SENSORVEILEDNING SØK1004 VÅREN 2015

Oppgave 1 (25%)

a)

i. De marginale sannsynlighetsfordelingene er gitt ved:

X	10	15	20	25	30
f(X)	0.2	0.1	0.15	0.2	0.35

Y	-10	0	100
g(Y)	0.45	0.25	0.3

ii. $E(X)=\sum X\cdot f(X)=22, Var(X)=\sum(X-E(X))^2 f(X)=E(X^2)-[E(X)]^2=542.5-(22)^2=58.5$

$E(Y)=\sum Y\cdot g(Y)=25.5, Var(Y)=E(Y^2)-[E(Y)]^2=3045-(25.5)^2=2394.75$

iii. Kovarians og korrelasjon er begge mål på graden av lineær samvariasjon. Positive verdier betyr positiv lineær samvariasjon, mens negative verdier betyr negativ lineær samvariasjon. Ingen lineær samvariasjon dersom kovarians og korrelasjon er lik null. Korrelasjonen er lettere å tolke fordi den er uavhengig av benevning og varierer mellom -1 (perfekt negativ lineær samvariasjon) og 1 (perfekt positiv lineær samvariasjon).

$Cov(X,Y)=E[(X-E(X)(Y-E(Y)]=E(XY)-E(X)E(Y)=232.5-22\cdot 25.5=-328.5$

$$\rho=Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(\mathrm{Y})}}=\frac{-328.5}{\sqrt{58.5}\sqrt{2394.75}}=-0.878$$

iv. La $Z=X+Y$ være produsentens totale profitt.

$E(Z)=E(X)+E(Y)=47.5$

$Var(Z)=Var(X)+Var(Y)+2Cov(X,Y)=1737.75$

b) Z er standard normalfordelt og $P(-1.52<Z<0)=0.4357$ (fra Tabell A1). Siden 0.4357>0.2 er k<0 og $P(k<Z<0)=0.4357-0.2=0.2357$. Dette gir k = -0.63.

Oppgave 2 (25%)

a) X er binomisk fordelt (to mulige utfall, konstant sannsynlighet for 'suksess', uavhengige forsøk) og sannsynlighetsfordelingen er gitt som:

$$P(X)=\frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$$

der π = 0.4 er sannsynligheten for suksess og n = 200 er antall uavhengige forsøk.

Forventning: $E(X)=n\cdot\pi=200\cdot 0.4=80$

Varians: $Var(X)=n\cdot\pi\cdot(1-\pi)=48$

b) Siden $n\pi>5$ og $n(1-\pi)>5$ er X tilnærmet normalfordelt med forventning $n\pi=80$ og varians $n\cdot\pi\cdot(1-\pi)=48$. Da er $P(60<X<80)=P(-2.89<Z<0)=0.4981$ og $P(X>90)=P(Z>1.44)=0.0749$.

c) Hypotesetest om populasjonsandel. $H_0:\pi=0.4$ og $H_A:\pi<0.4$. Utvalgsandelen er gitt som $p=\frac{63}{200}=0.315$. Testobservatoren $TS=\frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}=\frac{0.315-0.4}{\sqrt{0.4\cdot 0.6/200}}=-2.45$ er standard normalfordelt når H_0 er sann. Kritisk verdi 1% signifikansnivå = 2.33. Siden $TS=-2.45<-2.33$ forkastes H_0.

Oppgave 3 (20%)

a) Hypotesetest om populasjonsgjennomsnitt ved stort utvalg. $H_0:\mu=22$ og $H_A:\mu<22$. Testobservatoren $TS=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{19.5-22}{11/\sqrt{100}}=-2.27$ er standard normalfordelt når H_0 er sann. Kritisk verdi 5% signifikansnivå = 1.64. Siden $TS=-2.27<-1.64$ forkastes H_0.

b) Testens p-verdi: $P(Z<-2.27)=0.5-0.4884=0.0116$. Det laveste signifikansnivået som H_0 kan forkastes til er 1.16%.

Oppgave 4 (30%)

a) $H_0:\rho=0$ og $H_A:\rho>0$, der ρ er populasjonskorrelasjon. Testobservatoren $TS=\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ er t-fordelt med n-2 frihetsgrader når H_0 er sann. Utvalgskorrelasjonen finnes som $R=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2}\sqrt{\sum(Y_i-\bar{Y})^2}}=0.469$ som gir $TS=\frac{0.469\sqrt{87}}{\sqrt{1-0.469^2}}=4.95$.

Tabell A2 gir ikke kritisk verdi for 87 frihetsgrader. Vi benytter da 120 frihetsgrader og finner at kritisk verdi er henholdsvis 1.658 og 2.358 ved 5% og 1% signifikansnivå. H_0 forkastes med god margin både til 5% og 1% signifikansnivå.

b) Estimat for β: $b=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}=\frac{196664}{1971126}=0.0998\approx 0.1$. Tolkning: Dersom antall innbyggere per km^2 øker med 1, øker timelønna med 0.1 kr. Estimat for α:

$a = \bar{Y} - b\bar{X} = 290.2 - 0.1 \cdot 45 = 285.7$. Vi har altså $\hat{Y}_i = 285.7 + 0.1X_i$. Elastisitet av timelønn med hensyn på befolkningstetthet: $El_X Y = \frac{dY}{dX} \cdot \frac{\bar{X}}{\bar{Y}} = 0.1 \cdot \frac{45}{290.2} = 0.0155$. Dersom befolkningstettheten øker med 1%, øker timelønna med 0.0155%. Dvs: Dersom befolkningstettheten dobles (øker med 100%), øker timelønna med 1.55%.

c) $R^2 = \frac{b^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} = \frac{0.1^2 \cdot 1971126}{89054} = 0.22$. Modellens forklaringskraft sier at 22% av variasjonen i regionenes timelønn kan forklares med variasjon i befolkningstetthet.