Danmarks Tekniske Universitet *Side 1 af 20 sider.*

Skriftlig prøve: 26. maj 2019

Kursus navn og nr.: **Introduktion til Statistik (02323)**

Varighed: 4 timer

Tilladte hjælpemidler: Alle

Dette sæt er besvaret af

_____	_____	_____
(studienummer)	(underskrift)	(bord nr.)

Opgavesættet består af 30 spørgsmål af "multiple choice" typen, som er fordelt på 11 opgaver. For at besvare spørgsmålene skal du udfylde "multiple choice" svararket (6 separate sider) på CampusNet med numrene på de svarmuligheder, som du mener er de rigtige.

Der gives 5 point for et korrekt "multiple choice" svar og -1 point for et forkert svar. KUN følgende 5 svarmuligheder er gyldige: 1, 2, 3, 4 eller 5. Hvis et spørgsmål efterlades blankt eller et ugyldigt svar angives, gives der 0 point for spørgsmålet. Endvidere, hvis mere end et svar angives til det samme spørgsmål, hvilket faktisk er teknisk muligt i online-systemet, gives der 0 point for spørgsmålet. Det antal point der kræves, for at opnå en bestemt karakter eller for at bestå eksamen afgøres endeligt ved censureringen.

Den endelige besvarelse af opgaverne laves ved at udfylde og aflevere svararket online via CampusNet. Skemaet her er KUN et nød-alternativ til dette.
Husk at angive dit studienummer, hvis du afleverer på papir.

Opgave	I.1	I.2	II.1	II.2	III.1	III.2	III.3	IV.1	IV.2	V.1
Spørgsmål	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Svar										

Opgave	V.2	V.3	V.4	VI.1	VI.2	VII.1	VII.2	VII.3	VII.4	VIII.1
Spørgsmål	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Svar										

Opgave	IX.1	IX.2	IX.3	IX.4	X.1	X.2	X.3	X.4	XI.1	XI.2
Spørgsmål	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Svar										

Eksamenssættet består af 20 sider.

Fortsæt på side 2

Multiple choice opgaver: *Der gøres opmærksom på, at der i hvert spørgsmål er én og kun én svarmulighed, som er rigtig. Endvidere er det ikke givet, at alle de anførte alternative svarmuligheder er meningsfulde. Husk altid at afrunde dit eget resultat til antallet af decimaler givet i svarmulighederne før du vælger et svar.*

Opgave I

I en colasmagning har man 4 glas med cola. Hvert glas indeholder enten almindelig cola eller cola light. Man ved, at der er to glas af hver. En smager vælger tilfældigt to glas.

Spørgsmål I.1 (1)

Hvad er sandsynligheden for, at hun får almindelig i et af glassene og light i det andet?

1 ☐ $1/4$

2 ☐ $1/3$

3 ☐ $1/2$

4 ☐ $2/3$

5 ☐ $3/4$

Spørgsmål I.2 (2)

I et andet forsøg gives der et glas almindelig cola og et glas cola light til hver af 25 smagere. De får besked om at smage og svare på om de kan smage forskel på colaen i glassene. Svarene er uafhængige af hinanden.

Fra erfaring ved man, at det kan antages, at der er $p = 0.8$ sandsynlighed for at smagerne kan smage forskel. Lad X betegne antallet af de 25 smagere, som svarer at der er forskel. Hvad bliver variansen af X?

1 ☐ $\mathrm{V}(X) = 5$

2 ☐ $\mathrm{V}(X) = 4$

3 ☐ $\mathrm{V}(X) = 3$

4 ☐ $\mathrm{V}(X) = 2$

5 ☐ $\mathrm{V}(X) = 1$

Fortsæt på side 3

Opgave II

10 kvinder har målt deres morgentemperatur d. 1. juli og d. 1. december. Ud fra målingerne vil man gerne undersøge, om der er forskel på morgentemperaturen hos kvinder om sommeren i forhold til om vinteren. Det kan antages, at sommermålingerne er normalfordelte, og at vintermålingerne er normalfordelte.

Spørgsmål II.1 (3)

Hvilken analyse er den mest hensigtsmæssige?

1 ☐ Test for forskel mellem to andele

2 ☐ Regressionsanalyse

3 ☐ (Uparret) t-test

4 ☐ Parret t-test

5 ☐ Test i binomialfordelingen

Spørgsmål II.2 (4)

Når man udfører testet får man en p-værdi på 0.4. Det betyder, at:

1 ☐ Der er 40% sandsynlighed for, at der er forskel på morgentemperaturen om sommeren og morgentemperaturen vinteren.

2 ☐ Der er 0.4% sandsynlighed for, at der er forskel på morgentemperaturen om sommeren og morgentemperaturen vinteren.

3 ☐ Hypotesen kan ikke testes.

4 ☐ Der er helt sikkert forskel på morgentemperaturen om sommeren og morgentemperaturen om vinteren.

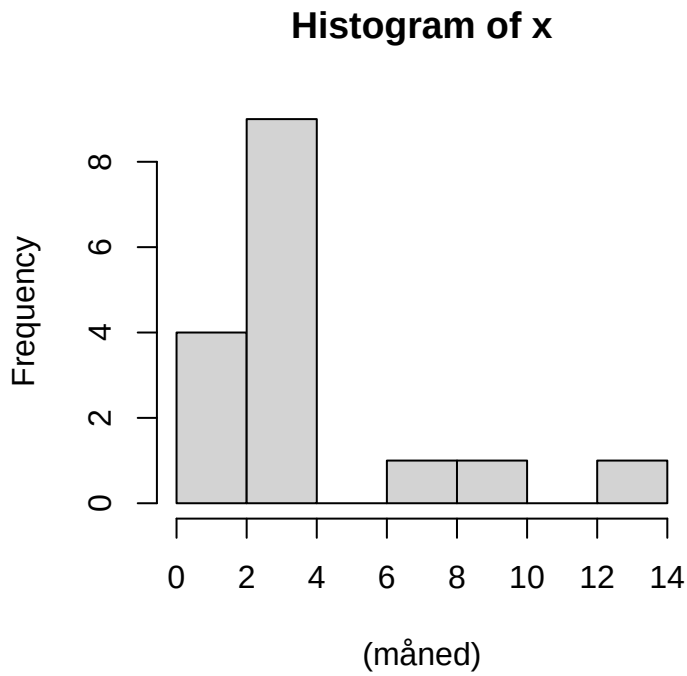
5 ☐ Under nulhypotesen er sandsynligheden for at få en teststørrelse, der er mindre ekstrem, end den man har fået lig 0.6.

Fortsæt på side 4

Opgave III

Et firma har indkøbt en ny 3D-printerteknologi og det ønskes undersøgt, om den kan bruges til at lave komponenter, der er holdbare nok til at indgå i et specifikt produkt.

Der er kørt et forsøg med komponenter printet med den nye teknologi. Komponenterne er blevet indsat i nogle test produkter og disse er blevet udsat for en test, der bestemmer deres levetid. Det antages at levetiden følger en eksponentiel fordeling, så lad $X \sim Exp(\lambda)$ betegne levetiden i måneder. Der er indsamlet en stikprøve for $n = 16$ produkter. Et histogram af stikprøven er:



De observerede levetider er gemt i vektoren `x` og følgende R kode er kørt:

```
## Number of simulations
k <- 10000
nx <- length(x)
## Simulate k times
simxsamples <- replicate(k, rexp(nx, 1/mean(x)))
## Calculate the sample mean
simmeans <- apply(simxsamples, 2, mean)
## Quantiles of the means
quantile(simmeans, c(0.005,0.995))

##  0.5% 99.5%
##  1.70  6.42

quantile(simmeans, c(0.025,0.975))
```

```
##  2.5% 97.5%
##  2.07  5.68

quantile(simmeans, c(0.05,0.95))

##   5%  95%
## 2.26 5.26
```

Spørgsmål III.1 (5)

Det var på forhånd planlagt, at man ville undersøge, om det kunne påvises på signifikansniveau $\alpha = 1\%$, at middellevetiden μ_X er over 2 mnd. for komponenterne. Kan det påvises på baggrund af den indsamlede stikprøve og ovenstående beregninger (både konklusion og argument skal være korrekt)?

1 ☐ Da 2 er indeholdt i det beregnede 99% konfidensinterval kan det ikke påvises.

2 ☐ Da 2 ikke er indeholdt i det beregnede 99% konfidensinterval kan det påvises.

3 ☐ Da 2 er indeholdt i det beregnede 95% konfidensinterval kan det ikke påvises.

4 ☐ Da 2 ikke er indeholdt i det beregnede 95% konfidensinterval kan det påvises.

5 ☐ Med de givne oplysninger kan man ikke svare på dette spørgsmål.

Spørgsmål III.2 (6)

Hvad er stikprøvegennemsnittet af den indsamlede stikprøve?

1 ☐ $\bar{x} = 3.40$

2 ☐ $\bar{x} = 3.76$

3 ☐ $\bar{x} = 3.875$

4 ☐ $\bar{x} = 4.06$

5 ☐ Med de givne oplysninger kan man ikke svare på dette spørgsmål.

Spørgsmål III.3 (7)

En ny stikprøve af levetider er indsamlet, hvor et nyt materiale er benyttet til print af komponenterne. De er efterfølgende udsat for samme tests og de observerede levetider er gemt i vektoren `y`. Der er $n_Y = 17$ observationer i den nye stikprøve.

Følgende R kode er derefter kørt:

```r
## Number of simulations
k <- 10000
nx <- length(x)
ny <- length(y)
## Simulate k times
simxsamples <- replicate(k, rexp(nx, 1/mean(x)))
simysamples <- replicate(k, rexp(ny, 1/mean(y)))
## Calculate the simulated statistics
simdifmeans <- apply(simysamples, 2, mean) - apply(simxsamples, 2, mean)
simdifmedians <- apply(simysamples, 2, median) - apply(simxsamples, 2, median)
## Quantiles of the simulated statistics
quantile(simdifmeans, c(0.025,0.975))

##  2.5% 97.5%
## 0.733 9.443

quantile(simdifmeans, c(0.05,0.95))

##   5%  95%
## 1.30 8.59

quantile(simdifmedians, c(0.025,0.975))

##   2.5%  97.5%
## -0.428  8.265

quantile(simdifmedians, c(0.05,0.95))

##     5%    95%
## 0.0837 7.3868
```

Hvilken af følgende konklusioner kan drages på baggrund af de udførte beregninger?

1 ☐ På $\alpha = 5\%$ signifikansniveau kan det konkluderes at 50% fraktilen af produktlevetiden er højere med komponenter at det nye materiale.

2 ☐ På $\alpha = 10\%$ signifikansniveau kan det konkluderes at 50% fraktilen af produktlevetiden er højere med komponenter at det nye materiale.

3 ☐ På $\alpha = 5\%$ signifikansniveau kan det konkluderes at der er mindst 50% sandsynlighed for at produktlevetiden er højere med komponenter af det nye materiale.

4 ☐ På $\alpha = 10\%$ signifikansniveau kan det konkluderes at der er mindst 50% sandsynlighed for at produktlevetiden er højere med komponenter af det nye materiale.

5 ☐ Med de givne oplysninger kan man ikke drage nogle konklusioner.

Fortsæt på side 7

Opgave IV

Antag at X er normalfordelt med middelværdi 10 og varians 4, Y er normalfordelt med middelværdi 20 og varians 25, samt at X og Y er uafhængige.

Spørgsmål IV.1 (8)

$2Y - 2X + 4$ har da følgende varians:

1 ☐ 36

2 ☐ 58

3 ☐ 84

4 ☐ 116

5 ☐ Ingen af ovenstående

Spørgsmål IV.2 (9)

Hvad er standardafvigelsen af $f(X, Y) = 2Y^2 + X^3/3$ (tips: hvis man løser det vha. simulering, så husk at have mange gentagelser, og regn med at resultatet er ca. $\pm$ 10 fra det opgivne tal i svaret)?

1 ☐ $\sigma_{f(X,Y)} \approx 100$

2 ☐ $\sigma_{f(X,Y)} \approx 250$

3 ☐ $\sigma_{f(X,Y)} \approx 350$

4 ☐ $\sigma_{f(X,Y)} \approx 450$

5 ☐ $\sigma_{f(X,Y)} \approx 5 \cdot 10^4$

Fortsæt på side 8

Opgave V

Sammenhængen mellem tryk (p) og dybde (h) i en åben væskebeholder kan teoretisk beskrives ved ligningen

$$p = p_0 + \rho g h\,,$$

hvor p_0 er lufttrykket, ρ er væskens densitet (massefylde) og g er tyngdeaccelerationen. Et forsøg blev udført med henblik på at bestemme densiteten af en særlig væske. Man har indsamlet 10 sammenhængende målinger af dybde (i m) og tryk (i Pa) i denne væske, som er blevet indlæst i R i to vektorer, hhv. `depth` og `pressure`. Desuden er følgende R-kode blevet kørt:

```
model1 <- lm(pressure ~ depth)
summary(model1)

##
## Call:
## lm(formula = pressure ~ depth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -119166  -73422   30513   53635  124689
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 1.019e+08  5.867e+04 1737.529  < 2e-16 ***
## depth       5.031e+03  9.455e+02    5.321 0.000711 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 85880 on 8 degrees of freedom
## Multiple R-squared:  0.7797,Adjusted R-squared:  0.7521
## F-statistic: 28.31 on 1 and 8 DF,  p-value: 0.0007105
```

Spørgsmål V.1 (10)

Angiv estimatet for lufttrykket under forsøget:

1 ☐ $5.031 \cdot 10^3$ Pa

2 ☐ $5.867 \cdot 10^4$ Pa

3 ☐ $9.455 \cdot 10^7$ Pa

4 ☐ $1.019 \cdot 10^8$ Pa

5 ☐ $1.025 \cdot 10^8$ Pa

Spørgsmål V.2 (11)

Man ønsker at teste en hypotese om, at det forventede lufttryk er $1.005 \cdot 10^8$ Pa under forsøgsforholdene. Angiv den sædvanlige teststørrelse, der benyttes til at teste denne hypotese:

1 ☐ $t_{\text{obs}} = 1738$

2 ☐ $t_{\text{obs}} = 5.321$

3 ☐ $t_{\text{obs}} = 23.86$

4 ☐ $t_{\text{obs}} = 28.31$

5 ☐ $t_{\text{obs}} = 0.000711$

Spørgsmål V.3 (12)

Angiv et 95% konfidensinterval for parameteren, der beskriver sammenhængen mellem dybde og tryk:

1 ☐ $1.019 \cdot 10^8 \pm 2.306 \cdot 85880/(10-2)$

2 ☐ $1.019 \cdot 10^8 \pm 2.306 \cdot 85880$

3 ☐ $5031 \pm 2.306 \cdot 85880$

4 ☐ $1.019 \cdot 10^8 \pm 2.306 \cdot 5.867 \cdot 10^4$

5 ☐ $5031 \pm 2.306 \cdot 945.5$

Spørgsmål V.4 (13)

Angiv et estimatet for væskens densitet under forsøget, når tyngdeaccelerationen g er 9.82 N/kg:

1 ☐ 512 kg/m^3

2 ☐ 1004 kg/m^3

3 ☐ 307 kg/m^3

4 ☐ 802 kg/m^3

5 ☐ 610 kg/m^3

Fortsæt på side 10

Opgave VI

Der er udtaget en stikprøve med uafhængige observationer fra en normalfordelt population. Man vil gerne teste hypotesen om, at middelværdien er nul mod alternativet, at den er forskellig fra nul. Teststørrelsen for testet følger en t-fordeling. Man får en p-værdi på 0.001.

Spørgsmål VI.1 (14)

Hvad ved man da om 99%-konfidensintervallet for middelværdien?

1 ☐ Det indeholder nul.

2 ☐ Det indeholder ikke nul.

3 ☐ Det indeholder nul, men ikke estimatet for middelværdien.

4 ☐ Der er ikke oplysninger nok til at vide noget specifikt om konfidensintervallet.

5 ☐ Det indeholder 0.01.

Spørgsmål VI.2 (15)

Hvis der var $n = 20$ observationer i stikprøven, hvad ved vi da om den observerede teststørrelse?

1 ☐ $t_{\text{obs}} = -1.33$ eller $t_{\text{obs}} = 1.33$

2 ☐ $t_{\text{obs}} = -1.73$ eller $t_{\text{obs}} = 1.73$

3 ☐ $t_{\text{obs}} = -3.55$ eller $t_{\text{obs}} = 3.55$

4 ☐ $t_{\text{obs}} = -3.58$ eller $t_{\text{obs}} = 3.58$

5 ☐ $t_{\text{obs}} = -3.88$ eller $t_{\text{obs}} = 3.88$

Fortsæt på side 11

Opgave VII

Fødevarestyrelsen ønsker at reducere andelen af resistente bakterier i grises tarmflora, da denne udgør en human risiko. qPCR er en mikrobiologisk metode til at tælle antallet af specifikke gener i en fæcesprøve. Nedenfor er vist tælletal for tre gener: 16S, som er et referencegen, og to gener, som koder for resistens mod tetracyklin (tetO og tetM). De fire prøver er udtaget til forskellige tidspunkter (først stikprøve 1, så 2, 3 og tilsidst 4) på den samme gård og man ønsker at undersøge om, der er sket ændringer.

	16S	tetO	tetM	Sum
Stikprøve 1	4675	171	76	4922
Stikprøve 2	2222	95	1	2318
Stikprøve 3	2750	49	2	2801
Stikprøve 4	2040	47	1	2088
Sum	11687	362	80	12129

Det ønskes at lave et χ^2-test for at undersøge, om andelene af resistente gener er ændret over tid.

Spørgsmål VII.1 (16)

Antallet af frihedsgrader i denne test er

1 ☐ 8

2 ☐ 12

3 ☐ 6

4 ☐ 9

5 ☐ Det giver ikke mening at lave en χ^2-test, når to af observationerne er 1.

Spørgsmål VII.2 (17)

Under nulhypotesen hvad er da det forventede antal kopier af tetM i stikprøve 4?

1 ☐ 20

2 ☐ 1

3 ☐ 13.77

4 ☐ 26.10

5 ☐ 696

Spørgsmål VII.3 (18)

Teststørrelsen beregnes til 132.3. Med hvilket kald i R findes den relevante p-værdi?

1 ☐ `1 - dchisq(132.3, df=6)`

2 ☐ `1 - pchisq(132.3, df=6)`

3 ☐ `qchisq(132.3, df=6)`

4 ☐ `pchisq(132.3, df=6)`

5 ☐ `qchisq(1/132.3, df=6)`

Spørgsmål VII.4 (19)

Man har på forhånd planlagt at undersøge om forekomsten af tetO har ændret sig mellem første stikprøve og fjerde stikprøve. Af årsager som ikke forklares her, skal man se bort fra observationerne af tetM i denne test. Man har kørt følgende kode med tilhørende output:

```
prop.test(x=c(171, 47), n=c(4675+171, 2040+47), correct=FALSE, conf.level=0.95)

##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(171, 47) out of c(4675 + 171, 2040 + 47)
## X-squared = 7.8067, df = 1, p-value = 0.005205
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.004550394 0.020982546
## sample estimates:
##     prop 1     prop 2
## 0.03528683 0.02252036
```

Der benyttes det sædvanlige $\alpha = 0.05$ signifikansniveau. Hvad bliver konklusionen (både konklusion og argumentation skal være korrekt)?

1 ☐ Der er ikke sket en signifikant ændring, da $0.02098 < 0.02252$.

2 ☐ Der er sket en signifikant ændring, da $0.0052 < 0.95$, men man kan ikke konkludere om forekomsten er faldet eller steget.

3 ☐ Der er sket en signifikant ændring, da $0.0052 < 0.05$, og forekomsten af tetO er steget.

4 ☐ Der er sket en signifikant ændring, da $0.0052 < 0.95$, og forekomsten af tetO er steget.

5 ☐ Der er sket en signifikant ændring, da $0.0052 < 0.05$, og forekomsten af tetO er faldet.

Fortsæt på side 14

Opgave VIII

Lad IQ af et tilfældigt valgt individ modelleres ved hjælp af en normalfordelt stokastisk variabel. 50% af befolkningen har en IQ over 100 (og 50% har en IQ under 100). Antag at 68% af befolkningen har en IQ i området 85-115.

Spørgsmål VIII.1 (20)

Hvilken procentdel af befolkning har en IQ på mindst 140 og betragtes dermed som genier ifølge denne model?

1 ☐ 0.01%

2 ☐ 1%

3 ☐ 4%

4 ☐ 0.4%

5 ☐ 0.06%

Fortsæt på side 15

Opgave IX

Man har indsamlet nedenstående data fra to grupper:

Gruppe 1: 10.5, 9.3, 10.7, 10.8, 11.2

Gruppe 2: 8.9, 9.5, 10.2, 9.8, 10.3

Alle målingerne antages at været taget uafhængige. Målingerne i gruppe 1 antages at stamme fra en normalfordeling, ligesom målingerne i gruppe 2 antages at stamme fra en normalfordeling. Desuden antages det, at varianserne i de to normalfordelinger er ens.

Spørgsmål IX.1 (21)

Hvad bliver stikprøvegennemsnittet af gruppe 2 stikprøven?

1 ☐ 9.74

2 ☐ 9.8

3 ☐ 10.2

4 ☐ 10.31

5 ☐ 48.5

Spørgsmål IX.2 (22)

Hvad bliver den numeriske værdi af teststørrelsen for det sædvanlige test af hypotesen om, at der ikke er forskel på middelværdien i de to grupper?

1 ☐ 0.8

2 ☐ 1.04

3 ☐ 1.86

4 ☐ 2.19

5 ☐ 2.55

Spørgsmål IX.3 (23)

Man vil gerne finde et 90%-konfidensinterval for middelværdien i gruppe 1. Dette bliver:

1 ☐ [9.61, 11.39]

2 ☐ [9.32, 11.68]

3 ☐ [8.92, 12.03]

4 ☐ [9.87, 12.03]

5 ☐ Ingen af ovenstående intervaller er korrekte.

Spørgsmål IX.4 (24)

Man vil designe et nyt forsøg, med henblik på at opnå en større styrke af det statistiske test vedrørende middelværdierne. Man vil stadig have lige mange observationer i hver gruppe. Man vil gerne have en styrke på 99% til at opdage en forskel i middelværdi på mindst 1 mellem de to grupper, ved signifikansniveau 1%. Som gæt på variansen bruger man det poolede variansestimat fra de to stikprøver givet i opgaven.

Hvad er det mindste antal observationer man skal have fra hver gruppe, for at ovenstående er opfyldt?

1 ☐ Mindst 4

2 ☐ Mindst 6

3 ☐ Mindst 12

4 ☐ Mindst 18

5 ☐ Mindst 22

Fortsæt på side 17

Opgave X

Hvor meget tøj en person har på (beklædningsniveauet) har stor indflydelse på det oplevede komfortniveau i kontorer. I tabellen herunder er vist stikprøver fra tre rum af det gennemsnitlige beklædningsniveau (på en skala 0 til 1):

	Room 1	Room 2	Room 3
	0.43	0.56	0.38
	0.36	0.71	0.39
	0.41	0.20	0.48
	0.42	0.57	0.52
	0.41	0.69	0.23
	0.54	0.55	0.37
	0.61	0.78	0.60
	0.53	0.42	0.46
	0.49	0.42	0.44
	0.69	0.59	0.44
Means	0.49	0.55	0.43

Som en indledende analyse laves en envejs variansanalyse, med rum som forklarende faktor. Resultatet er vist i R-outputtet herunder (hvor signifikanskoder dog er fjernet og enkelte tal er erstattet af bogstaver):

```
anova(lm(clo ~ room, data=Data))

## Analysis of Variance Table
##
## Response: clo
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## room       2 0.06963 0.034813     A   0.1385
## Residuals 27 0.44147 0.016351
```

Spørgsmål X.1 (25)

Hvad er værdien der skal stå i A (afrundet)?

1 ☐ $A = 1.07$

2 ☐ $A = 2.00$

3 ☐ $A = 2.13$

4 ☐ $A = 4.00$

5 ☐ $A = 4.26$

Spørgsmål X.2 (26)

Hvad er konklusionen (på signifikansniveau $\alpha = 0.05$) omkring forskellen i middel af beklædningsniveauet mellem de 3 rum (både konklusion og argument skal være korrekt)?

1 ☐ Der er kan påvises en signifikant forskel da $0.016351 < 0.05$.

2 ☐ Der er kan ikke påvises en signifikant forskel da $0.016351 < 0.05$.

3 ☐ Der er kan påvises en signifikant forskel da $0.1385 > 0.05$.

4 ☐ Der er kan ikke påvises en signifikant forskel da $0.1385 > 0.05$.

5 ☐ Der er kan påvises en signifikant forskel da $0.034813 < 0.05$.

Spørgsmål X.3 (27)

Hvad er et på forhånd planlagt 95%-konfidensintervallet for forskellen i middelværdi mellem rum 1 og rum 2 (dvs. det var inden stikprøven blev indsamlet planlagt kun at lave dette konfidensinterval)?

1 ☐ $[0.12, 0.45]$

2 ☐ $[0.03, 0.25]$

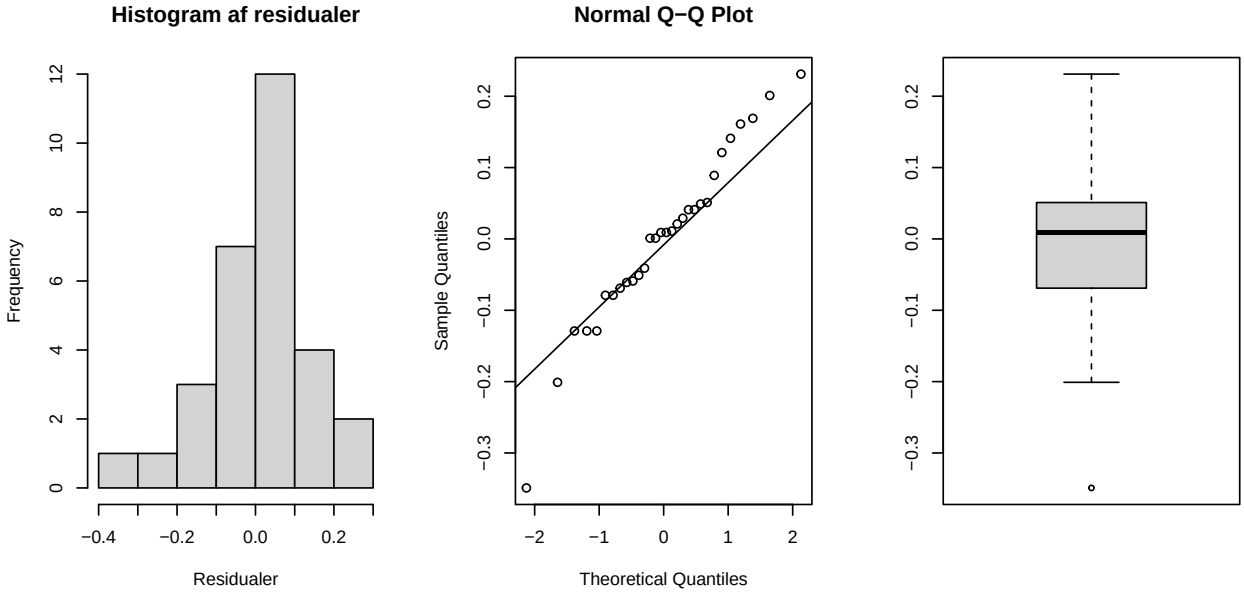
3 ☐ $[-0.17, 0.09]$

4 ☐ $[-0.06, 0.18]$

5 ☐ $[-0.30, 0.42]$

Spørgsmål X.4 (28)

Følgende histogram, normal qq-plot og box-plot er af residualerne:



Hvad kan med rette vurderes udfra disse med bogens definition af outliers?

1 ☐ At det er helt klart, at fordelingen af residualerne er venstre-skæv.

2 ☐ At residualerne ser normalfordelte ud, uden nogen outliers.

3 ☐ At residualerne ser normalfordelte ud, dog med en enkelt outlier.

4 ☐ At det er helt klart at fordelingen af residualerne er højre-skæv.

5 ☐ At residualerne ikke følger en normalfordeling.

Fortsæt på side 20

Opgave XI

Følgende stikprøve er sorteret:

$$10, 25, 25, 36, 37, 41, 54, 64, 68, 83$$

Spørgsmål XI.1 (29)

Hvad er medianen af stikprøven?

1 ☐ 37

2 ☐ 38

3 ☐ 39

4 ☐ 40

5 ☐ 41

Spørgsmål XI.2 (30)

Hvad er stikprøvevariansen?

1 ☐ $\mathrm{V}(x) = 22.60$

2 ☐ $\mathrm{V}(x) = 510.7$

3 ☐ $\mathrm{V}(x) = 1521$

4 ☐ $\mathrm{V}(x) = 1962$

5 ☐ $\mathrm{V}(x) = 2052$

SÆTTET ER SLUT. God sommer!