Løsningsforslag INT010 vår 2012

Oppgave 1

a) $Z = \left(\hat{p} - p/\sqrt{p(1-p)/n}\right) = \left(0{,}48 - 0{,}50/\sqrt{0{,}50(1-0{,}50)/505}\right) = \underline{0{,}90}$
Z er tilnærmet standard normalfordelt. Kritisk grense for tosidig test med 5% signifikansnivå er 1,96. Vi kan altså ikke forkaste en hypotese om at sann p er 0,50.

b) Konfidensintervallet er på formen $\hat{p} \pm k \cdot \sqrt{\hat{p}(1-\hat{p})/n}$.
Med 95 % konfidensnivå får vi da intervallet
$0{,}48 \pm 1{,}96 \cdot \sqrt{0{,}48(1-0{,}48)/505} = \ 0{,}48 \pm 1{,}96 \cdot 0{,}02 = \underline{0{,}48 \pm 0{,}04}$.

c) Vi skal ha at $1{,}96 \cdot \sqrt{\hat{p}(1-\hat{p})/n} = 0{,}01$. Vårt beste anslag for hva $\hat{p}$ vil bli er 0,48. Da blir ligningen $1{,}96 \cdot \sqrt{0{,}48(1-0{,}48)/n} = 0{,}01$. Det medfører at
$n = \left(\frac{1{,}96 \cdot \sqrt{0{,}48(1-0{,}48)}}{0{,}01}\right)^2 = \underline{9\,589}$. Å øke presisjonsnivået med 3 prosentpoeng vil altså være svært kostbart. Antall observasjoner må nesten 20-dobles.

d) I b) fant vi at feilmarginen var ±4 prosent. Hvor den sanne andelen ligger innenfor dette området ville neppe hatt avgjørende betydning for den etterfølgende debatten. [Det kan derimot ligge relevante metodeproblemer knyttet til spørsmålsstillingen. Hva legger respondentene i begrepene "ansvar" og "seksuelle overgrep" som på VGs førsteside ble til "skyld" og "voldtekt"? Mediekommentator Sven Egil Omdal skrev senere: "De som har et språk med flere nyanser enn det som selger løssalgsaviser, vet at den enes skyld ikke utelukker den andres ansvar. I denne saken er det for eksempel desken i VG som har skylden, mens redaktøren i VG har ansvaret."]

e) Vi oppgir presisjonen som $\pm k \cdot \sqrt{\hat{p}(1-\hat{p})/n}$ der *k* avhenger av valgt signifikansnivå.
Holder vi fast ved 95 prosent signifikansnivå får vi $\pm 1{,}96 \cdot \sqrt{\frac{0{,}10(1-0{,}10)}{505}} = \pm 0{,}03$. Vi ser altså at presisjonen øker når sann p endrer seg fra 0,48 til 0,10.

f) Spørsmålet er hvordan p(1-p) varierer med p. La f(p)=p(1-p). Da er f'(p) =1-2p, og f'(p)=0 gir løsningen p=0,5. Siden f''(p)=-2 er dette et maksimum. Følgelig er variansen til estimatet høyest, og dermed presisjonen lavest, når sann p=0,5.

g) La M= Respondenten *mener* Ja og S= Respondenten *svarer* Ja. Da er

$\mathbf{p} = P(S) = P(M) \cdot P(S|M) + P(M^c) \cdot P(S|M^c) = \mathbf{r} \cdot 5/6 + (1-\mathbf{r}) \cdot 1/6 = 1/6 + 2/3 \cdot \mathbf{r}$, som gir $\mathbf{r} = \frac{3}{2}\mathbf{p} - \frac{1}{4}$.

h) X er fordelt binomisk(n, **p**). Da er $\hat{p} = \frac{X}{n}$ en forventningsrett estimator for **p** og dermed $\hat{r} = \frac{3}{2}\hat{p} - \frac{1}{4} = \frac{3X}{2n} - \frac{1}{4}$ en forventningsrett estimator for **r**.

i) Variansen til $\hat{r}$ uttrykt ved $\mathbf{n}$ og $\mathbf{r}$ finner vi ved
$\mathrm{var}(\hat{r}) = \mathrm{var}\left(\frac{3}{2}\hat{p} - \frac{1}{4}\right) = \left(\frac{3}{2}\right)^2 \mathrm{var}(\hat{p}) = \frac{9}{4} \cdot \frac{\mathbf{p}(1-\mathbf{p})}{\mathrm{n}}$
$= \left(\frac{1}{4} + \mathbf{r}\right)\left(\frac{5}{4} - \mathbf{r}\right) \cdot \frac{1}{\mathrm{n}} = \left(\frac{5}{16} + \mathbf{r}(1-\mathbf{r})\right) \cdot \frac{1}{\mathrm{n}}.$

[Merk at dette uttrykket har sitt maksimum for $\mathbf{r}$=1/2, jfr. spørsmål f.]

Oppgave 2

a) Referansekategorien er enbedriftsforetak som er norskeid og privat.

b) Vi ser at produktivitetsnivået i hovedbedrifter ikke er signifikant forskjellig fra referansekategorien. Filialbedrifter har derimot signifikant lavere produktivitet. Siden koeffisienten på dummyvariabelen gir oss endring i log verdiskapning medfører estimatet at produktivitetsnivået er ca 20 prosent lavere i filialbedriftene. Vi ser videre at utenlandskeide bedrifter har ca 20 prosent høyere produktivitet enn referansekategorien, mens offentlig eide bedrifter har ca 10 prosent lavere produktivitet. Det siste estimatet er imidlertid for upresist til å være signifikant. Økt bruk av innsatsfaktorene arbeidskraft og kapital gir økt verdiskapingen slik man vil forvente. Med log-log funksjonsform har koeffisientene tolkning som elastisiteter. En 1 prosent økning i timeverk gir 0,91 prosent økning i verdiskapning, mens en 1 prosent økning i kapitalbruk gir en 0,05 prosent økning i verdiskapning. [Tilsynelatende har vi avtakende skalautbytte siden 0,91+0,05<1.] Forklaringsgraden til regresjonen målt ved justert R^2 er god. De seks forklaringsvariablene våre kan forklare 87 % av variasjonen i verdiskapningen i utvalget. [Det er ikke naturlig å gå inn på tolkning av konstantleddet, men merk at den logaritmiske transformasjonen påvirker en evt. tolkning. lnK og lnL er null for hhv K=1 og L=1. lnA negativ betyr at A er mellom null og 1.]

c) Forklaringsgraden er forklart variasjon over total variasjon, dvs SS(Regression)/SS(Total)=1586,50/1827,27=<u>86,82</u>.

d) Vi ser at vi har problemer med normalitet i feilleddene. Fordelingen har imidlertid klar klokkefasong, og antall observasjoner er så stort at det neppe er grunn til bekymring for at T-testene ikke skal være gyldige. "Residuals vs fits" indikerer heteroskedastisitet og at variansen til feilleddet faller med størrelsen på bedriftene. [Dette virker konterintuitivt og det visuelle bildet kan skyldes at vi har langt flere observasjoner – og dermed flere avvikende residualer – for små bedrifter.] Autokorrelasjon ("residual vs order") er ikke relevant da utvalget er et tverrsnitt.

e) Forskjellen i produktivitet mellom en privat, utenlandskeid bedrift og en offentlig, norskeid bedrift er 0,201-(-0,096)=0,297, dvs <u>29,7 prosent</u>.

f) Dette kan testes ved å innføre et interaksjonsledd: D^{UO} = *utenl*offentl*.

g) Et 95% konfidensintervall for kapitalelastisiteten, β er $\hat{\beta} \pm 1{,}96 \cdot SE(\hat{\beta}) = 0{,}050 \pm 0{,}027$.

h) H_0: β=0,25, H_A: β≠0,25. T=(0,05-0,25)/0,014 = −14,3. Vi ser at nullhypotesen forkastes for alle konvensjonelle signifikansnivåer.

Oppgave 3

a) Vi har at $\ln V_i = \ln A + \alpha \ln L_i + \beta \ln K_i + \varepsilon_i$ og $\ln K_i^* = \ln K_i + v_i$
Da er $\ln V_i = \ln A + \alpha \ln L_i + \beta \ln K_i^* + (\varepsilon_i - v_i)$
Siden v_i i feilleddet inngår i $\ln K_i^*$ blir feilleddet negativt korrelert med variabelen $\ln K_i^*$. Det er brudd på en av forutsetningene for minste kvadraters metode og trekker estimatet for β nedover mot null.

b) La $\ln K^B$ og $\ln K^I$ være de to kapitalmålene. De har samme målestøy, v, med varians σ^2. Variansen til gjennomsnittet er da

$\mathrm{var}(0{,}5\ln K^B + 0{,}5\ln K^I) = 0{,}25\mathrm{var}(\ln K^B) + 0{,}25\mathrm{var}(\ln K^I) = 0{,}25\mathrm{var}(v) + 0{,}25\mathrm{var}(v) = 0{,}25(\sigma^2 + \sigma^2) = \underline{0{,}5\,\sigma^2 < \sigma^2}.$

c) Vi ser at koeffisienten til gjennomsnittet av de to kapitalmålene (i kolonne 3) er større enn koeffisienten til de to kapitalmålene brukt hver for seg (i kolonne 1 og 2). Dette er konsistent med at gjennomsnittet er beheftet med minst målestøy, og at det dermed er minst negativ skjevhet i koeffisienten til gjennomsnittsverdien.
[Resultatene indikerer også at det kapitalmålet som er basert på investeringstall er beheftet med mindre målestøy enn det kapitalmålet som er basert på brannforsikringsverdier. Forklaringsgraden til regresjonene understøtter også at kvaliteten til variablene øker mot høyre i tabellen. Videre kan vi merke oss at koeffisienten til lnL påvirkes fordi L og sann K er korrelert. Stor målestøy i kapitalbruken fører til at effekten av kapital underestimeres og timeverksbruken plukker da opp noe av den "utelatte" effekten.]

Generelt: Det forventes ikke at studenten skal ha med poenger som er satt i klammeparentes.