

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1000 – Innføring i anvendt statistikk  
Eksamensdag: Onsdag 25. november 2020  
Tid for eksamen: 15:00–19:00  
Oppgavesettet er på 4 sider.  
Vedlegg: Ingen  
Tillatte hjelpemidler: Alle hjelpemidler er tillatt, men det er ikke tillatt å kommunisere eller samarbeide med andre.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgavesettet har fire oppgaver som til sammen består av elleve deloppgaver. Hver deloppgave teller likt.

### Oppgave 1

Du skal kjøpe julegave til nevøen din som nylig fylte syv år, og du har funnet en genser du tror han kommer til å like. Du kjenner ikke høyden til nevøen din. Du slår derfor opp høydekurvene fra Vekststudien i Bergen (2007) og finner at høyden til syv år gamle gutter har forventningsverdi  $\mu = 124,5$  cm og standardavvik  $\sigma = 4,8$  cm. Du antar videre at høyden til syv år gamle gutter er normalfordelt.

#### 1a

Regn ut sannsynligheten for at en tilfeldig valgt gutt på syv år er over 128 cm høy.

#### 1b

Genseren finnes i klesstørrelsene 116, 122, 128, 134, 140, 146 og 152. Du velger å kjøpe den minste klesstørrelsen som er slik at maksimalt 10% av syv år gamle gutter er høyere enn høydemålet angitt av klesstørrelsen. Hvilken klesstørrelse velger du?

### Oppgave 2

Vi undersøker en enhet av en væske som skal brukes til desinfisering. Produktet skal etter standarden for produktet ha konsentrasjon 0.60. Du

*(Fortsettes på side 2.)*

skal, ved å gjennomføre en statistisk hypotesetest, evaluere påstanden om at konsentrasjonen i enheten du undersøker (prøven) har konsentrasjon 0.60.

## 2a

Formuler hypotesene som ligger til grunn for hypotesetesten du vil gjennomføre.

## 2b

Du utfører to målinger av konsentrasjonen til væsken, og måleresultatene er 0.58 og 0.56. Forhåndskunnskap om målemetoden gir at du kan regne måleresultatene som normalfordelt  $N(\mu, \sigma = 0.022)$  der  $\mu$  er den sanne konsentrasjonen til prøven. Finn verdi for testobservator, og tilhørende P-verdi for hypotesetesten. Formuler en konklusjon på hypotesetesten.

## 2c

Regn ut et 95% konfidensintervall for forventet konsentrasjonen til prøven. Forklar kort hva resultatet betyr.

## Oppgave 3

60 marsvin er blitt gitt en daglig dose C-vitamin (enten 0.5, 1, eller 2 mg/dag) for å undersøke effekten på tannlengden deres. En enkel lineær regresjonsmodell for sammenhengen mellom responsvariabelen tannlengde ( $y_i$ ) og forklaringsvariabelen daglig dose C-vitamin ( $x_i$ ) er

$$y_i = \beta_0 + \beta_1 \cdot x_i + \epsilon_i, \quad i = 1, 2, \dots, 60. \quad (1)$$

Vi antar at  $\epsilon_i$ -ene er uavhengige normalfordelte  $N(0, \sigma)$ . R-utskriften for den tilhørende analysen er som følger:

Call:

```
lm(formula = toothlen ~ vitC_dose)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.4496	-2.7406	-0.7452	2.8344	10.1139

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.4225	1.2601	5.89	2.06e-07 ***
vitC_dose	9.7636	0.9525	?	1.23e-14 ***

---

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

(Fortsettes på side 3.)

Residual standard error: 4.601 on 58 degrees of freedom  
Multiple R-squared: 0.6443, Adjusted R-squared: 0.6382  
F-statistic: 105.1 on 1 and 58 DF, p-value: 1.233e-14

### 3a

Les av estimatene til de tre parameterne  $\beta_0$ ,  $\beta_1$  og  $\sigma$ , og forklar hva hver av disse parameterne beskriver. Les av  $R^2$ -verdien og forklar hva denne betyr.

### 3b

Skriv ned hypotesene som tilhører t-testen som er gjort for stigningstallet i R-utskriften. Beregn standardisert testobservator. Les av tilhørende P-verdi, og bruk denne til å formulere konklusjon på den statistiske hypotesetesten.

### 3c

Vi er spesielt interessert i den høyeste dosen, 2mg/dag. Beregn fra den estimerte lineærmodellen, forventet tannlengde når marsvinet har fått 2mg (dose) C-vitamin daglig.

### 3d

Angi 95% konfidensintervall for forventet tannlengde for et marsvin som har fått en daglig dose på 2mg C-vitamin. Angi også 95% prediksjonsintervall for tannlengde for et marsvin som har fått daglig dose på 2mg C-vitamin. Hva er ulikt i definisjonen av de to intervallene, og hvorfor ender de opp med ulik bredde?

```
> predict(mylin, newdata = data.frame(vitC_dose=2.0),
interval= 'confidence', level= 0.95)
      fit      lwr      upr
1      ?  24.96508  28.9342
> predict(mylin, newdata = data.frame(vitC_dose=2.0),
interval= 'predict', level= 0.95)
      fit      lwr      upr
1      ?  17.52801  36.37127
```

## Oppgave 4

Julenissen er veldig opptatt av at barn skal lære å se lyset når det er mørkt. Sannsynligheten for at et barn får lommelykt til jul kan modelleres ved logistisk regresjon med tre forklaringsvariabler:  $x_{i1}$ : barnets alder målt i antall år,  $x_{i2}$ : antall dager før julaften nissen pakker inn presangen,  $x_{i3}$ : om barnet har sendt ønskeliste til julenissen i år eller ikke (0=nei, 1=ja).

(Fortsettes på side 4.)

**4a**

Skriv ned den logistiske regresjonsmodellen for sammenhengen mellom responsvariabelen lommelykt til jul ( $y_i$ ) og forklaringsvariablene alder ( $x_{i1}$ ), forberedelsestid ( $x_{i2}$ ) og ønskeliste ( $x_{i3}$ ).

**4b**

Ta utgangspunkt i (en utvalgt del av) R-utskriften for analysen som gitt under. Konstruer et 95% konfidensintervall for odds-ratioen for lommelykt til jul for en økning i alder på ett år, når alle andre forklaringsvariabler holdes fast.

Call:

```
glm(formula = lommelykt ~ alder + ønskeliste + forberedelsestid,
family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9473	-0.5467	0.1785	0.3718	2.1605

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.93959	2.20015	-0.882	0.37801
alder	0.24758	0.09426	2.627	0.00862 **
ønskeliste	-0.94614	1.20481	-0.785	0.43227
forberedelsestid	-0.04807	0.22493	-0.214	0.83076

---

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

*For hver jul som kommer, er vi alle ett år eldre. God jul!*