Løsningsforslag INT010 høst 2010

Oppgave 1

a) Benytter t-test. Antall frihetsgrader er n-k-1 = 36-4-1 = **31**. Kritisk verdi er $t_{0,025,31} \approx t_{0,025,30} =$ **2,042**. Vi forkaster nullhypotesen for t-verdier større eller lik 2,042 (eller mindre eller lik -2,042). Dvs. alle koeffisienter er signifikante, bortsett fra YR (år) og YRSQ (år kvadrert).

b) 95% konfidensintervall for koeffisienten til "logPG": -0,164± 2,042·0,0187 = **[-0,168; -0,126]**, der vi har benyttet at $t_{0,025,31} \approx 2,042$.

c) Plottet kan minne om en **parabel**, og dermed vil en annenordens modell være passende, dvs. $y=a+ b\cdot x + c\cdot x^2 + \varepsilon$, der y er forbruk og x er år. Dette kan fange opp en **utflatingseffekt**, dvs. forbruket flater ut med årene. (Ut fra den bivariate sammenhengen mellom forbruk og tid vist i plottet vil vi forvente positiv b og negativ c.)

d) Modellen har **høy forklaringsgrad** (stor R^2), og F-verdien er stor, s.a. F-testen blir forkastet. Siden vi har en 'loglog' modell, vil koeffisientene til logPG (pris) og logI (inntekt) være **elastisiteten** til bensinforbruket. Koeffisienten til pris er -0,16, dvs. en 1% økning i prisen vil føre til **0,16% nedgang** i forbruket (som gir mening ift. økonomisk teori). Koeffisienten til inntekt er 1,26, dvs. en 1% økning i inntekt vil føre til en **1,26% økning** i bensinforbruket (som også gir mening). Vi ser at tidsvariablene YR og YRSQ begge er negative og ikke signifikante. Det er litt overraskende i lys av plottet vist i c, men må henge sammen med at regresjonen inneholder to andre variabler, pris og inntekt per kapita, som er positivt økende over tid og fanger opp forbruksveksten. En mulig tolkning av den negative effekten av tidsvariablene kan være at det etterspørres mindre bensin nå enn før *betinget på pris og inntekt per kapita* på grunn av fremskritt i energibesparende teknologi.

e) Grafene på venstre side viser at residualene er tilnærmet normalfordelte. Grafen øverst i høyre hjørne viser residualene plottet mot predikert produksjon, og det synes ikke å være problemer med heteroskedastisitet. Grafen nederst til høyre antyder tydelig (positiv) autokorrelasjon, spesielt tatt i betraktning at datasettet består av tidsrekker.

f) H_0: Ingen autokorrelasjon vs. H_1: Positiv autokorrelasjon. Durbin-Watsons testobservator er oppgitt i utskrift fra Minitab, d = **0,50**. Forkastningsområde er $d < d_L =$ **1,24** (fra tabell med n = 36, k =4 og signifikansnivå 5%). **Forkast dermed H_0.**

g) Autokorrelasjon medfører ikke forventningsskjeve koeffisienter, men den rapporterte inferensen er ikke gyldig. Videre vil ikke minste kvadraters metode være en effisient estimeringsmetode.

h) Modellen har **høy forklaringsgrad** (stor R^2), og F-verdien er stor, s.a. F-testen blir forkastet. P-verdiene til t-testene gir signifikante variable. Koeffisientene til logPG (pris) og logI (inntekt) er fremdeles **elastisiteten** til bensinforbruket, og har samme fortegn som før. Vi ser at koeffisienten til år kvadrert nå har blitt klart signifikant. Betinget på pris og inntekt faller altså etterspørselen over tid. Koeffisienten for lagget forbruk (Lagged logGpc) er positiv (0,7), dvs. at forrige års forbruk forklarer dette års forbruk i stor grad, noe som er rimelig. Vi inkluderer lagget pris (Lagged logPG) som en forklaringsvariabel da bensinforbruket ikke bare forklares ved dette års pris, men også forrige års pris, grunnet folks oppfatning om prisen er stigende eller synkende. Fra diagnoseplottet ser vi at **autokorrelasjonen** i residualene nå er kraftig **redusert** (riktignok har vi en outlier, som kanskje bør studeres ytterligere).

Oppgave 2

a) Test av en andel. Testobservator: $z = \dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ som er (tilnærmet) standard normalfordelt.

b) H_0: p = 0,1 mot H_1: p > 0,1. $\hat{p} = \dfrac{8}{50} = 0{,}16$. Da n er 50, blir z = (0,16-0,1)/√(0,1·(1-0,1)/50) = **1,41**. H_0 forkastes om $z > z_{0,05} = 1{,}64$ (fra tabell). Dvs. vi **beholder H_0**, og vi konkluderer med at bilprodusentens påstand holder.

c) P-verdi = P(z > 1,41) = 1 – P(z < 1,41) = 1 – 0,9207 = **0,0793** (der P(z < 1,41) er funnet i tabell).

Oppgave 3

a) X_1 = antall rosiner i 1. porsjon. Den vil være **binomisk fordelt(n,1/m)** fordi dette er en binomisk forsøksrekke med n forsøk (antall rosiner i pakken) og for hvert forsøk (hver rosin) er det 1/m sannsynlighet for at rosinen befinner seg i 1. porsjon (dvs. konstant suksess-sannsynlighet). Fordelingen til X_2 gitt at $X_1 = x_1$ er **binomisk (n-x_1,1/(m-1))** fordi det nå kun er igjen n-x_1 rosiner i pakken. Disse skal så uniformt fordeles på m-1 porsjoner.

b) Sannsynligheten for at den første porsjonen skal minst inneholde en rosin kan skrives: $P(X_1 \geq 1) = 1 - P(X_1 = 0) = 1 - (n!)/(0!(n-0)!) \cdot (1/m)^0 \cdot (1 - 1/m)^n = 1 - (1 - 1/m)^n$, siden X_1 er binomisk(n,1/m). Denne sannsynligheten skal være større enn β, dvs. $1 - (1 - 1/m)^n \geq \beta$ som medfører at $\mathbf{n \geq \ln(1-\beta)/\ln(1-1/m)}$.

c) X_1 og X_2 er **ikke uavhengige**, siden X_2 gitt X_1=x_1 avhenger av x_1. Korrelasjonen mellom X_1 og X_2 er negativ, dvs. $\boldsymbol{\rho(X_1, X_2) < 0}$, fordi om det er mange rosiner i 1. porsjon (X_1 stor) vil det sannsynligvis være få i 2. porsjon (X_2 liten).

d) Fra definisjon av forventning for en diskret fordeling finner vi $E(Y_1) = 1 \cdot P(X_1 = 0) + 0 \cdot P(X_1 > 0) = \mathbf{(1 - 1/m)^n}$, der vi benytter at X_1 er binomisk(n, 1/m) som før (se b)). Siden det ikke er noe spesielt med den 1. porsjonen vil alle Y_i ha forventning som over, dermed blir $E(W) = E(\sum Y_i) = \mathbf{m \cdot (1 - 1/m)^n}$.