Løsningsforslag INT010 vår 2011

Oppgave 1

a) F-test. Testobservator $F = s_1^2 / s_2^2 = 13^2 / 24^2 =$ **0.29**. Forkastningsområde på 5% signifikansnivå med 7-1 og 8-1 frihetsgrader: $F > F_{0,025,6,7} =$ **5,12** eller $F < 1/F_{0,025,7,6} = 1/5,70 =$ **0.18**. **Aksepter H_0** om lik varians.

b) **H_0: $\mu_1 - \mu_2 = 0$ mot H_1: $\mu_1 - \mu_2 \neq 0$. Testobservator $t = (x_1 - x_2) / \sqrt{[s_p^2 (1/n_1 + 1/n_2)]}$.** Kritisk grense på 5% nivå $t >$ **$t_{0,025,13} = 2,160$** eller $t <$ **$- t_{0,025,72} = - 2,160$**.

c) Forutsetninger: De to populasjonene (kunnskapsnivå) må være uavhengige og normalfordelte. Testobservator t = **-2,76 (Minitab)** og **p-verdi lik 0,016**. **Forkast H_0** og påstå at kunnskapsnivået er forskjellig i de to næringene.

d) Hypotesen **H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. H_1: minst en av forventningene er forskjellig** blir nå testet. Testobservatøren er F = 4,12 og p-verdien er 0,034 (fra Minitab), kan derfor **forkaste H_0**, og påstå at kunnskapsnivået i de tre næringene er signifikant forskjellig.

e) **H_0: Plasseringen ('location') til de tre populasjonene er den samme vs. H_1: Minst to populasjonsplasseringer er forskjellige.** Kritisk grense på 5% nivå $H > \chi_{\alpha,k-1} = \chi_{0,05,2} =$ **5,99. Behold H_0** pga. p-verdi lik 0,057 eller pga. testobservator H = 5,74. Her er resultatet motsatt i forhold til d). Men **ser vi på p-verdiene**, ser vi at forskjellen ikke er større enn 0,023, slik at **resultatene** faktisk **ikke er så langt fra hverandre**.

Oppgave 2

a) F-test: F = MSR/MSE = 1346/21,4 = **62,9**. Kritisk grense $F > F_{0,05,2,523-2-1} \approx$ **$F_{0,05,2,\infty} = 3$. Forkast H_0.**

b) 95% konfidensintervall for koeffisienten til "Educ": $0,913 \pm 1,96 \cdot 0,08219 =$ **[0,752;1,074]**, der vi har benyttet at $t_{0,025,\infty} = 1,96$.

c) Fra resultatet i a) og en rimelig høy R^2 konkluderer vi med at modellen har en relativt **god forklaringsgrad.** Koeffisientene til utdanning og arbeidserfaring er begge positive, dette virker rimelig. **Dess høyere utdanning og arbeidserfaring, dess høyere lønn.**

d) Grafene på venstre side viser at residualene avviker noe fra normalfordelingen, spesielt med tunge haler i fordelingen. Grafen øverst i høyre hjørne viser residualene plottet mot predikert lønn, og det synes å være problemer med **heteroskedastisitet.** Vi ser økende varians i residualene. Grafen nederst til høyre antyder ingen autokorrelasjon, noe som er rimelig, da vi ikke har tidsrekkeobservasjoner.

e) $W = n \cdot R^2 =$ **$523 \cdot 0,0215 = 11,24$**. Kritisk grense er $\chi^2_{0,05,5} =$ **11,1** siden vi har k lik 5 (antall forklaringsvariable). Vi **forkaster derfor H_0** (dog med liten margin) og påstår at vi har heteroskedastisitet.

f) Alle koeffisienter er signifikante (både t-tester og F-tester), og med en relativt høy R^2 har modellen god forklaringsgrad. Koeffisientene til utdanning og arbeidserfaring er begge fremdeles positive, dette virker rimelig. Da det nå er gjennomført en **semilogaritmisk transformasjon,** representerer koeffisientene til utdanning og arbeidserfaring semielastisiteter, dvs. **ett års lengre utdanning** fører til at **lønnen går opp** med om lag **9,4%**, tilsvarende vil **ett års lengre arbeidserfaring** føre til at **lønnen går opp** med om lag **1,0%.**

g) Vi ser nå fra plottene på venstre side at residualene er **normalfordelt** (vesentlig mer så enn i d). Tilsvarende er tyder plottet øverst til høyre på **ingen heteroskedastisitet**.

Oppgave 3

a) Prisen ved tidspunkt n er prisen ved tidspunkt n-1 multiplisert med (1+a) om oppgang (da er $I_n = 1$) eller multiplisert med (1-b) om nedgang (da er $1-I_n = 1$). Ved så å innsette tilsvarende uttrykk for prisen ved tidspunkt n-1 (P_{n-1}) til tidspunkt 1 (P_1) fremkommer det endelige uttrykket. Siste overgang er kun en enkel omskriving.

b) $$EP_n = E\left[P_0(1-b)^n\left(\frac{1+a}{1-b}\right)^{X_n}\right] = P_0(1-b)^n E\left[\left(\frac{1+a}{1-b}\right)^{X_n}\right] = P_0(1-b)^n E\left[z^{X_n}\right] = \text{(bruker Hint)}$$

$$= P_0(1-b)^n\left[1-p+pz\right]^n = P_0(1-b)^n\left[1-p+p\left(\frac{1+a}{1-b}\right)\right]^n = P_0\left[(1-b)-p(1-b)+p\left(\frac{1+a}{1-b}\right)(1-b)\right]^n$$

$$= P_0\left[1-b-p-pb+p+ap\right]^n = P_0\left(1-b(1-p)+ap\right)^n \quad \text{q.e.d.}$$

c) $EP_n = P_0 \cdot (1-0{,}1\cdot(1-0{,}4) + 0{,}2\cdot 0{,}4)^n = P_0 \cdot 1{,}02^n$

$EP_2 = P_0 \cdot 1{,}0404$
$EP_{10} = P_0 \cdot 1{,}219$