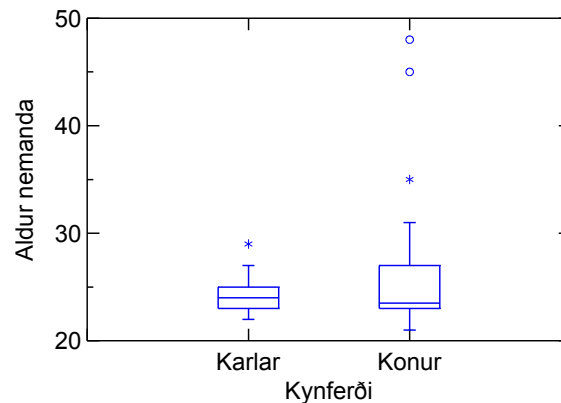


Eftirfarandi er yfirlit yfir spurningar sem notaðar hafa verið í prófum námskeiðinu.

Verkefni sem eru þess eðlis að þau myndu aðeins birtast í heimaprófi eru auðkennd með H. Almennt eru þetta þau verkefni sem krefjast aðgangs að tölfræðiforritum eða mikilla útreikninga.

1. Hvaða upplýsingar gefur  $R^2$  í aðfallsgreiningu?
2. Hver er munurinn á leif (*residual*) og villu (*error*) í aðfallsgreiningu? Hvaða kröfur eru gerðar til villunnar? Hvaða kröfur eru gerðar til leifarinnar?



3. Kassaritið sýnir aldursdreifingu í námskeiði á öðru ári í sálfræði.

Gerðu ráð fyrir því að þú viljir kanna aldursmun karla og kvenna meðal nemenda í sálfræði með viðeigandi marktektarprófi. Þessi mynd er fyrsta skrefið í úrvinnslunni. Lýstu henni, þ.e. öllum mikilvægustu atriðum hennar, með hliðsjón af því að þú vilt kanna hvort meðalaldur sé ólíkur eftir kyni.

4. „Staðalfrávik er breytilegt milli úrtaka og því fáum við aðeins ónákvæmt mat á staðalvillunni. Við bregðumst við með því að nota  $t$ -dreifingu til að lýsa dreifingu meðaltala ...“ Útskýrðu þetta fyrir mér, þ.e. hvers vegna er staðalfrávik breytilegt, hvaða afleiðingar hefur það fyrir staðalvilluna, hvers vegna og hvernig er  $t$ -dreifing ólík normaldreifingu? Ég einfaldlega *skil þetta alls ekki!*
5. Í úrtaki 40 karla og 56 kvenna sögðust 67% kvennanna stunda líkamsrækt en aðeins 61% karlanna.

Prófaðu þá tilgátu að ástundun líkamsræktar sé ólík eftir kyni. Gefðu upp niðurstöðu viðeigandi tölfræðiprófs, gefðu til kynna sennilegan mun á kynjunum og tilgreindu efnislega heildarniðurstöðu um þetta úrlausnarefni.

6. Geirhart geðhjúkrunarfræðingur gerði könnun á áhrifum gönguferða á depurð. Hann fékk 20 þunglyndissjúklinga til þátttöku í rannsókn. Helmingur þeirra (valdir með tilviljunaraðferð) fóru í röska göngutúra en hinir horfðu á gamla þætti af Staupasteini (*Cheers, a popular sitcom, aired between 1982 and 1993*)—allt undir strangri leiðsögn hans. Að loknu sex vikna rannsóknartímabili lagði Geirhart depurðarpróf fyrir báða hópa og bar niðurstöður saman.

Gögnin litu vel út, meðferðarhópurinn (gönguferðir) var með minna þunglyndi að meðferð lokinni en samanburðarhópurinn (sjónvarpsgláp) en ekki nógu mikið til að  $t$ -prófið væri marktækt. Niðurstæða hans var því sú að

gönguferðir væru ekki álitlegt meðferðarúrræði við depurð eða a.m.k. engu betri en áhorf á andstyggilegheit (*nastiness*) vinsælla sjónvarpsstjarna.

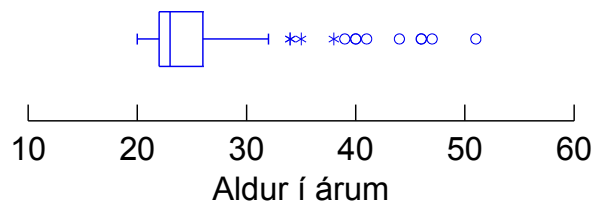
Ef aðaltilgátan er að depurð sé ólík eftir hópum en núlltilgátan að depurðin sé jafnmikil hjá báðum hópum, er ályktun Geirharts rétt? Rökstyddu nákvæmlega.

7. Í óformlegri skoðanakönnun meðal nemenda félagsvísindadeildar sögðust 10 nemendur af 245 myndu kjósa F-lista. Gerðu ráð fyrir því að þessir 245 séu tilviljunarúrtak úr nemendum deildarinnar.

Gefðu upp mat þitt á fylgi listans í þýði miðað við 95% öryggi með bæði normalnággun og leiðréttu (*modified Wald; Wilson estimate*) öryggisbili. Þú færð ólíkar niðurstöður eftir aðferðum. Í hverju er munurinn fólgin nákvæmlega (gefðu upp bæði tölulegt og efnislegt svar) og hvor aðferðin er réttari í þessu tilviki? Rökstyddu.

8. „Marktektarpróf byggist á aðaltilgátu,  $H_a$ . Sú tilgáta tilgreinir engin sérstök þýðisgildi og er því ekki prófanleg.“ *Ekki* svara með annarri orðréttri tilvitnun í glæru; segðu mér frekar hvað þetta þýðir. Hvernig má það vera að aðaltilgátan tilgreini ekki sérstakt þýðisgildi og hvers vegna leiðir af því að hún sé ekki prófanleg?
9. Fróði kannaði greindarfar kven- og karlnema í sálfræði við Háskóla Íslands skólaárið 2004–2005. Hann reiknaði  $t$ -próf sem reyndist ómarktækt. Þegar hann athugaði 95% öryggisbil, kom í ljós að það var mjög þröngt; heildarbreidd þess var aðeins um 4 greindarvísitölustig.

Hvaða ályktun getur Fróði dregið um greindarfar kvenna samanborið karla meðal nemenda í sálfræði? Rökstuddu nákvæmlega.



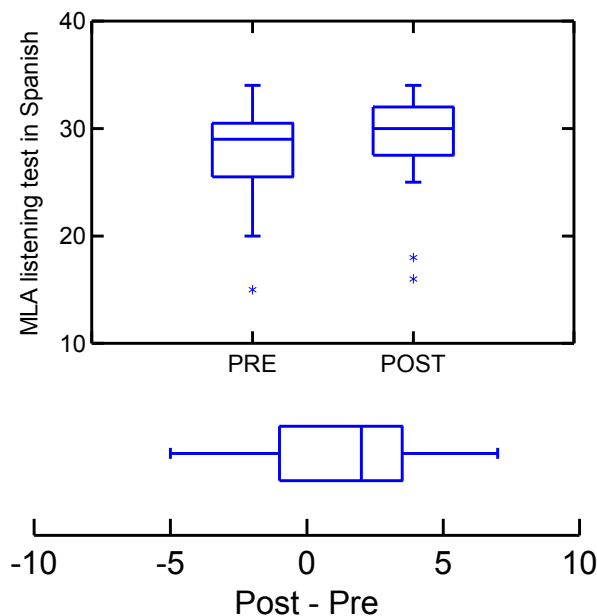
10. Kassaritið sýnir aldursdreifingu í Tölfræði II vorið 2005. Gerum ráð fyrir því að þú viljir athuga hvort það sé munur á aldri karla og kvenna í námskeiðinu. Hvort myndi henta betur að nota  $t$ -próf í tveimur hópum eða Wilcoxon Mann-Whitney prófið? Rökstyddu nákvæmlega með tilvísun til myndarinnar, tilgreindu m.a. hvaða atriði ráða úrslitum varðandi ákvörðun þína og hvaða sjónarmið liggja þar að baki. Gættu þess sérstaklega að hafa hliðsjón af umræðu um kosti og galla þessara prófa.
11. Á glærunni Takmarkanir tilgátuprófunar í Inngangi að ályktunum segir: „Ef núlltilgátu er hafnað ber að líta svo á að aðaltilgátan sé rétt. Það þýðir ekki að hún sé raunverulega rétt né að mikil frávik séu frá  $H_0$ .“ Útskýrðu muninn á því að (a) geta hafnað núlltilgátu og (b) því að hún sé raunverulega röng. Hvers vegna fer þetta ekki saman. Útskýrðu einnig (c) mikilvægi þess að frávik frá rangri núlltilgátu gæti verið óverulegt.
12. „Marktektarpróf byggist á aðaltilgátu,  $H_a$ . Sú tilgáta tilgreinir engin sérstök þýðisgildi og er því ekki prófanleg.“ *Ekki* svara með annarri orðréttri tilvitnun í glæru; segðu mér frekar hvað þetta þýðir. Hvernig má það vera að aðaltilgátan tilgreini ekki sérstakt þýðisgildi og hvers vegna leiðir af því að hún sé ekki prófanleg?

13. Lestu vel dæmi 7.20 á bls. 501–502 í 5. útgáfu kennslubókarinnar (bls. 539–541 í 4. útgáfu bókarinnar). Notaðu afkastareikni UCLA til að reikna afköst prófsins í dæminu en hann finnurðu á slóðinni <http://calculators.stat.ucla.edu/powercalc/normal/n-2-equal/>. H

Gerðu ráð fyrir að lyfleysuhópurinn hafi engum framförum tekið en hjá kalsíumhópnum hafi blóðþrýstingur lækkað að meðaltali um 5 einingar. Gerðu einnig ráð fyrir að sameiginlegt staðalfrávik sé 7,5. Settu þessar upplýsingar inn í afkastareikninn; gættu þess að setja spurningamerki inn fyrir Power og velja tvíhliða próf. Reiknaðu svo afköstin fyrir nokkrar úrtaksstærðir: Byrjaðu á 10 þátttakendum í hvorum hópi, hækkaðu síðan um 5 í hverju skrefi þar til þú ert komin upp í 55 þátttakendur í hvorum hópi. Skráðu hjá þér niðurstöðurnar og birtu myndrænt með fjölda í hvorum hópi á láréttum en afköstin á lóðréttum ási myndrísins. (Dæmi um svona myndrit er undir *power* í 5. útgáfu Orðgnóttar og einnig á [http://www.indiana.edu/~statmath/stat/all/power/power\\_plot.jpg](http://www.indiana.edu/~statmath/stat/all/power/power_plot.jpg).) Ef þú treystir þér ekki til að gera myndritið, settum við okkur við APA töflu með sömu upplýsingum.

Endurtaktu nú leikinn miðað við að blóðþrýstingslækkunin sé aðeins 3 einingar (en ekki 5 eins og áður) hjá kalsíumhópnum.

- (a) Birtu niðurstöðurnar myndrænt eða í töflu fyrir hvort tilvik fyrir sig.
- (b) Lýstu afköstum *t*-prófsins eins og þau birtast á myndritunum tveimur. Útskýrðu muninn á niðurstöðunum.
- (c) Tilgreindu hvað þú þarft marga þátttakendur til að ná 80% afköstum miðað við hvort tilviki fyrir sig. Athugaðu að þetta gæti krafist viðbótarúrvinnslu.



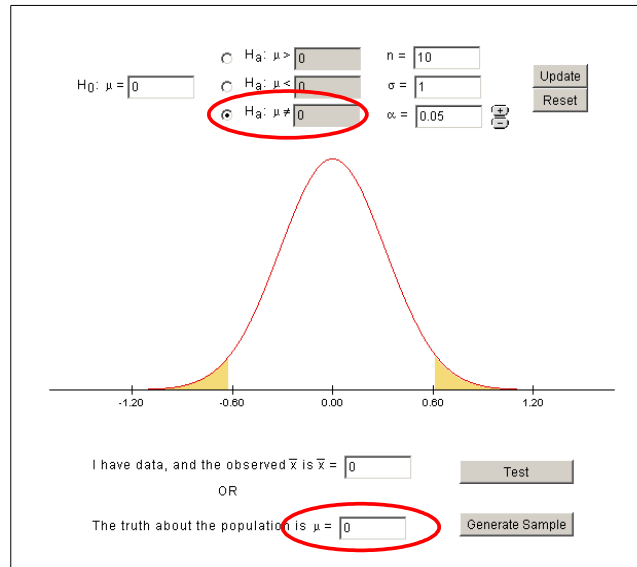
14. Fyrri myndin sýnir kassarit fyrir forpróf og eftirpróf tungumálaprófs sem lýst er í dæmi 7.42 á bls. 483 í 5. útgáfu kennslubókarinnar (bls. 522 í 4. útgáfu bókarinnar). Seinni myndin sýnir kassarit fyrir framfarirnar milli for- og eftirprófsins hjá þessum tuttugu kennurum.

- (a) Lýstu myndunum hvorri fyrir sig og gættu þess að gera grein fyrir öllum þeim atriðum sem almennt gætu skipt máli til að meta forsendur *t*-prófa,

einnig þeirra sem hugsanlega skipta litlu eða engu máli í samhengi við kringumstæðurnar sem lýst er í dæmi 7.42. Gættu þess að tiltaka nákvæmlega mikilvæga þætti sem sjást á kassaritum, skipta almennt máli þegar  $t$ -próf á í hlut og eru hugsanlega í lagi á þessum myndum, þ.e. frávik sem *ekki* eru fyrir hendi á myndinni.

(b) Kynntu þér vel þær forsendur sem gilda um *parað*  $t$ -próf sérstaklega og gerðu skýra og skilmerkilega grein fyrir því hvort þær eru eða eru ekki til staðar í þessu tilviki. Gættu þess að rökstyðja mál þitt allnákvæmlega og vísa meðal annars til myndanna.

15. Í hvaða tilvikum er betra að nota Wilcoxon Mann-Whitney í stað  $t$ -prófs í tveimur óháðum hópum? Hvers vegna?



16. Farðu í marktektarforritlingin á vefsvæði kennslubókarinnar á slóðinni [http://bcs.whfreeman.com/ips5e/content/cat\\_010/applets/statsig.html](http://bcs.whfreeman.com/ips5e/content/cat_010/applets/statsig.html).

Veldu tvíhliða próf eins og sést á myndinni og smelltu nokkrum sinnum á neðsta takkann sem er auðkenndur með „Generate Sample.“ Útskýrðu hvað er að gerast: Fyrir hvað stendur normalferillinn á myndinni, hvað tákna skyggðu svæðin í báðum hölum hans, fyrir hvað stendur bláa línan og bláu punktarnir sem birtast og nákvæmlega hvað þýðir það þegar stendur neðst ýmis „Significant ...“ eða „Not significant ...“ ýmist með rauðu eða grænu letri fyrir neðan myndina? H

17. Farðu aftur í marktektarforritlinginn sem þú notaðir í spurningu 16 en ritaðu núna 0.5 (passaðu þig á að nota *punktur* en ekki kommu) þar sem stendur „The truth about the population is  $\mu =$ “. Smelltu svo nokkrum sinnum á „Generate Sample.“ Hvað hefur breyst? H

(a) Smelltu 20 sinnum með  $\mu = 0.5$  og síðan 20 sinnum með  $\mu = 0$ . Teldu hversu oft niðurstaðan er marktæk í hvoru tilviki fyrir sig og gefðu upp niðurstöðuna. Hver er munurinn og nákvæmlega hvers vegna kemur hann fram?

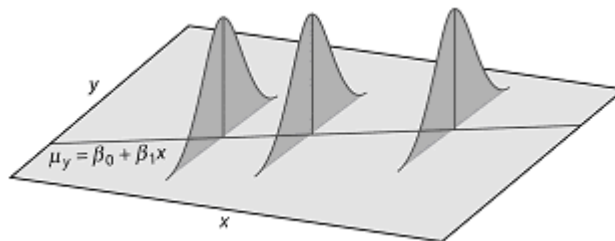
(b) Hvað myndi gerast ef við hefðum annað  $\mu$ , annað  $n$ , annað  $\sigma$  eða annað  $\alpha$ ? Tilgreindu stuttlega áhrif hvers af þessum fjórum þáttum fyrir sig.

18. Jónmundur kannaði biturleika hjá 4 þunglyndum einstaklingum og 3 gláðsinna sálfræðinemum. Hann komst að því að þótt mun minni biturleika gætti hjá sálfræðinemunum heldur en hjá þeim þunglyndu, var munurinn ekki marktækur samkvæmt  $t$ -prófi. Steinunn benti honum á að afköstin (*power*)

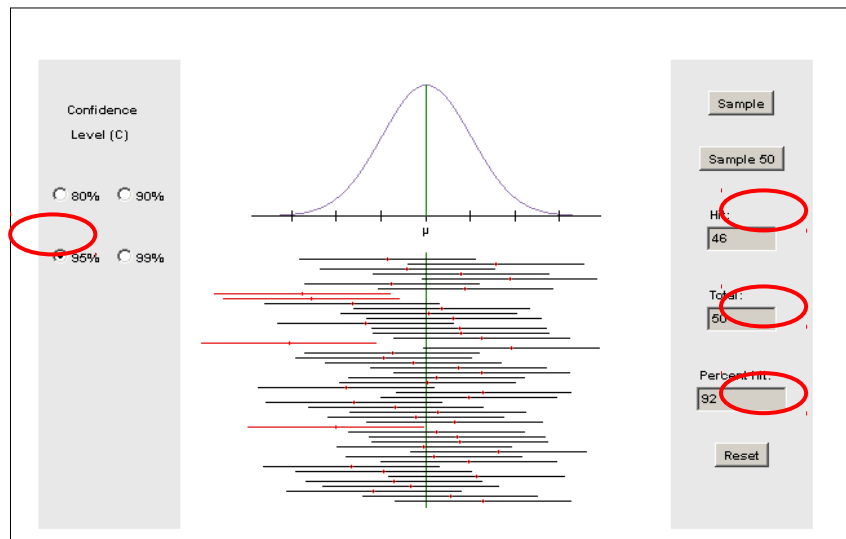
væru of lítil. Hann endurtók því rannsóknina með 40 þunglyndum og 30 sálfræðinemum. Þeim til mikillar undrunar var munurinn samt sem áður ómarktækur.

Útskýrðu tvennt: (a) Hafði Steinunn rétt fyrir sér og (b) hvernig stóð á því að niðurstaðan var ekki marktæk í síðara skiptið—höfðu afköstin kanski ekkert aukist þrátt fyrir allt?

19. Yfirgreining (*metaanalysis*) gefur yfirlit yfir niðurstöður þar sem margar rannsóknir eru teknar saman. Hvaða vanda í tengslum við tilgátuprófun (*hypothesis test*) leysir þetta? Útskýrðu nákvæmlega.
20. Þegar *t*-próf er framkvæmt þarf að reikna eitthvað sem er kallað staðalvilla meðaltala (*standard error of the mean*). Hvað er þetta nákvæmlega, hvaða stærð í þýði er verið að meta (*estimate*) og hvaða fyrirbæri eða eiginleika í þýði mælir sú stærð? Útskýrðu nákvæmlega.
21. Þegar á allt er litið, þá er mun betra að nota Wilcoxon Mann-Whitney prófið heldur en hefðbundið *t*-próf! Fjallaðu gagnrýnið um þessa fullyrðingu með því að tiltaka kosti—og eftir atvikum ókosti—Wilcoxon Mann-Whitney prófsins samanborið við *t*-próf.
22. Hver er núlltilgátan í kíкваðratprófi (*chi square test*) fyrir krosstöflur og hvernig nákvæmlega birtast frávik frá núlltilgátunni?
23. Ef ég vil prófa tilgátu þarf ég að setja fram núlltilgátu. Núlltilgátunni get ég hafnað en ekki staðfest.
  - (a) Útskýrðu hvers vegna ég þarf að setja fram núlltilgátu, (b) hvernig ég get hafnað henni og (c) hvers vegna það er ekki hægt að staðfesta hana.



24. Myndin sýnir einfalda aðfallsgreiningu (*simple regression*) þar sem *x* er eina frumbreytan (*independent variable*) og *y* er fylgibreytan (*dependent variable*). Jafnan (*formula*) neðst í vinstra horninu ( $\mu_y = \beta_0 + \beta_1 x$ ) lýsir tengslum frumbreytunnar við fylgibreytunnar. Útskýrðu tvennt: (a) Hvaða tvær mikilvægar forsendur (*assumptions*) aðfallsgreiningar sýnir myndin? (b) Hvaða merkingu hafa táknið  $\beta_0$  og  $\beta_1$  í formúlunni?



25. Farðu í öryggisbilaforritlingin á vefsvæði kennslubókarinnar á slóðinni: [bcs.whfreeman.com/ips5e/content/cat\\_010/applets/confidenceinterval.html](https://bcs.whfreeman.com/ips5e/content/cat_010/applets/confidenceinterval.html) H

Stilltu forritlingin á 95% öryggi með því að merkja í réttan reit í vinstri jaðri forritlingsins. Smelltu síðan á **Sample 50** takkann. Útskýrðu nákvæmlega það sem er að gerast:

(a) Hvers vegna verða sum bilin rauð en önnur svört, (b) fyrir hvað stendur græna lóðrétta línan á myndinni og gríski stafurinn  $\mu$ , (c) nákvæmlega fyrir hvað stendur normalferillinn efst á myndinni, (d) hvaða tölur færð þú í Hit, Total og Percent hit textareitina og hvað tákna þær, (e) hversu oft ættu bilin að vera rauð, (f) er það í samræmi við töluna í Percent hit og (g) ef ekki hvers vegna? [Ef það er samræmi milli e- og f-liða, ýttu þá á **Reset** og **Sample 50** á víxl þar til ósamræmi myndast.]

26. Ýttu á **Reset**. Smelltu síðan til skiptis á **Sample** og **Reset** þar til myndin sýnar eitt stakt rautt bil. H

Túlkaðu þetta staka öryggisbil eins og það væri þetta eina reiknaða öryggisbilið fyrir meðaltal sem þú færð í einni stakri rannsókn.

Tóti tölvísi túlkar niðurstöðuna svona: „Það eru 95% líkur á því að meðaltalið sé á þessu tiltekna bili.“ Svaraðu því nú ótvírætt hvort þetta sé (a) rétt túlkun og þá hvers vegna eða (b) röng og þá hvað nákvæmlega sé rangt við það sem Tóti fullyrðir.

Ýttu þá á **Reset** og **Sample 50** á víxl og skráðu hjá þér töluna í Percent hit fyrir hvert skipti og endurtaktu þetta 20 sinnum. Birtu niðurstöðurnar í einfaldri töflu með meðaltali yfir skiptin 20 nedst. Útskýrðu hvers vegna niðurstöðurnar eru breytilegar og hvers vegna meðaltalið er nær réttri tölu heldur en hver og ein af tölunum 20.

27. Skoðaðu vel glæruna Ákvörðunarreglan í fyrirlestrinum Inngangi að ályktunum.

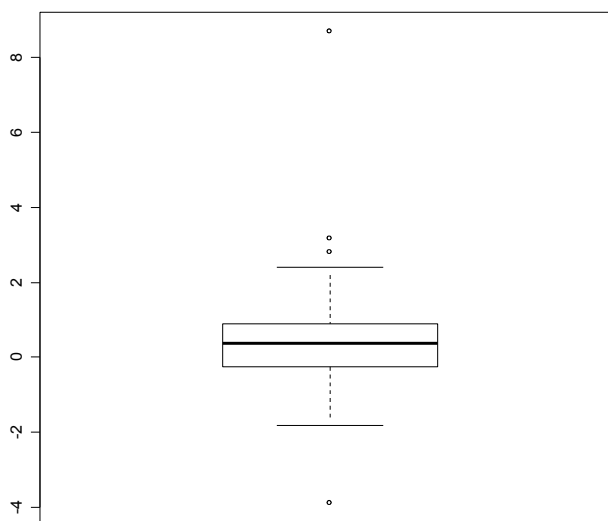
Útskýrðu fyrir mér með *eigin orðum* (a) hvað talan  $\alpha$  stendur nákvæmlega fyrir, (b) hvað sé átt við þegar sagt að hún sýni áhættuna sem er tekinn undir núlltilgátunni og (c) hvers vegna við getum hafnað núlltilgátunni þegar líkurnar á viðkomandi fráviki eða meira fráviki frá núlltilgátunni er jafnt og eða minna en  $\alpha$ ?

Komdu með raunhæft dæmi úr daglegu lífi—ekki úr kennslustund, kennslubók eða af netinu—þar sem sams konar ályktunarreglu er beitt. Útskýrðu hvers vegna dæmið er sambærilegt.

28. Á Túlkun öryggisbils ef ekki marktækt í  $t$ -prófum á meðaltöl stendur eftirfarandi: „Ef prófið er marktækt, get ég *ekki* hafnað núlltilgátunni.“

Mér skilst að með marktækt próf, geti ég hafnað núlltilgátunni og tekið upp aðaltilgátuna. En hvers vegna get ég ekki staðfest núlltilgátuna? Þarf þetta að virkilega að vera svona flókið og þarf ég alltaf að segja „get ekki hafnað“ þegar ég myndi vilja geta *staðfest* núlltilgátuna?

Útskýrðu nákvæmlega (a) hvers vegna ekki er hægt að staðfesta núlltilgátuna, (b) hvað felist nákvæmlega í því að geta hvorki hafnað henni né staðfest hana og (c) hvernig öryggisbil (*confidence interval*) hjálpar mér í svona tilvikum?



29. Myndin sýnir kassarit af einu 39 manna úrtaki úr þýði. Ætlunin er að nota eins hóps  $t$ -próf og reikna öryggisbil.

Lýstu myndinni nákvæmlega með sérstöku tilliti til þeirra atriða sem skipta mestu ef beita á  $t$ -prófi á gögnin. Fram þurfa að koma *öll* atriði sem skipta máli þegar meta þarf hvort  $t$ -próf sé viðeigandi.

Hentar  $t$ -próf fyrir þetta gagnasafn? Rökstyddu nákvæmlega með tilvísun til myndarinnar, tilgreindu m.a. hvaða atriði ráða úrslitum varðandi ákvörðun þína og hvaða sjónarmið liggja þar að baki. Gættu þess sérstaklega að hafa hliðsjón af umræðu um kosti og galla þessa prófs.

30. Óformleg könnun var gerð í kennslustund á notkun nagladekkja. Í ljós kom að af 60 nemendum sem tóku þátt og voru á bíl voru 26 á nagladekkjum. Einnig kom í ljós að af 20 körlum voru 10 á nagladekkjum en 15 af 40 konum.

Hvort bæri að nota normalnálgun hér eða leiðrétt öryggisbil? Rökstyddu nákvæmlega, tilgreindu þau atriði sem þarf að skoða og hvernig þú metur þau hvert fyrir sig.

Reiknaðu öryggisbil bæði fyrir heildina og fyrir mismun eftir kynjum. Gerðu þetta bæði með normalnálgun og með leiðréttingu. [Samtals eru þetta fjögur öryggisbil, tvö af hvoru.]



Berðu öryggisbil normalnálgunar við samsvarandi leiðrétt öryggisbil. Eru þau ólík og þá hvers vegna eða eins og tilgreindu þá ástæðu þess. Hvara nálgunina væri eðlilegra að nota?

31. Lestu vel dæmi 7.20 á bls. 501–502 í 5. útgáfu kennslubókarinnar (bls. 539–541 í 4. útgáfu bókarinnar). Notaðu afkastareikni [Rollins Brants](#) til að reikna afköst prófsins í dæminu en hann finnurðu á slóðinni <http://stat.ubc.ca/~rollin/stats/ssize/n2.html>. H

Gerðu ráð fyrir að lyfleysuhópurinn hafi engum framförum tekið en hjá kalsíumhópnum hafi blóðþrýstingur lækkað að meðaltali um 4,5 einingar. Gerðu einnig ráð fyrir að sameiginlegt staðalfrávik sé 7,3. Settu þessar upplýsingar inn í afkastareikninn; gættu þess að velja Calculate Power (for specified Sample Size) og velja tvíhliða próf.

Reiknaðu svo afköstin fyrir nokkrar úrtaksstærðir: Byrjaðu á 10 þátttakendum í hvorum hópi, hækkaðu síðan um 5 í hverju skrefi þar til þú ert komin upp í 55 þátttakendur í hvorum hópi. Skráðu hjá þér niðurstöðurnar og birtu myndrænt með fjölda í hvorum hópi á láréttum en afköstin á lódréttum ási myndrísins. (Dæmi um svona myndrit er undir [power](http://www.indiana.edu/~statmath/stat/all/power/power_plot.jpg) í 5. útgáfu Orðgnóttar og einnig á [http://www.indiana.edu/~statmath/stat/all/power/power\\_plot.jpg](http://www.indiana.edu/~statmath/stat/all/power/power_plot.jpg).) Ef þú treystir þér ekki til að gera myndritið, sættum við okkur við APA töflu með sömu upplýsingum.

Endurtaktu nú leikinn miðað við að blóðþrýstingslækkunin sé aðeins 2,9 einingar (en ekki 4,5 eins og áður) hjá kalsíumhópinum.

- (a) Birtu niðurstöðurnar myndrænt eða í töflu fyrir hvort tilvik fyrir sig.  
 (b) Lýstu afköstum  $t$ -prófsins eins og þau birtast á myndritunum tveimur. Útskýrðu muninn á niðurstöðunum.  
 (c) Tilgreindu hvað þú þarft marga þátttakendur til að ná 80% afköstum miðað við hvort tilviki fyrir sig. Athugaðu að þetta gæti krafist viðbótarúrvinnslu.

32. Í kafla 6.1 í kennslubókinni (bls. 356–372 í 6. útgáfu) er fjallað um bilspá (*interval estimation*) og öryggisbil þýðismedaltals (*confidence intervals for a population mean*). Útskýrðu með *eigin orðum* helstu eiginleika öryggisbila. Gættu að því að þú þarft að orða svarið þannig að það sé eftir þinni eigin hugsun og líkist ekki orðalagi kennslubókarinnar, Orðgnóttar, verkefnum á vefnum eða öðrum álíka heimildum. H

Gættu þess að svar þitt feli í sér skýr svör við eftirfarandi atriðum. Þú mátt gjarnan skipta svarinu í liði í samræmi við neðangreint.

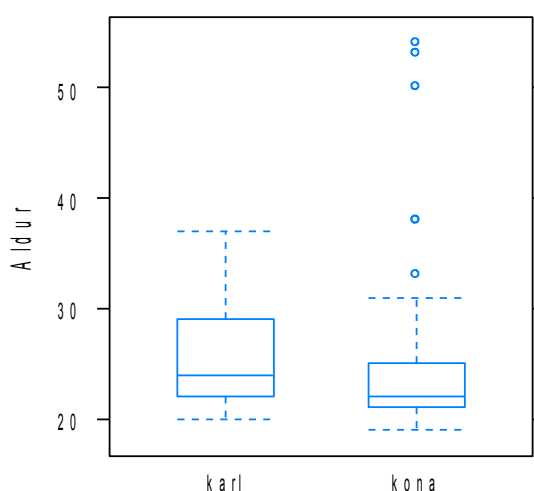
- (a) Hver er munurinn á bilspá og punktspá (*point estimation*), (b) hverjar eru líkurnar á því að eitt tiltekið (*one specific*) 95% öryggisbil innihaldi viðkomandi þýðismedaltal, (c) hvaða upplýsingar gefur öryggi (*confidence*) 95% bills, (d) hvaða áhrif (*influence*) nákvæmlega hafa úrtaksstærð (*sample size*) og breytileiki (*variability*) úrtaks á öryggisbilið og hvers vegna, (e) nákvæmlega hvaða óvissu (*villu; error*) gefur öryggisbil til kynna, (f) hvað væri dæmi um áhrifaþætti sem ekki er tekið tillit til og (g) á hvaða forsendum (*assumptions*) byggja útreikningar öryggisbils fyrir þýðismedaltal?



33. Í kennslubókinni (6. útgáfa, inngangur að kafla 6, bls. 355) er eftirfarandi útskýring á öryggisbilum (*confidence intervals*) og marktæktarprófum (*tests of significance*): H

Báðar tegundir af ályktunum byggja á úrtakadreifingu úrtakstalna. Í þessu felst að í báðum tilvikum fást líkindi sem segja til um hvað myndi gerast ef ályktunaraðferðin væri notuð mörgum sinnum. Slíkar fullyrðingar um líkindi einkenna hefðbundnar tölfræðilegar ályktanir.

Útskýrðu hvernig þetta á við um tilgátuprófun (*hypothesis testing*) og marktæktarpróf (*statistical test*). Gættu þess að í svári þínu komi m.a. svar við eftirfarandi atriðum: (a) Hvernig tengist marktæktarpróf úrtakadreifingu (*sampling distribution*), (b) hver nákvæmlega eru líkindin (*probability*) sem um ræðir, (c) í hvaða skilningi er eitthvað gert *oft* og (d) nákvæmlega hvaða ályktun er hægt að draga?



34. Myndin sýnir aldursdreifing eftir kyni í úrtaki nemenda í námskeiðinu Tölfræði II. Gerum ráð fyrir því að þú viljir athuga hvort meðalaldur sé ólíkur eftir kynjum. Eftirfarandi tafla gefur yfirlit yfir niðurstöður í úrtakinu. H

Kyn	Meðaltal	Staðalfrávik	Fjöldi
Karlar	25,1	4,0	28
Konur	25,0	7,8	55
Heild	25,1	5,9	83

- (a) Lýstu nákvæmlega þeim atriðum sem skipta máli fyrir forsendur  $t$ -prófs og sambærilegra úrvinnsluáðferða. Gættu þess að tilgreina bæði þau atriði sem eru í lagi og þau sem gætu verið í ólagi.
- (b) Taktu afstöðu til þess hvort Wilcoxon's Mann-Whitney prófið myndi henta betur en  $t$ -próf. Rökstyddu nákvæmlega með hliðsjón af myndinni og upplýsingum í tölflunni.
- (c) Reiknaðu  $t$ -próf á grundvelli upplýsinganna, birtu niðurstöðuna og túlkadu hana bæði tölfræðilega og efnislega. Gefðu einnig upp öryggisbil og túlkadu það efnislega og tölfræðilega. (Við mælum með því að helstu útreikningar fylgi svo hægt sé að taka tillit til klaufavillna.)

35. Lestu glærurnar **Er hægt að sanna núlltilgátuna?**, **Takmarkanir tilgátu-prófunar** og **Takmarkanir stakra rannsókna** í fyrirlestrinum **Inngangi að ályktunum** og glæruna **Túlkun öryggisbils ef ekki marktækt í *t*-prófum á meðaltöl**. Lestu einnig athugasemd Douglas G. Altman og J. Martin Bland, [Absence of evidence is not evidence of absence](#), í BMJ í ágústmánuði 1995. Athugaðu að það eru fleiri greinar með sama nafni, gættu þess að lesa rétta grein! H

Lestu það sem höfundar segja um hugsanleg tengsl málsins antimons (*antimony*) í ungbarnadýnum við vöggudaða (*sudden infant death*) og umræðu þeirra um sama efni í greininni allri.

Rökstyddu nákvæmlega (a) hvað sé rangt við það að álykta að fyrst marktæktarpróf er ómarktækt séu engin tengsl milli antimons og vöggudaða og (b) nákvæmlega hvernig gætu öryggisbil hjálpað í þessu efni.

36. Í kennslubókinni (bls. 363 í 7. útgáfu) er eftirfarandi fullyrðing:

Marktæktarpróf er aðferð til að meta hversu markverðan vitnisburð gögnin veita gegn núlltilgátunni.

Útskýrðu þetta nákvæmlega með því að útlista þá hugsun sem liggur að baki marktæktarprófi á meðaltal. Gættu þess að í svari þínu komi fram eftirfarandi atriði. (a) Hvað er núlltilgáta og hvað er aðaltilgáta og hvernig tengjast þær, (b) ef núlltilgátan er rétt, hvaða upplýsingar gefur það annars vegar um þýðismedaltalið og hins vegar um úrtaksmedaltalið og (c) í nákvæmlega hvaða tilvikum (og í hvaða skilningi) gefur niðurstaðan í úrtakinu skýran vitnisburð gegn núlltilgátunni.

Athugaðu að hér ertu beðin um að skýra nákvæmlega hugsunina á bakvið marktæktarpróf og rökstyðja vandlega svar þitt.

37. *t*-próf er traust (*robust*) gagnvart ýmsum brotum á forsendum (*assumptions*) sínum. (a) Útskýrðu nákvæmlega hvað forsenda er í þessu samhengi og gerðu stuttlega grein fyrir þeim eins og þær birtast bæði í þýði (*population*) og í úrtaki (*sample*), (b) útskýrðu nákvæmlega hvað felst í traustleika (*robustness*) *t*-prófs og (c) gerðu grein markgildissetningunni (*central limit theorem*) og mikilvægi hennar fyrir traustleika prófsins.
38. Eigandi smábíls mældi eyðslu bílsins í tuttugu skipti samtals, í hvert skipti sem hann fyllti á tankinn. Hann gætti þess að keyra u.þ.b. jafnlangt milli áfyllinga og svo vel vildi til að á þessu tímabili var hann alltaf í þessu venjulega snatti til og frá skóla, í leikskóla og vikulega í matvörubúð. Eftirfarandi tuttugu tölur er eyðsla (l/100km) bílsins þessi tuttugu skipti.

5,7; 4,6; 6,4; 6,3; 6,9; 5,2; 4,9; 5,4; 4,9; 5,6;  
5,4; 5,3; 4,9; 5,1; 5,0; 6,0; 6,3; 5,4; 5,3; 5,4

(a) Reiknaðu út meðaleyðslu bílsins og staðalvillu þessa meðaltals. Gerðu ráð fyrir því að staðalfrávik eyðslunnar í þýði ( $\sigma_x$ ) sé 0,67. (b) Teiknaðu normalrit af eyðslumælingunum og leggðu mat á það hvort forsendan um normaldreifingu sé fyrir hendi. (c) Rökstyddu hvort lögum dreifingarinnar leyfi eða leyfi ekki að öryggisbil sé reiknað og túlkað. (d) Reiknaðu 95% öryggisbil og túlkaðu nákvæmlega hvað það merkir fyrir eyðslu bílsins. (e) Endurreiknaðu öryggisbilið miðað við að þú þekkir ekki staðalfrávik þýðis. (f) Hvernig breytist öryggisbilið og hvers vegna?

39. *t*-próf hefur ákveðnar forsendur (*assumptions*) og er talið traust (*robust*) gagnvart þeim.

Útskýrðu (a) hvað átt er við með forsendu og hvað leiðir af því að forsenda sé fyrir hendi eða að hún sé ekki fyrir hendi og (b) útskýrðu hugtakið traustleika (*robustness*) og tengsl þess við markgildissetninguna (*central limit theorem*).

40. Útskýrðu stuttlega eftirfarandi atriði og mikilvægi þeirra í tengslum við takmarkanir og misnotkun (*limitations and abuse*) á marktektarprófum (*statistical tests*).

(a) Marktektarpróf byggjast á tvíkostaákvörðunum en vitneskjan (*evidence*) sem við höfum gegn núlltilgátunni kemur í samfellu (*continuous*) frá því að vera veik upp í að vera sterk. (b) Ef núlltilgátu er hafnað, þarf það ekki að þýða að það sé munur í þýði sem skipti umtalsverðu máli. (c) Ef ekki er hægt að hafna núlltilgátunni er hæpið og hættulegt að álykta sem svo að þá hljóti hún að vera rétt.

41. Úrtaksmeðaltal hefur ákveðna þekkta úrtakadreifingu (*sampling distribution*) ef það er reiknað fyrir tilviljunarúrtak úr normaldreifðu þýði. (a) Útskýrðu nákvæmlega hvað hér er átt við og gættu þess að útskýra vel öll hugtök.

Ef það eru frávillingar (*outliers*) í þýðinu, er úrtakadreifingin breytt. (b) Hvert er mikilvægi þess fyrir marktektarpróf, hvers vegna breyttist hún og hvernig myndir þú gera ráð fyrir að hún breyttist.

Ef ég sé áberandi frávilling í úrtaki sem er að öðru leyti virðist sæmilega normaldreift, gæti ég fjarlægt hann og lagað þannig dreifinguna í úrtakinu. (c) Hvers vegna er þetta slæm hugmynd og hvaða áhrif hefði þetta á úrtakadreifingu úrtaksmeðaltalsins.

*Aukaliður sem getur komið til hækkunar:* Ef ég þekki staðalfrávik þýðis og get þannig reiknað staðalfrávik úrtaksmeðaltals í úrtökum af ákveðinni stærð, hvaða ályktun gæti ég dregið almennt um staðalfrávik (þ.e. staðalvillu) þýðismeðaltalsins?

42. Mikið hefur verið rætt um þá staðreynd að klukkan á Íslandi er u.þ.b.  $1\frac{1}{2}$  klukkustund á eftir sólar tíma í Reykjavík. Oft er erfitt fyrir nemendur að vakna á morgnanna og því eru margir þeirra hlynntir því að seinka klukkunni um eina klukkustund. Nemi á 2. misseri í sálfræði ákvað að kanna hve margir væru hlynntir slíkri breytingu.

Hann fékk að senda könnun á 300 sálfræðinema þar sem spurt var hvort þeir væru hlynntir því að seinka klukkunni um eina klukkustund. Svörun var dræm en af þeim 186 sem svöruðu voru 56,3% hlynnt þessari breytingu.

Neminn reiknaði öryggisbil fyrir niðurstöðuna. Staðalvillan var um 3,6% og því reiknaði hann 95% öryggisbil sem [49,2%; 63,4%].

(a) Útskýrðu niðurstöðuna þannig að upplýstur almenningur skilji. Gættu þess m.a. að það komi fram fyrir hvað þessi 95% standa, hvaða upplýsingar talnabil öryggisbilsins veitir og hversu margir sálfræðinemar eru hlynntir þessari seinkun á klukkunni. Rökstyddu vel og ítarlega.

(b) Miðað við þessar talnalegu niðurstöður, er hægt að álykta um það hvort meirihluti sálfræðinema vilji seinka klukkunni? Rökstyddu afar vandlega.

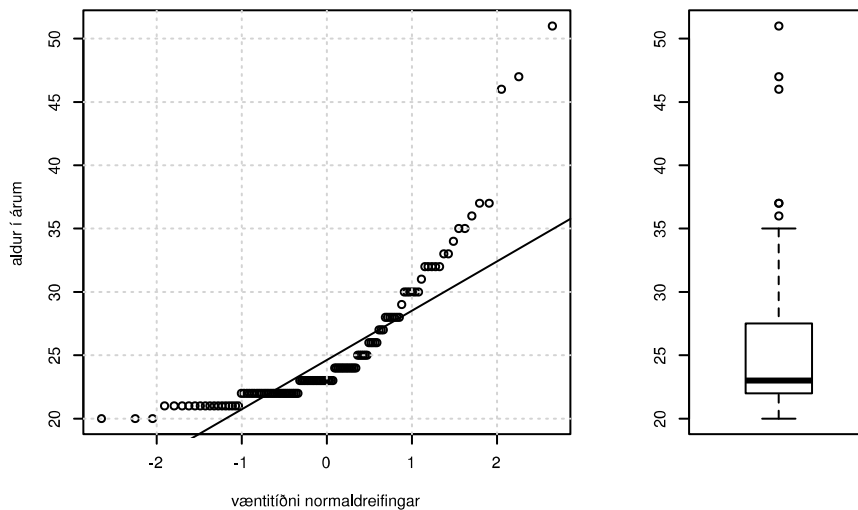
(c) Hvaða atriði gætu ógnað þessum talnalegu niðurstöðum? Er einhverju slíku ábótavant í þessu dæmi og hvaða áhrif hefur það þá á túlkun og niðurstöður?

43. Útskýrðu hugtökin tilgátu (*hypothesis*), marktækt (*statistical significance*) og tölfræðipróf (*statistical test*). Gættu þess að svara m.a. eftirfarandi spurningum. Miðaðu við að þú sért að rannsaka hvort það sé munur á meðalfjölda klukkustunda sem nemandi á fyrsta misseri varði í námið eftir því hvort viðkomandi komst síðan upp á annað misseri (SAT-reglan) eða ekki.

(a) Hvað nákvæmlega er aðaltilgáta (*alternative hypothesis*) og hvers vegna þarf að setja fram núlltilgátu (*null hypothesis*)? Settu núlltilgátuna í ofangreindu dæmi fram með viðeigandi táknum.

(b) Nákvæmlega hvaða hlutfall eða líkur felast í marktæktinni, þ.e.  $p$ -gildinu (*significance; p-value*)?

(c) Ef marktæktin ( $p$ -gildið) er lágt, hvaða upplýsingar gefur það okkur um núlltilgátuna? Hversu miklar líkur eru á því að núlltilgátan sé rétt ef við fáum  $p=0,01$ ? Rökstyddu svarið vel.



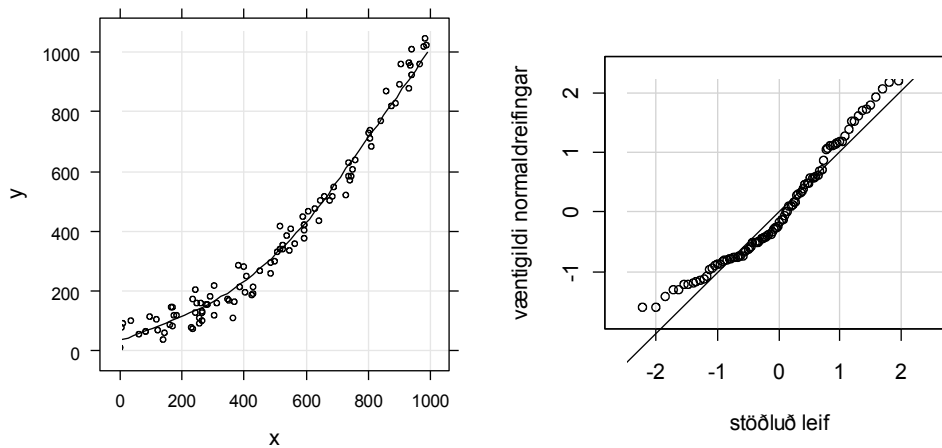
44. Myndin hér fyrir ofan sýnir dreifingu aldurs í úrtaki 124 nemenda í sálfræði við Háskóla Íslands. Gerum nú ráð fyrir því að þú viljir kanna hvort meðalaldur nemenda sé ólíkur eftir kynferði.

(a) Gerðu grein fyrir helstu forsendum  $t$ -prófs í tveimur hópum. (b) Lýstu myndunum með sérstöku tilliti til þessara forsendna. Gættu þess að það komi skýr fram hvort viðkomandi forsendna sé eða sé ekki fyrir hendi. (c) Rökstyddu hvort  $t$ -próf hentar fyrir þessi gögn. Þú þarft tilgreina nákvæmlega hvort það henti eða ekki og einnig að rökstyðja nákvæmlega ákvörðun þína og greina frá því á hverju hún byggir nákvæmlega.

45. Karl vildi kanna hvort sá tími sem háskólanemar á aldrinum 20-25 ára verja að meðaltali í nám sitt væri meiri en vinnutími jafnaldra á almennum vinnumarkaði. Prófið sem hann notaði reyndist vera ómarktækt.

(a) Hvaða ályktanir er hægt að draga af svona ómarktæku marktæktarprófi? Reyndu að vera nákvæm og rökvis. (b) Get ég dregið einhverjar ályktanir um muninn á meðaltölum hópanna? (c) Bætir öryggisbil einhverju við niðurstöðu marktæktarprófsins og þá nákvæmlega hverju?

46. Hverjir eru helstu kostir (*advantages*) Wilcoxon Mann-Whitney prófsins umfram  $t$ -próf?



47. Myndirnar sýna niðurstöður einfaldrar aðfallsgreiningar (*simple linear regression*) þar sem X er eina frumbreytan og Y er fylgibreytan.
- Greindu frá forsendum (*assumptions*) aðfallsgreiningar (*linear regression*).
  - Hvaða forsendur er hægt að meta (*assess*) á fyrri myndinni og hvaða forsendur er hægt að meta á grundvelli seinni myndarinnar? (*Which assumptions is possible to assess based on the first graph and which can be assessed by inspecting the second graph.*)
  - Eru þessar forsendur, teknar hver fyrir sig, uppfylltar eða ekki uppfylltar miðað við þær upplýsingar sem myndirnar gefa? (*Are those assumptions fulfilled each taken individually, judged by the information provided by the two graphs?*)
48. Við tilgátuprófun (*hypothesis testing*) er aðaltilgáta (*alternative hypothesis*) tekin upp ef núlltilgáta (*null hypothesis*) er hafnað.
- Útskýrðu (a) í hvaða tilvikum hægt er að hafna (*reject*) núlltilgátu, (b) með hvaða rökum það er gert og (c) hversu góð sönnun (*proof*) þetta er á aðaltilgátunni.
49. Útskýrðu kíkvadrátpróf (*chi-square test*) fyrir krosstöflur (*two-way tables*). Tilgreindu (a) hvaða tilgátur (aðal- og núlltilgáta; *alternative and null hypothesis*) eru prófaðar, (b) hvaða ályktun má draga af marktæku prófi (*significant test*) og (c) hvaða þrjár aðferðir má nota við að túlka niðurstöðurnar efnislega (*substantive interpretation*).
50. Útskýrðu mælitöluna Cohens *d*.
- Hvernig er þessi mælitala reiknuð, (b) hvernig er niðurstaðan túlkuð og (c) hvers konar viðbótarupplýsingar gefur Cohens *d* umfram sjálft marktæktargildið (*significance value*) annars vegar þegar tölfræðipróf er marktækt (*statistically significant*) og hins vegar þegar það er ómarktækt (*not statistically significant*)?