

Skriftlig prøve: 15. december 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 12 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.**

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

<b>Opgave</b>	VI.1	VI.2	VII.1	VII.2	VII.3	VII.4	VII.5	VII.6	VIII.1	VIII.2
<b>Spørgsmål</b>	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
<b>Svar</b>										

<b>Opgave</b>	VIII.3	IX.1	IX.2	X.1	X.2	XI.1	XI.2	XII.1	XII.2	XII.3
<b>Spørgsmål</b>	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
<b>Svar</b>										

Eksamenssættet består af 21 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

En biodynamisk gård ønsker at nedbryde sine biomasserester til biobrændstof, der skal bruges til vedvarende energiproduktion. I et forsøg anvendte landmændene 10 liters reaktionsbeholdere til at vurdere effektiviteten af biomasseomdannelsen. Varierende mængder af en enzymatisk cocktail blev tilsat hver af beholderne, og blandingerne blev efterladt til tre dages reaktionstid. Derefter blev mængderne af produceret biovæske bestemt.

En simpel lineær regressionsmodel på formen  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  blev opstillet, med henblik på at undersøge forholdet mellem tilsat enzymmængde (**enzyme**, i ml) og biovæske udbytte (**liquid**, i dl). R-outputtet fra model blev fitted ses nedenfor:

```
##
## Call:
## lm(formula = liquid ~ enzyme)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.8103 -4.3885 -0.0775  4.3672  9.2489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.325      2.291    4.070 0.000653 ***
## enzyme         1.956      0.196    9.982 5.42e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.438 on 19 degrees of freedom
## Multiple R-squared:  0.8398, Adjusted R-squared:  0.8314
## F-statistic: 99.64 on 1 and 19 DF,  p-value: 5.419e-09
```

### Spørgsmål I.1 (1)

Baseret på R-udskriften ovenfor, hvad er stikprøvestørrelsen  $n$ ?

- 1  20
- 2  19
- 3  21

4  1

5  Stikprøvestørrelsen kan ikke bestemmes ud fra dette R-output.

### Spørgsmål I.2 (2)

I forsøget var den gennemsnitlige mængde enzymatisk cocktail, som blev anvendt i en reaktionsbeholder,  $\bar{x} = 10$  ml. Beregn det gennemsnitlige biovæske udbytte,  $\bar{y}$ .

1   $\bar{y} = 28.9$  dl

2   $\bar{y} = 9.3$  dl

3   $\bar{y} = 2.0$  dl

4   $\bar{y} = 19.6$  dl

5   $\bar{y} = 24.1$  dl

### Spørgsmål I.3 (3)

Hvilket udsagn nedenfor repræsenterer ikke en nødvendig antagelse for en simpel lineær regressionsmodel?

1  Fejlene  $\varepsilon_i$  er uafhængige.

2  Fejlene  $\varepsilon_i$  er identisk fordelte.

3  De afhængige variable  $Y_i$  er identisk fordelte.

4  De afhængige variable  $Y_i$  er uafhængige.

5  De afhængige variable  $Y_i$  og fejlene  $\varepsilon_i$  har samme varians.

Fortsæt på side 4

## Opgave II

I forbindelse med eksamen i et introducerende statistikkursus ønskes det undersøgt om studerende, som har været indskrevet på uddannelsen i ét år, klarer sig anderledes end studerende der har været indskrevet i to år. Scoren til eksamen regnes ud som et tal mellem -30 og 150, idet der:

- er i alt 30 spørgsmål,
- gives -1 point for et forkert svar,
- gives 5 point for et korrekt svar,
- kun kan afgives ét svar til hvert spørgsmål.

Der er udtaget to tilfældige stikprøver af scoren: En for studerende der er på deres første år ( $x$ ), og en for studerende der er på deres andet år ( $y$ ).

Stikprøverne er hver på 50 observationer, og deres gennemsnit er henholdsvis  $\bar{x} = 84.0$  og  $\bar{y} = 86.6$ . Følgende simulationer og beregninger laves i R:

```
k <- 10000

simxsamples <- replicate(k, sample(x, replace = TRUE))
simysamples <- replicate(k, sample(y, replace = TRUE))
simmeandifs <- apply(simxsamples, 2, mean) - apply(simysamples, 2, mean)

quantile(simmeandifs, c(0.05, 0.95))

##      5%      95%
## -15.12   9.87

quantile(simmeandifs, c(0.025, 0.975))

##   2.5%  97.5%
## -17.26  12.42

quantile(simmeandifs, c(0.005, 0.995))

##   0.5%  99.5%
## -22.04  17.32
```

### Spørgsmål II.1 (4)

Nulhypotesen

$$H_0 : \mu_X = \mu_Y$$

ønskes testet på signifikansniveau  $\alpha = 5\%$ , uden antagelse om scorens fordeling i de to stikprøver. Hvilken af følgende svarmuligheder er korrekt? (Både konklusioner og argument skal være rigtige).

- 1  Nulhypotesen kan ikke afvises, da  $0 \in [-17.26, 12.42]$ . Der kan således ikke påvises en signifikant forskel.
- 2  Nulhypotesen kan ikke afvises, da  $2.6 \in [-15.12, 9.87]$ . Der kan således ikke påvises en signifikant forskel.
- 3  Nulhypotesen afvises, da  $0 \in [-17.26, 12.42]$ . Det kan således påvises, at studerende på første år klarer sig bedre end studerende på andet år.
- 4  Nulhypotesen afvises, da  $0 \notin [-22.04, 17.32]$ . Det kan således påvises, at studerende på første år klarer sig bedre end studerende på andet år.
- 5  Nulhypotesen kan ikke afvises, da  $2.6 \in [-22.04, 17.32]$ . Det kan således påvises, at studerende på andet år klarer sig bedre end studerende på første år.

Fortsæt på side 6

### Opgave III

På et hospital udtages tilfældigt en gruppe af patienter, der får et spørgeskema om hospitalets service. Hver patient udfylder spørgeskemaet både når de indlægges og når de udskrives. I begge spørgeskemaer bedes patienterne angive deres tilfredshed med hospitalets service på en kontinuert skala fra 0 til 1. Ved den efterfølgende analyse af data viser det sig, at begge serier af målinger af servicetilfredsheden kan antages at være normalfordelte. Hvilken af 5 nedenstående tests er mest passende til sammenligning af hvordan servicen vurderes hhv. ved indlæggelsen og ved udskrivningen?

#### Spørgsmål III.1 (5)

- 1  Et  $\chi^2$ -test i en antalstabel
- 2  En en-vejs variansanalyse
- 3  Et  $t$ -test med to uafhængige stikprøver
- 4  Et parret  $t$ -test
- 5  En regressionsanalyse

Fortsæt på side 7

## Opgave IV

### Spørgsmål IV.1 (6)

Antag at den stokastiske variabel  $X \in [0, 1]$  følger en fordeling med tæthedsfunktion  $f(x) = 2x$  for  $x \in [0, 1]$ , og at den dermed har fordelingsfunktionen  $F(x) = x^2$ . Hvilket af følgende stykker R kode simulerer udfald af den stokastiske variabel  $X$ ?

1  `2 * runif(k)`

2  `rchisq(k, df = 1)`

3  `runif(k)^2`

4  `rchisq(k, df = k - 1)`

5  `sqrt(runif(k))`

Fortsæt på side 8

## Opgave V

En plastikproducent ønsker at finde ud af, om der er forskel på kvaliteten af plastik produceret med materialer fra forskellige leverandører (**Supplier**). I fremstillingen måles en bestemt variabel  $Y$  ( $y$ ), som man ved, bestemmer kvaliteten af plastikken. Jo højere værdi af  $Y$  der måles, des højere bliver plastikkens kvalitet. Tabellen nedenfor viser værdier af  $Y$  indsamlet fra uafhængige produktionskørsler med materialer fra 5 forskellige leverandører. Efterfølgende vises output fra den analyse, som firmaets ingeniører har kørt i R.

Leverandør A	Leverandør B	Leverandør C	Leverandør D	Leverandør E
9.9	8.7	8.3	10.4	7.7
10.5	10.3	10.7	12.1	11.7
8.2	6.1	8.7	11.5	10.1
7.7	7.6	9.5	11.2	9.0

```
anova(lm(y ~ Supplier))

## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## Supplier   4   20.9    5.23    2.77  0.066 .
## Residuals 15   28.3    1.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Spørgsmål V.1 (7)

Givet den model, som er anvendt i analysen, hvad er da estimatet for den forventede værdi  $E(Y_{D,i})$  for leverandør D?

- 1  10.3
- 2  10.5
- 3  10.8
- 4  11.3
- 5  11.5

### Spørgsmål V.2 (8)

Hvilken af følgende svarmuligheder beskriver mest nøjagtigt den hypotese, som testes i R outputtet ovenfor?



- 1  Hypotesen  $\alpha_i = 1$  for alle  $i = 1, 2, 3, 4, 5$ , i en model af formen  $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ .
- 2  Hypotesen  $\alpha_i = 0$  for alle  $i = 1, 2, 3, 4, 5$ , i en model af formen  $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ .
- 3  Hypotesen  $\beta_0 = 1$  i en model af formen  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ .
- 4  Hypotesen  $\beta_1 = 0$  i en model af formen  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ .
- 5  Ingen af ovenstående svarmuligheder beskriver den hypotese, som bliver testet.

### Spørgsmål V.3 (9)

Benyt signifikansniveau  $\alpha = 5\%$ . Er der en signifikant forskel på kvaliteten af plastikken produceret med materialerne fra de 5 leverandører? (Både konklusion og argument skal være rigtige).

- 1  Der kan ikke påvises en signifikant forskel på kvaliteten, da  $p$ -værdien er over signifikansniveauet.
- 2  Der kan påvises en signifikant forskel på kvaliteten, da  $p$ -værdien er under signifikansniveauet.
- 3  Der kan ikke påvises en signifikant forskel på kvaliteten, da  $p$ -værdien er under signifikansniveauet.
- 4  Der kan påvises en signifikant forskel på kvaliteten, da  $p$ -værdien er over signifikansniveauet.
- 5  Ingen af ovenstående konklusioner er korrekte.

### Spørgsmål V.4 (10)

Hvor stor en andel af den totale variation kan ikke forklares af modellen?

- 1   $\frac{16.7}{28.3+20.9} = 33.9\%$
- 2   $\frac{20.9}{28.3+20.9} = 42.5\%$
- 3  6.6%
- 4   $\frac{28.3}{28.3+20.9} = 57.5\%$
- 5   $\frac{1.89}{1.89+5.23} = 26.5\%$

Fortsæt på side 10

## Opgave VI

Man har målt værdien af  $X$  for hhv. 5 individer i gruppe 1 og 10 individer i gruppe 2. Det kan antages, at observationerne i begge grupper er normalfordelte, og at alle observationerne er indbyrdes uafhængige. Varianserne i de to grupper tillades at være forskellige. Man ønsker at teste hypotesen om at de to grupper har samme middelværdi (mod alternativet at middelværdierne er forskellige). Testet foretages på et 5% signifikansniveau.

### Spørgsmål VI.1 (11)

Ved en sammenligning med den sædvanlige teststørrelse, hvilken af følgende fraktiler kan nemt bruges til at afgøre, om der er signifikant forskel på de to middelværdier?

- 1  0.025 fraktilen i den relevante  $t$ -fordeling.
- 2  0.05 fraktilen i den relevante  $t$ -fordeling.
- 3  0.95 fraktilen i standard normalfordelingen.
- 4  0.90 fraktilen i standard normalfordelingen.
- 5  0.50 fraktilen i standard normalfordelingen.

### Spørgsmål VI.2 (12)

Stikprøvemiddelværdien og  $s$ -spredningen i gruppe 1 er hhv.  $\bar{x}_1 = 1.99$  og  $s_1 = 0.58$ , mens de tilsvarende størrelser for gruppe 2 er hhv.  $\bar{x}_2 = 1.14$  og  $s_2 = 0.84$ . Det antages, at varianserne i de to grupper er forskellige. Teststørrelsen for ovenstående test bliver da:

- 1   $t_{\text{obs}} = 4.3$
- 2   $t_{\text{obs}} = 1.9$
- 3   $t_{\text{obs}} = 2.3$
- 4   $t_{\text{obs}} = 6.2$
- 5  Ingen af ovenstående muligheder.

Fortsæt på side 11

## Opgave VII

Et forskningsprojekt går ud på at indsamle insekter ved at køre foruddefinerede ture med et net på taget af en bil. Efter turen indsendes de indsamlede insekter til universitetet, hvor de tælles.

### Spørgsmål VII.1 (13)

Hvilken af følgende fordelinger er antageligvis bedst til at beskrive antallet af insekter i et net?

- 1  En eksponentialfordeling
- 2  En binomialfordeling
- 3  En normalfordeling
- 4  En hypergeometrisk fordeling
- 5  En Poissonfordeling

### Spørgsmål VII.2 (14)

Man planlægger nu at køre i alt fire ture på den samme vejstrækning og antager, at variansen af antallet af insekter,  $\sigma^2$ , er den samme på hver af de fire ture. Desuden antager man, at udfaldene fra de fire ture er uafhængige. Hvad er variansen af det samlede antal indfangede insekter på de fire ture?

- 1   $16\sigma^2$
- 2   $4\sigma^2$
- 3   $\sigma^2/4$
- 4   $4\sigma$
- 5   $\sigma^2/2$

Man gennemfører nu de fire planlagte ture og opdeler de indfangede insekter i to typer: små og store insekter. Resultatet af optællingen ses i antalstabellen herunder.

	Tur 1	Tur 2	Tur 3	Tur 4	Samlet
Små insekter	178	242	126	87	633
Store insekter	26	59	30	8	123
Samlet	204	301	156	95	756

### Spørgsmål VII.3 (15)

Hvis man ser på det samlede resultat (alle fire ture samlet), hvilket af følgende er så et 95% konfidensinterval for andelen af store insekter?

- 1  [0.14; 0.19]
- 2  [0.81; 0.87]
- 3  [0.16; 0.23]
- 4  [0.09; 0.23]
- 5  [0.81; 0.86]

### Spørgsmål VII.4 (16)

Der er specielle grunde til at undersøge, om andelen af store insekter kan antages at være den samme på tur 1 og tur 2. Hvad bliver  $p$ -værdien og konklusionen på signifikansniveau  $\alpha = 5\%$ , ved et test for om der er forskel på andelen af store insekter på tur 1 og tur 2?

- 1   $P$ -værdien er 0.043, og der kan derfor påvises en forskel.
- 2   $P$ -værdien er 0.03, og der kan derfor påvises en forskel.
- 3   $P$ -værdien er 0.060, og der kan derfor påvises en forskel.
- 4   $P$ -værdien er 0.043, og der kan derfor ikke påvises en forskel.
- 5   $P$ -værdien er 0.060, og der kan derfor ikke påvises en forskel.

I de følgende spørgsmål ser vi på data fra alle fire ture, med henblik på at teste om fordelingen mellem store og små insekter kan antages at være ens på alle ture.

### Spørgsmål VII.5 (17)

For at udføre det statistiske test skal det forventede antal insekter i hver celle udregnes under nulhypotesen. Hvad er det forventede antal store insekter på tur 3?

- 1  130.6
- 2  4.9
- 3  105.5
- 4  25.4

5  32.1

**Spørgsmål VII.6 (18)**

Den sædvanlige teststørrelse for om der er forskel mellem fordelingen af antallet af insekter på de fire ture er udregnet til 9.6127. Hvad bliver  $p$ -værdien og den tilsvarende konklusion på signifikansniveau  $\alpha = 5\%$ ?

- 1   $P$ -værdien bliver 0.022, så der kan ikke påvises en forskel.
- 2   $P$ -værdien bliver 0.087, så der er forskel.
- 3   $P$ -værdien bliver 0.087, så der kan ikke påvises en forskel.
- 4   $P$ -værdien bliver 0.022, så der er forskel.
- 5   $P$ -værdien bliver 0.045, så der er forskel.

Fortsæt på side 14

## Opgave VIII

Mængden af vaskemiddel, der er nødvendigt til vask af vasketøj, afhænger typisk af flere faktorer. I denne sammenhæng skal forholdet mellem vaskeeffektivitet (**efficiency**), vandhårdhed (**hardness**) og den mængde vaskemiddel, der benyttes (**detergent**) undersøges ved anvendelse af følgende multiple lineære regressionsmodel:

$$efficiency_i = \beta_0 + \beta_1 \cdot hardness_i + \beta_2 \cdot detergent_i + \varepsilon_i,$$

hvor  $\varepsilon_i$ 'erne er uafhængige og  $N(0, \sigma^2)$ -fordelte. R-output fra modellen er vist nedenfor:

```
##
## Call:
## lm(formula = efficiency ~ hardness + detergent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9022 -1.4491 -0.5854  1.4225  5.3286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.5892     3.6590  -0.434   0.6695
## hardness     -2.1981     0.8958  -2.454   0.0252 *
## detergent     3.0239     0.4961   6.095 1.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.496 on 17 degrees of freedom
## Multiple R-squared:  0.7322, Adjusted R-squared:  0.7006
## F-statistic: 23.23 on 2 and 17 DF,  p-value: 1.371e-05
```

### Spørgsmål VIII.1 (19)

Se på R-outputtet ovenfor. Hvilket af de følgende udsagn er korrekt, givet et signifikansniveau på  $\alpha = 1\%$ ?

- 1  Vandets hårdhed har tilsyneladende en signifikant effekt på vaskeeffektiviteten, men mængden af vaskemiddel, der benyttes, har ikke.
- 2  Effekten af vandhårdhed på vaskeeffektivitet er ikke signifikant, idet  $p$ -værdien er større end 0.01.
- 3  Både vandhårdhed og mængden af vaskemiddel, der benyttes, er signifikante, idet  $p$ -værdierne er mindre end 0.05.
- 4  Hverken vandhårdhed eller mængden af vaskemiddel, der benyttes, synes at være signifikant, idet  $p$ -værdierne er mindre end 0.05.

5  Modellens intercept (skæring) er signifikant, idet  $p$ -værdien på 0.6695 er større end 0.01.

### Spørgsmål VIII.2 (20)

Se på samme R-output ovenfor. Hvilken effekt har en stigning på to enheder vaskemiddel på den forventede vaskeeffektivitet? Antag at vandhårdhed er konstant.

- 1  Den forventede vaskeeffektivitet stiger med 3.02 enheder.
- 2  Den forventede vaskeeffektivitet falder med 2.20 enheder.
- 3  Den forventede vaskeeffektivitet falder med 4.40 enheder.
- 4  Den forventede vaskeeffektivitet stiger med 6.05 enheder.
- 5  Den forventede vaskeeffektivitet forbliver konstant.

### Spørgsmål VIII.3 (21)

Giv et estimat af variansen  $\sigma^2$  ud fra ovenstående R-output.

- 1   $\hat{\sigma}^2 = 23.23$
- 2   $\hat{\sigma}^2 = 0.7006$
- 3   $\hat{\sigma}^2 = 2.496$
- 4   $\hat{\sigma}^2 = 0.7322$
- 5   $\hat{\sigma}^2 = 6.230$

Fortsæt på side 16

## Opgave IX

Man har målt temperaturen i et køleskab kl. 12 på tilfældigt valgte dage i løbet af juli måned. Man fik følgende observationer (målt i grader celsius), som er indlæst i R i vektoren  $x$ :

```
x <- c(6.5, 5.7, 1.2, 0.2, 7.0, 3.3)
```

Det antages, at observationerne følger en normalfordeling, og at de er uafhængige.

### Spørgsmål IX.1 (22)

Beregn den sædvanlige teststørrelse for et test af hypotesen om at middelværdien er 3.0 grader.

1   $t_{\text{obs}} = 0.84$

2   $t_{\text{obs}} = 0.20$

3   $t_{\text{obs}} = 2.41$

4   $t_{\text{obs}} = 3.01$

5   $t_{\text{obs}} = 1.99$

### Spørgsmål IX.2 (23)

Bestem et 90% konfidensinterval for variansen af køleskabstemperaturen.

1  [2.9, 51.5]

2  [3.2, 49.2]

3  [3.7, 35.7]

4  [3.9, 33.7]

5  [4.1, 31.7]

Fortsæt på side 17



## Opgave X

Brændstofforbruget for to forskellige traktorer blev undersøgt ved 8 forskellige arbejdsopgaver, og man fik følgende målinger (i liter per hektar). Brændstofforbruget kan antages normalfordelt for hver traktor:

	Opg. 1	Opg. 2	Opg. 3	Opg. 4	Opg. 5	Opg. 6	Opg. 7	Opg. 8
Traktor A	10.8	8.2	8.7	12	6.2	11.2	8.6	5.5
Traktor B	8.4	8.1	9.4	12.9	10.1	10.4	10.2	11.8

### Spørgsmål X.1 (24)

Hvilken af følgende metoder er bedst egnet til at analysere, om der er forskel på de to traktors brændstofforbrug til de udførte opgaver?

- 1  Test i en multipel lineær regressionsmodel.
- 2   $\chi^2$ -test i en antalstabel.
- 3  Ikke-parret  $t$ -test med to stikprøver.
- 4  En-vejs variansanalyse
- 5  Parret  $t$ -test med to stikprøver.

### Spørgsmål X.2 (25)

Median brændstofforbruget (udtrykt ved liter per hektar) for traktor A er:

- 1  8.7
- 2  12
- 3  8.9
- 4  8.65
- 5  12

Fortsæt på side 18

## Opgave XI

Lad  $X \sim N(0, \sigma^2)$  og definer den stokastiske variabel  $Y$  ved  $Y = e^X$ .

### Spørgsmål XI.1 (26)

Hvad er  $P(Y > 1)$ ?

1  0.84

2  0.5

3  0.025

4  0.16

5  0.95

### Spørgsmål XI.2 (27)

Hvad er variansen af  $Y$ ?

1   $e^{\sigma^2/2}$

2   $e^{2+\sigma^2}(e^{1/2} - 1)$

3   $e^{\sigma^2}(e^{\sigma^2} - 1)$

4   $e^{2+\sigma^2}(e^{\sigma^2} - 1)$

5   $e^{\sigma^2}$

Fortsæt på side 19

## Opgave XII

En fabrikant vil gerne undersøge kvaliteten af deres produktionsanlæg. Man har taget en tilfældig stikprøve fra produktionsanlægget med 20 målinger af den tid der går ind imellem at der produceres et fejlbehæftet element. Værdierne er i timer og indlæst i R med følgende kode:

```
x <- c(39.5, 59.7, 42.1, 13, 3.6, 10.9, 61.6, 1, 17.8, 5,  
      24.3, 21, 4.2, 21.1, 78.9, 11.1, 6.6, 0.3, 9.2, 10.4)
```

### Spørgsmål XII.1 (28)

Benyt bogens definition af stikprøve fraktiler til at bestemme stikprøvens IQR (*“Inter Quartile Range”*)

- 1  IQR = 69.6
- 2  IQR = 26.1
- 3  IQR = 58.35
- 4  IQR = 6.25
- 5  IQR = 16.05

### Spørgsmål XII.2 (29)

Man har besluttet, at man vil stoppe anlægget og reparere det, hvis tiden mellem fejlene bliver for lav. Man vil undlade at antage hvilken fordeling tiden mellem fejlene har, og derfor lave et ikke-parametrisk 95% bootstrap konfidensinterval for medianen. Hvilken af følgende R-koder beregner dette interval korrekt?

- 1  `simsamples <- replicate(10000, sample(x, replace = TRUE))`  
`quantile(apply(simsamples, 2, mean), c(0.05, 0.95))`
- 2  `simsamples <- replicate(10000, sample(x, replace = FALSE))`  
`quantile(apply(simsamples, 2, mean), c(0.025, 0.975))`
- 3  `simsamples <- replicate(10000, sample(x, replace = FALSE))`  
`quantile(apply(simsamples, 2, median), c(0.05, 0.95))`
- 4  `simsamples <- replicate(10000, sample(x, replace = TRUE))`  
`quantile(apply(simsamples, 2, median), c(0.025, 0.975))`
- 5  `simsamples <- replicate(10000, sample(x, replace = TRUE))`  
`quantile(apply(simsamples, 2, median), c(0.005, 0.995))`

### Spørgsmål XII.3 (30)

Efter en reparation af anlægget tages en ny stikprøve, og den indlæses i R med nedenstående kode:

```
y <- c(15.3, 28.2, 53.3, 42, 28.5, 45.3, 40.3, 32.3, 81.1, 29.3,  
      82.9, 38.7, 131.5, 24.7, 5.7, 104.3, 30, 31.8, 46.9, 34.9)
```

Derefter udføres følgende simulationer og beregninger:

```
simXsamples <- replicate(10000, rexp(length(x), 1/mean(x)))  
simYsamples <- replicate(10000, rexp(length(y), 1/mean(y)))  
simDiff <- apply(simXsamples, 2, median) - apply(simYsamples, 2, median)  
  
quantile(simDiff, c(0.005,0.995))  
  
##          0.5%          99.5%  
## -50.595024    7.893082  
  
quantile(simDiff, c(0.025,0.975))  
  
##          2.5%          97.5%  
## -42.009475    2.646692  
  
quantile(simDiff, c(0.05,0.95))  
  
##          5%          95%  
## -37.50105759  -0.01677407
```

Hvilken af følgende konklusioner er korrekt på baggrund af R-outputtet i dette spørgsmål?

- 1  På  $\alpha = 1\%$  signifikansniveau konstateres der en signifikant forskel i median, når der ikke laves antagelser om tidernes fordelinger.
- 2  På  $\alpha = 5\%$  signifikansniveau konstateres der ikke en signifikant forskel i median under antagelse af, at tiderne i begge stikprøver er eksponentialfordelte.
- 3  På  $\alpha = 10\%$  signifikansniveau konstateres der ikke en signifikant forskel i middelværdi under antagelse af, at tiderne i begge stikprøver er eksponentialfordelte.
- 4  På  $\alpha = 10\%$  signifikansniveau konstateres der en signifikant forskel i middelværdi under den antagelse, at tiderne i begge stikprøver er normalfordelte.
- 5  På  $\alpha = 1\%$  signifikansniveau konstateres der en signifikant forskel i middelværdi uden antagelse om tidernes fordelinger.

Fortsæt på side 21

Eksamenssættet er slut. God juleferie!