Løsningsforslag eksamen høst 2023

Statistikkdelen ØKA120

Desember 2023

Oppgave 1

a)

$$H_0 : \mu = 200 \quad (\text{evt. } \mu \leq 200)$$
$$H_A : \mu > 200$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{206{,}22 - 200}{50/\sqrt{50}} = \frac{6{,}22}{7{,}071} = 0{,}88$$

Kritisk verdi: $kv = z_{0{,}05} = 1{,}645$
Vi beholder H_0. Vi har ikke funnet støtte for at boligene er større enn 200 kvm.

b)

$$H_0 : \mu = 4\ 370\ 018 \quad (\text{evt. } \mu \geq 4\ 370\ 018)$$
$$H_A : \mu < 4\ 370\ 018$$

$$T = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} = \frac{3\ 609\ 600 - 4\ 370\ 018}{1\ 560\ 400/\sqrt{50}} = \frac{-760\ 418}{220\ 673{,}88} = -3{,}45$$

Kritisk verdi: $kv = -t_\alpha(n-1) = -t_{0{,}05}(49) = -1{,}677$. (*Eventuelt kan man argumentere for å bruke normalfordelingen for å finne kritisk verdi her siden* $n > 30$. *Dette gir* $kv = -z_{0{,}05} = -1{,}645$)
Vi forkaster nullhypotesen ved et 5% signifikansnivå. Boligprisene i november er lavere enn i oktober.

c)

- Siden $\sigma > s_Y$ vil testobservatoren Z være mindre i absoluttverdi enn testobservatoren T vi fant i b). Dette fører til lavere sannsynlighet for å forkaste nullhypotesen siden testobservatoren beveger seg nærmere null.
- Kritisk verdi er lavere i absoluttverdi når vi bruker normalfordelingen enn ved bruk av t-fordelingen med 49 frihetsgrader, $|z_{0{,}05}| < |t_{0{,}05(49)}|$. Dette fører til større sannsynlighet for å forkaste nullhypotesen siden kritisk verdi beveger seg nærmere null. *Merk at det går an å bruke normalfordelingen for å finne kritisk verdi i oppgave b) dersom man argumenterer for at det er et relativt stort utvalg (n>30).*

Vi får (ikke nødvendig å regne ut dette):

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} = \frac{3\ 609\ 600 - 4\ 370\ 018}{3\ 500\ 000/\sqrt{50}} = \frac{-760\ 418}{494\ 974{,}75} = -1{,}54$$

$$kv = -z_{0{,}05} = -1{,}645$$

Vi beholder nå nullhypotesen.

Oppgave 2

a)

- $\hat{\beta}_0 = 1428{,}45$: En bolig som har en størrelse på null kvadratmeter forventes å selges for kr 1 428 450. Dette gir liten mening, siden boliger ikke kan være 0 kvm store. Eventuelt kan dette representere pris for en tomt. Konstantleddet brukes her mest for å plassere regresjonslinja best mulig.
- $\hat{\beta}_1 = 10{,}577$: En bolig som er 1 kvm større enn en annen forventes å selges til kr 10 577 mer.

b)

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

$$T = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{10{,}577}{1{,}600} = 6{,}61$$

Kritiske verdier:

$$kv = \pm t_{\alpha/2}(n-k-1) = \pm t_{\alpha/2}(n-k-1) = \pm t_{0{,}005}(50-1-1) = \pm t_{0{,}005}(48) = \pm 2{,}682$$

Testobservator er i forkastningsområdet, så vi forkaster nullhypotesen. Vi har funnet støtte for påstanden om antall kvadratmeter har en signifikant effekt på boligprisen.

c)
Prisen for boligen predikeres til

$$\hat{Y}_{22} = \hat{\beta}_0 + \hat{\beta}_1 X_{22} = 1428{,}45 + 10{,}577 \cdot 300{,}45 = 4606{,}31$$

Altså kr 4 606 310.
d)
Prediksjonsfeilen blir da

$$\hat{u}_{22} = Y_{22} - \hat{Y}_{22} = 4550 - 4606{,}31 = -56{,}31$$

Altså har modellen overestimert med kr 56 310.

e)

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{62\ 461\ 357}{119\ 306\ 469} = 1 - 0{,}524 = 0{,}476$$

47,6% av variasjonen i salgsprisene forklares av modellen (antall kvadratmeter).
Evt. kan vi regne ut at $ESS = TSS - RSS = 119\ 306\ 469 - 62\ 465\ 406 = 56\ 841\ 063$ og beregne $R^2 = ESS/TSS = 56\ 841\ 063/119\ 306\ 469 = 0{,}476$

Oppgave 3

a)

$$\begin{aligned}
P(X > 300) &= 1 - P(X \leq 300) = 1 - P\left(Z \leq \frac{300-\mu}{\sigma}\right) \\
&= 1 - P\left(Z \leq \frac{300-200}{50}\right) = 1 - P\left(Z \leq \frac{100}{50}\right) \\
&= 1 - P(Z \leq 2) = 1 - 0{,}9772 = 0{,}0228
\end{aligned}$$

Sannsynligheten for at en tilfeldig valgt bolig er større enn 300 kvm er 2,28%.

b)

$$\begin{aligned}
P(250 \leq X \leq 300) &= P(X \leq 300) - P(X \leq 250) \\
&= P\left(Z \leq \frac{300-\mu}{\sigma}\right) - P\left(Z \leq \frac{250-\mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{300-200}{50}\right) - P\left(Z \leq \frac{250-200}{50}\right) \\
&= P(Z \leq 2) - P(Z \leq 1) = 0{,}9772 - 0{,}8413 = 0{,}1359
\end{aligned}$$

Sannsynligheten for at en tilfeldig valgt bolig er mellom 250 og 300 kvm stor er 13,59%. Pluss for diskusjon om $<$ og $\leq$ ved en kontinuerlig variabel.

c)

$$\begin{aligned}
P(A) &= 0{,}36 \\
P(R) &= 0{,}20 \\
P(A \cap R) &= 0{,}05
\end{aligned}$$

d)

$$P(A \cup R) = P(A) + P(R) - P(A \cap R) = 0{,}36 + 0{,}20 - 0{,}05 = 0{,}51$$

$$P(A|R) = \frac{P(A \cap R)}{P(R)} = \frac{0{,}05}{0{,}20} = 0{,}25$$

e)

- $P(A \cup R) = 0{,}51$: Sannsynligheten for at en bolig har sjøutsikt eller basseng, eller begge deler, er 51%
- $P(A|R) = 0{,}25$: Sannsynligheten for at en bolig har sjøutsikt dersom vi fra før vet at den har basseng, er 25%

f)
Hvis vi kaller gruppen med sjøutsikt for gruppe A og gruppen uten sjøutsikt for gruppe B, så har vi

$$\bar{X}_A = 285 \quad s_A = 118 \quad n_A = 18$$
$$\bar{X}_B = 162 \quad s_B = 56 \quad n_B = 32$$

Vi ønsker å teste

$$H_0 : \mu_A = \mu_B$$
$$H_A : \mu_A > \mu_B$$

der μ_A og μ_B er populasjonsgjennomsnittet til størrelsen på boliger med og uten sjøutsikt.

$$T = \frac{\bar{X}_A - \bar{X}_B}{S\sqrt{1/n_A + 1/n_B}}$$

der

$$S = \sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A+n_B-2}} = \sqrt{\frac{17 \cdot 118^2 + 31 \cdot 56^2}{18+32-2}} = \sqrt{\frac{333\ 924}{48}} = \sqrt{6956{,}75} = 83{,}407$$

Vi får testobservator

$$T = \frac{\bar{X}_A - \bar{X}_B}{S\sqrt{1/n_A + 1/n_B}} = \frac{285-162}{83{,}407 \cdot \sqrt{1/18 + 1/32}} = \frac{123}{24{,}575} = 5{,}01$$

Kritisk verdi:

$$t_\alpha(n_A + n_B - 2) = t_{0{,}05}(48) = 1{,}677$$

Vi forkaster nullhypotesen. Vi har funnet støtte for påstanden om at boliger med sjøutsikt er større enn boliger uten sjøutsikt.

Det er også OK å bruke $\bar{X}_B - \bar{X}_A$ for å beregne testobservator. Husk da å bruke $-t_\alpha(n_A + n_B - 2)$ som kritisk verdi.