

Tentti
23.10.2019

Tentissä on kolme tehtävää, joissa on yhteensä 18 alatehtävää. Kunkin alatehtävän lopussa kerrotaan, kuinka monta pistettä oikeasta vastauksesta saa. Vastaa kuhunkin tehtävään lyhyesti ja perustele vastauksesi täsmällisesti. Huomaa, että pitkästä vastauksesta ei saa lisäpisteitä, merkitystä on vain vastauksen sisällöllä.

1. Tarkastellaan lineaarista regressiomallia $Y_i = \beta_0 + \beta_1 X_i + u_i$, jossa Y_i on työntekijän tuntipalkka (dollareina) ja X_i on työntekijän koulutus (vuosina). Pienimmän neliösumman (PNS; OLS) menetelmä, perustuen 1748 miehen ja 1202 naisen satunnaisotokseen, tuottaa seuraavat tulokset:

$$\hat{Y}_i = -3.00 + 2.00 \times X_i, R^2 = 0.13, SER = 8.8,$$

jossa β_0 :n ja β_1 :n estimaattorien varianssit ovat 1.5^2 ja 0.5^2 .

- Tulkitse tuloksia lyhyesti (β_0 :n ja β_1 :n estimaatit, niiden keskivirheet, R^2 ja SER). (2p.)
 - Aseta nollahypoteesi ja vaihtoehtoinen hypoteesi regressiomallin kulmakertoimelle ja perustele lyhyesti, miten päädyit asettamiisi hypoteeseihin. (2p.)
 - Laske t -testisuureen (t -statistic) arvot asettamillesi nollahypoteeseille ja kerro, mitä jakaumaa t -testisuureet suurin piirtein noudattavat (perustele vastauksesi lyhyesti). (1p.)
 - Selosta lyhyesti, miten t -testisuureen jakauma liittyy β_1 :n estimaattorin jakaumaan. (1p.)
 - Arvioi, mikä on kulmakerrointa koskevaan hypoteesiin liittyvä p -arvo (p -value) ja selosta lyhyesti, mikä p -arvo on. (2p.)
 - Mikä on kahden koulussa vietetyn lisävuoden vaikutus palkkaan? Laske ”ennusteelle” 95% luottamusväli (confidence interval). (1p.)
 - β_0 :n ja β_1 :n estimaattorien varianssit ovat heteroskedastisuuskorjattuja. Selosta lyhyesti, miksi heteroskedastisuuskorjattujen keskivirheiden (varianssien) käyttö on suositeltavaa. (1p.)
2. Tarkastellaan lineaarista regressiomallia $Y_i = \beta_0 + \beta_1 X_i + u_i$, jossa Y_i on työntekijän tuntipalkka (dollareina) ja X_i on binäärimuuttuja (binary variable) joka saa arvon 0, jos henkilö on mies, ja arvon 1 kaikissa muissa tapauksissa. Pienimmän neliösumman (PNS; OLS) menetelmä perustuen 1748 miehen ja 1202 naisen satunnaisotokseen tuottaa seuraavat tulokset:

$$\hat{Y}_i = 20.00 - 4.00 \times X_i, R^2 = 0.013, SER = 9.3,$$

jossa $\beta_0:n$ ja $\beta_1:n$ estimaattorien varianssit ovat 5.0^2 ja 2.0^2 .

- a) Mikä on miesten ja naisten välinen palkkaero (gender gap)? (0.5p.)
 - b) Luennoitsija väittää, että palkkaero ei ole tilastollisesti merkitsevä. Aseta sopiva nollahypoteesi (ja vaihtoehtoinen hypoteesi) testataksesi luennoitsijan väitettä. (0.5p.)
 - c) Laske t -testisuureen arvot ja niihin liittyvät p -arvot (arvio riittää) asettamillesi nollahypoteeseille. Onko palkkaero tilastollisesti merkitsevä (statistically significant)? (0.5p.)
 - d) Muodosta 95% luottamusväli miesten ja naisten väliselle palkkaerolle. (0.5p.)
 - e) Mikä on miesten keskipalkka otoksessa? Entä naisten? (1p.)
 - f) Luennoitsija väittää, että palkkaeron estimaatti on harhainen, koska naiset jäävät huomattavasti useammin vanhempainvapaalle kuin miehet ja koska työkokemus vaikuttaa palkkaan. Tarkastele luennoitsija väitettä pienimmän neliösumman estimaattorin oletusten (the least square assumptions) avulla. (2p)
 - g) Oletetaan, että luennoitsija on oikeassa, ja palkkaeron estimaattori on harhainen. Selosta lyhyesti, miten ratkaisisit ongelman? (1p.)
3. Olennaisen selittävän muuttujan pois jättämisestä regressiomallista $Y_i = \beta_0 + \beta_1 X_i + u_i$ aiheutuu ongelmia, jotka voidaan matemaattisesti tiivistää seuraavan kaavan avulla:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \frac{\sigma_u}{\sigma_X},$$

jossa $\rho_{Xu} = \text{corr}(X_i, u_i)$, $\sigma_u^2 = \text{var}(u_i)$ ja $\sigma_X^2 = \text{var}(X_i)$. Käytä kyseistä kaavaa hyväksesi vastatessasi seuraaviin tehtäviin:

- a) Selosta lyhyesti, mitä puuttuvan muuttujan harhalla tarkoitetaan (Omitted variable bias). (2p.)
- b) Selosta lyhyesti, mistä puuttuvan muuttujan harha aiheutuu. (2p.)
- c) Selosta lyhyesti, miten ongelma voidaan välttää (anna esimerkit ideaalisesta keinosta ja käytännössä usein käytetystä keinosta). (2p.)
- d) Selosta lyhyesti, miten pois jätetyn muuttujan harha liittyy kausaalisuuspäätelyyn. (2p.)