Skriftlig prøve: 3. Juni 2024

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 14 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.

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

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

Opgave	XI.2	XI.3	XII.1	XII.2	XIII.1	XIII.2	XIII.3	XIV.1	XIV.2	XIV.3
Spørgsmål	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Svar										

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

Tag følgende hypotesetest i betragtning:

$H_0 : \mu = 18$
$H_1 : \mu \neq 18$

En stikprøve med $n = 48$ observationer gav en stikprøvegennemsnit $\bar{x} = 17$ og en prøvestandardafvigelse $s = 4.5$. Det antages, at stikprøven er indsamlet fra normalfordelt population.

Spørgsmål I.1 (1)

Hvad er værdien af teststatistikken (t_{obs}) og hvad er de kritiske værdier med signifikansniveau $\alpha = 0.05$ (begge svar skal være rigtige)?

1 ☐ $t_{\text{obs}} = -1.54$. $t_{0.025} = -2.012$ og $t_{0.975} = 2.012$

2 ☐ $t_{\text{obs}} = -2.01$. $t_{0.025} = -2.685$ og $t_{0.975} = 2.685$

3 ☐ $t_{\text{obs}} = -1.96$. $t_{0.025} = -1.678$ og $t_{0.975} = 1.678$

4 ☐ $t_{\text{obs}} = -1.89$. $t_{0.025} = -2.012$ og $t_{0.975} = 2.012$

5 ☐ Kan ikke udregnes uden mere information.

Spørgsmål I.2 (2)

Fra en anden stikprøve med $n = 45$ observationer, er værdien af teststatistikken (t_{obs}) -1.74. Beregn p-værdien og drag en konklusion, der bruger signifikansniveau $\alpha = 0.05$ (begge skal være korrekte).

1 ☐ p-værdien er 0.0889. Vi afviser ikke nulhypotesen.

2 ☐ p-værdien er 0.0805. Vi afviser nulhypotesen.

3 ☐ p-værdien er 0.0560. Vi afviser ikke nulhypotesen.

4 ☐ p-værdien er 0.1339. Vi afviser ikke nulhypotesen.

5 ☐ Vi kan ikke drage en konklusion.

Fortsæt på side 3

Opgave II

Der indsamles en tilfældig stikprøve på $n = 30$ observationer, stikprøvegennemsnittet estimeres til at være $\bar{x} = 1.01$ og stikprøvens standardafvigelse til $s = 0.09$.

Spørgsmål II.1 (3)

Hvad er 95% konfidensintervallet for standardafvigelsen?

1 ☐ [0.06, 0.13]

2 ☐ [0.24, 0.40]

3 ☐ [0.09, 0.13]

4 ☐ [0.05, 0.09]

5 ☐ [0.072, 0.121]

Fortsæt på side 5

Opgave III

Udfaldet af et eksperiment beskrives med den tilfældige variabel X, hvor X har den følgende fordelingsfunktion:

x	0	1	2	3	4
$f(x)$	0.17	0.22	0.28	0	0.33

Spørgsmål III.1 (4)

Hvad er sandsynligheden $P(X < 3)$

1 ☐ Dette kan ikke afgøres ud fra den givne information

2 ☐ 0.28

3 ☐ 0.33

4 ☐ 0.40

5 ☐ 0.67

Spørgsmål III.2 (5)

Gennemsnittet af X er 2.10. Hvad er variansen af X?

1 ☐ 1.49

2 ☐ 2.10

3 ☐ 2.21

4 ☐ 4

5 ☐ 6.62

Fortsæt på side 6

Opgave IV

Der er gennemført en undersøgelse for at vurdere koncentrationen (vi ser bort fra enheden) af en kemisk forbindelse i jorden fire forskellige steder. De målte koncentrationer fra undersøgelsen er angivet i nedenstående tabel:

Lokalitet 1	Lokalitet 2	Lokalitet 3	Lokalitet 4
253.7	261.1	257.9	244.1
241.2	244.2	263.5	244.9
255.8	250.5	258.6	243.9
249.3	264.9		247.1
	259.3		

Den samlede gennemsnitlige koncentration fundet på tværs af alle fire lokaliteter er 252.5, og gennemsnits koncentrationerne for hver lokalitet er angivet nedenfor:

Lokalitet	1	2	3	4
Gennemsnits koncentration	250.0	256.0	260.0	245.0

Vi estimerer nu en envejs ANOVA-model på formen

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

hvor fejlene, ε_{ij}, er uafhængige og følger normalfordelinger med middelværdi 0 og varians σ^2. Den overordnede middelværdi er μ, og effekten af lokalitet i er α_i. ANOVA tabellen for modellen viser, at den samlede sum af kvadrater afvigelser, SST, er 915.92 og summen af kvadrater afvigelser for lokaliteter, $SS(\mathit{Lokalitet})$, er 480.00.

Spørgsmål IV.1 (6)

Hvad er estimatet af effekten af Lokalitet 1?

1 ☐ -2.5

2 ☐ 2.5

3 ☐ 6.0

4 ☐ 250.0

5 ☐ 252.5

Spørgsmål IV.2 (7)

Hvad er estimatet af standardafvigelsen for residualerne?

1 ☐ $\hat{\sigma} = 6.027$

2 ☐ $\hat{\sigma} = 6.325$

3 ☐ $\hat{\sigma} = 7.814$

4 ☐ $\hat{\sigma} = 20.879$

5 ☐ $\hat{\sigma} = 36.327$

Fortsæt på side 8

Opgave V

12 observationer af perfluoroctansulfonsyre (PFOS) i seks forskellige koncentrationer blev analyseret ved hjælp af en ny eksperimentel testmetode SPETT.

PFOS-koncentrationer er målt i mg/kg. Dataene var læst i R af:

```
# PFOS concentrations
x <- c(0, 0, 2, 2, 4, 4, 6, 6, 8, 8, 12, 12)
# SPETT values
y <- c(16, 116, 1170, 841, 2287, 2012, 2653, 3333, 4270, 3999, 5750, 5407)
```

Den lineære regressions model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ hvor } \varepsilon_i \sim N(0, \sigma^2) \text{ og i.i.d. for } i = 1, \ldots, 12,$$

blev sat op, hvor Y_i er SPETT-værdien og x_i PFOS-koncentrationen af den i'th observation.

Bemærk, at i den resterende del af opgaven er normalfordelings og i.i.d. antagelserne af residualerne implicit (derfor ikke skrevet med modellen).

Spørgsmål V.1 (8)

Hvad er estimatet af β_1?

1 ☐ 160.7

2 ☐ 467.6

3 ☐ 511.0

4 ☐ 1020.7

5 ☐ 259.0

Spørgsmål V.2 (9)

Vi ønsker at teste hypotesen $H_0 : \beta_0 = 0$, da dette ville indikere om den forventede SPETT-værdi er nul for en PFOS-koncentration på nul. Vi bruger et signifikansniveau på $\alpha = 0.05$.

Hvilket af følgende udsagn er korrekt?

1 ☐ Vi afviser ikke nulhypotesen, da $|\hat{\beta}_0| > 1.96$, hvor 1.96 er 95% fraktilen i en normalfordeling.

2 ☐ Vi afviser ikke nulhypotesen, da p-værdien er 0.23.

3 ☐ Vi afviser ikke nulhypotesen, da p-værdien er 0.0027.

4 ☐ Vi afviser nulhypotesen, da p-værdien er 0.0027.

5 ☐ Vi afviser nulhypotesen, da p-værdien er 0.23.

Spørgsmål V.3 (10)

Forskere vil gerne kende usikkerheden af SPETT-værdien for en ny observation med PFOS-koncentration 7 mg/kg. Hvad er 95% prædiktionsintervallet ved denne koncentration?

1 ☐ [2829, 4039]

2 ☐ [3253, 3615]

3 ☐ [487, 522]

4 ☐ [1898, 3099]

5 ☐ [2653, 4270]

Fortsæt på side 10

Opgave VI

På en papirfabrik foregår produktionen i partier. I et eksperiment for kvalitetskontrol, blev partier tilfældigt udvalgt, dvs. kvaliteten var uafhængig mellem de valgte partier. I eksperimentet blev 20 ud af 85 partier fundet til ikke at leve op til kvalitetskravet.

Spørgsmål VI.1 (11)

Hvad er konfidensintervallet beregnet på et signifikansniveau $\alpha = 10\%$ for andelen som ikke lever op til kvalitetskravene (bemærk, bogens formel giver det præcis rigtige svar)?

1 ☐ $[0.084, 0.387]$

2 ☐ $[0.128, 0.342]$

3 ☐ $[0.153, 0.342]$

4 ☐ $[0.160, 0.311]$

5 ☐ $[0.176, 0.294]$

Spørgsmål VI.2 (12)

Der er planlagt et nyt kvalitetstjek. Fabrikken ønsker et 95% konfidensinterval med en forventet bredde på 0.1. Som et gæt på populationsandelen, anvendes den observerede andel fra det første kvalitetscheck (dvs. 20 ud af 85 partier).

Hvad er minimumsantallet af partier, der skal udtages for at den nye kontrol opnår denne nøjagtighed?

1 ☐ $n = 18$

2 ☐ $n = 70$

3 ☐ $n = 195$

4 ☐ $n = 277$

5 ☐ $n = 385$

Fortsæt på side 11

Opgave VII

Hos en dansk virksomhed er der lavet et studie for at undersøge, om online træningsmoduler til at lære om bæredygtighed forbedrer elevers viden derom. Eleverne deltog i en quiz før og efter at have gennemført aktiviteterne i online træningsmodulet. De to quizzer omtales som pre-test og post-test. Seks elevers scores i pre-testen (betegnet `pre`) og post-testen (betegnet `post`) er vist i tabellen nedenfor.

Elev	1	2	3	4	5	6
pre	41	46	35	49	33	42
post	42	47	43	55	28	49

Følgende kode køres nu i R for at indlæse data:

```
pre <- c(41,46,35,49,33,42)
post <- c(42,47,43,55,28,49)
```

Spørgsmål VII.1 (13)

Forudsat at observationerne stammer fra normalfordelte populationer og forbedringen måles ved en forskel i middelværdien af populationer, hvilken af følgende koder vil resultere i den ønskede hypotesetest efter indlæsning af data?

1 ☐ `t.test(mean(pre), mean(post))`

2 ☐ `t.test(post, pre, var.equal=TRUE)`

3 ☐ `t.test(pre, post)`

4 ☐ `t.test(pre-post)`

5 ☐ `t.test(sd(pre), sd(post))`

Fortsæt på side 12

Opgave VIII

Det antages at der er 2.5 skovbrande på en gennemsnitlig varm sommerdag, samt at antallet af skovbrande følger en poissonfordeling.

Spørgsmål VIII.1 (14)

Hvad er sandsynligheden for at der er mindst fem skovbrande på en varm sommerdag?

1 ☐ 0.004

2 ☐ 0.067

3 ☐ 0.109

4 ☐ 0.175

5 ☐ 0.762

Spørgsmål VIII.2 (15)

Lad en tilfældig variabel X angive antallet af skovbrande i en periode a syv på hinanden følgende varme sommerdage. Hvilken fordeling følger X?

1 ☐ En poissonfordeling med middelværdi 17.5.

2 ☐ En binomialfordling med $n = 49$ og $p \approx 0.36$.

3 ☐ En normalfordeling med middelværdi 17.5 of standardafvigelse 7.

4 ☐ En exponentiel fordeling med rate ≈ 0.57.

5 ☐ Ingen af de ovenstående

Spørgsmål VIII.3 (16)

Det estimeres at 78% af skovbrande kunne have været undgået, hvis man havde overholdt nogle simple forholdsregler. Hvis brandvæsenet en dag rapporterer fem skovbrande, hvad er så sandsynligheden for at alle disse skovbrande kunne have været undgået, hvis man havde overholdt disse forholdsregler?

1 ☐ 0.0005

2 ☐ 0.156

3 ☐ 0.175

4 ☐ 0.289

5 ☐ 0.711

Fortsæt på side 13

Opgave IX

30 sommerfugle er blevet indfanget og man har målt længden på vingerne. Diverse statistiske oplysninger om stikprøven (som er gemt i `length_cm`) vises herunder.

```
round(summary(length_cm),2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.06    3.72    3.93    4.03    4.32    5.37

round(sd(length_cm),2)

## [1] 0.49
```

Spørgsmål IX.1 (17)

Udfør en parametrisk bootstrap og simulér 10000 stikprøver under den antagelse at stikprøven stammer fra en normalfordelt population. Hvilken af de følgende er det korrekte 99% konfidensinterval for medianen af sommerfuglenes vingelængde?

HINT: Hvis du følger "standard" kodningen fra bogen, vil du få det korrekte resultat med `set.seed(2023)`.

1 ☐ [3.46 cm, 3.93 cm]

2 ☐ [3.38 cm, 4.00 cm]

3 ☐ [2.81 cm, 5.48 cm]

4 ☐ [3.76 cm, 4.31 cm]

5 ☐ [3.90 cm, 4.17 cm]

Spørgsmål IX.2 (18)

Den følgende kode er kørt:

```
median_cm <- 3.93
mean_cm <- 4.03
sd_cm <- 0.49
n <- length(length_cm)
k <- 10000
```

Hvilken af de følgende svarmuligheder benytter et ikke-parametrisk bootstrap til at estimere et 95% konfidensinterval for den første kvartil (25. percentil) af vingelængde målt i cm?

1 ☐
```r
fun <- function(x) quantile(x, 0.25, type=2)
sim_samples <- replicate(k, sample(length_cm, n, replace = TRUE))
sim_stats <- apply(sim_samples, 2, fun)
quantile(sim_stats, c(0.005, 0.995))
```

2 ☐
```r
fun <- function(x) quantile(x, 0.25, type=2)
sim_samples <- replicate(k, rnorm(n, mean_cm, sd_cm))
sim_stats <- apply(sim_samples, 2, fun)
quantile(sim_stats, c(0.025, 0.975))
```

3 ☐
```r
fun <- function(x) quantile(x, 0.75, type=2)
sim_samples <- replicate(k, sample(length_cm, n, replace = TRUE))
sim_stats <- apply(sim_samples, 2, fun)
quantile(sim_stats, c(0.01, 0.99))
```

4 ☐
```r
fun <- function(x) quantile(x, 0.75, type=2)
sim_samples <- replicate(k, rnorm(n, mean_cm, sd_cm))
sim_stats <- apply(sim_samples, 2, fun)
quantile(sim_stats, c(0.01, 0.99))
```

5 ☐
```r
fun <- function(x) quantile(x, 0.25, type=2)
sim_samples <- replicate(k, sample(length_cm, n, replace = TRUE))
sim_stats <- apply(sim_samples, 2, fun)
quantile(sim_stats, c(0.025, 0.975))
```

Fortsæt på side 16

Opgave X

En undersøgelse involverer to populationer: en population af systemanalytikere, som bruger en aktuel teknologi, og en population af systemanalytikere, der bruger en ny softwarepakke. Antagelsen om normalfordelt population er opfyldt for begge populationer. Med hensyn til tiden der kræves for at fuldføre et systemdesign projekt befolkningsmidler er som følger:

μ_1 er middel af projektgennemførelsestid for systemanalytikere ved hjælp af den nuværende teknologi.
μ_2 er middel af projektgennemførelsestid for systemanalytikere ved hjælp af den nye softwarepakke.

Forskeren med ansvar for det nye softwarepakke vil gerne vide, om den nye softwarepakke vil give en kortere middel projektgennemførelsestid. Hypotesen som skal testes er

H_0: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 \neq 0$

Følgende opsummerende statistikker er givet:

- Stikprøvestørrelse: $n_1 = 12$, $n_2 = 12$
- Stikprøvegennemsnit: $\bar{x}_1 = 325$ timer, $\bar{x}_2 = 286$ timer
- Stikprøvestandardafvigelse: $s_1 = 40$, $s_2 = 44$

Spørgsmål X.1 (19)

Hvad er værdien af t-teststatistikken (t_{obs}) og antal frihedsgrader (begge svar skal være rigtige)?

1 ☐ $t_{\text{obs}} = 1.80$ og $df = 16.8$

2 ☐ $t_{\text{obs}} = 2.27$ og $df = 21.8$

3 ☐ $t_{\text{obs}} = 2.91$ og $df = 20.8$

4 ☐ $t_{\text{obs}} = 3.12$ og $df = 19.8$

5 ☐ $t_{\text{obs}} = 2.02$ og $df = 21.8$

Fortsæt på side 17

Opgave XI

I et laboratorie ønsker man at fortynde en lagerbeholdningen af opløsning A med koncentration C_A til en opløsning B med koncentration C_B. Den følgende regel anvendes:

$$C_B = \frac{C_A V_A}{V_B}$$

Her angiver V_A og V_B volumenerne af hhv. opløsning A og opløsning B. C_A, C_B, V_A og V_B er alle tilfældige variable.

Spørgsmål XI.1 (20)

Hvilken af de følgende udtryk kan bruges til at approximere standard afvigelsen af opløsning B's koncentration (σ_{C_B}) via den ikke-lineære regel for udbredning af fejl?

1 ☐ $\sigma_{C_B} = \sqrt{\frac{\left(\frac{\partial C_B}{\partial C_A}\sigma_{C_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_A}\sigma_{V_A}\right)^2}{\left(\frac{\partial C_B}{\partial V_B}\sigma_{V_B}\right)^2}}$

2 ☐ $\sigma_{C_B} = \left(\frac{\partial C_B}{\partial C_A}\sigma_{C_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_A}\sigma_{V_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_B}\sigma_{V_B}\right)^2$

3 ☐ $\sigma_{C_B} = \sqrt{\left(\frac{\partial C_B}{\partial C_A}\sigma_{C_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_A}\sigma_{V_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_B}\sigma_{V_B}\right)^2}$

4 ☐ $\sigma_{C_B} = \frac{\left(\frac{\partial C_B}{\partial C_A}\sigma_{C_A}\right)^2 + \left(\frac{\partial C_B}{\partial V_A}\sigma_{V_A}\right)^2}{\left(\frac{\partial C_B}{\partial V_B}\sigma_{V_B}\right)^2}$

5 ☐ $\sigma_{C_B} = \left(\frac{\partial C_B}{\partial C_A}\sigma_{C_A}\right) + \left(\frac{\partial C_B}{\partial V_A}\sigma_{V_A}\right) + \left(\frac{\partial C_B}{\partial V_B}\sigma_{V_B}\right)$

Spørgsmål XI.2 (21)

Lad X_i være en tilfældig variabel. Den følgende kode er kørt i R for at trække k realisationer af X_i fra en distribution.

```
x <- rnorm(k)^2 + rnorm(k)^2
```

Hvilken af de følgende udsagn er korrekt?

1 ☐ X_i følger en χ^2-fordeling med 1 frihedsgrad, og derfor er $q_{0.25} = 0.102$.

2 ☐ X_i følger en standard normalfordeling med middelværdi 0 og varians 1, og derfor er $q_{0.25} = -0.674$.

3 ☐ X_i følger en χ^2-fordeling med 2 frihedsgrader, og derfor er $q_{0.25} = 0.575$.

4 ☐ X_i følger en χ^2-fordeling med 3 frihedsgrader, og derfor er $q_{0.25} = 1.213$.

5 ☐ X_i følger en normalfordeling med middelværdi 0 og varians 3, og derfor er $q_{0.25} = -1.168$.

Spørgsmål XI.3 (22)

Hvilken af følgende udsagn er IKKE korrekt?

1 ☐ Ikke-parametrisk bootstrap er en re-sampling teknik som bruges til at estimere spredningen af en parameter uden at lave antagelser om den underliggende population.

2 ☐ Ikke-parametrisk bootstrap involverer gentagen stikprøvetagning med tilbagelægning fra en original stikprøve, for at generere mange nye simulerede stikprøver af samme størrelse som den originale stikprøve.

3 ☐ Ikke-parametrisk bootstrap foretrækkes fremfor parametrisk bootstrap i den situation hvor man kender fordelingen i populationen.

4 ☐ Ikke-parametrisk bootstrap kan anvendes til at estimere 95% konfidensinterval for stikprøvegennemsnittet.

5 ☐ Ikke-parametrisk bootstrap kan bruges til at estimere konfidensintervaller for en parameter samt til at teste hypoteser.

Fortsæt på side 19

Opgave XII

En bestemt industriel proces afhænger af pH-værdien og mængden af katalysator i en opløsning. Forholdet mellem output pr. time (kg/t) (`effektivitet`), pH (`ph`) og mængden af anvendt katalysator (`katalysator`) skal undersøges ved hjælp af følgende multipel lineære regressionsmodel

$$effektivitet_i = \beta_0 + \beta_1 \cdot ph_i + \beta_2 \cdot katalysator_i + \varepsilon_i \,,$$

hvor ε_i er uafhængige og $N(0, \sigma^2)$-fordelte. R output fra estimation af modellen med de tilgængelige data er vist nedenfor:

```
##
## Call:
## lm(formula = effektivitet ~ ph + katalysator)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.6360  -2.4232  -0.5899   1.7566  16.8564
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.660     14.652   2.707  0.01496 *
## ph            -4.059      1.532  -2.649  0.01687 *
## katalysator    4.593      1.247   3.683  0.00184 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.34 on 17 degrees of freedom
## Multiple R-squared:  0.5492,Adjusted R-squared:  0.4962
## F-statistic: 10.36 on 2 and 17 DF,  p-value: 0.001145
```

Spørgsmål XII.1 (23)

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

1 ☐ Effekten af katalysator på effektivitet er ikke signifikant, da p-værdien er mindre end 0.01.

2 ☐ Mængden af katalysator har en signifikant effekt på effektivitet, mens pH ikke har.

3 ☐ Både mængden af katalysator og pH er signifikant, da p-værdierne er mindre end 0.05.

4 ☐ Hverken mængden af katalysator eller pH er signifikant, da p-værdierne er mindre end 0.05.

5 ☐ Modellens skæringspunkt med y-aksen er signifikant, da p-værdien på 0.0150 er større end 0.01.

Spørgsmål XII.2 (24)

Se på det samme R-output som ovenfor. Hvilken effekt har en stigning på to enheder katalysator på det forvented output pr. time, under antagelse af at pH-niveauet holdes konstant på 4?

1 ☐ Den forventede produktion stiger med 3.87 kg pr. time.

2 ☐ Den forventede produktion falder med 4.06 kg pr. time.

3 ☐ Den forventede produktion stiger med 4.59 kg pr. time.

4 ☐ Den forventede produktion stiger med 9.19 kg pr. time.

5 ☐ Den forventede produktion forbliver konstant.

Fortsæt på side 21

Opgave XIII

I et forsøg blev fem forskellige kemikalier blandet med jord i lige høje koncentrationer. Den samme type planter blev dyrket ved siden af hinanden i hver af de fem blandinger, således var planterne udsat for de samme vækstbetingelser bortset fra de forskellige kemiske stoffer i jorden.

Da planterne var modnet, blev de høstet, og for hver af de fem forskellige blandede jorde blev der talt hvor mange af planterne, som havde et spor af kemikalierne.

De registrerede antal for hvert af kemikalierne var:

	A	B	C	D	E
Spor	12	14	18	5	15
Ikke spor	38	35	35	42	35

R-koden til indlæsning af data er:

```
tbl <- matrix(c(12, 14, 18, 5, 15,
                38, 35, 35, 42, 35), nrow=2, byrow=TRUE)
```

Spørgsmål XIII.1 (25)

Under nulhypotesen om ingen forskel i andel af planter, der ikke har spor af kemikalier

$$H_0 : p_i = p, i = 1, \ldots, 5$$

Hvad er det forventede antal planter uden spor for kemikalie C?

1 ☐ 12.1

2 ☐ 12.6

3 ☐ 39.4

4 ☐ 50.2

5 ☐ 53.1

Spørgsmål XIII.2 (26)

Hvad er konklusionen på hypotesetesten på signifikansniveau $\alpha = 0.05$ for nulhypotesen, at der ikke er nogen forskel i andel af planter

$$H_0 : p_i = p, i = 1, \ldots, 5$$

med spor af kemikalie (både argument og konklusion skal være korrekte)?

1 ☐ Den relevante p-værdi er 0.041, dog er reglen for gyldighed ikke opfyldt. Derfor kan der ikke drages nogen konklusion.

2 ☐ Den relevante p-værdi er 0.041 og reglen for gyldighed kontrolleres ok. Derfor er en signifikant forskel i optagelsen af fem kemikalier påvist.

3 ☐ Den relevante p-værdi er 0.083, dog er reglen for gyldighed ikke opfyldt. Derfor kan der ikke drages nogen konklusion.

4 ☐ Den relevante p-værdi er 0.083 og reglen for gyldighed kontrolleres ok. Det må derfor accepteres, at der ikke er fundet forskel i optag af de fem kemikalier.

5 ☐ Den relevante p-værdi er 0.021 og reglen for gyldighed kontrolleres ok. Derfor er en signifikant forskel i optagelsen af fem kemikalier påvist.

Spørgsmål XIII.3 (27)

I et andet forsøg blev frø fra 200 forskellige plantetyper udvalgt. Til hver plantetype blev der foretaget en særlig genmodifikation af nogle frø. Det regelmæssige og de modificerede frø blev dyrket parvis ved siden af hinanden og dermed udsat for samme forhold. Efter at planterne var modnet, blev deres salgskvalitet vurderet. De resulterende optællinger var:

	Lav	Mellem	Høj
Lav	25	9	19
Mellem	23	21	23
Høj	28	33	19

hvor de almindelige frøkategorier er rækkerne og det modificerede frø kategorier er kolonnerne.

Det ønskes at udføre en test for uafhængighed af de to variable på signifikansniveau $\alpha = 0.05$. Testens gyldighed er blevet kontrolleret og fundet i orden. Hvad er resultatet af den sædvanlige test (både konklusion og argumenter skal være korrekte)?

1 ☐ Den relevante p-værdi er 0.048, så det accepteres at variablerne er uafhængige.

2 ☐ Den relevante p-værdi er 0.048, således konkluderes det at variablerne ikke er uafhængige.

3 ☐ Den relevante p-værdi er 0.063, således konkluderes det at variablerne ikke er uafhængige.

4 ☐ Den relevante p-værdi er 0.063, så det accepteres at variablerne er uafhængige.

5 ☐ Ingen af ovenstående svar er korrekte.

Fortsæt på side 23

Opgave XIV

Nogle forskere undersøger virkningerne af forskellige typer software og forskellige typer hardware på en computers samlede køretid af et program. Forskerne har anvendt en tovejs ANOVA-model og har observeret følgende ANOVA-tabel, hvor nogle tal er erstattet med bogstaver:

Source	DF	Sum Sq	Mean Sq	Test statistic	p-value
Software	5	$SS(S)$	$MS(S)$	F_1	0.432
Hardware	4	$SS(H)$	$MS(H)$	F_2	0.036
Residual	20	SSE	MSE		

Spørgsmål XIV.1 (28)

Hvad kan udledes om forholdet mellem F_1 og F_2 fra de givne oplysninger i ANOVA-tabellen ovenfor?

1 ☐ $F_1 > F_2$

2 ☐ $0.036 \cdot F_1 = 0.432 \cdot F_2$

3 ☐ $F_1 = F_2$

4 ☐ $F_1 < F_2$

5 ☐ Der kan ikke udledes noget om forholdet mellem F_1 og F_2 med de givne oplysninger.

Spørgsmål XIV.2 (29)

Hvilket af følgende udsagn er korrekt?

1 ☐ Ved et signifikansniveau på 1% finder vi, at både softwaretypen og hardwaretypen har en signifikant effekt på computerprogrammets køretid.

2 ☐ Ved et signifikansniveau på 5% finder vi, at både softwaretypen og hardwaretypen har en signifikant effekt på computerprogrammets køretid.

3 ☐ Ved et signifikansniveau på 10% finder vi, at hverken softwaretypen eller hardwaretypen har en signifikant effekt på computerprogrammets køretid.

4 ☐ Ved et signifikansniveau på 5% finder vi, at softwaretypen har en signifikant effekt på computerprogrammets køretid, mens hardwaretypen ikke har det.

5 ☐ Ved et signifikansniveau på 1% finder vi, at hverken softwaretypen eller hardwaretypen har en signifikant effekt på computerprogrammets køretid.

Spørgsmål XIV.3 (30)

Forskerne ønsker nu at udføre post-hoc analyse og har valgt et signifikansniveau α. Forskerne ønsker at foretage parvise sammenligninger af middel kørselstid for alle de forskellige hardwaretyper, og for at kontrollere type I fejlraten, ønsker forskerne at anvende Bonferroni-korrektion. Hvilket Bonferroni-korrigeret signifikansniveau skal forskerne bruge?

1 ☐ $\alpha/4$

2 ☐ $\alpha/5$

3 ☐ $\alpha/6$

4 ☐ $\alpha/10$

5 ☐ $\alpha/15$

SÆTTET ER SLUT. God sommer!