

Háðar mælingar og t-próf milli óháðra úrtaka

Ásdís Bergþórsdóttir

Sjálfstætt starfandi

Netfang: asb23@hi.is

Eitt vinsælasta tölfræðipróf, sem sálfræðingar nota, er t-próf á milli óháðra úrtaka. Sálfræðingar og aðrir grípa oft til þess við skýrslugerð og athuganir en ekki bara í formlegum rannsóknum. Það er vel skiljanlegt enda er það afkastamikið. T-próf milli óháðra úrtaka er þýðisbundið próf og því þurfa forsendur þýðisbundinna prófa að vera til staðar til að beita meg t-prófi. Ein af þeim forsendum er að mælingarnar þurfa að að vera óháðar (Moore, McCabe og Craig, 2012). Til þess að tryggja að mælingarnar séu óháðar er oft tekið tilviljunarúrtak. Hins vegar er það stundum illgerlegt vegna þess að afköst eru háð stærð úrtaks og því getur hreint tilviljunarúrtak verið of lítið. Hvað gerist ef við tökum úrtak þar sem mælingar eru háðar? Skólaumhverfi er dæmi um umhverfi þar sem mælingar geta verið háðar. Svipaðir nemendur úr sama skóla mynda klasa, það er að segja þeir hópa sig oft normalt dreift í kringum meðaltal vegna þess að þeir deila skólaumhverfi, deila kennurum, nemendur hafa áhrif hvorn á annan, koma úr sama hverfi og þar með úr svipuðu félagslegu umhverfi (Snijders og Bosker, 2004). Séu nemendur í sama bekk deila þeir einnig sömu kennurum, sömu kennslustundum og sama bekkjaranda.

Ef ég gef mér að einkunnir barna í sama skóla eða sama bekk séu óháðar þá gef ég mér að engin tengsl séu á milli skólans og námsframmistöðu nemenda (Snijders og Bosker, 2004). Í rannsókn er ekki eðlilegt að gefa sér slíka forsendu. Ef gögnin eru hins vegar háð og einkunn eins nemenda tengist að einhverju leyti einkunn annars nemenda þá er fylgni milli einkunna þeirra. Þessi fylgni er kölluð innri fylgni (e.

intraclass correlation) táknud með bókstafnum ρ . Hún er skilgreind sem breytileiki milli hópa deilt með heildarbreytileika. Hér má sjá formúluna fyrir ρ þar sem σ_m er breytileiki milli klasa en σ_i er breytileiki innan klasa.

$$\rho = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_i^2}$$

Ef ρ er 1 er enginn breytileiki innan klasa (allir nemendur í sama bekk eru með sömu einkunn) en ef ρ er 0 er enginn breytileiki milli klasa (nemendur í sama bekk fá ekki svipaðar einkunnir) og þá eru mælingarnar óháðar. Ef innri fylgni er 0 má taka t-próf. Ef við tökum t-próf á háð gögn þá aukast líkur á því að fá marktækt fyrir tilviljun ef munur í þýði er ekki til staðar („type I error“).

Hedges (2007) setti fram formúlur til að leiðrétta p-gildi úr t-prófum þar sem innri fylgni er til staðar. Dorman (2008) notaði formúlur hans til að sýna fram á að aðeins lítil innri fylgni getur haft mikil áhrif á p-gildið. Hann gaf sér dæmi um samanburð á milli tveggja hópa barna þar sem hvor hópur samanstóð af nokkrum 20 barna bekkjum og innri fylgnin var 0,20. Hann sýndi fram á að ef hóparnir voru bornir saman með t-prófi, sem leit fram hjá því að gögn væru háð, og t-prófið gæfi p-gildið 0,37 þyrfti að leiðrétta það í 0,05 miðað við $\alpha = 0,05$. Dorman (2008) skoðaði líka hvað skekkir p-gildin mest þegar t-próf er tekið á háð gögn. Eins og við var að búast skiptir innri fylgnin máli en hins vegar hefur fjöldi klasa takmörkuð áhrif. Ef til vill

kemur mest á óvart að fjöldi í hverjum klasa skiptir máli. Ef nemendur eru 25 í hverjum bekk, innri fylgni 0,20 og t-próf gefur $p=0,418$ þarf að leiðrétta það í 0,05 miðað við $\alpha=0,05$.

Þannig er ljóst að í rannsóknum þarf annað hvort að leiðrétta p-gildið úr t-prófi eða beita öðrum tölfræðiaðferðum en t-prófi ef innri fylgni er til staðar. Tölfræðiaðferðin, sem mælt er með í stað t-prófs, er Margstigagreining (e. *multilevel analysis*) (Olthof, 2015; Snijders og Bosker 2004).

Olthof (2015) skoðaði hversu miklu getur munað á á margstigagreiningu og t-prófi. Hún skoðaði mállega getu barn í tveimur mismunandi tegundum af skólum, almennum skólum ($M=37,1$) og mótmælendaeinkaskólum ($M=42,7$). Þegar hún tók t-próf fékk hún út $t(250)=-4,71$ og $p<0,00001$. Þegar innri fylgni bekkja var skoðuð kom í ljós að hún var 0,29. Þegar Olthof notaði margstigagreiningu kom í ljós að einfalt líkan, þar sem gert var ráð fyrir að bekkirnir dreifðust um meðaltalið 39,4 með staðalfrávik 5,0, útskýrði best dreifingu mállegrar getu. Áhrifin af skólagerð voru ekki marktæk ($p=0,06$) og slíkt líkan var ómarktækt ($p=0,052$).

Að öllu töldu er ljóst að það að taka próf á háð gögn veldur því að ef enginn munur er á hópunum aukast líkur á því að fá marktækni fyrir tilviljun. Þó svo að hér hafi verið fjallað um nemendur og bekki þá á umfjöllunin við um öll tilfelli þar sem gögnin innihalda klasa. Ætíð þarf að skoða hvort innri fylgni sé til staðar. Sé hún til staðar er best að beita margstigagreiningu. Lesum skýrslur og rannsóknir með þetta í huga og höfum fyrirvara á niðurstöðum ef menn eru að taka t-próf á gögn þar sem klasar geta verið til staðar og rökstyðja ekki þá ákvörðun með því að sýna fram á að mælingar séu óháðar. Sýnum gott fordæmi, reynum að beita þeim aðferðum sem heppilegastar eru taldar.

Heimildir

- Dorman, J. P. (2008). Conducting statistical tests with data from clustered school samples. *International Journal of Research & Method in Education*, 31(2), 113–124.
- Hedges, L. V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, 32(2), 151–179.
- Moore, D. S., McCabe, G. P. og Craig, B. A. (2012). *Introduction to the Practice of Statistics* (6. útgáfa). New York, NY: W.H. Freeman & Company.
- Olthof, M. C. (2015). The analysis of clustered data. Í H. Adèr og G. Mellenbergh (ritstjórar), *Advising on research methods: Selected topics 2014*. Huizen: Johannes van Kessel Advising.
- Snijders, T. A. og Bosker, R. (2004). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (5. útgáfa). Trowbridge, Wiltshire: Cromwell Press Limited.