

i Forside

Eksamen i: **STK1000 - Innføring i anvendt statistikk.**

Eksamensdag: **Mandag 2. desember 2019.**

Tid for eksamen: **14:30-18:30 (fire timer)**

Tillatte hjelpemidler:

- **Lærebok: Moore, McCabe & Craig: Introduction to the practice of statistics.** Alle utgaver av læreboken er tillatt (eldre utgaver har kun Moore & McCabe som forfattere). Det er tillatt å notere for hånd i læreboka, men det er ikke lov å klistre inn notater.
- **Kapittel 14 fra læreboken, som ligger som lenke under oppgavelinjen.**
- **Godkjent kalkulator**
- **Ordlister for STK1000.**

Lenker til ordliste, tabellene fra læreboken og kapittel 14 fra læreboken er gitt under oppgavelinjen.

I dette oppgavesettet skal du besvare alle oppgavene (1-4) med digital håndtegning. Du bruker skisseark du får utdelt. Det er anledning til å bruke flere ark per oppgave. Se instruksjon for utfylling av skisseark i lenken under oppgavelinjen.

Det blir IKKE gitt ekstratid for å fylle ut informasjonsboksene på skissearkene (engangskoder, kand.nr. o.l.).

Alle deloppgavene teller like mye ved bedømmelsen av besvarelsen din.

1 Oppgave 1

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark i lenken under oppgavelinjen.

I en liten innsjø kan det antas at lengden på fisk (X) er normalfordelte med forventning $\mu = 42.5$ cm og standardavvik $\sigma = 4.8$ cm, dvs $X \sim N(42.5, 4.8)$.

a) Finn sannsynligheten for at en fisk i dette vannet er kortere enn 40 cm.

b) Du fisker 5 fisk i denne innsjøen. Du kan anta at lengdene til disse fiskene er uavhengige av hverandre. Hva er sannsynligheten for at alle disse 5 fiskene er kortere enn 40 cm?

2 Oppgave 2

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark i lenken under oppgavelinjen.

Mennesker har en dominant hånd og en ikke-dominant hånd. Det mest vanlige er å være høyrehendt, dvs at den høyre hånden er den dominante. For venstrehendte er den venstre hånden den dominante. En forsker ønsket blant annet å undersøke om reaksjonstiden for individers dominante hånd er kortere enn for deres ikke-dominante hånd. Forskeren målte tiden (i sekunder) det tok for 12 tilfeldig utvalgte studenter å fange en fallende pinne både med deres dominante hånd og med deres ikke-dominante hånd. For hver student ble det tilfeldig bestemt om reaksjonstiden til den dominante eller den ikke-dominante hånden skulle måles først. Pinnen ble sluppet på lik måte hver gang. Dataene kan sees i tabellen under

Student nr	Dominant hånd	Ikke-dominant hånd	Differanse
1	0.177	0.179	-0.002
2	0.21	0.202	0.008
3	0.186	0.208	-0.022
4	0.189	0.184	0.005
5	0.198	0.215	-0.017
6	0.194	0.193	0.001
7	0.16	0.194	-0.034
8	0.163	0.16	0.003
9	0.166	0.209	-0.043
10	0.152	0.164	-0.012
11	0.19	0.21	-0.02
12	0.172	0.197	-0.025

Det er naturlig å pare observasjonene, dvs analysere differansene for hver student mellom reaksjonstid for den dominante og den ikke-dominante hånden. Anta at disse differansene kommer fra en normalfordeling med forventning μ_D og standardavvik σ_D . Under ser du deler av en utskrift fra en analyse av dataene i R for å teste om reaksjonsevnen for individers dominante hånd er kortere enn for deres ikke-dominante hånd.

One Sample t-test

```
data: diff
t = -2.7759, df = 11, p-value = 0.009017
alternative hypothesis: true mean is less than 0
```

```
sample estimates:
mean of x
-0.01316667
```

a) Bruk R-utskriften til å finne estimater for forventningen μ_D og standardavviket σ_D .

b) Konstruer på bakgrunn av R-utskriften et 95% konfidensintervall for forventet forskjell i reaksjonstid mellom individers dominante og ikke-dominante hånd (μ_D).

c) Formuler hypotesene som ligger til grunn for testen du ser i R-utskriften. Formuler en kort og presis konklusjon på testen.

3 Oppgave 3

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark i lenken under oppgavelinjen.

En medisinsk studie ønsket å undersøke sammenhengen mellom spedbarns blodtrykk ('SBT', målt i mm Hg) og deres vekt ('vekt', målt i kg) og alder ('alder', målt i dager). Data ble samlet inn på 25 tilfeldig valgte spedbarn. En multippel lineær regresjonsmodell for sammenhengen mellom responsvariabelen SBT (y_i) og forklaringsvariablene alder (x_{i1}) og vekt (x_{i2}) til spedbarn nr i er

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

for $i = 1, \dots, 25$. Vi antar at ε_i -ene er uavhengige og normalfordelte med forventning 0 og standardavvik σ .

Utskrift av resultatene fra R for den tilhørende analysen for disse dataene kan sees helt til slutt i oppgaven.

a) Hvilken parameter i modellen beskriver økning i forventet blodtrykk for en dags økning i alder, dersom vekt er konstant? Finn deretter estimater til alle parametrene $\beta_0, \beta_1, \beta_2$ og σ fra R-utskriften.

b) Sett opp hypotesene som hører til t-testen for β_1 . Finn tilhørende p-verdi. Bruk denne p-verdien til å si om et 95%-konfidensintervall for β_1 vil inneholde 0 eller ikke.

c) Beregn et estimat for det forventede blodtrykket til fire dager gamle spedbarn som veier 3.00 kg. Finn deretter et tilhørende 95% konfidensintervall fra R-utskriften.

d) Finn et 95% prediksjonsintervall for blodtrykket til et fire dager gammelt spedbarn som veier 3.00 kg fra R-utskriften. Forklar kort hvorfor dette intervallet har ulik bredde sammenlignet med intervallet i Oppgave c).

```
Call:
lm(formula = SBT ~ alder + vekt)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-4.1779	-1.2224	0.2005	1.5164	4.5465

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	57.2644	3.7986	15.075	4.44e-13	***
alder	5.8041	0.6415	9.048	7.22e-09	***
vekt	3.3162	1.5522	2.136	0.044	*

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.454 on 22 degrees of freedom
```

```
Multiple R-squared:  0.9199, Adjusted R-squared:  0.9126
```

```
F-statistic: 126.3 on 2 and 22 DF,  p-value: 8.696e-13
```

```
> predict(lm(SBT~alder+vekt),newdata=data.frame(alder=4,vekt=3.00),
interval='confidence',level=0.95)
```

	fit	lwr	upr
1	?	88.9649	91.89418

```
> predict(lm(SBT~alder+vekt),newdata=data.frame(alder=4,vekt=3.00),
interval='prediction',level=0.95)
```

	fit	lwr	upr
1	?	85.13406	95.72503

4 Oppgave 4

I denne oppgaven kan du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark i lenken under oppgavelinjen.

Risikoen for jordskred på en bestemt lokasjon kan modelleres ved logistisk regresjon med mange forklaringsvariable. En slik forklaringsvariabel er høyde over havet i meter (x_1). Modellen med m forklaringsvariable for sannsynligheten p for jordskred er

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$$

Maximum likelihood-estimering fra data gir at estimatet for β_1 i denne logistiske modellen er $b_1 = -0.71$ og tilhørende standardfeil er $SE_{b_1} = 0.05$.

Finn et 95% konfidensintervall for odds-ratioen for jordskred for en økning i høyde over havet på 1 meter, når alle andre forklaringsvariable holdes fast.