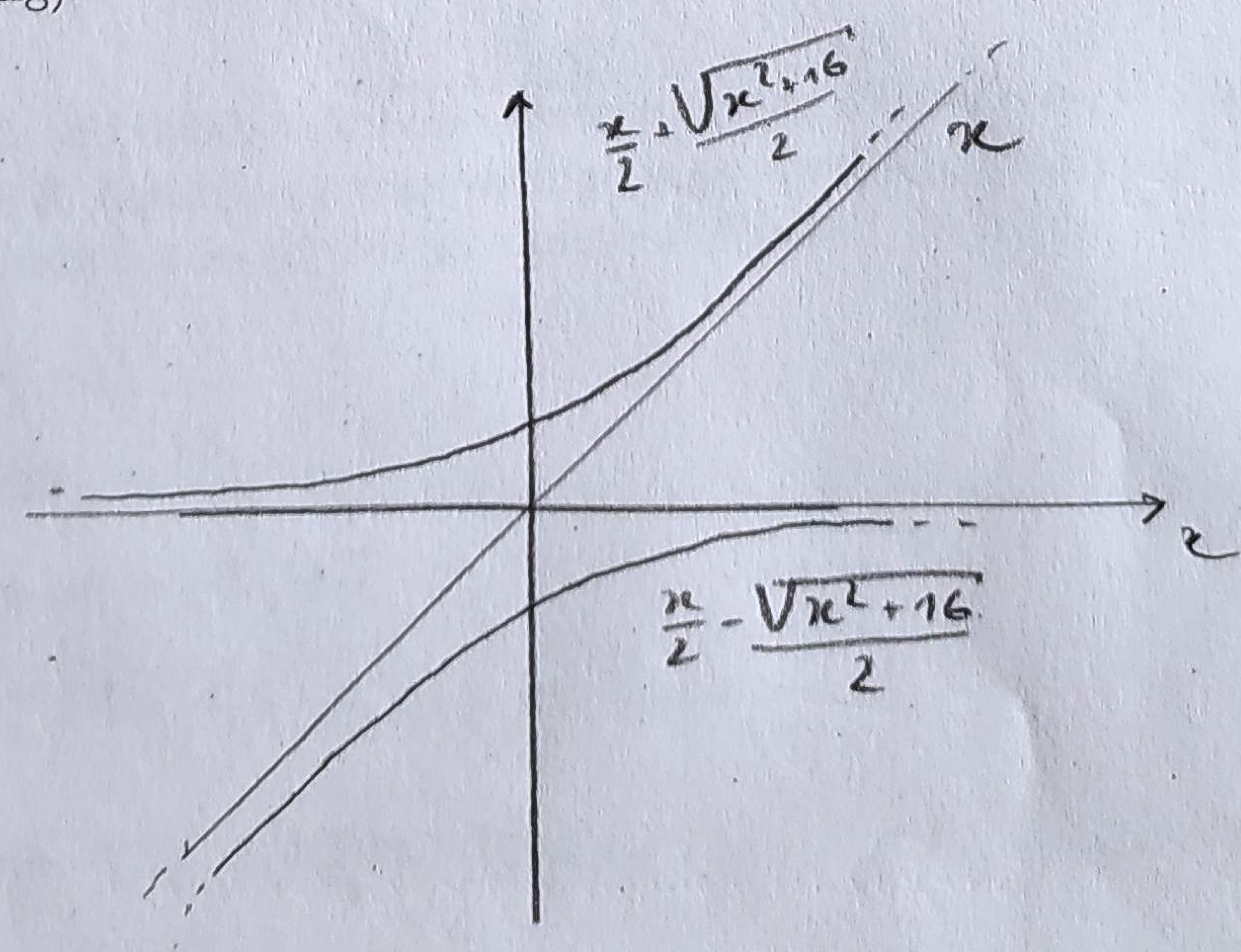
Zo ja, welke? $x =$ ....

(numeriek, 4 cijfers na de komma)

Verklaar grafisch jouw antwoord door een tekening te maken van de eigenwaarden in functie van x (op 1 tekening).

```
syms x
A=[0 2 0;2 x 0; 3 x x]
D= eig(A)
solve([D(1)==D(2), D(2)==D(3)],x)
hold on
fplot(D(1))
fplot(D(2))
fplot(D(3))
```

/5

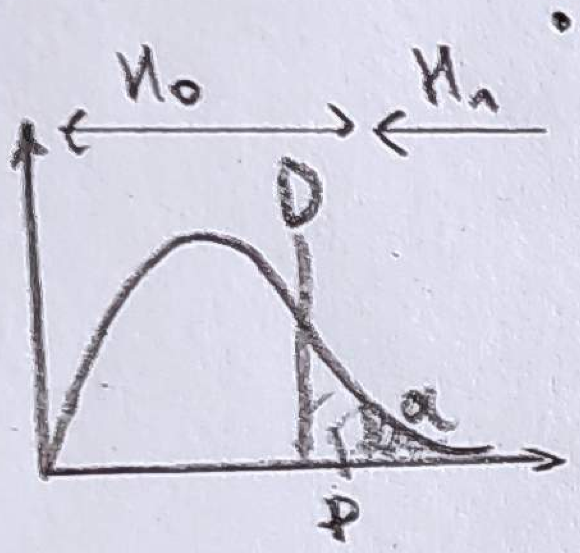


5. Formuleer in deze oefening telkens H_0 en H_1 en maak een schets met alle informatie indien van toepassing; vermeld de berekende waarden gegenereerd door MATLAB (geen code).
De hemoglobinewaardes van patiënten die aan drie ziekten lijden, worden vergeleken in het labo van het ziekenhuis. De hemoglobinewaarde (uitgedrukt in g/dl) wordt bij elk van de ziektes voor 7 patiënten gemeten: /6

Ziekte A:	19	22	18	19	22	24	18
Ziekte B:	16	22	17	16	15	12	17
Ziekte C:	10	14	16	15	14	13	16

Is er een significant verschil tussen de gemiddelde hemoglobinewaardes op populatieniveau voor de 3 ziekten? Zo ja, waartussen? Leg je besluitvorming uit en ga de nodige veronderstellingen na voor de gebruikte test.

Hier gebruiken we one-way ANOVA. Twee voorwaarden:



• Groepen zijn normaal verdeeld

KS-test: H_0 = gegevens komen uit normaal verdeelde populatie
$H1$ = gegevens komen niet uit.

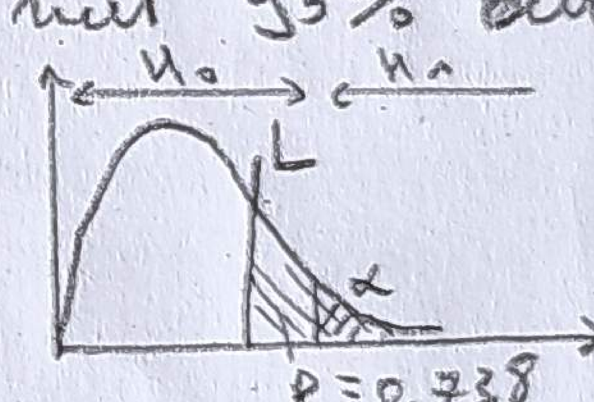
$p = \begin{cases} 0{,}5870 \\ 0{,}5440 \\ 0{,}8439 \end{cases} > \alpha = 0{,}05 \Rightarrow H_0$ aanvaarden

Gegevens normaal verdeeld met 95% betr.

• groepen hebben gelijke variantie

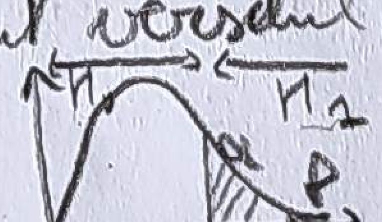
Levene test: $H_0 = \sigma_A^2 = \sigma_B^2 = \sigma_C^2$
H_1: niet alle σ_i^2 zijn gelijk

$p = 0{,}7438 > \alpha \Rightarrow H_0$ aanvaarden
Var. gelijk met 95% betrouwbaarheid



One way ANOVA: H_0: $\mu_A = \mu_B = \mu_C$
H_1 = niet alle μ_i zijn gelijk

$p = 0{,}0007 < \alpha = 0{,}05$ ⇝ H_0 aanvaarde: er is met 95% betrouwbaarheid een significant verschil tussen de hemoglobinewaardes.



Subhypotheses met 95% betr.:
$\mu_1 \neq \mu_2$
$\mu_1 \neq \mu_3$
$\mu_2 = \mu_3$

H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 \neq \mu_2$, $p = 0{,}0258 < \alpha$ ⇝ H_0 verworpen
H_0: $\mu_1 = \mu_3$, H_1: $\mu_1 \neq \mu_3$, $p = 0{,}0005 < \alpha$ ⇝ H_0 verwerpen
H_0: $\mu_2 = \mu_3$, H_1: $\mu_2 \neq \mu_3$, $p = 0{,}1939 > \alpha$ ⇝ H_0 accepteren

Bereken de vijf belangrijkste beschrijvende statistieken van de boxplot van de gegevens van ziekte B. Bereken de voorwaarden voor eventuele uitschieters en noteer de formules die je gebruikt. Zijn er hier uitschieters? Welke? Verklaar op een wiskundige manier!

1e kwartiel = 15,25
3e kwartiel = 17
interkwartielafstand = 1,75
ondergrens: $15{,}25 - 1{,}5 \cdot 1{,}75 = 12{,}6250$
bovengrens: $17 + 1{,}5 \cdot 1{,}75 = 19{,}6250$ ⇝ 22 is een uitschieter