Skriftlig prøve: 13. december 2016

Kursus navn og nr: **Introduktion til Statistik (02323 og 02402)**

Tilladte hjælpemidler: Alle

Dette sæt er besvaret af

(studienummer)	(underskrift)	(bord nr)

Opgavesættet består af 30 spørgsmål af "multiple choice" typen fordelt på 18 opgaver. Besvarelserne af "multiple choice"spørgsmålene anføres i det i CampusNet uploadede svarark, med numrene på de svarmuligheder, du mener er de korrekte.

Der gives 5 point for et korrekt "multiple choice" svar og -1 for et ukorrekt svar. KUN følgende 5 svarmuligheder er gyldige: 1, 2, 3, 4 eller 5. Hvis et spørgsmål efterlades blankt eller andet type svar angives, tæller det ikke med i besvarelsen. Envidere, hvis mere end et svar angives, hvilket faktisk er teknisk muligt i online-systemet, så tæller det ikke med (dvs. giver "0 point"). Det antal point, der kræves for, at et sæt anses for tilfredstillende besvaret, afgøres endeligt ved censureringen.

Den endelige besvarelse af opgaverne gøres ved at udfylde og online-aflevere svararket via CampusNet. Skemaet her er KUN et nød-alternativ til dette (husk at angive dit studienummer på din besvarelse, hvis du afleverer skemaet).

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

Opgave	VIII.2	IX.1	IX.2	IX.3	IX.4	X.1	XI.1	XI.2	XI.3	XI.4
Spørgsmål	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Svar										

Opgave	XI.5	XII.1	XII.2	XIII.1	XIII.2	XIV.1	XV.1	XVI.1	XVII.1	XVIII.1
Spørgsmål	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Svar										

Sættet består af 25 sider.

Fortsæt på side 2

Multiple choice opgaver: *Der gøres opmærksom på, at ideen med opgaverne er, at der er ét og kun ét rigtigt svar på de enkelte spørgsmål. Endvidere er det ikke givet, at alle de anførte alternative svarmuligheder er meningsfulde.*

Opgave I

Archaeopteryx er en uddød fugleslægt, der danner en af de vigtigste fossile mellemformer mellem krybdyr og fugle. Det er den tidligste kendte og mest primitive urfugl. Antag, at man har data fra 6 fossile fund af Archaeopteryx, og har målt længe af lårbenet (femur) og overarmsbenet (humerus), som vist i nedenstående tabel.

Femur	38	46	56	59	64	74
Humerus	41	50	63	70	71	76

Data kan indlæses i R ved:

```
femur = c(38,46,56,59,64,74)
humerus = c(41,50,63,70,71,76)
```

Spørgsmål I.1 (1)

Arkæologer har længe ment, at der bør være en lineær sammenhæng mellem længde af femur og humerus for uddøde dyr som Archaeopteryx. Hvilken konklusion kommer man frem til ved at analysere ovenstående data når signifikansniveau $\alpha = 0.05$ anvendes?

1 ☐ Der er grund til at antage en lineær sammenhæng, da længde af ben i dyr altid er positivt korreleret.

2 ☐ Der er grund til at antage en lineær sammenhæng, idet p-værdien for det relevante test er 0.0013.

3 ☐ Der er grund til at antage en lineær sammenhæng, idet p-værdien for det relevante test er 0.0780.

4 ☐ Der er ikke grund til at antage en lineær sammenhæng, idet p-værdien for det relevante test er 0.0033.

5 ☐ Der er ikke grund til at antage en lineær sammenhæng, idet p-værdien for det relevante test er 0.0780.

Fortsæt på side 3

Opgave II

Man planlægger et studie hvor man vil undersøge om indtagelse af et naturprodukt har betydning for vægt. I forsøget skal der indgå 10 forsøgspersoner (mænd med nogenlunde ens vægt), og man vil registrere vægtændringen D_i $(i = 1, \ldots, 10)$ efter brug af naturproduktet i 1 måned. Man er interesseret i at teste om middelændringen kan tænkes at være lig nul, dvs. hypotesen $H_0 : \mu_D = 0$ mod alternativet $H_1 : \mu_D \neq 0$. Man beslutter sig for signifikansniveauet $\alpha = 0.05$.

Spørgsmål II.1 (2)

Hvis man antager, at standardafgivelsen for vægtændringen er $\sigma = 1$ kg, hvor stor styrke (eng: power) har man så for at opdage en faktisk vægtændring på mindst 1 kg? (hint: man kan med fordel anvende funktionen `power.t.test` i R.)

1 ☐ 50.0%

2 ☐ 69.3%

3 ☐ 80.3%

4 ☐ 89.7%

5 ☐ 99.3%

Fortsæt på side 4

Opgave III

Det antages, at kolesterolindhold i hønseæg, X, er normalfordelt med middelværdi $\mu = 200$ mg og standardafgivelse $\sigma = 15$ mg, dvs. $X \sim N(200, 15^2)$.

Spørgsmål III.1 (3)

Hvor stor en andel af hønseæg har et kolesterolindhold, der er højere end 205 mg?

1 ☐ $P(X > 205) = 0.631$

2 ☐ $P(X > 205) = 0.369$

3 ☐ $P(X > 205) = 0.491$

4 ☐ $P(X > 205) = 0.394$

5 ☐ $P(X > 205) = 0.605$

Spørgsmål III.2 (4)

Til industrikøkkener sælges kartoner med æg, hvor indholdet, Y, i en karton svarer til indholdet af 100 æg, dvs. det samlede indhold af kolesterol i en karton er $Y = \sum_{i=1}^{100} X_i$. Indholdet i æggene kan antages uafhængige fra hinanden.

Du køber en karton med æg svarende til at købe 100 æg. Hvilken af nedenstående R kommandoer giver sandsynligheden for, at det samlede kolesterolindhold, Y, er højere end 20.5 g (bemærk at 200 mg er 0.2 g)?

1 ☐ `pnorm(q=100*0.205, mean=100*0.200, sd=100*0.015)`

2 ☐ `1-pnorm(q=100*0.200, mean=100*0.205, sd=sqrt(100*0.015*0.015))`

3 ☐ `pnorm(q=100*0.205, mean=100*0.200, sd=100*100*0.015)`

4 ☐ `1-pnorm(q=100*0.205, mean=100*0.200, sd=sqrt(100*0.015*0.015))`

5 ☐ `pnorm(q=100*0.205, mean=100*0.200, sd=sqrt(100*0.015*0.015))`

Fortsæt på side 5

Opgave IV

Vi betragter en binomialfordelt stokastisk variabel Y hvor $n = 100$ og $p = 0.45$.

Spørgsmål IV.1 (5)

Beregn nu $P(Y > 40)$:

1 ☐ 0.183

2 ☐ 0.971

3 ☐ 0.420

4 ☐ 0.817

5 ☐ 0.866

Spørgsmål IV.2 (6)

Vi definerer en ny stokastisk variabel, X, ved $X = k \cdot Y$, hvor konstanten k er givet ved $k = 2$ og Y er binomial fordelt med $n = 100$ og $p = 0.45$. Angiv nu variansen for den stokastiske variabel X:

1 ☐ $\mathrm{Var}(X) = \mathrm{Var}(k \cdot Y) = k + n \cdot p(1-p) = 26.75$

2 ☐ $\mathrm{Var}(X) = \mathrm{Var}(k \cdot Y) = k^2 \cdot n^2 \cdot p^2(1-p)^2 = 49.50^2$

3 ☐ $\mathrm{Var}(X) = \mathrm{Var}(k \cdot Y) = k^2 \cdot n^2 \cdot p(1-p) = 9900$

4 ☐ $\mathrm{Var}(X) = \mathrm{Var}(k \cdot Y) = k^2 \cdot n \cdot p(1-p) = 99.00$

5 ☐ $\mathrm{Var}(X) = \mathrm{Var}(k \cdot Y) = k \cdot n \cdot p(1-p) = 49.50$

Fortsæt på side 6

Opgave V

Vi betragter en eksponentialt fordelt stokastisk variabel X med parameter β, og hvor fordelingsfunktionen er givet ved $P(X \leq x) = 1 - e^{-x/\beta}$, hvor $x > 0$ og og $\beta > 0$. Det oplyses, at middelværdien for X er lig β.

Spørgsmål V.1 (7)

Angiv nu medianen for X:

1 ☐ Medianen for X bliver $0.5 \cdot 2 \cdot \beta$

2 ☐ Medianen for X bliver $0.5^2 \cdot \beta$.

3 ☐ Medianen for X bliver $\log(2) \cdot \beta$ (hvor log er den naturlige logaritme)

4 ☐ Medianen for X bliver $\log(\frac{1}{2}) \cdot \beta^2$ (hvor log er den naturlige logaritme)

5 ☐ Medianen for X bliver $2 \cdot \beta$

Fortsæt på side 7

Opgave VI

En biolog er interesseret i at undersøge effekten af 3 forskellige fodertyper (A, B, C) til opdræt af tigerrejer. Der indkøbes 24 ensartede larver fra et klækkeri til eksperimentet. Hver larve placeres i sin egen beholder, og ved lodtrækning bestemmes tildeling af fodertype således at hver fodertype bliver afprøvet på 8 larver. Larverne vokser sig til tigerrejer, og efter endt forsøg måles vægten af rejerne, Y_{ij} (i gram). Da vægten kan antages normalfordelt, vælger man at analysere data ud fra følgende variansanalysemodel

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

I modellen angiver α_i effekten af fodertype i ($i = 1, 2, 3$). Endelig er μ gennemsnittet og ε_{ij} er modellens afvigelser, der antages normalfordelt med middelværdi 0 og standardafvigelse σ_ε.

En variansanalyse for ovenstående model er givet nedenfor, og det ses, at fodertype er statistisk signifikant.

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
Fodertype  2  44.67 22.3350  8.9221 0.001568 **
Residuals 21  52.57  2.5033
```

Spørgsmål VI.1 (8)

På forhånd var der en interesse i at sammenligne middelværdi for fodertype A og C, der har estimeret middelværdi på hhv. $\hat{\mu}_A = 12.725$ og $\hat{\mu}_C = 15.725$. Angiv nu et 95% konfidensinterval for middelforskellen mellem fodertype A og C:

1 ☐ $-3.000 \pm 2.119\sqrt{52.488^2(\frac{1}{12} + \frac{1}{12})}$

2 ☐ $-3.000 \pm 1.960\sqrt{64.624(\frac{1}{12} + \frac{1}{12})}$

3 ☐ $-3.000 \pm 1.960\sqrt{44.670(\frac{1}{3} + \frac{1}{3})}$

4 ☐ $-3.000 \pm 2.080\sqrt{2.503(\frac{1}{8} + \frac{1}{8})}$

5 ☐ $-3.000 \pm 2.080\sqrt{52.570(\frac{1}{4} + \frac{1}{4})}$

Fortsæt på side 8

Opgave VII

Opgaven er ikke længere del af pensum

Spørgsmål VII.1 (9)

Spørgsmålet er ikke længere del af pensum

Fortsæt på side 9

Opgave VIII

I et forsøg vil man sammenligne slidstyrken af to forskellige slags gummi (A og B), der anvendes som materiale i skosåler. I forsøget indgår 100 skolebørn i alderen 8-10 år. Man udfører forsøget ved at hvert barn får udleveret et par sko, hvor sålen på den ene sko er lavet i materiale A, mens sålen på den anden sko er materiale B. For hvert par sko har man trukket lod om hvorvidt materiale A skal være på skoen til højre eller til venstre. Børnene anvender skoene hver dag i 3 måneder, og efter endt forsøg, måles slid (i mm) på hver sko.

Spørgsmål VIII.1 (10)

Idet man kan antage at det målte slid er kontinuert og normalfordelt for hvert slags gummi, skal man angive hvilken af nedenstående analyser man bør anvende, såfremt man ønsker at teste om materiale A og B er ens mht. slidstyrke:

1 ☐ En test i en antalstabel

2 ☐ En F test for sammenligning af to varianser

3 ☐ En almindelig (ikke-parret) t-test

4 ☐ En parret t-test

5 ☐ En ensidet variansanalyse

Spørgsmål VIII.2 (11)

Det viser sig, at man må acceptere nul-hypotesen om at de to materialer er lige slidstærke. I stedet beregner man for hvert barn i forsøget det gennemsnitlige slid for et par sko. Idet der indgik 50 piger og 50 drenge i forsøget, vil man nu undersøge ved brug af et almindeligt t-test, om drenge og piger slider skoene lige meget, eller der alternativt er forskel på slid for køn (tosidet test).

Idet slidet kan antages normalfordelt indenfor hvert køn med ens varians, fås den sædvanlige teststørrelse til $t_{obs} = 2.23$ med 98 frihedsgrader. Angiv nu p-værdien for testet:

1 ☐ p-værdi bliver 0.014

2 ☐ p-værdi bliver $2 \cdot 0.014$

3 ☐ p-værdi bliver $2 \cdot 0.05$

4 ☐ p-værdi bliver 0.23

5 ☐ p-værdi bliver $1 - 0.23$

Fortsæt på side 10

Opgave IX

Et kursus på en højere læreranstalt bliver udbudt hvert semester, typisk med flere end 300 studerende, der går til eksamen. Eksamensresultaterne, for 280 studerende, der har bestået kurset ved den sidste eksamen, er gengivet i nedenstående tabel. Eksempelvis ses, at 24 studerende fik karakteren 12. Fordelingen af de 280 karakterer indgår i de 4 næste spørgsmål.

Karakter	02	4	7	10	12	I alt
Antal	22	78	84	72	24	280

Data kan indlæses i R ved:

```
karakterer = rep(x=c(2,4,7,10,12), times=c(22,78,84,72,24))
```

Spørgsmål IX.1 (12)

Benyt den centrale grænseværdisætning til at bestemme et 95% konfidensinterval for middelkarakteren baseret på de studerende, der har bestået eksamen (Det er vigtigt i dette spørgsmål, at karakterene opfattes numerisk, fx. svarer 02 til tallet 2 osv.).

1 ☐ [6.51 ; 7.43]

2 ☐ [6.62 ; 7.32]

3 ☐ [4 ; 10]

4 ☐ [5.12 ; 8.67]

5 ☐ [5.99 ; 8.72]

Fortsæt på side 11

Spørgsmål IX.2 (13)

Man ønsker nu at teste om andelen af studerende, der har bestået eksamen med karakteren '7' eller højrere kan antages at være 65%, hvilket har været en målsætning i konstruktionen af karakterskalaen. Vi kalder denne andel p_{7+}. Fra tabellen i forrige spørgsmål ses, at 180 studerende ud af 280 fik karakteren '7' eller højere.

Angiv p-værdien når vi tester $H_0 : p_{7+} = 0.65$ mod $H_1 : p_{7+} \neq 0.65$

1 ☐ 0.8021

2 ☐ $1.745 \cdot 10^{-6}$

3 ☐ $8.725 \cdot 10^{-7}$

4 ☐ 0.5989

5 ☐ 0.4011

Spørgsmål IX.3 (14)

En studerende synes det er interessant at analysere data nærmere, og kører nedenstående kode, idet de 280 karakterer er gemt i vektoren `karakterer`

```
k = 100000
samples = replicate(k, sample(karakterer, replace = TRUE))
simval = apply(samples, 2, sd)
resultater = quantile(simval, c(0.025,0.975))
```

Hvilket numerisk resultat er blevet beregnet i vektoren `resultater`

1 ☐ Et 95% konfidensinterval for standard afvigelsen af karaktererne (ikke-parametrisk bootstrap)

2 ☐ Et 95% konfidensinterval for karakterfordelingen (parametrisk bootstrap)

3 ☐ Et 95% prædiktionsinterval for medianen af karaktererne (ikke-parametrisk bootstrap)

4 ☐ Et 95% prædiktionsinterval for standard error af karaktererne (parametrisk bootstrap)

5 ☐ Et 95% konfidensinterval for 75% fraktilen af karaktererne (parametrisk bootstrap)

Fortsæt på side 12

Spørgsmål IX.4 (15)

Man ønsker nu at undersøge om karakterfordelingen er ens for mænd og kvinder. Karakterfordelingen efter køn er vist i nedenstående tabel.

Karakter	02	4	7	10	12	I alt
Mænd	14	47	59	47	18	185
Kvinder	8	31	25	25	6	95

Angiv nu det forventede antal for mænd med karakteren '7', såfremt karakterfordelingen er ens for mænd og kvinder (dvs. under antagelse af nul-hypotesen).

1 ☐ 55.5

2 ☐ 59

3 ☐ 47.57

4 ☐ 28.5

5 ☐ 42

Fortsæt på side 13

Opgave X

En diskret stokastisk variabel, X, anvendes til at beskrive antal hændelser per tidsinterval. X har tæthedsfunktion på den velkendte form: $P(X = x) = \frac{2^x}{x!} e^{-2}$, for $x \geq 0$.

Spørgsmål X.1 (16)

Hvad er middelværdien af X?

1 ☐ $\frac{1}{2}$

2 ☐ $\log(2)$ (hvor log er den naturlige logaritme)

3 ☐ 2

4 ☐ π

5 ☐ 2^2

Fortsæt på side 14

Opgave XI

I et studie sammenligner man kognetive evner hos 3 grupper af børn. Grupperne består af a) børn med Tourette Syndrom (TS), b) børn med ADHD og c) børn uden nogen af disse diagnoser (Kontrol).

I undersøgelsen blev hvert barn bedt om at løse en sekvens af opgaver på en computer, og man målte den gennemsnitlige reaktionstid, R_i, (i millisekunder) for det enkelte barn. I undersøgelsen indgik 17 børn med TS, 13 med ADHD og 20 Kontrol, dvs. i alt $n = 50$ børn.

Ved analyse af data fra forsøget har man antaget, at reaktionstiden R_i kan antages normalfordelt for hver gruppe med konstant varians, σ_E^2. Med henblik på at undersøge om middelreaktionstiden kan antages ens for de tre grupper (TS, ADHD og Kontrol) er følgende variansanalysetabel beregnet.

```
Analysis of Variance Table

Response: reaktionstid
          Df Sum Sq Mean Sq F value    Pr(>F)
gruppe     A 485848  242924 D          .976e-07 ***
Residuals  B 542563  C
---
```

Det ses dog, at ikke alle tal er oplyst i variansanalysetabellen, men er i stedet givet ved symbolerne A, B, C og D. Disse 4 symboler indgår i løsningen af det næste spørgsmål.

Spørgsmål XI.1 (17)

Såfremt det gælder, at middelreaktionstiden er ens for de tre grupper (TS, ADHD og Kontrol), hvilken fordeling vil den beregnede værdi D i tabellen så følge?

1 ☐ $F(A, A+B)$ dvs. en F-fordeling med frihedsgrader A og $A+B$ fundet i variansanalysetabellen.

2 ☐ $F(C, B)$ dvs. en F-fordeling med frihedsgrader C og B fundet i variansanalysetabellen.

3 ☐ $F(A, B)$ dvs. en F-fordeling med frihedsgrader A og B fundet i variansanalysetabellen.

4 ☐ $F(A, C)$ dvs. en F-fordeling med frihedsgrader A og C fundet i variansanalysetabellen.

5 ☐ $F(B, A)$ dvs. en F-fordeling med frihedsgrader B og A fundet i variansanalysetabellen.

Fortsæt på side 15

Spørgsmål XI.2 (18)

Hvad bliver konklusionen på analysen givet i variansanalysetabellen, såfremt signifikansniveau $\alpha = 0.05$ anvendes?

1 ☐ Man må afvise hypotesen om at middelreaktionstiden for kontrolbørn er som for børn med TS eller ADHD.

2 ☐ Man må afvise hypotesen om at variansen for reaktionstiden for kontrolbørn er som for børn med TS eller ADHD.

3 ☐ Man kan påvise at variansen for reaktionstiden er ens for de tre grupper, idet p-værdien er lig $0.976 \cdot 10^{-7}$.

4 ☐ Man må afvise hypotesen om at middelreaktionstiden er ens for de tre grupper, idet p-værdien er lig $0.976 \cdot 10^{-7}$.

5 ☐ Man kan ikke afvise hypotesen om at middelreaktionstiden er ens for de tre grupper.

Spørgsmål XI.3 (19)

Efterfølgende ville man undersøge hvorvidt børnenes alder $x_{1,i}$ havde indflydelse på reaktionstiden, Y_i. Man udførte derfor et nyt forsøg, hvor i $n = 12$ børn uden diagnose (Kontrol), men med forskellige alder løste sekvensen af opgaver og den gennemsnitlige reaktionstid blev målt.

Man anvender modellen $Y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \varepsilon_i$ for at undersøge sammenhæng mellem alder og reaktionstid, hvor afvigelserne antages i.i.d. og normalfordelt med konstant varians, altså $\varepsilon_i \sim N(0, \sigma^2)$. Man får følgende output for de nye data:

```
Call:
lm(Reaktionstid ~ Alder)

Residuals:
    Min      1Q  Median      3Q     Max
-54.520 -35.522   4.268  27.160  51.949

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   933.15     153.23   6.090 0.000117 ***
Alder         -41.05      15.36  -2.672 0.023400 *
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 37.72 on 10 degrees of freedom
```

```
Multiple R-squared:  0.4166,Adjusted R-squared:  0.3583
F-statistic: 7.141 on 1 and 10 DF,  p-value: 0.0234
```

Ved brug af tallene i output for modellen skal teststørrelsen for hypotesen $H_0 : \beta_1 = 0$ angives

1 ☐ -2.672

2 ☐ 153.23

3 ☐ -41.05

4 ☐ 37.72

5 ☐ 0.4166

Spørgsmål XI.4 (20)

Ved brug af resultat af analysen i forrige spørgsmål, find da estimatet af korrelationen, $\hat{\rho}$, mellem Reaktionstid (Y_i) og Alder ($x_{1,i}$):

1 ☐ $\hat{\rho} = -\sqrt{0.3583}$

2 ☐ $\hat{\rho} = -\sqrt{0.4166}$

3 ☐ $\hat{\rho} = \sqrt{0.3583}$

4 ☐ $\hat{\rho} = 0.4166$

5 ☐ $\hat{\rho} = -\sqrt{0.3583/0.4166}$

Fortsæt på side 17

Spørgsmål XI.5 (21)

En kritik af modellen $Y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \varepsilon_i$ i forrige spørgsmål er, at der ikke tages højde for hvorvidt der svares korrekt eller ej på de enkelte spørgsmål, men at blot reaktionstiden registreres.

Man ønsker derfor en udvidet model på formen: $Y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \varepsilon_i$, hvor $x_{2,i}$ nu er antal korrekte svar i sekvensen af spørgsmål (øvrige variable er som i forrige spørgsmål). Baseret på et nyt forsøg med 12 børn får man resultatet

```
Call:
lm(Reaktionstid ~ Alder + Korrekte)

Residuals:
    Min      1Q  Median      3Q     Max
-39.958 -24.407   6.917  12.897  42.297

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  892.633    123.203   7.245 4.84e-05 ***
Alder        -48.104     15.081  -3.190   0.0110 *
Korrekte       5.310      1.765   3.009   0.0147 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 28.73 on 9 degrees of freedom
Multiple R-squared:  0.5908,Adjusted R-squared:  0.4999
F-statistic: 6.498 on 2 and 9 DF,  p-value: 0.01793
```

Angiv nu estimaterne $\hat{\beta}_1$, $\hat{\beta}_2$ og $\hat{\sigma}^2$:

1 ☐ $\hat{\beta}_1 = -48.104$, $\hat{\beta}_2 = -5.310$ og $\hat{\sigma}^2 = 28.73$

2 ☐ $\hat{\beta}_1 = 892.633$, $\hat{\beta}_2 = -48.104$ og $\hat{\sigma}^2 = 28.73$

3 ☐ $\hat{\beta}_1 = 892.633$, $\hat{\beta}_2 = -48.104$ og $\hat{\sigma}^2 = 5.310$

4 ☐ $\hat{\beta}_1 = -48.104$, $\hat{\beta}_2 = 5.310$ og $\hat{\sigma}^2 = 892.633$

5 ☐ $\hat{\beta}_1 = -48.104$, $\hat{\beta}_2 = 5.310$ og $\hat{\sigma}^2 = 28.73^2$

Fortsæt på side 18

Opgave XII

En ingeniør analyserer en process Y, der kan udtrykkes ved $Y = U/B$. Det kan antages, at U og B er uafhængige. Ingeniøren har 20 sammenhørende målinger af U og B gemt som vektorer i statistikprogrammet R og disse er benævnt `uobs` og `bobs`.

Spørgsmål XII.1 (22)

Ingeniøren vil gerne beregne et 95% konfidensinterval for variansen for Y, dvs. σ_Y^2 ved brug af ikke-parametrisk bootstrap. Hvilket af nedenstående forslag (kode i R) er mest hensigtsmæssigt?

1 ☐
```
samples = replicate(10000,sample(uobs/bobs,replace=FALSE))
results = apply(samples,1, var)
quantile(results, c(0.025, 0.975))
```

2 ☐
```
samples = replicate(10000,sample(uobs/bobs,replace=TRUE))
results = apply(samples,2, var)
quantile(results, c(0.025, 0.975))
```

3 ☐
```
samples = replicate(10000,sample(var(uobs)/var(bobs),replace=TRUE))
results = apply(samples,2, var)
quantile(results, c(0.025, 0.975))
```

4 ☐
```
samples = replicate(10000,sample(uobs/bobs,replace=TRUE))
results = apply(samples,1, var)
quantile(results, c(0.95))
```

5 ☐
```
samples = replicate(10000,sample(uobs/bobs,replace=FALSE))
results = apply(samples,2, var)
quantile(results, c(0.025, 0.975))
```

Fortsæt på side 19

Spørgsmål XII.2 (23)

Vi fortsætter med problemstillingen fra forrige opgave, dvs. vi analyserer en process Y, der kan udtrykkes ved $Y = U/B$.

Hvis vi antager, at $U \sim N(\mu = 35, \sigma^2 = 10^2)$ og $B \sim N(\mu = 50, \sigma^2 = 10^2)$, hvad bliver så sandsynligheden for at Y overstiger 1, dvs. angiv $P(Y > 1)$:

1 ☐ < 0.001

2 ☐ 0.1444

3 ☐ 0.4701

4 ☐ 0.5298

5 ☐ 0.8556

Fortsæt på side 20

Opgave XIII

Et analyseinstitut vil gerne angive et 95% konfidensinterval for den sande andel, p, af forbrugere, der bevidst går efter at købe økologiske fødevarer når de handler. Analyseinstituttet planlægger at stille n forbrugere spørgsmålet: "Går De bevidst efter at købe økologiske fødevarer når De handler?". Svarmulighederne til dette skal være "Ja"eller "Nej".

Spørgsmål XIII.1 (24)

Hvor mange uafhængige forbrugere n skal svare på undersøgelsen for at 95% konfidensinterval for den sande andel, p, højst bliver 0.04 bredt (Hint: Som udgangspunkt for beregningerne kan det antages, at 50% af forbrugerne bevidst går efter at købe økologiske fødevarer når de handler)?

1 ☐ $n = \frac{1.96 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0.01} = 49$

2 ☐ $n = (\frac{1.96 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0.02})^2 = 600.25$ dvs. mindst 601

3 ☐ $n = (\frac{1.96^2 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0.02^2}) = 2401$

4 ☐ $n = (\frac{2 \cdot 1.96 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0.02})^2 = 1250.50$ dvs. mindst 1251

5 ☐ $n = (\frac{1.96 \cdot \sqrt{\frac{1}{2} \cdot \frac{1}{2}}}{0.01})^2 = 9604$

Spørgsmål XIII.2 (25)

Spørgsmålet er ikke længere en del af pensum

Fortsæt på side 21

Opgave XIV

Antag, at antal forsøg til køreprøven (før den er bestået) i en bestemt kommune kan beskrives ved modellen $Y = X + 1$, hvor X er en poissonfordelt stokastisk variabel med middelværdi $\lambda = 0.4$, dvs. $X \sim Pois(\lambda = 0.4)$.

Spørgsmål XIV.1 (26)

Vi betragter nu antal forsøg blandt 100 tilfældigt udvalgte personer, der skal bestå køreprøven. Hvad bliver middelværdi, μ, og varians, σ^2, for det samlede antal forsøg, $\sum_{i=1}^{100} Y_i$ der skal gennemføres?

1 ☐ $\mu = 140$ og $\sigma^2 = \sqrt{40}$

2 ☐ $\mu = 140$ og $\sigma^2 = \sqrt{140}$

3 ☐ $\mu = 140$ og $\sigma^2 = 140$

4 ☐ $\mu = 140$ og $\sigma^2 = 40$

5 ☐ $\mu = 140$ og $\sigma^2 = 40^2$

Fortsæt på side 22

Opgave XV

Antag at antal færdselsuheld per dag, X, følger en poissonfordeling. Fra 200 uafhængige observationer har man estimeret raten λ til $\hat{\lambda} = 1.2$.

Spørgsmål XV.1 (27)

Angiv nu et 95% konfidensinterval for den sande rate λ:

1 ☐ $[1.2; 4.8]$

2 ☐ $[0; 4]$

3 ☐ $1.2 \pm 1.96 \cdot \sqrt{\frac{1.2}{200}}$

4 ☐ $1.2 \pm 1.96 \cdot \frac{1.2}{200}$

5 ☐ $1.2 \pm 1.96 \cdot \frac{1.2^2}{200^2}$

Fortsæt på side 23

Opgave XVI

To typer receptpligtig medicin (A og B) til at sænke kolesterol i blodet, bliver sammenlignet i et klinisk studie. Ved analyse af data, estimeres hvor meget medicin A reducerer kolesteroltallet, benævnt Δ_A, og tilsvarende hvor meget medicin B reducerer kolesteroltallet, benævnt Δ_B (positive værdier af Δ angiver reduktion og man fandt, at både A og B reducerede kolesteroltallet i middel.). Endelig finder man et 95% konfidensinterval for forskellen i reduktion $(\Delta_A - \Delta_B)$. Dette interval bliver $[0.24; 0.50]$ mmol/L.

Spørgsmål XVI.1 (28)

Hvilket af følgende udsagn er en rimelig konklusion på undersøgelsen?

1 ☐ Medicin A reducerer kolesterol med 0.24 mmol/L mens medicin B reducerer kolesterol med 0.50 mmol/L

2 ☐ Der er 95% sandsynlighed for at medicin A er bedre til at sænke kolestroltallet end medicin B for en vilkårlig person

3 ☐ Der er 95% sandsynlighed for at medicin A vil sænke kolestroltallet med mindst 39 mmol/L i forhold til medicin B for en vilkårlig person

4 ☐ Der er mindst 95% sikkerhed for at medicin A virker bedre end medicin B

5 ☐ Ingen af ovenstående

Fortsæt på side 24

Opgave XVII

En ingeniør planlægger at tage en stikprøve fra en population. Vi betragter følgende 3 udsagn:

I. Såfremt stikprøven har varians nul, så er variansen i populationen også nul.

II. Såfremt populationen har varians nul, så er variansen i stikprøven også nul.

III. Såfremt stikprøven har varians nul, så bliver middelværdi og median ens i stikprøven.

Spørgsmål XVII.1 (29)

Hvilke af de 3 ovenstående udsagn er korrekt(e)?

1 ☐ Kun I. og II.

2 ☐ Kun I. og III.

3 ☐ Kun II. og III.

4 ☐ I., II., og III. er alle korrekte

5 ☐ Ingen af ovensteånde

Fortsæt på side 25

Opgave XVIII

Det er velkendt, at gøge lægger deres æg i andre fuglearters rede, og dermed overlader opgaven med at udruge og opfostre deres unger til værtsfuglen. Envidere er der en teori om, at gøge er i stand til at tilpasse størrelsen af deres æg, afhængig af hvor stor værtsfuglen er.

For at undersøge denne teori, har en ornitolog over en periode registreret størrelsen (længde af ægget) af 10 gøgeæg i hver af 2 forskellige værtsfugles reder, her benævnt værtsfugl A og B, altså samlet 20 æg. Hun regner sig frem til, at standardafgivelsen for størrelsen er 2 mm for både værtsfugl A og B.

Spørgsmål XVIII.1 (30)

Det viser sig nu, at den observerede forskel i størrelse, $\bar{X_A} - \bar{X_B}$, mellem æggene er 1 mm. Hvilken konklusion kommer man frem til, idet man ønsker at teste hypotesen $H_0 : \mu_A = \mu_B$ mod $H_1 : \mu_A \neq \mu_B$ med anvendelse af et almindelig t-test og signifikansniveau $\alpha = 5\%$?

1 ☐ Forskel i størrelse på æggene er statistisk set signifikant

2 ☐ Forskel i størrelse på æggene er statistisk set ikke-signifikant

3 ☐ Man kan ikke konkludere noget uden oplysning om den faktiske størrelse på æggene for A og B

4 ☐ Man kan ikke konkludere noget uden kendskab til populationsstørrelsen

5 ☐ Det er ikke relevant at anvende et almindelig t-test her

SÆTTET ER SLUT. Hav en god juleferie!