

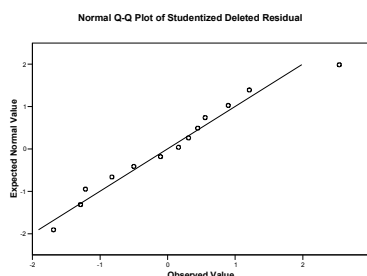
Eftirfarandi er yfirlit yfir spurningar sem notaðar hafa verið í prófum námskeiðinu.

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.

Þessum prófspurningum er ætlað að gefa þér gott yfirlit yfir það hvernig spurt er á prófum námskeiðsins. Allar þessar spurningar hafa verið notaðar og gætu verið notaðar í prófum námskeiðsins því sjálfur leita ég stundum fanga í þessum lista.

Ef þú sérð einhverja spurningu sem þú telur að eigi ekki heima hér, sé í ósamræmi við núverandi lesefni eða villandi að öðru leyti, vinsamlega hafðu samband svo ég get tekið afstöðu til þess hvort eigi að fjarlægja hana.

1. Hvaða eiginleika metur summa kvaðrata í dreifigreiningartöflu?
2. Hvaða stærð metur meðalsumma kvaðrata leifar, MSE, í dreifigreiningartöflu?
3. Notaðu gagnasafnið ex10_013.sav. **H**
 - i. Breyttu hallanum fyrir Skakka turninn úr tíundu hlutum úr millimetra í sentímetra umfram 2,9 metra.
 - ii. Notaðu aðfallsgreiningu til að spá fyrir um halla turnsins fyrir árin 1990, 2010, 2030 og 2050. Gefðu upp viðeigandi öryggisbil. Túlkaðu niðurstöðurnar og tilgreindu þá heildarmynd sem niðurstöður og öryggisbil gefa.
4. Hver er munurinn á leif (*residual*) og villu (*error*)? Hvaða kröfur eru gerðar til villunnar í aðfallsgreiningu? Hvaða kröfur eru gerðar til leifarinnar?



5. Hvaða frávik—ef einhver—frá forsendum aðfallsgreiningar sýnir þessi mynd?

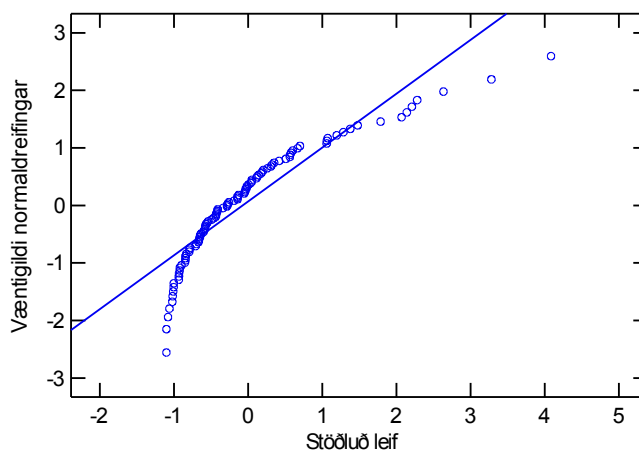
ANOVA

Model	Sum of Squares	ddf	Mean Square	F	Sig.
1 Regression	1302,08	1
Residual	961,00	36		
Total	37		

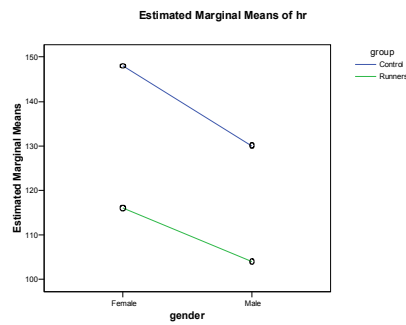
6. Í þessa dreifigreiningartöflu vantar mikið af upplýsingum. Settu réttar upplýsingar í alla reitina sem hafa punktalínur í stað talna.
7. Gerðu ráð fyrir að þú viljir athuga tengsl hjúskaparstöðu (einhleypur, í sambúð) og starfsaldurs (í árum) við heildarlaun starfsmanna stórs sveitarfélags. Útskýrðu með tilbúnum dæmum sem tilgreina launamun eftir hjúskaparstöðu, starfsaldri eða eftir báðum breytum samtímis hvernig *línuleg tengsl* þeirra við heildarlaun ættu að birtast.
8. Í skilaverkefni 2 var hallastuðullinn fyrir heilastærð (MRIPix) jafnt og 0,00012 þegar hún var eina frumbreytan en hækkaði upp í 0,00021 þegar líkamshæð var einnig inni í líkaninu. Hvað er á seyði: (a) Hvers vegna er hallastuðullinn tvöfalt hærri í

síðara tilvikinu og (b) nákvæmlega hvaða merkingu hefur hallastuðullinn í hvoru tilviki fyrir sig?

9. Hvaða áhrif hefur það réttmætri niðurstaðna aðfallsgreiningar ef frumbreyturnar eru allar með miklum gólfhrifum (*ceiling effect*), dreifing þeirra mikið skekkt og með mörg fráviksgildi (frávillinga)?
10. Hversu hátt þarf R^2 að til að eitthvað sé að marka niðurstöðu aðfallsgreiningar?
11. Hvaða núlltilgátu athugar Próf Levenes, nákvæmlega hvaða forsendu dreifigreiningar tengist hún og hvaða aðrar aðferðir er mælt með að nota til að kanna þá forsendu?
12. Á hvern hátt gefur F -próf dreifigreiningar takmarkaðar upplýsingar um meðaltöl hópanna í þýði? Hvernig má bregðast við því?
13. Hver er meginmunurinn á fyrirfram (*a priori*) og eftir á (*post hoc; multiple comparisons*) samanburðum? Hvaða áhrif hefur þessi munur á tíðni höfnunarmistaka (*Type I error rate*)?
14. Í hverju er aðferð Bonferronis fólgin og hvaða vanda er henni ætlað að leysa?
15. Hvað nákvæmlega felst í samvirkni (*interaction*) tveggja frumbreyta? Komdu einnig með skýrt dæmi um samvirkni þar sem þú skýrir nákvæmlega hvers vegna þetta er samvirkni en ekki meginhrif (*main effect*).
16. Hvað nákvæmlega er fylgnileif (*residual correlation*) í þáttgreiningu og að hvaða leyti og hvernig er hægt að styðjast við hana við úrvinnsluna?



17. Hvaða frávik—ef einhver—frá forsendum dreifigreiningar sýnir þessi mynd?
18. Sæktu skrána eg13_011.por í heimaprófsmöppuna í Uglu. Þetta eru gögnin sem voru notuð í dæmi 13.8 í bókinni. Lestu vel lýsinguna á gögnunum í bókinni (bls. 813). Lestu gögnin inn í SPSS með því að nota File/ Open / Data... Þar undir Files of type velur þú SPSS Portable (*.por). **H**



- i. Endurgerðu myndina sem er hér fyrir ofan.
 - ii. Túlkaðu niðurstöðurnar sem sjást á myndinni í samræmi við lýsinguna á gagnasafninu í bókinni.
19. Hvaða upplýsingar gefur R^2 í aðfallsgreiningu?
 20. Nefndu þrjár mikilvægar forsendur aðfallsgreiningar og gerðu grein fyrir því hvernig sé best að kanna hvort viðkomandi forsendur séu fyrir hendi?
 21. Hvað nákvæmlega er misleitni (*heteroscedasticity*) og hvaða aðferðir henta best til að meta hana í tvíhliða dreifingreiningu?
 22. Hvaða hlutverki gegnir snúningur (*rotation*) í þáttgreiningu? Hver er munurinn á hornréttum (*orthogonal*) og hornskökum (*oblique; nonorthogonal*) snúningi.
 23. Sæktu skrána CSData.por í heimaþrófsmöppuna í Uglu. Lestu vel lýsinguna á gögnunum í bókinni (bls. D2). Lestu gögnin inn í SPSS með því að nota File/ Open / Data.... Þar undir Files of type velur þú SPSS Portable (*.por). **H**

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26,303	2	13,151	26,626	,000 ^a
	Residual	109,160	221	,494		
	Total	135,463	223			

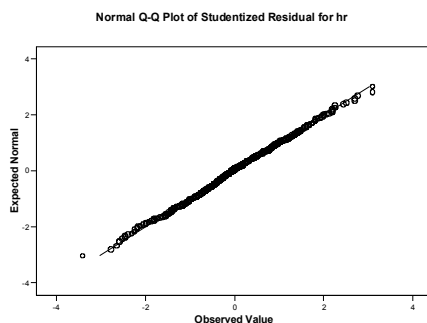
a. Predictors: (Constant), satm , hsm
 b. Dependent Variable: gpa

- i. Framkvæmdu aðfallsgreiningu þar sem GPA er fylgibreytan og SATM og HSM eru frumbreytur. Þú átt að fá m.a. töfluna hér fyrir ofan. Berðu þínar niðurstöður saman við hana og gakktu úr skugga um að þú hafir framkvæmt úrvinnsluna rétt.
 - ii. Gefðu upp aðfallsjöfnuna eins og hún er metin af forritinu.
 - iii. Túlkaðu alla hallastuðla. (a) Segðu hvað þeir merkja nákvæmlega svona tölfræðilega séð. (b) Útskýrðu með mannamáli hvað niðurstöðurnar þýða.
 - iv. Reiknaðu 95% öryggisbil fyrir hallastuðlana. (a) Hvaða ályktanir getur þú dregið um áhrif HSM? (b) En um SATM?
 - v. Segðu á einföldu mannamáli hvað niðurstöðurnar þýða? Miðaðu við að amma þín myndi skilja þig.
24. Á glærunni **Túlkun hallatalna í Marghliða aðfallsgreiningu** stendur eftirfarandi: „Áhrif einkunna í vísindum og ensku eru ekki marktæk. Þau eru einnig lítil, jafnvel þótt horft sé til efri marka öryggisbilanna.“ Hvað í ósköpunum er átt við?
 - i. Halltölurnar eru ómarktækar og því skyldi ég þá skoða öryggisbilið?
 - ii. Hvaða máli skiptir það að efri mörk öryggisbilanna séu lág?

25. Á glærunni **Mynstur meðaltala í Einhliða dreifgreiningu** stendur eftirfarandi: „Allsherjarprófið í dreifgreiningu segir ekkert um mynstur meðaltala, þ.e. hver þeirra séu ólík og hver eins.“ Þetta er nú allt frekar dularfullt.

- i. Hvaða allsherjarpróf er þetta?
- ii. Hvers vegna segir það ekkert til um mynstur meðaltala?
- iii. Hvað getum við gert ef við viljum vita hvert mynstrið er?

26. Sæktu skrána eg13_011.por í heimaprófsmöppuna í Uglu. Þetta eru gögnin sem voru notuð í dæmi 13.11 í bókinni. Lestu vel lýsinguna á gögnunum í bókinni (bls. 784). Lestu gögnin inn í SPSS með því að nota File/ Open / Data.... Þar undir Files of type velur þú SPSS Portable (*.por). **H**



- i. Framkvæmdu tvíhliða dreifgreiningu og endurgerðu myndina hér fyrir ofan.
- ii. Túlkaðu myndina með sérstakri áherslu á það hvort forsendur úrvinnslunnar séu til staðar.

27. Sæktu skrána CSData.por í heimaprófsmöppuna í Uglu. Lestu vel lýsinguna á gögnunum í bókinni (bls. D3). Lestu gögnin inn í SPSS með því að nota File/ Open / Data.... Þar undir Files of type velur þú SPSS Portable (*.por). **H**

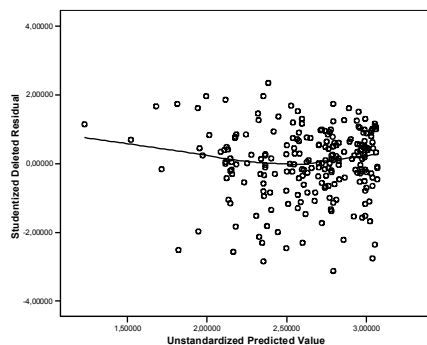
ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26,303	2	13,151	26,626	,000 ^a
	Residual	109,160	221	,494		
	Total	135,463	223			

a. Predictors: (Constant), satm, hsm

b. Dependent Variable: gpa

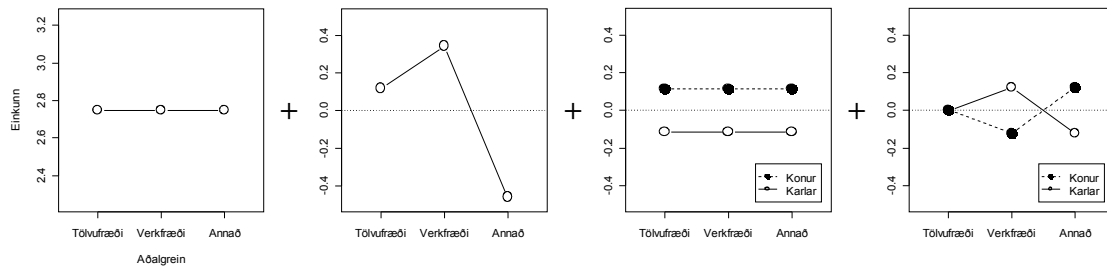
- i. Framkvæmdu aðfallsgreiningu þar sem GPA er fylgibreytan og SATM og HSM eru frumbreytur. Þú átt að fá m.a. töfluna hér fyrir ofan. Berðu þínar niðurstöður saman við hana og gakktu úr skugga um að þú hafir framkvæmt úrvinnsluna rétt.



- ii. Gerðu mynd eins og þessa hérna. Línan er búin til með því að biðja um Loess-línu með 80% glugga.

28. Í fyrirlestrinum **Marghliða aðfallsgreining** er á glærum 1–4 og 14–15 rætt um sérhæf og ósérhæf áhrif og skilyrta og óskilyrta hallastuðla. Það er einnig rætt stuttlega um þetta efni á bls. 695 í 5. útgáfu kennslubókarinnar. **H**
- Útskýrðu með *eigin orðalagi* hvers vegna óskilyrtur hallastuðull HSS á glæru 14 er 0,15 meðan skilyrti hallastuðullinn er 0,03.
 - Útskýrðu með eins skýru orðalagi og þú getur hvað felst í því að áhrif sjálfsmyndar sé *leiðrétt* fyrir greind í líkani 2 á glæru 15. Gættu þess að nota þitt eigið orðalag.
29. Í fyrirlestrinum **Einhliða dreifigreiningu** er fjallað um misleitni (*heteroscedasticity*) á glærunni **Misleitni** og gerð grein fyrir forsendum dreifigreiningar á glærunni **Einhliða dreifigreining**. **H**
- Ef alvarleg misleitni væri til staðar, hvernig myndir þú búast við að myndin á **Misleitni** myndi líta út? Hvers vegna?
 - Nákvæmlega hvaða forsendu dreifigreiningar skv. **Einhliða dreifigreiningu** er brotin ef misleitni er til staðar? Tilgreindu nákvæmlega *hvað* á glærunni gefur til kynna forsenduna um einsleitni (*homoscedasticity*).
30. Lestu vel lýsingu á rannsókninni á öryggi á vinnustað (*study of workplace safety*) sem lýst er í sýnidæmum (*example*) 12.3–12.21. Útskýrðu síðan eftirfarandi: **H**
- Nákvæmlega hvaða spurningar prófa samanburðirnir (*contrasts*) tveir sem lýst er í sýnidæmi 12.17? Svarið þarf að vera á einföldu og skýru *mannamáli* (*ordinary everyday language, please*)!
 - Hvaða ályktun get ég dregið um ψ_2 á grundvelli tölfræðiprófsins í sýnidæmi 12.20 og öryggisbilsins í sýnidæmi 12.21? Notaðu þitt eigið orðalag, *ekki* éta upp úr bókinni!
 - Hvaða kosti hefur það að nota slíka samanburði fremur en að byggja eingöngu á allsherjarprófinu (*omnibus test*) sem lýst er í upphafi sýnidæmis 12.14?
31. Í hvaða tilvikum notum við aðferð Bonferronis og nákvæmlega hvaða vanda er henni ætlað að leysa?
32. Í hverju felst skriðupróf Cattells (*Cattell's scree test*), hvernig er það notað og hvaða afleiðingar hefur það ef dregnir eru annað hvort færri eða fleiri þættir (*factors*) heldur en skriðuprófið gefur til kynna?
33. Hverjar eru helstu forsendur (*assumptions*) dreifigreiningar og hvernig er best að kanna hvort þær séu fyrir hendi?
34. Í hverju nákvæmlega felst aðferð Scheffé við eftir á samanburði? Fyrir hvers konar samanburði hentar aðferðin?
35. Í rannsókn á 282 konum var athugað hvaða þættir tengdust ánægju við líkamsrækt. Fylgibreytan, *Ánægja*, gat tekið gildin 18 til 136. Frumbreyturnar voru þrjár: Smekkur fyrir tónlistinni (*Tónlistarsmekkur*) var á kvarðanum 4 til 28; *Ánægja* með leiðbeinandann var á bilinu 6 til 42; *Samsemd*, þ.e. hvort viðkomandi fann sig í líkamsræktinni.
- Fylgni á milli breytanna fjögurra var umtalsverð og í öllum tilvikum marktæk með $p < 0,05$. Hallatalan fyrir Tónlistarsmekk var 1,02, hún var 0,96 fyrir *Ánægju* með leiðbeinandann og 0,30 fyrir *Samsemd*.
- Í niðurstöðum rannsóknarinnar kom eftirfarandi einnig fram: $R^2 = 0,33$.
- Geturðu gefið upp aðfallsjöfnuna? Ef svarið er jákvætt, skaltu gefa jöfnuna sem svar. Ef svarið er neikvætt, þarftu að útskýra hvers vegna.
 - Allar fylgnitölur voru margtækar. Þýðir það að hallastuðlarnir séu marktækir líka? Rökstyddu svarið nákvæmlega.

- iii. Núlltilgátan var að allar hallatölur væru núll. Niðurstaða tölfræðiprófsins var $F=45,64$. Gefðu upp frígráður prófsins, marktækt þess og túlkaðu niðurstöðuna.
- iv. Hve hátt hlutfall af dreifingu fylgibreytunnar má skýra með frumbreytunum þremur?
36. Sæktu skrána `spurning2.sav` í heimaprófsmöppuna í Uglu. Þessi skrá innheldur sömu gögn—`csdata.sav`—og eru notuð í fyrirlestrinum [Marghliða aðfallsgreiningu](#) nema hvað fylgibreytan er ný og heitir núna `newy`. **H**
- Framkvæmdu aðfallsgreiningu þar sem `newy` er fylgibreytan og HSS, HSE og HSM eru frumbreytur. Birtu hallastuðlana og dreifigreiningartöfluna.
 - Endurgerðu helstu myndritin í fyrirlestrinum [Marghliða aðfallsgreiningu](#) með þessum nýju gögnum: Kassarit af leif, normalrit af leif og leifarritið eins og á glæru 20.
 - Tilgreindu fyrir hverja mynd hvaða eiginleikar villunnar hægt er að meta með þeirri mynd og lýstu því hvort gögnin standist forsendur dreifigreiningar eins og þeim er lýst t.d. á [Marghliða aðfallsgreiningu](#) og á bls. 685 í 5. útgáfu kennslubókarinnar (ISP) undir fyrirsögninni *Multiple linear regression model*.
37. Skoðaðu vel mynd 12.6 á bls. 728 í bókinni (einnig birt á glæru 1 í [Einhliða dreifigreiningu](#)) með hliðsjón af forsendunni um einsleitni (*homoscedasticity*). **H**
- Útskýrðu nákvæmlega (a) hvað átt er við með einsleitni, sbr. umræðu á bls. 728–729 og (b) tilgreindu nákvæmlega hvernig einsleitnin birtist á mynd 12.6.
 - Lýstu þremur megináferðum við að meta einsleitni og útskýrðu hvernig nákvæmlega misleitni (*heteroscedasticity*) sést með áferðunum hverjum fyrir sig.
38. Skoðaðu dæmi 12.18 á bls. 740 í bókinni. Þú gætir þurft að lesa blaðsíðurnar og dæmin á undan og eftir einnig. **H**
- Gættu þín á villu í bókinni. Samanburðurinn á að vera svona:
- $$c_1 = \bar{X}_{SU} - \frac{1}{2}(\bar{X}_{UN} + \bar{X}_{SK})$$
- Þeir sem gerðu rannsóknina tóku eftir því að verkstjórar voru með hæstu niðurstöðuna, öruggasta vinnuumhverfið. Þeir ákváðu því að *endurtaka* rannsóknina og gera fyrirfram samanburð þar sem verkstjórar (*supervisors*) væru bornir saman við hina. Þetta rökstuddu þeir með því að verkstjórar væru með allt annað vinnuumhverfi en faglærðir og ófaglærðir, bæði fylgdust þeir með og stjórnðu í stað þess að vinna verkin sjálfur auk þess sem umtalsverður hluti vinnutímans eru fundir og skrifborðsvinna.
- Niðurstaða samanburðarins var tæpum 7,9 stig verkstjórum í hag. Þeir birtu niðurstöðu sína svona: „Verkstjórar mátu vinnuumhverfi sitt öruggara ($M=78,8$) heldur en aðrir starfsmenn ($M=71,0$), $t(587)=3,04$, $p<0,01$.“ Þeir birtu ekki allsherjarprófið (*omnibus test*).
- Útskýrðu hvers vegna þeir máttu nota fyrir fram (*a priori; contrasts*) prófi í þessari nýju rannsókn. Gefðu skýrt og nákvæmt svar.
 - Hverjir eru kostir fyrir fram prófs og hverjir eru ókostir eftir á (*post hoc; multiple comparisons*) prófs í þessu samhengi?
 - Hvernig stendur á því að þeir birta ekki allsherjarprófið, er það réttmætt?

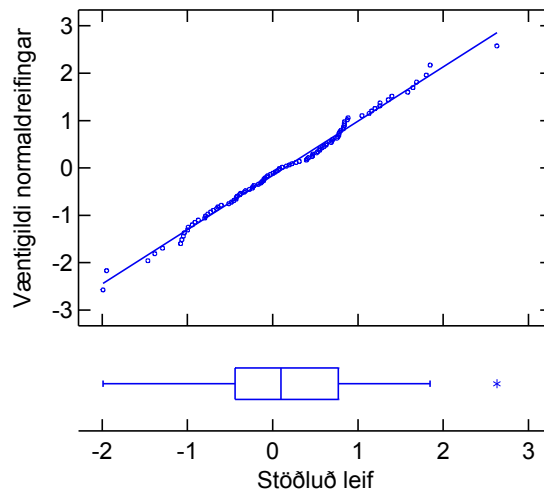


39. Skoðaðu vel ofangreinda mynd sem er sýnidæminu úr fyrirlestrinum [Tvíhliða dreifigreiningu](#) og sýnir sundurliðun meðaltala af glærunni [Mat á samvirkni](#), sbr. neðri töfluna á glærunni.
- Útskýrðu hverja mynd fyrir sig, þ.e. nákvæmlega hvað hún sýnir.
 - Berðu aðra mynd frá vinstri saman við myndina á glærunni Samvirkni. Í báðum tilvikum er hægt að sjá meginhrif. Að hvaða leyti er myndin hér skýrari heldur en myndin á glærunni?
 - Berðu síðustu myndina saman við myndina á glærunni Samvirkni. Í báðum tilvikum sést samvirkni. Að hvaða leyti er myndin hér skýrari heldur en myndin á glærunni eða er það svo að þessi mynd sé skýrari?
40. Hvað er áhættuhlutfall (*odds ratio*) og nákvæmlega hvaða upplýsingar gefur það okkur um tengsl frum- og fylgibreytu?
41. Marghliða aðfallsgreining (*multiple regression*) gefur sérhæf áhrif frumbreytu á fylgibreytu þar sem leiðrétt er fyrir (*controlled for*) áhrif annarra frumbreyta sem eru í líkaninu.
Hvað nákvæmlega felst í þessum sérhæfu eða leiðréttu áhrifum? Hvernig fæ ég óleiðrétt (ósérhæf) áhrif og hvernig yrðu þau öðru vísi, þ.e. stærri eða minni, og hve miklu munar? Skýrðu nákvæmlega m.a. með því að taka ímyndað dæmi.
42. Athuguð voru tengsl einkunnna þar sem einkunn í Tölfræði II var fylgibreytan en einkunn í Tölfræði I frumbreyta. Í ljós kom að hallastuðullinn var 0,57. Þegar meðaleinkunn á stúdentsprófi var bætt við sem frumbreytu lækkaði hallastuðull fyrir Tölfræði I verulega eða niður í 0,36. [Þetta eru skáldaðar tölur!]
- Túlkaðu fyrri hallastuðulinn, þ.e. hallastuðul fyrir Tölfræði I þegar sú einkunn er eina frumbreyta líkansins.
 - Túlkaðu seinni hallastuðulinn, þ.e. hallastuðul fyrir Tölfræði I þegar bæði sú einkunn og stúdentsprófseinkunn eru frumbreytur í líkaninu.
 - Útskýrðu (a) hvers vegna seinni hallastuðullinn er lægri og (b) hvers vegna meðaleinkunn á stúdentsprófi breytir hallastuðli einkunnar í Tölfræði I.
43. Sæktu skrána `spurning2.sav` í heimaprófsmöppuna í Uglu. Þessi skrá á uppruna sinn í töflu 1.9 í kennslubókinni en gögnum hefur verið lítillega breytt. **H**
- Framkvæmdu aðfallsgreiningu þar sem `gpa` er fylgibreytan og `iq`, `gender` og `concept` eru frumbreytur. Birtu hallastuðlana og dreifigreiningartöfluna.
 - Taktu afstöðu til þess hvort forsendur aðfallsgreiningar standist með því að skoða leifina. Birtu kassarit af leif, normalrit af leif og leifarrit og túlkaðu þessi rit með hliðsjón af forsendunum.
 - Reiknaðu spágildi fyrir einkunn (`gpa`) miðað við ólík gildi sjálfsmýndar (`concept`). Gerðu þetta sérstaklega fyrir karla (hafa gildið 2 á `gender`) og fyrir konur (hafa gildið 1) og miðaðu við að í báðum tilvikum sé greindarvísitalan (`iq`) nákvæmlega 100,0. Birtu niðurstöðurnar myndrænt, þ.e. ein mynd þar sem sjálfsmýnd er á lárétta ásnum (20–80), einkunn á lóðrétta ásnum (4,0–7,0) og

spágildi fyrir konur eru sýnd með rauðri línu en spágildi fyrir karla með blárri (eða með sambærilegri aðgreiningu eftir kyni).

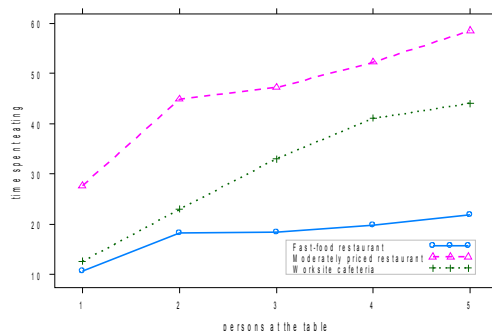
Levene's Test of Equality of Error Variances			
Dependent variable: Tekjur síðasta sumar			
F	df1	df2	Sig.
4,803	3	45	0,005

44. Skoðaðu vel töfluna með prófi Levenes hér fyrir ofan og svaraðu eftirfarandi spurningum.
- Hvað er verið að prófa með þessu prófi og hver er núlltilgáta þess.
 - Hvaða ályktun megum við draga af annars vegar marktækri og hins vegar ómarktækri niðurstöðu?
 - Hvaða aðrar aðferðir höfum við til að meta það sem verið er að prófa og hvernig samhfum við niðurstöður þeirra við þessar?
45. Á bls. 739 í kennslubókinni segir: „A contrast expresses an effect in the population as a combination of population means. To estimate the contrast, form the corresponding **sample contrasts** by using sample means in place of population means. Under the ANOVA assumptions, a sample contrast is a linear combination of independent normal variables and therefore has a normal distribution.“
- Útskýrðu ofangreint með þínum eigin orðum. Gættu þess að skýrt komi fram (a) hvað felist í fyrirfram samanburði, (b) hver er munurinn á samanburði í úrtaki og samanburði í þýði, (c) í hvaða skilningi samanburðir eru línulegir og (d) hvernig *samanburðir* geta haft normaldreifingu.
45. Brussella gerði dreifigreiningu á meðaleinkunnum í samræmdum prófum í 170 íslenskum grunnskólum. Þar sem hún vildi vita hverjir þeirra væru ólíkir, bar hún þá saman tvo og tvo og reiknaði *t*-próf milli meðaltalanna. Hún fann 20 þör skóla sem voru með marktækt ólíkar meðaleinkunnir og í sumum tilvikum var munur mjög mikill.
- Hægfinnur var ekki hrifinn af niðurstöðunum og hafnaði þeim alfarið. Hann sagði að Brussella hafði ranglega notað LSD-próf (sjá bls. 744) og niðurstaðan gæti verið fyrir tilviljun eina. Nota þyrfti leiðréttingu eins og t.d. Bonferroni eða Scheffé.
- Brusella snöggreiddist: „Þetta eru sko engar ofs kynjanir hjá mér og það væri fráleitt að fara að leiðrétta niðurstöðuna því þá myndi marktæktin hverfa. Sjálfur ert þú bara aumkunarverður hasshaus.“
- Útskýrðu hvað Hægfinnur á við með sínum athugasemdum.
- Hvernig getur þetta verið fyrir tilviljun þrátt fyrir marktæktina?
 - Hvað kemur fíkniefnið LSD málinu við eða hvað er maðurinn að tala um í sambandi við LSD?
 - Hver er munurinn á villutíðninni α (alfa) sem *t*-próf miðast við og þeirri villutíðni sem í reynd er stuðst við ef leiðrétt er með Bonferroni eða Scheffé (sjá bls. 742–748)?
46. Í hverju felast eftir á samanburðir (*post hoc comparisons; multiple comparisons*) og hvaða vandi fylgir því að gera marga eftir á samanburði þar sem hver og einn miðast við α ?



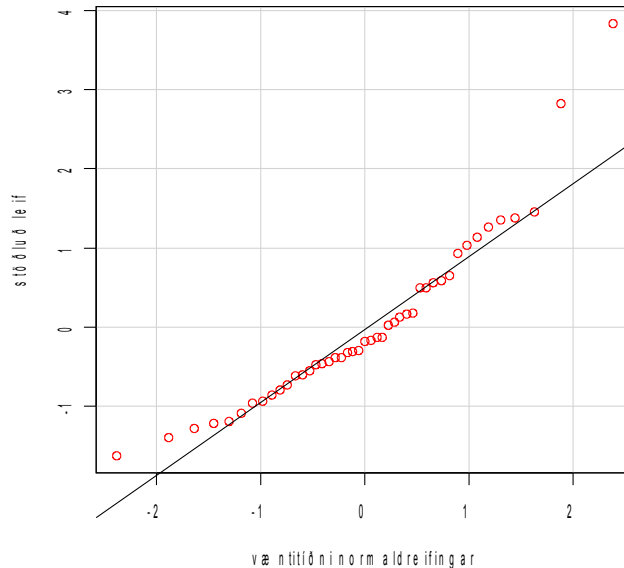
47. Myndin sýnir normalrit (*probability plot*) og kassarit (*box plot*) fyrir staðlaða leif (*standardized residual*) aðfallsgreiningar (*regression*). Segðu mér tvennt: (a) Hvaða frávik (*deviation*) frá forsendum (*assumptions*) aðfallsgreiningar sjást almennt (*can in general be detected*) á normal- og kassaritum? (b) Sjást einhver þessara frávik (*can any such deviations be seen*) á myndinni? Rökstyddu og vertu nákvæm og skýr í framsetningu.
48. Sæktu skrána `happiness.sav` í heimaprófsmöppuna í Uglu. Þessi skrá er einnig að finna á geisladiskinum sem fylgir kennslubókinni og er notuð í dæmum 11.31–11.33. Lestu vel lýsinguna á gagnasafninu í viðauka D4 í kennslubókinni og kynntu þér upplýsingar sem koma fram í dæmunum sem nefnd eru hér að ofan. **H**
- Gerðu marghliða aðfallsgreiningu þar LSI er fylgibreytan en DEMOCRAC, GINI og LIFE eru frumbreytur. Túlkaðu niðurstöðuna með tilliti til hallastuðlanna.
 - Gerðu aðra aðfallsgreiningu þar sem breytan CORRUPT bætist við sem frumbreyta. Túlkaðu hallastuðlana hérna einnig.
 - Reiknaðu út hver meðallífshamingja þjóðar ætti að vera samkvæmt líkönunum tveimur miðað við að DEMOCRAC taki gildin 0, 1, 2, 3, 4, 5 og 6 en aðrar breytur séu settar jafnar meðaltölum sínum. Birtu þessar niðurstöður sem mynd þar sem stig lýðræðis (DEMOCRAC) er á lárétta ásnum en spáð lífshamingja (LSI) er á lóðrétta ásnum. Myndin ætti að hafa eina línu fyrir fyrra líkanið og aðra fyrir seinna líkanið.
 - Útskýrðu (a) hvernig línurnar eru ólíkar og (b) hvers vegna hallastuðullinn fyrir DEMOCRAC breytist milli líkana. Athugaðu að það er ætlast til þess að þau skiljir í hvaða grundvallaratriðum líkönin eru ólík og sért fær um að túlka þann mun á skýran og skiljanlegan hátt.
49. Útskýrðu hvaða upplýsingar fást með R^2 í aðfallsgreiningu. Miðaðu við að svara eftirfarandi spurningum skýrt og afdráttarlaust.
- Hvaða er þessi stuðull að mæla nákvæmlega?
 - Hvað þarf R^2 að vera hátt til þess að það sé að marka niðurstöðu aðfallsgreiningar? Gættu þess að ræða þetta og rökstyðja ítarlega!
50. Á einni glæru námskeiðsins stendur eftirfarandi: „Dreifing miðar við að villan hafi sama staðalfrávik í öllum hópum.“
- Þýðir þetta að það verði að vera sama staðalfrávik í öllum hópum rannsóknarinnar í úrtakinu? Útskýrðu og rökstyddu vandlega!
 - Teiknaðu (handteiknuð mynd er fullnægjandi) kassarit þar sem þessi forsenda (*assumption*) er augljóslega brotin.

- iii. Hvaða aðferðir höfum við til að meta þessa forsendu og hvernig samhæfum við niðurstöður þeirra?



51. [Bell og Pliner \(2003\)](#) gerðu athugun á því hvernig fjöldi þeirra sem situr við borð hefur áhrif á hversu lengi er setið við máltíð.
- Þeir fundu meginhrif fyrir fjölda þeirra sem matast var með, $F(4, 352) = 58,0, p < 0,001$, þannig að matartíminn varð stystur (17 mínútur) þegar aðeins einn sat til borðs, tæplega 33 mínútur þegar þrír sátu til borðs og rúm 41 mínúta þegar fimm (eða fleiri) sátu til borðs. Þeir fengu einnig meginhrif fyrir þann stað sem matast var á, $F(2, 352) = 232,6, p < 0,001$, þar sem matartíminn varð lengstur (46 mínútur) þegar matast var á miðlungsdýrum veitingarstað, næstlengstur (31 mín.) í mötuneyti og stystur (18 mín.) þegar matast var á skyndibitastað.
- Það flækta málið eilítið að í ljós kom samvirkni ofangreindra tveggja breyta, $F(8, 352) = 7,7, p < 0,001$. Myndin sýnir samvirkni breytanna.
- Gerðu nákvæma grein fyrir vandanum við að meta meginhrif (*main effect*) þegar samvirknihrif (*interaction effect*) eru til staðar.
 - Í hverju eru samvirknihrifin nákvæmlega fólgin? Gættu þess að í skýringu þinni komi skýrt fram bæði almenn útskýring á samvirknihrifum og útlitun á samvirknihrifunum eins og þau birtast á myndinni hér að ofan.
 - Er í þessu tiltekna tilviki merkingarbært að túlka meginhrifin? Útskýrðu nákvæmlega og gættu þess sérstaklega að túlka hvor meginhrif fyrir sig ef þú telur þau túlkanleg eða tilgreina að öðrum kosti nákvæmlega í hvoru tilviki fyrir sig í hvaða skilningi þau eru ótúlkanleg.
 - Ef þú telur meginhrifin vera túlkanleg, rissaðu þá upp samsvarandi mynd þar sem meginhrifin væru ótúlkanleg. Ef þú hins vegar telur meginhrifin vera ótúlkanleg, rissaðu þá upp samsvarandi mynd sem sýnir samsvorandi túlkanleg meginhrif.
52. Í marghliða aðfallsgreiningu (*multiple linear regression*) eru hallastuðlar (*regression coefficients*) sérhæfir (*specific*) þannig að leiðrétt (*controlled for*) er fyrir aðrar frumbreytur í líkaninu.
- Hvað nákvæmlega felst í þessu sérhæfi og leiðréttingu fyrir aðrar breytur?
 - Útskýrðu með eigin dæmi úr daglegu lífi (*example from everyday life*) hrif (effect) sem eru ýmist sérhæf eða ekki sérhæf. Gættu að því að þetta verður að vera eigið dæmi sett fram með eigin orðalagi. Sérstaklega þarf að varast að nota dæmi úr fyrirlesturum, glærum eða úr sambærilegum áttum.
 - Oft er fullyrt að konur séu betri öikumenn en karlar. Ef ég hins vegar segi að konur séu ekkert betri öikumenn en karlar, vegna þess að færri óhöpp kvenna skýrist af því að þær keyri að jafnaði minna en karlar, hvort er ég þá að vísa til sérhæfra eða ósérhæfra hrifa? Rökstyddu vandlega.
53. Ef allsherjarpróf (*omnibus test*) einhliða dreifigreiningar er marktækt, þá má ég hafna núlltilgátunni (null hypothesis).

- i. Útskýrðu nákvæmlega hver núlltilgátan er.
- ii. Ef ég hafna núlltilgátunni, nákvæmlega hvaða ályktun get ég dregið?
- iii. Ef allsherjarprófið er ómarktækt, get ég þá staðfest núlltilgátuna? Útskýrðu nákvæmlega og tilgreindu hvaða ályktun ég get dregið í þessu tilviki.



54. Skoðaðu vel þessa mynd sem sýnir normalrit (*normality plot; quantile plot*) af leif (residual) í dreifgreiningu (ANOVA). Athugaðu vel hvernig myndin snýr: Leifin er á lóðrétta ásnum en væntitiðni normaldreifingar (*expected normal value*) á lárétta ásnum!
 - i. Lýstu myndinni allnákvæmlega þannig að fram komi helstu einkenni á lögum dreifingarinnar (*shape of distribution*).
 - ii. Hvaða frávik (*deviations*)—ef einhver—frá forsendum dreifgreiningar (assumptions of ANOVA) sýnir þessi mynd?
 - iii. Rissaðu upp (*draw*) kassarit (box plot) sem sýnir sömu dreifingu, þ.e. sýnir öll helstu frávik sem er að sjá á normalritinu.

55. Í sýnidæmi (example) 13.10 á bls. 693 í 6. útgáfu kennslubókarinnar (bls. 781 í 5. útgáfu) er sýnd samvirkni milli þjóðernis hlustenda og hvaða landi tónlist tengist. Kynntu þér sýnidæmið vel og umræðu um samvirkni (interaction) í þessum kafla kennslubókarinnar, á glærum og umfjöllun í kennslustund. **H**
 - i. Útskýrðu nákvæmlega hvað samvirkni er bæði með því að (a) setja fram eins konar skilgreiningu (með þínum eigin orðum) og (b) útskýra með tilbúnu frumsömdu dæmi.
 - ii. Útskýrðu vandlega hvers vegna það er ekki hægt að túlka meginhrif í sýnidæminu. Svo virðist af meðaltölum að Túnisar séu með betri takt en Frakkar, hvað er rangt við að fullyrða það?
 - iii. Rissaðu upp tvær sams konar myndir sem sýna Frakka og Túnisa, franska tónlist og túniska tónlist og færni í því að klappa í takt. (a) Önnur myndin þarf að sýna auðtúlkaleg meginhrif þrátt fyrir samvirkni en (b) hin meginhrif sem eru þannig að það gæti verið mjög villandi að túlka þau. Útskýrðu í hvoru tilviki fyrir sig hvers vegna meginhrif eru túlkaleg eða villandi. [Svona riss getur hæglega verið frihendis með blýanti eða penna. Ef þú hefur reynt af því, gætirðu einnig teiknað þetta í ritvinnsluforriti með tækjum á Drawing tólastikunni, en þá ertu kanski komin í óþarflega mikið föndur. Svo ef einhver endilega vill hafa mikið fyrir þessu, þá

má auðvitað teikna þetta í töflureikni eða í sérhæfðu teikniforriti. Það er samt vandfundið öflugra verkfæri en velyddaður blýantur.]

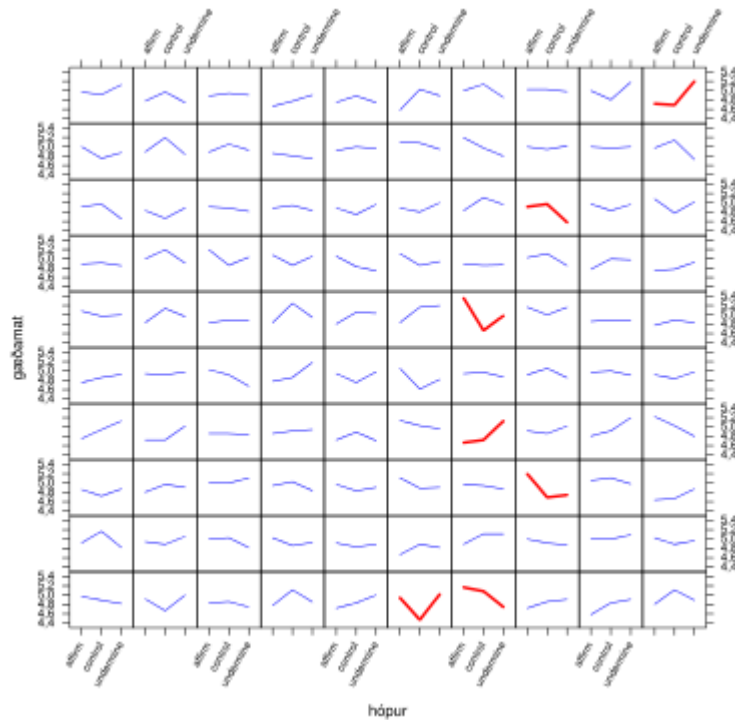
56. Þegar unnið er með tvískipta (*binominal; dichotomous*) fylgibreytu kemur til álita að nota venjulega línulega aðfallsgreiningu (*linear regression*) í stað aðfallsgreiningar hlutfalla (*logistic regression*).
- Hvaða tvö atriði mæla eindregið gegn því að nota línulega aðfallsgreiningu og hvers vegna?
 - Hvernig er fylgibreytan meðhöndluð (*modified*) í aðfallsgreiningu hlutfalla til að komast á móts við ofangreind vandamál?
57. Sólveig vildi athuga með línulegri aðfallsgreiningu (*linear regression*) hvaða breytur tengdust niðurstöðu á spurningalista um depurð (þunglyndi; *depression*) ætluðum fyrir börn og unglinga. Fylgibreytan var niðurstaðan á spurningalistanum en meðal frumbreyta var kynferði, aldur, félagsleg leikni og gæði samskipta við foreldra. Allsherjar (*omnibus*) F -prófið gaf eftirfarandi niðurstöðu: $F(4, 176) = 2,5, p = 0,04$.
- Hvaða ályktun getur Sólveig dregið um það hvort frumbreyturnar tengist niðurstöðunni á spurningalistanum? Nánar tiltekið: Er henni stætt á því að fullyrða að líkanið skýri depurð barnanna eins og hún birtist á spurningalistanum? Útskýrðu nákvæmlega hvers vegna hún getur það eða, eftir atvikum, getur það ekki.
 - Hver nákvæmlega er núlltilgáta (*null hypothesis*) allsherjarprófsins? Útskýrðu hana bæði í orðum og ritaðu hana með réttum táknum.
 - Setjum nú sem svo að allsherjarprófið sé marktækt. Hvaða ályktun nákvæmlega getur Sólveig dregið um einstaka hallastuðla (*regression coefficients*)?
58. Hver er munurinn á fyrirfram samanburði (*a priori contrast*) og eftir á samanburði (*post hoc comparison*) með aðferð Scheffé?
- Hvað felst í því nákvæmlega að samanburður sé eftir á í stað fyrirfram?
 - Hvers vegna þarf ég að leiðrétta eftir á samanburði?
 - Í hverju felst leiðréttingin þegar Scheffé á í hlut, hvernig fer leiðréttingin fram?
 - Í hvaða tilvikum ætti ég að nota aðferð Scheffé fremur en aðferð Bonferroni? Komdu með skýrt dæmi.
59. Tveir nemendur báru saman BA-verkefnið sín. Annar, Sæunn, kannaði áhrif efnahagshruns og skuldastöðu á tilfinningalega líðan fólks meðan hinn, Pétur, þróaði stytta útgáfu af algengu greindarprófi. Bæði notuðu línulega aðfallsgreiningu (*linear regression*) og R^2 var 0,18 hjá Sæunni en miklu hærri hjá Páli. Páli fannst þetta vera mjög lélegur árangur hjá Sæunni.
- Útskýrðu hvað R^2 er nákvæmlega að meta. Vertu nákvæm í útskýringu þinni.
 - Hvað þarf R^2 að vera hátt til þess að það sé nægjanlegt. Útskýrðu nákvæmlega.
 - Er rannsókn Sæunnar alveg vonlaus og í hverju gætu þá mistök hennar verið fölgin?
60. Hver eru helstu viðmið eða aðferðir við að velja fjölda þátta (*factors*) í þáttagreiningu (*factor analysis*)? Útskýrðu hverja um sig (*each and every one*) nákvæmlega.
61. Hvað er átt við þegar talað er um allsherjarpróf (*omnibus test*) dreifigreiningar? Hvaða núlltilgáta (*null hypothesis*) er prófuð og hvaða ályktanir má draga ef prófið er marktækt (*significant*) eða ómarktækt? Hver er meginveikleiki (*main weakness*) allsherjarprófsins?
62. Gerðu grein fyrir því (a) hvaða forsendu (*assumption*) misleitni (*heteroscedacity; unequal variance*) brýtur í dreifigreiningu, (b) hvernig skuli kanna forsenduna myndrænt og (c) hvaða þumalputtareglu megi hafa um það hvort misleitni sé til staðar eða ekki.

63. Í sýnidæmi 12.25 á bls. 627 í 7. útgáfu kennslubókarinnar er fjallað um áhrif augnlitar á viðbrögð við auglýsingu. Lestu lýsinguna vel svo þú sért vel heima í efninu. Þú getur fundið upprunalegu greinina með [leit í Scholar](#) þótt það sé óvíst að hún hjálpi umtalsvert. **H**

Framkvæmdu einhliða dreifigreiningu þar sem rannsóknarspurningin er hvort augnlitur hafi áhrif á viðhorf þátttakenda. Augnlitur var ýmist blár, brúnn, grænn eða enginn, þ.e. ekki sást í augu fyrirsætunnar þar sem hún leit niður.

Gagnaskrána `eyes.sav` eða `eyes.txt` finnur þú í heimaprófsmöppunni.

- i. Settu fram núlltilgátu og aðaltilgátu í samræmi við ofangreinda rannsóknarspurningu.
 - ii. Birtu töflu með meðaltölum, staðalfrávikum og fjölda í hverju hólf sniðsins. Taflan þarf að vera á APA formi sbr. t.d. töfluna á glærunni **Samanburðum með Bonferroni** í fyrirlestrinum **Einhliða dreifigreiningu**.
 - iii. Framkvæmdu dreifigreiningu og birtu niðurstöður hennar sem hluta af samfelldu máli og að öllu leyti í samræmi við formreglur APA. Gættu að því að þú svarir nákvæmlega eftirfarandi spurningum: (a) Hverjar voru niðurstöður tölfræðiprófsins, (b) hver var núlltilgátan, (c) hver var aðaltilgátan og (d) hvaða upplýsingar gefa niðurstöðurnar um viðkomandi þýði?
 - iv. Athugaðu hvort forsendur úrvinnslunnar standist með því að skoða (a) normalrit og kassarit af leif og (b) skoða staðalfrávik, próf Levenes og kassarit af fylgibreytunni eftir augnlit. Gerðu nákvæma grein fyrir því hvaða forsendur eru metnar í hvoru tilviki fyrir sig (liðir a og b) og hvort þú telur þær hvora um sig fyrir hendi eða ekki.
64. Súsanna vildi kanna hversu oft á viku foreldrar lásu fyrir börn sín fyrir svefninn. Rannsóknin tók til tveggja kynja, þ.e. mæðra og fedra, og þriggja aldurshópa barnanna, 2-5 ára, 6-9 ára og 10-13 ára. Hún fékk 24 þátttakendur í hvert hólf sniðsins, þ.e. 24 fedur 2-5 ára barna, 24 mæður 2-5 ára barna o.s.frv. Hún gætti þess vandlega að aðeins eitt foreldri hvers barns tæki þátt í rannsókninni og að hún næði til eins tiltekens barns þess foreldris.
- Athugaðu að hér ert þú ekki beðin um að vinna úr neinum gögnum og því fylgir engin gagnaskrá þessari spurningu.
- Leystu úr eftirfarandi viðfangsefnum:
- i. Teiknaðu myndrit sem sýnir niðurstöður þar sem eru tvö meginhrif, meginhrif kyns foreldris og meginhrif fyrir aldur barns.
 - ii. Skilgreindu í hverju samvirkni væri nákvæmlega fólgin með hliðsjón af þessu dæmi. Gættu þess að skilgreining þín sé tæknilega alveg rétt.
 - iii. Teiknaðu myndrit sem sýnir samvirkni milli kyns foreldris og aldurs barns. Gættu þess að meginhrifin séu túlkanleg þrátt fyrir samvirknina. Útskýrðu hvers vegna hægt sé að túlka meginhrifin og í útskýrðu í hvaða tilvikum þau væru ekki túlkanleg.
 - iv. Gefðu upp (a) núlltilgátuna fyrir samvirkni kyns foreldra og aldurs barns, (b) gefðu upp niðurstöður sem hluta af samfelldum texta miðað við að niðurstaða F -prófsins hafi verið 3,1 og (c) gættu þess að þú túlki niðurstöðuna. Í ofangreindu þarftu að gæta þess að fylgja formreglum APA út í yst æsar.
65. Hvaða eiginleika leifar (*residual*) er rétt að skoða þegar gerð er dreifigreining? Tilgreindu hvern eiginleika, forsenduna sem er prófuð og aðferðirnar sem best er að nota.



Mynd 1. Niðurstöður allsherjarprófs í 100 tilviljunarúrtökum dregnum úr þýði þar sem núlltilgátan er rétt.

66. Á mynd 1 er sýnd niðurstaða allsherjarprófs (*omnibus test*) í 100 úrtökum dregnum úr þýði. Myndin er sambærileg við myndina á glærinni **Hermun (simulation)** í fyrirlestrinum **Einhliða dreifigreiningu**. Línan er rauð og þykkari í þeim úrtökum þar sem allsherjarprófið var marktækt en blá þar sem prófið var ómarktækt.
- Í rannsókn safnar þú gögnum í úrtaki úr þýði og úrtakið er eina úrtakið sem þú hefur aðgang að. Hugsaðu um hvern reit fyrir sig (*taking each panel separately*) og gerðu grein fyrir þeirri ályktun sem þú gætir dregið ef reiturinn sýndi þetta eina úrtakið miðað við að allsherjarprófið reynist (a) ómarktækt og (b) ómarktækt. Rökstyddu.
 - Nú vill svo til að úrtökin voru öll dregin úr þýði þar sem núlltilgáta allsherjarprófsins var rétt. Það eru því viðbótarupplýsingar sem þú byggir á í þessum og seinni liðum spurningarinnar. Nú er prófið marktækt í sjö af hundrað úrtökum úr þessu þýði, samræmist það því að alfa sé 0,05? Útskýrðu og rökstyddu annað hvort (a) að þetta sé eðlilegt og hvers vegna eða (b) að þetta sé óeðlilegt og hvaða rök séu fyrir því.
 - Í þessum lið áttu að bæta við þeirri forsendu (*assumption*) að úrtökin séu dregin með tilviljunaraðferð (*random sampling*). Við skulum byrjum neðst vinstra megin og teljum reitina frá vinstri til hægri og neðan frá og upp. Þá er allsherjarprófið marktækt í sjötta úrtakinu og meðaltölin hæst fyrir *affirm* og *undermine*, í sjöunda úrtakinu eru þau hæst fyrir *affirm* og *control*, o.s.frv. fyrir hin úrtökin með marktæk próf. Gerðu (a) grein fyrir því hversu ólík túlkunin verður eftir því hvert úrtakið er og (b) útskýrðu hvaða vanda þetta felur í sér við túlkun stakra rannsókna.
67. Útskýrðu (a) hvað nákvæmlega felst í eftir á (*post hoc*) prófum í dreifigreiningu, (b) hvaða vandi fylgir þeim og (c) hvernig megi bregðast við þeim vanda. Gættu þess að í svari þínu komi fram og þú útskýri m.a. hugtökin alfa (*alpha*), höfnunarmistök (*Type I error*), eftir á próf (*post hoc text*), fyrirfram próf (*a priori test; contrast*), marktækt (*statistical significance*) og núlltilgáta (*null hypothesis*).
Útskýrði einnig (d) fyrirfram samanburði og útskýrðu hvers vegna þessi vandi fylgir þeim ekki.

66. Í dæmi 13.8 á bls. 654 í sjöundu útgáfu kennslubókarinnar eru sýnd meðaltöl eftir kyni og þjóðerni fyrir spurningu sem metur hópþrýsting (*peer pressure*) í tengslum við frammistöðu í stærðfræði (*mathematics achievement*). Einnig fylgja upplýsingar um marktækt í dreifigreiningu en þó engin F -próf. **H**

Ef þú vilt, getur þú nálgast [upprunalegu greinina](#) á vefnum en gættu þess að herma ekki eftir framsetningunni þar.

Við höfum rætt ítarlega í kennslustund um framsetningu myndrita, túlkun samvirknihrifa og framsetningu niðurstaðna í rituðu máli.

- i. Teiknaðu mynd af niðurstöðunum og gættu þess að hún sé alveg smásmugulega í samræmi við stílreglur APA.
 - ii. Lýstu niðurstöðunum í samfelldum texta í samræmi við myndritið og aðrar upplýsingar sem þú hefur úr dæminu í bókinni. Gættu þess að þú hefur ekki F -prófin og því búumst við ekki nákvæmum talnategum tölfræðiupplýsingum í svarinu. Að öðru leyti þarf textinn að fylgja APA stílreglum til hlítar og vera að öllu leyti eins og samsvarandi texti í niðurstöðukafla fræðigreinar.
 - iii. (a) Útskýrðu nákvæmlega hvað felst í samvirkni (*interaction*), vertu nákvæm og ögn tæknileg í þessu (*specific and technical exposition*). (b) Gerðu grein fyrir því í hverju samvirknin felst í þessu tiltekna tilviki. (c) Útskýrðu hvort hvor meginhrif um sig séu túlkanleg í þessu tilviki og þá hvers vegna eða hvers vegna ekki.
67. Í þáttagreiningu er lausninni snúið (*rotation*) og þannig leitast við að ná einfaldri formgerð (*simple structure*). Hvert er mikilvægi einfaldrar formgerðar og í hverju er hún fólgin í aðalatriðum?

© 2004–2013 Guðmundur B. Arnkelsson (síðast breytt 9. desember 2013)