

Skriftlig prøve: 17. December 2022

Kursus navn og nr.: **Introduktion til Statistik (02402)**

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 10 opgaver. For at besvare spørgsmålene skal du udfylde “multiple choice” siderne på eksamen.dtu.dk.

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 online. Skemaet her er KUN et nød-alternativ til dette. Husk at angive dit studienummer, hvis du afleverer på papir.**

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

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

<b>Opgave</b>	VIII.2	VIII.3	IX.1	IX.2	IX.3	IX.4	X.1	X.2	X.3	X.4
<b>Spørgsmål</b>	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
<b>Svar</b>										

Eksamenssættet består af 24 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. Husk også, at der kan forekomme små afvigelser mellem resultatet af bogens formler og tilsvarende indbyggede funktioner i R.

### Opgave I

Lad  $X$  og  $Y$  være uafhængige stokastiske variable, hvor  $X$  har middelværdi 2 og varians 2, mens  $Y$  har middelværdi -1 og varians 3.

#### Spørgsmål I.1 (1)

Hvad er middelværdien af  $2X + Y$ ?

- 1  0
- 2  2
- 3  3
- 4  11
- 5  Vi har ikke tilstrækkeligt information til at kunne bestemme middelværdien af  $2X + Y$ .

Fortsæt på side 3

## Opgave II

Målinger af serumkolesterol (mg/100 ml),  $x$  og arteriel calciumaflejring (mg/100g tørvægt af væv),  $y$ , blev foretaget på tolv dyr. Dataene blev læst ind i R:

```
y <- c(59, 52, 42, 59, 24, 24, 40, 32, 63, 55, 34, 24)
x <- c(298, 303, 270, 287, 236, 245, 265, 233, 286, 290, 264, 239)
```

Betragt følgende simple lineære regressionsmodel.

$$Y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \text{ hvor } \varepsilon_i \sim N(0, \sigma^2)$$

### Spørgsmål II.1 (2)

Beregn variationskoefficienten ( $R^2$ ) og vælg det rigtige svar nedenfor:

- 1  0.9129
- 2  0.8334%
- 3  0.8334
- 4  91.29%
- 5  0.8168%

### Spørgsmål II.2 (3)

Følgende kodelinje er blevet kørt i R.

```
fit <- lm(y~x)
```

Hvilken af følgende kommandoer kan bruges som en del af modelvalideringen, dvs. til at kontrollere om normalitetsantagelserne er opfyldt?

- 1  `qqnorm(fit$fitted.values)`  
`qqline(fit$fitted.values)`

2  `qqnorm(fit$residuals)`  
`qqline(fit$residuals)`

3  `qqnorm(y)`  
`qqline(y)`

4  `qqnorm(residuals)`  
`qqline(residuals)`

5  `qqnorm(lm$residuals)`  
`qqline(lm$residuals)`

### Spørgsmål II.3 (4)

Modelresuméet for en simpel lineær regressionsmodel er vist nedenfor.

```
##  
## Call:  
## lm(formula = y ~ x)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -9.2249 -3.4900 -0.8876  2.1968 10.9510   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -102.3218    20.5319  -4.984 0.000551 ***   
## x              0.5398     0.0763   7.074 3.4e-05 ***   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 6.358 on 10 degrees of freedom  
## Multiple R-squared:  0.8334, Adjusted R-squared:  0.8168   
## F-statistic: 50.04 on 1 and 10 DF,  p-value: 3.401e-05
```

Hvilket af følgende udtryk beregner 95% konfidensintervallet for kolesterollædningen ( $\hat{\beta}_1$ )?

1   $0.5398 \pm 1.9600 \cdot 0.0763$

$$2 \square 0.5398 \pm 2.1788 \cdot 0.0763$$

$$3 \square 0.5398 \pm 2.2281 \cdot 0.0763$$

$$4 \square -102.3218 \pm 1.9600 \cdot 20.5319$$

$$5 \square -102.3218 \pm 2.2281 \cdot 20.5319$$

Fortsæt på side 6

### Opgave III

En person overvejer at købe en elbil. For at træffe et velinformeret valg finder han rækkevidden for en fuldt opladet bil ( $km$ ) og batteristørrelse ( $kWh$ ) for forskellige bilmodeller, som angivet af bilfabrikanterne.

I første omgang overvejer den potentielle bilejer elbilernes effektivitet, det vil sige rækkevidde pr.  $kWh$ . Så han eksekverede følgende R-kode (hvor `range1` er rækkevidden, og `batteri` er batteristørrelsen angivet af bilproducenten):

```
t.test(range1 / battery)

##
## One Sample t-test
##
## data:  range1/battery
## t = 45.117, df = 34, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  6.170481 6.752586
## sample estimates:
## mean of x
##  6.461533
```

#### Spørgsmål III.1 (5)

I første omgang ønsker den potentielle bilejer at teste hypotesen

$$H_0 : \text{Middel effektiviteten er } 6 \text{ km/kWh}$$

Hvad er konklusionen ved brug af signifikansniveau  $\alpha = 0,05$  (alle dele af svaret skal være korrekt)?

- 1  Effektiviteten er signifikant forskellig fra  $6km/kWh$ , da  $p$ -værdien fra outputtet ovenfor er mindre end  $2,2 \cdot 10^{-16}$
- 2  Effektiviteten er ikke statistisk forskellig fra  $6km/kWh$  som  $6 < 6,17$ .
- 3  Effektiviteten er mindre end  $6km/kWh$  da  $6 < 6,17$ .
- 4  Effektiviteten er større end  $6km/kWh$  da  $6 < 6,17$ .
- 5  Effektiviteten er ikke signifikant forskellig fra 6, da  $p$ -værdien er mindre end  $2,2 \cdot 10^{-16}$

#### Spørgsmål III.2 (6)

Baseret på analysen ovenfor, hvad er et 99% konfidensinterval for elbilers effektivitet?

- 1  [5.67, 7.25]
- 2  [5.93, 6.99]
- 3  [6.07, 6.85]
- 4  [6.11, 6.81]
- 5  [6.17, 6.75]

Den potentielle bilejer beslutter sig for at beregne et konfidensinterval for log effektivitet, som en hjælp til analysen gives følgende R-kode med output

```
mean(log(range1 / battery))  
## [1] 1.857769  
  
var(log(range1 / battery))  
## [1] 0.01643549
```

### Spørgsmål III.3 (7)

Hvad er 95% konfidensintervallet for log-effektivitet for elbiler?

- 1  [1.66, 2.06]
- 2  [1.71, 2.00]
- 3  [1.75, 1.96]
- 4  [1.81, 1.90]
- 5  [1.85, 1.86]

Et bilmagasin lavede en uafhængig test af de samme bilmodeller, og den potentielle bilejer vil nu sammenligne den effektivitet, som fabrikanterne angiver, med den som bilmagasinet har fundet.

### Spørgsmål III.4 (8)

I den følgende R-kode angiver `range2` rækkevidden af et fuldt batteri som fundet af bilmagasinet. Hvilken af følgende kodestykker tester, om der er en signifikant forskel mellem effektiviteten givet af bilfabrikanterne og effektiviteten fundet af bilmagasinet?

1  `t.test(log(range1), log(range2), mu = 1, paired = TRUE)`

2  `t.test(log(range1 / battery), log(range2 / battery), mu = 1)`

3  `t.test(log(range1), log(range2))`

4  `t.test(log(range1), log(range2), paired = TRUE)`

5  `t.test(range1 / battery, range2 / battery, mu = 1)`

### Spørgsmål III.5 (9)

Hvis standardafvigelsen for elbilers effektivitet antages at være  $0.8 \text{ km}/(\text{kWh})$ , hvor mange biler skal så testes for at få en fejlmargen (Margin of error) på 0.1?

1  16

2  61

3  157

4  246

5  492

Fortsæt på side 9



## Opgave IV

I en æbleplantage blev der udført et forsøg, hvor forskellige dyrkningsbetingelser blev afprøvet. Træer blev inddelt i 4 grupper, som hver blev dyrket under forskellige forhold. Efter vækstsæsonen blev æblerne plukket og vejet for hvert træ. De målte værdier er vist i kg i følgende tabel opdelt for hver gruppe:

Forhold A	Forhold B	Forhold C	Forhold D
14.9	14.3	14.6	14.0
15.9	16.1	12.7	15.1
16.2	14.9	13.6	13.3
15.9	16.4	13.7	17.3
15.9	15.7		12.9
	14.9		13.1
	15.7		

For at undersøge, om dyrkningsforholdene førte til signifikant forskellige middelvægte af æblerne, blev data analyseret ved hjælp af en ANOVA. Resultaterne er opgjort med følgende ANOVA tabel:

```
anova(lm(Vaegt ~ Forhold))

## Analysis of Variance Table
##
## Response: Vaegt
##           Df Sum Sq Mean Sq F value Pr(>F)
## Forhold    3  14.163    4.721  4.1852 0.02061 *
## Residuals  18  20.305    1.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Spørgsmål IV.1 (10)

Hvad er den totale variation  $SST$  for disse data?

- 1  1.128
- 2  5.85
- 3  14.2
- 4  34.5
- 5  40.3

### Spørgsmål IV.2 (11)

Hvad er det korrekte udsagn baseret på resultaterne fra ANOVA-tabellen (både konklusion og argument skal være korrekt)?

- 1  På signifikansniveau 1% er der påvist en signifikant forskel i vægt som følge af dyrkningsforholdene, da  $1.128 < 4.721$ .
- 2  På signifikansniveau 5% er der ikke påvist en signifikant forskel i vægt som følge af dyrkningsforholdene, da  $4.1852 < 5$ .
- 3  På signifikansniveau 5% er der ikke påvist en signifikant forskel i vægt som følge af dyrkningsforholdene, da  $4.721 > 1.128$ .
- 4  På signifikansniveau 5% er der påvist en signifikant forskel i vægt som følge af dyrkningsforholdene, da  $0.02061 < 0.05$ .
- 5  Ingen af ovenstående udsagn er korrekte.

Fortsæt på side 11

## Opgave V

På DTU er 35% af nye studerende kvinder. En gruppe på seks nye studerende blev tilfældigt udvalgt til en fokusgruppe.

### Spørgsmål V.1 (12)

Hvad er sandsynligheden for at mindst to kvinder udvælges?

- 1  0.243
- 2  0.319
- 3  0.328
- 4  0.647
- 5  0.681

### Spørgsmål V.2 (13)

Hvis vi lader den stokastiske variabel  $X$  angive antallet af kvinder i fokusgruppen, hvad er da variansen af  $X$ ?

- 1  0.23
- 2  0.74
- 3  1.17
- 4  1.37
- 5  2.10

### Spørgsmål V.3 (14)

En af de studerende tager bussen når han/hun skal til DTU. Antag at busser ankommer tilfældigt til busstoppestedet med gennemsnitligt én bus hvert femtende minut. Hvad er sandsynligheden for at den studerende må vente mindst 20 minutter på bussen?

- 1  1.8%
- 2  26.4%
- 3  35.1%

4  52.8%

5  73.7%

**Spørgsmål V.4 (15)**

Det betragtes som en uheldig dag hvis den tid, man venter på bussen, er over 90%-fraktilen af ventetidsfordelingen. Hvor længe skal man vente før at ens dag bliver uheldig?

1  13.5 min

2  26.9 min

3  34.5 min

4  46.1 min

5  135 min

Fortsæt på side 13

## Opgave VI

### Spørgsmål VI.1 (16)

Vi ønsker at simulere 50 tilfældige udfald fra en uniform fordeling, hvor 0 og 100 definerer grænsen af mulige udfald.

1  `runif(50, 0, 100)`

2  `dnorm(50, 0, 100)`

3  `dunif(0, 100, 50)`

4  `runif(0, 100, 50)`

5  `rnorm(50, 0, 100)`

Fortsæt på side 14

## Opgave VII

Vi er interesserede i at studere det systoliske blodtryk  $y$  i forhold til vægt  $x_1$  (lb (pund)) og alder  $x_2$  (år) hos en gruppe mænd med omtrent samme højde.

### Spørgsmål VII.1 (17)

En multipel lineær regressionsmodel af følgende form er blevet etableret.

$$Y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \varepsilon_i \text{ hvor } \varepsilon_i \sim N(0, \sigma^2)$$

Modelresuméet er givet nedenfor.

```
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.6447 -2.0191 -0.0607  2.1331  6.0856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.48354    26.36399   0.967 0.356537
## x1           0.62049     0.13513   4.592 0.000992 ***
## x2           0.04803     0.12948   0.371 ?
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.286 on 10 degrees of freedom
## Multiple R-squared:  0.7957, Adjusted R-squared:  0.7549
## F-statistic: 19.48 on 2 and 10 DF,  p-value: 0.0003555
```

Beregn den udeladte  $p$ -værdi for hypotesen  $H_0 : \beta_2 = 0$ . Hvilket af følgende svar er korrekt?

- 1  0.0500
- 2  0.3592
- 3  0.3596
- 4  0.7184
- 5  0.7192

### Spørgsmål VII.2 (18)

Se på modelresuméet fra spørgsmålet ovenfor. Hvor mange observationer ( $n$ ) blev målt i dette datasæt?

- 1   $n = 9$
- 2   $n = 10$
- 3   $n = 11$
- 4   $n = 12$
- 5   $n = 13$

### Spørgsmål VII.3 (19)

Se modelresuméet ovenfor. Hvilket af følgende udsagn er sandt ved brug af signifikansniveau  $\alpha = 0.05$  (både konklusion og argument skal være korrekte)?

- 1   $p$ -værdien for vægt ( $x_1$ ) er mindre end 0.05, derfor er der en signifikant sammenhæng mellem blodtryk og vægt.
- 2   $p$ -værdien for vægt ( $x_1$ ) er mindre end 0.05, derfor er der ikke en signifikant sammenhæng mellem blodtryk og vægt.
- 3   $t$ -teststørrelsen for vægt ( $x_1$ ) er større end  $t_{crit} = 1.96$ , derfor er der ikke en signifikant sammenhæng mellem blodtryk og vægt .
- 4   $t$ -teststørrelsen for vægt ( $x_1$ ) er større end  $t_{crit} = 1.96$ , derfor er der en signifikant sammenhæng mellem blodtryk og vægt.
- 5   $t$ -teststørrelsen for vægt ( $x_1$ ) er større end  $t_{crit} = 0.05$ , derfor er der ikke en signifikant sammenhæng mellem blodtryk og vægt.

Fortsæt på side 16

## Opgave VIII

En fiskeopdrætforskningsgruppe gennemførte et forsøg for at finde ud af, om der er en signifikant forskel mellem fem udvalgte fiskefoder (behandlinger), samt på placeringerne af opdrættet. Fiskene blev opdrættet tre forskellige placeringer, hvor de på hvert sted blev inddelt i grupper, som hver fik et af de fem forskellige fiskefoder. Vægten af hver gruppe blev målt efter en periode, her præsenteret i kg:

	Placering 1	Placering 2	Placering 3
Behandling 1	232.8	225.3	226.1
Behandling 2	201.6	214.9	205.3
Behandling 3	189.6	193.3	180.8
Behandling 4	184.5	209.3	179.0
Behandling 5	270.1	244.2	207.1

En tovejs ANOVA blev anvendt, og følgende resultat blev indsamlet (bemærk, at nogle værdier er blevet erstattet af et 'X'):

```
anova(lm(Vaegt ~ Placering + Behandling))  
  
## Analysis of Variance Table  
##  
## Response: y  
##           Df Sum Sq Mean Sq F value    Pr(>F)  
## Placering   2  959.1   479.54   2.1576      X  
## Behandling  4 6329.2  1582.31   7.1192      X  
## Residuals   8 1778.1   222.26  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Spørgsmål VIII.1 (20)

På signifikansniveau  $\alpha = 5\%$ , hvilke konklusioner kan drages (begge konklusioner og argumenter skal være korrekte)?

- 1  Der er signifikant effekt af både Placering og Behandling, da de relevante  $p$ -værdier er henholdsvis 1.55% og 1.67%.
- 2  Der er ingen signifikant effekt af hverken Placering eller Behandling, da de relevante  $p$ -værdier er henholdsvis 15.5% og 16.7%.
- 3  Der er signifikant effekt af Placering, men ingen signifikant effekt af Behandling, da de relevante  $p$ -værdier er henholdsvis 1.55% og 16.7%.
- 4  Der er signifikant effekt af både Placering og Behandling, da de relevante  $p$ -værdier er henholdsvis 1.78% og 0.95%.



- 5  Der er ingen signifikant effekt af Placering, men der er signifikant effekt af Behandling, da de relevante  $p$ -værdier er henholdsvis 17.8% og 0.95%.

### Spørgsmål VIII.2 (21)

Den anvendte model er

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

hvor  $Y_{ij}$  repræsenterer vægten af den  $(i, j)$ 'te gruppe,  $i$  angiver Placering og  $j$  angiver Behandling.

Hvilket af følgende udsagn er ikke korrekt om modellen og ovenstående analyse?

- 1  Der er i alt femten  $Y_{ij}$  og de er stokastiske variable.
- 2  Afvigelserne er  $\varepsilon_{ij}$  og de antages i.i.d. for alle  $i$  og  $j$ .
- 3  De testede hypoteser i analysen er  $H_0 : \alpha_i = 0$  og  $H_0 : \beta_j = 0$  for henholdsvis alle  $i$  og  $j$ .
- 4  Modellens prædiktioner er  $\mu + \alpha_i + \beta_j + \varepsilon_{ij}$ .
- 5   $\sigma^2$  er estimeret med  $MSE$ .

### Spørgsmål VIII.3 (22)

Hvis testen for signifikant effekt af behandling i to-vejs ANOVAen udføres på signifikansniveau 5%, hvilket signifikansniveau skal da bruges til Bonferroni-korrigeret post-hoc sammenligninger, når alle de mulige behandlingssammenligninger foretages?

- 1  0.2%
- 2  0.3125%
- 3  0.5%
- 4  1.875%
- 5  2.5%

Fortsæt på side 18

## Opgave IX

Mængden af gluten i havremel er vigtig, hvis man vil sælge melet som glutenfrit. I et eksperiment målt mængden af gluten i en stikprøve af havremel. Havremellet blev opblandet i en speciel mikser (kaldet A) for at gøre glutenindholdet så homogent som muligt. 10 små prøver udtoges fra opblandingen, og glutenindholdet målt i hver prøve med ELISA tests. Resultaterne er gemt i `glutenA`. Måleenheden er ppm (parts per million/milliontedele).

### Spørgsmål IX.1 (23)

Hvilket af følgende stykker R kode udregner et 95% konfidensinterval for standardafvigelsen af målt glutenindhold ved ikke-parametrisk bootstrap?

1 

```
simsamples <- replicate(10000, sample(glutenA, replace = FALSE))
simmeans <- apply(simsamples, 2, mean)
quantile(simmeans, c(0.025, 0.975))
```

2 

```
simsamples <- replicate(10000, sample(glutenA, replace = FALSE))
simmeans <- apply(simsamples, 2, sd)
quantile(simmeans, c(0.025, 0.975))
```

3 

```
simsamples <- replicate(10000, sample(glutenA, replace = TRUE))
simmeans <- apply(simsamples, 2, sd)
quantile(simmeans, c(0.05, 0.95))
```

4 

```
simsamples <- replicate(10000, sample(glutenA, replace = TRUE))
simmeans <- apply(simsamples, 2, sd)
quantile(simmeans, c(0.025, 0.975))
```

5 

```
simsamples <- replicate(10000, sample(glutenA, replace = FALSE))
simmeans <- apply(simsamples, 2, sd)
quantile(simmeans, c(0.05, 0.95))
```

### Spørgsmål IX.2 (24)

Vi vil nu antage at det målte glutenindhold følger en normalfordeling. Hvilket af følgende stykker R kode udregner et 95% konfidensinterval for standardafvigelsen af målt glutenindhold ved parametrisk bootstrap?

```
1  simsamples <- replicate(10000, rnorm(10,mean(glutenA),sd(glutenA)))
simsds <- apply(simsamples, 2, mean)
quantile(simsds, c(0.025, 0.975))
```

```
2  simsamples <- replicate(10000, rnorm(10,mean(glutenA),sd(glutenA)))
simsds <- apply(simsamples, 2, sd)
quantile(simsds, c(0.025, 0.975))
```

```
3  simsamples <- replicate(10000, rnorm(10,mean(glutenA),var(glutenA)))
simsds <- apply(simsamples, 2, sd)
quantile(simsds, c(0.025, 0.975))
```

```
4  simsamples <- replicate(10000, sample(glutenA,replace = FALSE))
simsds <- apply(simsamples, 2, sd)
quantile(simsds, c(0.025, 0.975))
```

```
5  simsamples <- replicate(10000, sample(glutenA,replace = TRUE))
simsds <- apply(simsamples, 2, sd)
quantile(simsds, c(0.025, 0.975))
```

Fortsæt på side 20

### Spørgsmål IX.3 (25)

Vi betragter nu en ny portion mel fra samme forsyning af havremel. Dette mel blev opblandet med en anden type mikser, B. 10 små prøver udtoges fra opblandingen, og glutenindholdet målte i hver prøve med ELISA tests. Resultaterne er gemt i `glutenB`.

Vi ønsker nu at sammenligne de to miksere ud fra hvor homogent glutenindholdet i melet er efter opblandingen. Vi sammenligner derfor forskellen i standardafvigelserne for glutenindholdet i melprøver opblandet med hhv. mikser A og B.

Hvilket af følgende stykker R-kode udregner et 95% parametrisk konfidensinterval for forskellen i standardafvigelserne mellem de to sæt målinger?

1 

```
simAsamples <- replicate(10000, rnorm(10,mean(glutenA),sd(glutenA)))
simBsamples <- replicate(10000, rnorm(10,mean(glutenB),sd(glutenB)))
simDifsds <- apply(simAsamples,2,sd) - apply(simBsamples,2,sd)
quantile(simDifsds, c(0.025, 0.975))
```

2 

```
simsamples <- replicate(10000, rnorm(10,mean(glutenA)-mean(glutenB),
sd(glutenA)-sd(glutenB)))
simDifsds <- apply(simsamples,2,sd)
quantile(simDifsds, c(0.025, 0.975))
```

3 

```
simAsamples <- replicate(10000, rnorm(10,mean(glutenA),sd(glutenA)))
simBsamples <- replicate(10000, rnorm(10,mean(glutenB),sd(glutenB)))
simDifmeans <- apply(simAsamples,2,mean) - apply(simBsamples,2,mean)
quantile(simDifmeans, c(0.025, 0.975))
```

4 

```
simAsamples <- replicate(10000, rnorm(10,mean(glutenA),var(glutenA)))
simBsamples <- replicate(10000, rnorm(10,mean(glutenB),var(glutenB)))
simDifsds <- apply(simAsamples,2,sd) - apply(simBsamples,2,sd)
quantile(simDifsds, c(0.025, 0.975))
```

5 

```
simsamples <- replicate(10000, rnorm(10,mean(glutenA-glutenB),
sd(glutenA-glutenB)))
simDifsds <- apply(simsamples,2,sd)
quantile(simDifs, c(0.025, 0.975))
```

### Spørgsmål IX.4 (26)

Som resultat af forrige spørgsmål fik man konfidensintervallet  $[-4.43; 1.53]$ . Hvilket af følgende udsagn er korrekt (både konklusion og argument skal være korrekt)?

- 1  Eftersom konfidensintervallet indeholder 0, kan vi konkludere at standardafvigelserne er signifikant forskellige. Vores eksperiment indikerer derfor at mikser B er bedre end mikser A.
- 2  Eftersom konfidensintervallet indeholder 0, kan det ikke afvises at standardafvigelserne er ens. Vi konkluderer derfor at det gennemsnitlige glutenindhold er ens i de to melprøver.
- 3  Eftersom konfidensintervallet indeholder 0, kan vi konkludere at det gennemsnitlige glutenindhold er ens i de to melprøver.
- 4  Eftersom konfidensintervallet indeholder 0, kan vi konkludere at standardafvigelserne er signifikant forskellige. Vores eksperiment indikerer derfor at mikser A er bedre end mikser B.
- 5  Eftersom konfidensintervallet indeholder 0, kan det ikke afvises at standardafvigelserne er ens. Vi kan derfor ikke konkludere at den ene mikser er bedre end den anden.

Fortsæt på side 22

## Opgave X

Et palæstinensisk universitet gennemførte en undersøgelse for at undersøge årsagerne til at holde webcam slukket under online læring. 1268 elever svarede. Følgende antalstabel viser svarene på udsagnet ”Lærere beder os ikke om at tænde for webcam”.

Køn	Enig	Uenig	Ligegyldig	Række i alt
Mand	339	42	35	416
Kvinde	746	61	45	852
Kolonne i alt	1085	103	80	1268

### Spørgsmål X.1 (27)

Hvad er 95% konfidensintervallet for andelen af ”kvindelige” studerende baseret på dataene givet ovenfor?

- 1  [0.285, 0.340]
- 2  [0.0.302, 0.353]
- 3  [0.561, 0.615]
- 4  [0.646, 0.698]
- 5  [0.659, 0.715]

### Spørgsmål X.2 (28)

Hvad er 95% konfidensintervallet for forskellen i andelen af mandlige og kvindelige elevers svar, der vælger ”enig” ( $p_{\text{enig}} \text{ når kvinde} - p_{\text{enig}} \text{ når mand}$ )?

- 1  [0.003, 0.194]
- 2  [0.013, 0.113]
- 3  [0.017, 0.104]
- 4  [0.284, 0.357]
- 5  [0.336, 0.414]

### Spørgsmål X.3 (29)

Hvad er det forventede antal elever med køn ”mand” og ”enig” under nulhypotesen om uafhængighed mellem køn og overenskomst?

- 1  111.22
- 2  227.78
- 3  233.08
- 4  266.20
- 5  355.96

**Spørgsmål X.4 (30)**

Nulhypotesen om uafhængighed mellem køn og overensstemmelse med udsagnet skal testes med  $\chi^2$ -test.

Hvad er den relevante kritiske værdi at bruge til at teste, om der er en signifikant kønsforskel i elevernes svar på udsagnet (med signifikansniveau  $\alpha = 0,05$ )?

- 1  3.841
- 2  5.991
- 3  7.815
- 4  9.210
- 5  12.59

Fortsæt på side 24

SÆTTET ER SLUT. God jul!