

Kursus:	M4STI1 Statistik for Ingeniører		
Eksamensdato:	08.06.2021		
Eksamenstermin:	Forår 2021		
Underviser:	Allan Leck Jensen og Ulla-Lisbeth Hoffmann		

Praktiske informationer:

Digital eksamen

Opgaven tilgås og afleveres gennem den digitale eksamensportal. Håndskrevne dele af opgavebesvarelsen skal digitaliseres og afleveres i den digitale eksamensportal. **Opgavebesvarelsen skal afleveres i PDF-format.**

Husk angivelse af navn og studienummer på alle sider, samt i dokumenttitel/ filnavn.

Husk at uploade og aflevere i Digital eksamen. Du vil modtage en elektronisk afleveringskvittering, straks du har afleveret.

Husk at aflevere til tiden, da der ellers skal indsendes dispensationsansøgning.

Hjælpemidler:

Alle hjælpemidler må benyttes, herunder internettet som opslagsværktøj, men det er **IKKE** tilladt at kommunikere med andre digitalt.

Særlige bemærkninger:

Det er kun muligt at aflevere elektronisk via Digital Eksamen portalen.

Ingeniørhøjskolen Aarhus Universitet Maskinteknik Eksamenstermin: Forår 2021

Prøve i:

M4STI1 Statistik for Ingeniører

Bemærk følgende:

- Decimaltal i opgaverne er angivet med engelsk decimalseparator (.). Desuden anvendes engelsk tusindtalsseparator (,).
- Alle data fra opgaverne kan downloades fra Digital Eksamen portalen i et regneark med navnet: Data_M4STI1_2021F. I regnearket angiver kolonnenavnet, hvilken opgave data hører til.
- Nogle delopgaver benytter resultatet fra en tidligere delopgave. Hvis du ikke kunne løse den, kan du blot antage en værdi for resultatet og regne videre med det.
- Selv om der er brugt MATLAB i undervisningen, er det tilladt at bruge andre værktøjer. I nogle delopgaver kan det være enklere at bruge f.eks. Excel.

Eksamenstermin: Forår 2021

Prøve i:

M4STI1 Statistik for Ingeniører

Opgave 1

En virksomhed producerer havregryn i pakker, der ifølge deklarationen hver indeholder 950 g havregryn. Produktionsanlæggets vægt er indstillet, så der i gennemsnit kommer 950 g havregryn i pakkerne, men vejninger har vist, at den faktiske masse af havregryn er normalfordelt med en middelværdi på 950 g og en standardafvigelse på 20 g.

- a. Bestem sandsynligheden for at massen af havregryn er mindre end 900 g.
- **b.** Bestem sandsynligheden for at massen af havregryn ligger i intervallet fra 900 g til 1010 g.

Virksomheden vil sikre sig, at højst 3.0 % af havregrynspakkerne vejer under de tilsigtede 950 g.

c. Hvilken middelværdi skal vægten indstilles på for at sikre, at højst 3.0 % af havregrynspakkerne vejer under 950 g, idet det fortsat antages, at standardafvigelsen er 20 g?



Eksamenstermin: Forår 2021

Prøve i:

M4STI1 Statistik for Ingeniører

Opgave 2

En virksomhed ønsker at sammenligne komponent-tykkelsen fra to producenter A og B.

Der udtages en stikprøve på 12 komponenter fra producent A og en stikprøve på 15 komponenter fra producent B. De målte komponent-tykkelser i mm er indført i nedenstående tabeller:

Producent A

22.32	21.71	21.54	22.09	22.03	22.08
21.85	21.64	22.33	21.79	22.29	21.93

Producent B

22.24	22.49	22.50	22.47	22.36	22.10
22.44	22.65	22.68	22.54	22.61	22.28
22.19	22.30	22.37			

Opgaven drejer sig om at undersøge ved hjælp af en hypotesetest, om der er forskel på komponenttykkelsen fra de to producenter, når der vælges et signifikansniveau på 5 %.

- **a.** Opstil nulhypotese og alternativ hypotese for hypotesetesten.
- **b.** Opstil en formel for teststørrelsen (teststatistikken), og angiv hvilken fordeling den følger.
- c. Bestem det kritiske område for testen.
- **d.** Beregn teststørrelsens (teststatistikkens) værdi, idet nødvendige mellemregninger medtages.
- **e.** Konkluder på hypotesetesten.
- **f.** Hvilke antagelser er der foretaget for at udføre hypotesetesten? Er disse antagelser rimelige? I vurderingen af data fra de to producenter skal der indgå normalfordelingsplot og f.eks. Bartlett's test eller boksplot.

Ingeniørhøjskolen Aarhus Universitet Eksamenstermin:

Maskinteknik Forår 2021 M4STI1 Statistik for Ingeniører

Prøve i:

Opgave 3

Danske forskere har undersøgt kvaliteten af den antigen-test, der i daglig tale kaldes 'lyntest' eller 'hurtigtest'. Hurtigtesten bruges som supplement til den mere sikre, men langsommere PCR-test.

Man beskriver kvaliteten af en test med de to begreber *sensitivitet* og *specificitet*. *Sensitiviteten* af en test angiver procentdelen af sandt positive, d.v.s. testens evne til at 'fange' de personer, der har sygdommen. *Specificiteten* af en test angiver procentdelen af sandt negative, d.v.s. testens evne til at bedømme raske personer korrekt (som raske).

Forskerne har fundet ud af, at hurtigtesten har en sensitivitet på 70 % og en specificitet på 99.5 %.

Lad os bruge *A* og *B* til at betegne følgende hændelser for en tilfældig person, der får foretaget hurtigtesten: *A* angiver at personen er Covid-19 smittet. *B* angiver at hurtigtesten bedømmer, at personen er smittet.

- **a.** Oplys følgende sandsynligheder for hurtigtesten: P(B|A), $P(B^c|A)$, $P(B^c|A^c)$ og $P(B|A^c)$.
- **b.** Angiv følgende:
 - Testens procentdel af falsk positive
 - Testens procentdel af falsk negative
 - Testens power
- **c.** En given dag testes 10,000 tilfældigt udvalgte personer fra samme region med hurtigtesten. Antag at procentdelen af smittede i regionen er 20 % på den pågældende dag. Beregn forventet:
 - Antal falsk positive tests
 - Procentdelen af alle positive tests, der er falske
 - Procentdelen af alle negative tests, der er falske
- **d.** I en anden region er kun 1 % af befolkningen smittet. Her testes også 10,000 tilfældigt udvalgte personer med hurtigtesten på en given dag.

Her fås følgende resultater:

- Antal falsk positive tests: 50
- Procentdelen af alle positive tests, der er falske: 41 %
- Procentdelen af alle negative tests, der er falske: 0.30 %

Giv en forklaring på forskelle i resultaterne i denne region i forhold til regionen med mange smittede i spørgsmål **c.**

e. I en tredje region tester man ligeledes 10,000 tilfældigt udvalgte personer med hurtigtesten. De personer, der testes positiv med hurtigtesten, kontroltestes med den næsten sikre PCR-test, som vi i denne sammenhæng kan antage er 100 % sikker. Det viser sig, at 43 af dem ikke var smittet, så hurtigtesten tog fejl. Beregn procentdelen af smittede i den pågældende region, *P*(*A*).

Opgave 4

En forskergruppe har foretaget sammenhørende målinger af parametrene x_1 , x_2 og y, som vist i tabellen herunder:

X ₁	X2	y
1.0	110	166.6
1.0	150	148.2
1.0	190	110.2
1.0	230	84.2
2.0	110	153.8
2.0	150	178.2
2.0	190	93.7
2.0	230	133.5
3.0	110	178.8
3.0	150	152.6
3.0	190	117.3
3.0	230	114.0
4.0	110	229.1
4.0	150	258.8
4.0	190	148.4
4.0	230	196.1
5.0	110	448.7
5.0	150	365.0
5.0	190	519.8
5.0	230	357.2

Forskergruppen forventer at kunne udvikle en lineær regressionsmodel på baggrund af de målte observationer: $y = f(x_1, x_2)$.

- **a.** Lav et scatter plot for y mod hver af regresssorvariablene x_1 og x_2 . Diskutér, om der lader til at være korrelation mellem den enkelte regressorvariabel og responsvariablen.
- **b.** Lav en multipel lineær regressionsmodel, der beskriver y som funktion af x_1 og x_2 . Skriv funktionsudtrykket op.
- **c.** Forklar vha. regressionsanalysens statistikker (f.eks. R-squared, F og p-value), om modellen beskriver observationerne godt.
- **d.** Bliver modellen bedre, hvis du inddrager interaktion mellem x_1 og x_2 ?
- **e.** Forskergruppen undersøger to muligheder for at forbedre modellen gennem transformationer af data:

Transformeret model 1: $y = b_0 + b_1 \exp(x_1) + b_2 x_2$

Transformeret model 2: $\ln(y) = b_0 + b_1 x_1 + b_2 x_2$

hvor b_0 , b_1 og b_2 er konstanter. Hvilken af de to modeller er bedst? Begrund dit valg. Bemærk at i MatLab skrives den naturlige logaritme ln som log.

f. Undersøg på baggrund af din valgte model, om der er unormale observationer i datasættet (outliers, løftestangs-punkter eller indflydelsespunkter).