

Inngangur að ályktunum

Fyrirlestur í Tölfræði II (SÁL203G)

Spá eða tilgátuprófun

Í spá (estimation) höfum við úrtak sem er rétt dregið úr þýði og viljum álykta um eiginleika þess í þýði. Við notum úrtakstölur (statistic) sem spátölur (*estimate*) til að álykta um þýðistölurnar (*parameter*).

Ef meðalhæð í úrtaki 20 kvenna er 172,3 cm, get ég ályktað að meðaltal allra kvenna sé u.þ.b. 172,3 cm. Það er örugglega ekki nákvæmlega rétt en það er sennilegasta þýðisgildið miðað við þær upplýsingar sem ég hef.

Tilgátuprófun felst í því að leita svara við tilteknum spurningum. Þá er sett fram tilgáta og mælitölur úrtaksins notaðar til að ákvarða hvort tilgátan sé rétt eða ekki. Þetta felur *ekki* í sér spá.

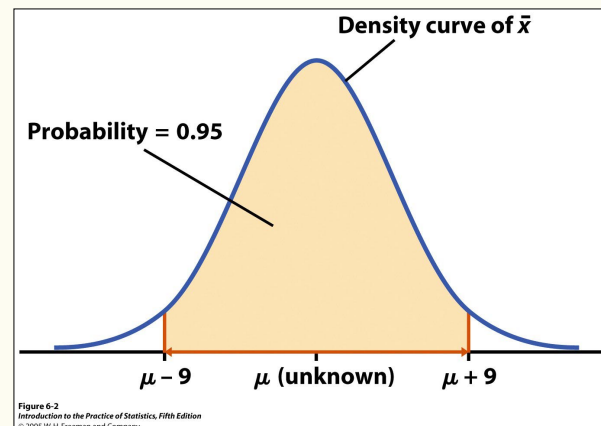
Ef ég vil vita hvort líkamshæð sé ólík eftir kynjum, set ég það fram á formi tilgátu. $H_1: \mu_{kk} \neq \mu_{kvk}$ Ef meðalhæð er mjög ólík eftir kynjum í úrtakinu, get ég ályktað—samkvæmt sérstökum reglum—að tilgátan sé rétt. Ég veit þó ekki hversu röng hún er; ég þarf spá til að meta hve mikill munurinn er.

Bilspá

Í punktspá (*point estimate*) finnum við sennilegustu þýðistöluna. Það er villandi þar sem óvissan getur verið veruleg.

Í bilspá (*interval estimate*) tilgreinum við talnabil sem sennilega inniheldur þýðistöluna. Ef bilið er vítt, er mikil óvissa en lítil, ef bilið er þröngt. Bilspá gefur því fyllri upplýsingar um þýðistöluna heldur en punktspá.

Öryggisbil (*confidence interval*) gefur bilspá miðað við ákveðna tilgreinda óvissu, svonefnt öryggi.



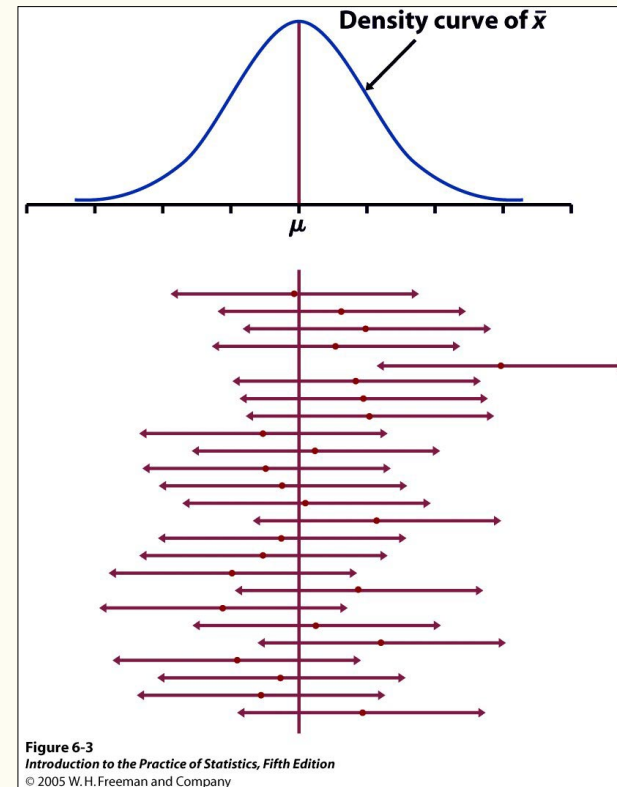
Úrtaksmeðaltöl 500 manna úrtaka þar sem þýðisstaðalfrávikikið er 100 víkja í 95% tilvika 9 stig eða minna frá þýðismeðaltalinu—eða þýðismeðaltalið frá úrtaksmeðaltölunum, það er sami hluturinn! Talnabilið $\bar{X} - 9$ til $\bar{X} + 9$ inniheldur því þýðimeðaltalið í 95% af úrtökunum.

Túlkun öryggisbila

Myndin sýnir 95% öryggisbil í 25 úrtökum úr sama þýði; öll nema eitt öryggisbil, 94% bilanna, innihalda þýðistöluna.

Öryggið vísar til líkinda yfir mörg úrtök, 95% öryggisbilanna innihalda þýðismeðaltali. Þetta eru ekki líkindi þess að *eitt tiltekið* bil innihaldi þýðismeðaltalið. Við vitum heldur ekki *hvar* í tilteknu öryggisbili þýðismeðaltalið er.

Líkindin vísa því til *aðferðarinnar* sem notuð er, ekki til tiltekinna bila.



Eiginleikar öryggisbila

Öryggisbil gefur bæði upplýsingar um þýðið og þá óvissu sem er til staðar. Þau eru því skýrustu upplýsingarnar sem við fáum.

Bilin miðast við það öryggi sem við kjósum. Algengast er að miða við 95% en einnig er hægt að miða við 99%. Það er mikilvægt að tilgreina öryggið með skýrum hætti.

Bilið stækkar við aukna óvissu, þ.e. við minni úrtaksstærð og stærra úrtaksstaðalfrávik.

Stök öryggisbil má ekki túlka út frá líkindum. Við vitum *ekki* líkindi þess að eitt tiltekið öryggisbil innihaldi þýðistöluna. Við vitum heldur ekki *hvar* í talnabilinu þýðistalan liggur líklegast. Best er að hugsa um tiltekin öryggisbil sem sennileika- eða trúverðugleikabil—það er *trúlegt* að þýðistalan sé innan öryggisbilsins.

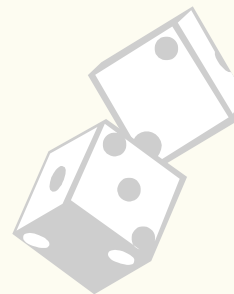
Það eru 95% líkindi—reiknað yfir mörg úrtök—að *aðferðin* skili bili sem innihaldi þýðistöluna.

Marktektarpróf

Marktektarpróf leitast við að svara tilteknum spurningum um þýðið. Þau eru því eðlisólík öryggisbilum sem reyna að lýsa eiginleikum þýðisins.

Við marktektarprófum búum við til líkan af aðstæðum í þýði og berum það saman við niðurstöður úrtaksins. Ef úrtaksniðurstöður víkja mikið frá þýðislíkaninu, minnkar trúverðugleiki líkansins.

Við höfnum líkaninu ef úrtakið væri mjög óvenjulegt samkvæmt líkaninu.



Gerum ráð fyrir að teningur í Lúdó sé réttur, þ.e. sömu líkur séu á að fá allar sex tölur. Líkurnar á því að fá 6 eru því $1/6$ eða $0,17$.

Tóti vill nota sinn eigin tening og fær sexu tvisvar í röð. Líkurnar á því eru tæp 3% og því óvenjulegt. Ef hann hefði fengið þrjár sexur í röð, væru líkurnar um 0,4% og grunsemdir gætu vaknað. Líkurnar á fjórum sexur í röð eru aðeins um 0,08% og því rétt að krefjast þess að Tóti noti sama tening og þú.

Tilgátur

Marktektarprófun byggist á því að hægt sé að mynda líkindalíkan sem lýsir ímynduðum aðstæðum í þýði. Flestar tilgátur eru þannig að þetta er ekki hægt.

Tilgátur sem á að prófa, aðaltilgátur, tilgreina að það sé munur á einhverju tvennu eða einhver mæling sé ólík einhverri stærð. Við viljum vita hvort það sé launamunur á kynjunum, hvort meðferð *bæti* líðan, hvort viðhorf til útlendinga sé *neikvætt*, hvort örvun auki þroska barna, o.s.frv.

Tilgáta sem t.d. segir tvö meðaltöl ólík, lýsir ekki tilteknu ástandi í þýði. Líkindalíkan krefst ákveðins ástands í þýði; að þýðistölur séu tilgreindar.

Tilgátan $\mu_{kk} \neq \mu_{kvk}$ verður því ekki prófuð beint. Andhverfan, $\mu_{kk} = \mu_{kvk}$, þ.e. $\mu_{kk} - \mu_{kvk} = 0$, tilgreinir hins vegar tiltekið ástand í þýði og því prófanleg.

Ef andhverfan, núlltilgátan, er röng, hlýtur aðaltilgátan að vera rétt. Ef við getum hrakið núlltilgátunni, höfum við því staðfest aðaltilgátuna.

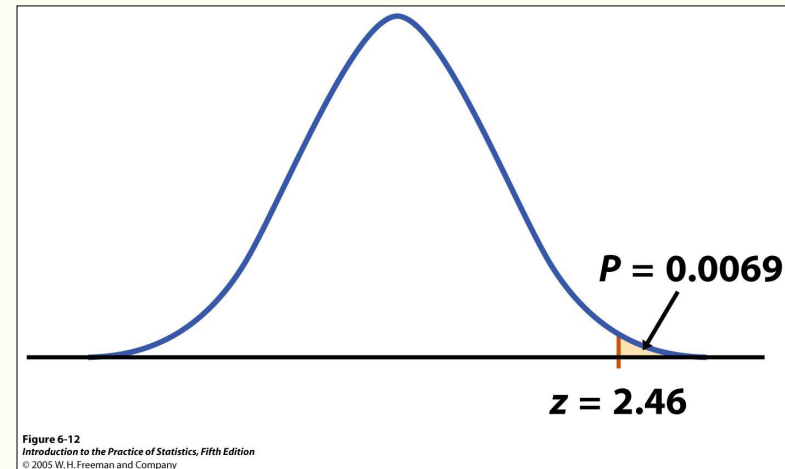
Marktektarprófun

Marktektarpróf byggist á aðaltilgátu, H_a . Sú tilgáta tilgreinir engin sérstök þýðisgildi og er því ekki prófanleg.

Andhverfan, núlltilgátan, gefur upp eitt gildi og er því prófanleg.

Við búum til líkindalíkan og berum úrtaksmeðaltalið saman við það.

Ef úrtaksmeðaltalið er ólíklegt samkvæmt núlltilgátunni, veikir það „trú“ okkar á H_0 og eykur trúleika aðaltilgátunnar.



$$H_0 : \mu = 450; H_a : \mu > 450$$

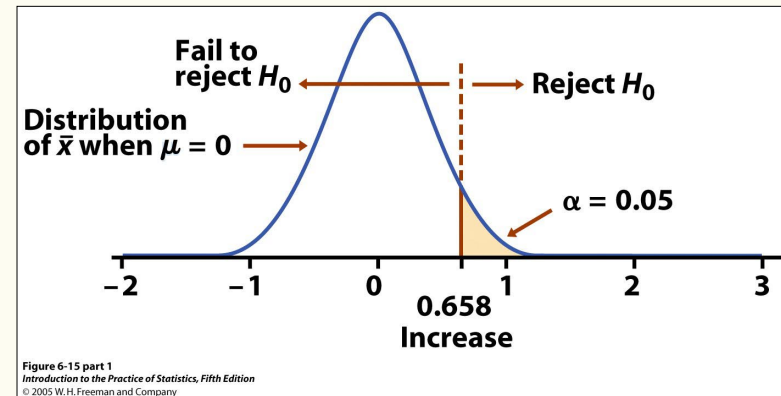
Líkurnar á meðaltalinu 461 eða meira fráviki í þá átt frá H_0 eru um 0,69%.

Ákvörðunarreglan

Núlltilgátan stendur fyrir ímyndað ástand í þýði—andhverfu H_a .

Við ákvörðum α , áhættuna sem við viljum taka undir H_0 . Þar með höfum við ákvarðað hvað frávikið frá H_0 þarf að vera mikið til að við höfnum henni og tökum upp H_a .

Ef frávikið er minna eru líkurnar á niðurstöðunni *meiri* en α og við höfnum ekki H_0 ; ef frávikið er meira, eru líkurnar minni en α og við höfnum núlltilgátunni.



$$H_0 : \mu = 0; H_a : \mu > 0$$

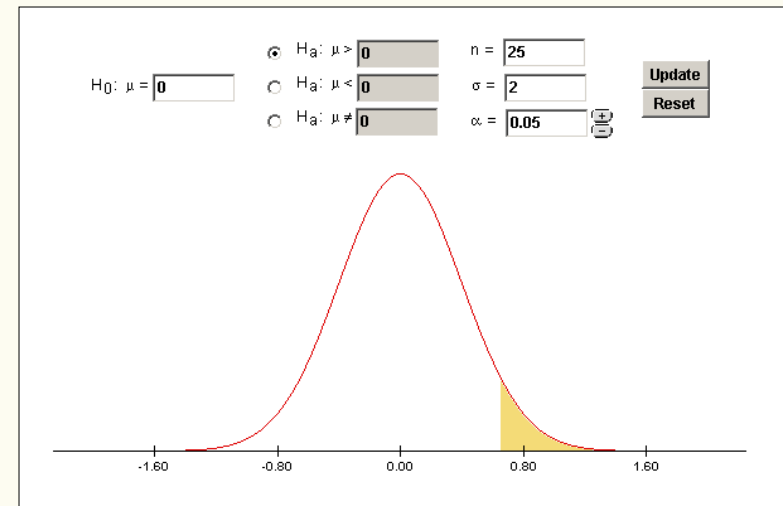
Hér erum við með stefnutilgátu og því eru líkurnar aðeins reiknaðar í efri enda ferilsins. Við höfnum H_0 í hvert sinn sem $M \geq 0,658$.

Marktektarforritlingur

Marktektarforritlingur á heimasvæði bókarinnar gerir okkur kleift að prófa okkur áfram með marktekt.

Ef við stækkum úrtakið (n), minnkum staðalfrávikinu (σ) eða hækkum α (t.d. úr 0,01 í 0,05), minnkar það frávik frá H_0 sem þarf til að geta hafnað henni.

Við veljum sjálf α og ákvörðum þannig þá áhættu sem við viljum taka þegar H_0 er sönn. Stærð úrtaksins ræðst af hönnun rannsóknarinnar en σ af breytileika og mæliaðferðum.



http://bcs.whfreeman.com/ips5e/content/cat_010/applets/statsig.html

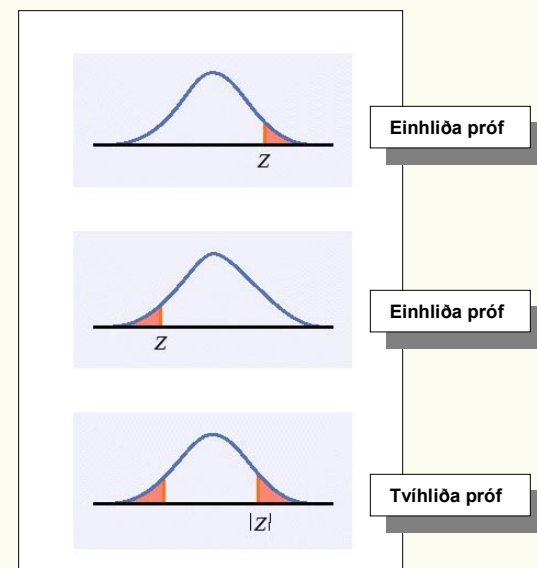
z-próf

z-próf er einfaldasta marktektarprófið og byggist á því að við þekkjum staðalfrávik þýðis.

Prófið getur ýmist verið einhliða eða tvíhliða eftir því hvort aðaltilgátan tilgreinir frávik upp á við, niður á við eða í hvora átt sem er.

Ef prófniðurstaðan samsvarar því að líkurnar á þetta miklu eða meira frávik frá núlltilgátunni séu litlar, t.d. minni en 0,05, má hafna núlltilgátu og taka upp aðaltilgátuna.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$



z-gildi	Líkindi undir H_0	
	Einhliða	Tvíhliða
1,64	0,05	0,10
1,96	0,025	0,05
2,33	0,01	0,02
2,58	0,005	0,01

Tölfræðileg marktekt

Áður en við gerum marktektarpróf ákveðum við hve sterkan vitnisburð við viljum áður en við höfnum H_0 . Algengast er að miða við frávik frá núlltilgátu sem er svo mikið að aðeins í 5% tilvika fáist svona mikil eða meiri frávik frá H_0 í úrtaki. Þessi líkindi *ákveðum* við fyrirfram og eru táknuð með α , marktektarmörkunum.

Alfagildið samsvarar þeirri áhættu sem við viljum taka *ef* H_0 er rétt, þ.e. hve oft við erum tilbúin til að hafna H_0 í slíku tilviki.

Prófið gefur p -gildi sem eru líkur þess að fá svona mikil eða meira frávik frá H_0 . Ef p er minna eða jafnt og α , má hafna núlltilgátunni, annars getum við ekki hafnað henni.

Marktektarprófið miðast við ákveðnar forsendur. Ef þær eru réttar, gefur p upp réttar líkur fyrir frávikið og tíðni höfnunarmistaka samsvarar því α . Ef forsendur standast ekki getur p -gildi prófsins of- eða vanmetið líkindin þannig að tíðni höfnunarmistaka samsvari ekki lengur α .

Er hægt að sanna núlltilgátuna?

Ef p -gildið er hærra en α , höfum við ekki næg rök fyrir því að hafna H_0 . Niðurstöðurnar eru þá í samræmi við núlltilgátuna.

Þetta væri eins og að vera í Lúdó og fá stundum sexu, sjaldnast tvær í röð og nánast aldrei þrjár sexur í röð. Við höfum því enga ástæðu til að efast um það að teningurinn sé réttur.

Ef tilgátan er $\mu_{kk} \neq \mu_{kvk}$ þá samræmist ómarktæk niðurstaða núlltilgátunni

$$\mu_{kk} = \mu_{kvk}.$$

Þótt niðurstaðan sé í samræmi við H_0 , getum við **ekki** staðfest hana. Ef niðurstaða samræmist $\mu_{kk} - \mu_{kvk} = 0$, er mjög líklegt að hún sé í samræmi við fjölmargar aðrar líkar tilgátur, t.d. $\mu_{kk} - \mu_{kvk} = 0,1$ eða $\mu_{kk} - \mu_{kvk} = -0,6$.

Til að „staðfesta“ H_0 þurfum við því að fá vitneskju um að frávikið frá H_0 sé það lítið að það skipti ekki máli. Teningurinn í Lúdó er kanski ekki *alveg* réttur en nánast réttur. Þessar upplýsingar gefa marktæktarpróf ekki.

Marktektarpróf og öryggisbil

Marktektarpróf veita upplýsingar sem eru ólíkar öryggisbilum en aðferðirnar tvær tengjast samt náíð.

Marktektarpróf prófar núlltilgátu og getur hafnað henni miðað við ákveðið α . Líkindin vísa ekki til vissu eða hversu rétt ákvörðunin er.

Öryggisbil gefur bilspá, talnabil sem sennilega inniheldur þýðistöluna. Öryggið vísar ekki til viðkomandi bils heldur til þess hversu oft *aðferðin* skilar bili sem inniheldur þýðistöluna.

Þýðistölum sem liggja utan $1-\alpha$ bils myndi vera hafnað af marktektarprófi miðað við α og tvíhliða tilgátu. Við getum því lesið niðurstöðu tvíhliða marktektarprófs út úr öryggisbilinu.

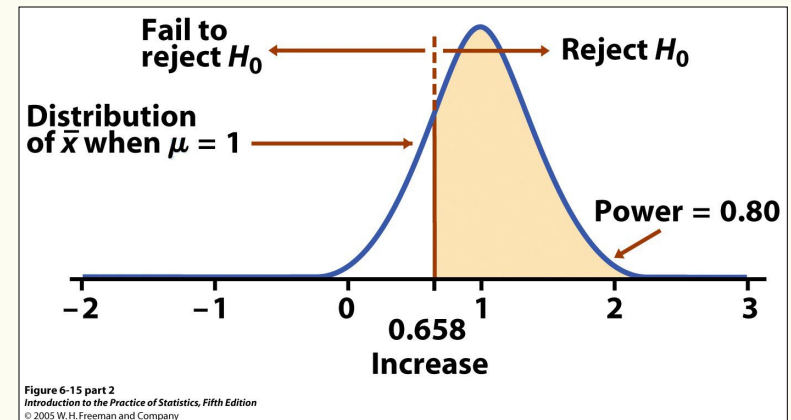
Þetta þýðir *ekki* að samsvarandi próf sé 95% öruggt (miðað við $\alpha=0,05$). Í öryggisbili vísar öryggið til þess að í 95% tilvika skilar aðferðin talnabili sem inniheldur þýðistöluna. Þegar H_0 er rétt, verður próf marktækt í α og ómarktækt í $1-\alpha$ tilvika. 95% birtast því aðeins þegar H_0 er rétt.

Dæmi um ástand í þýði

Ef við gefum okkur tiltekið meðaltal í þýði, getum við reiknað út hversu oft niðurstaða í úrtaki verður marktæk. Þetta getum við aðeins reiknað fyrir tiltekið þýðismeðaltal.

Ef úrtaksmeðaltalið verður 0,658 eða hærra, höfnum við H_0 . Ef $\mu = 1,0$, gerist þetta í 80% tilvika.

Afköst (*power*) fyrir þetta tiltekna þýðismeðaltal ($\mu = 1,0$) er því 0,80. Afköstin eru ýmist meiri eða minni eftir því fráviki frá H_0 sem er skoðað.



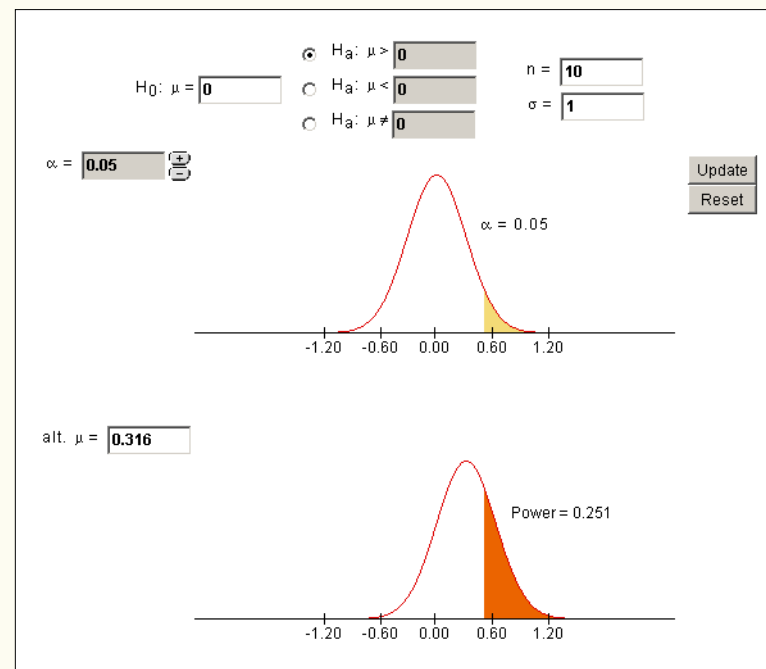
Þessir afkastareikningar miðast við núlltilgátuna— $H_0: \mu = 0$ —tiltekið staðalfrávik, úrtaksstærð og það frávik frá H_0 sem við viljum finna.

Afkastareiknir

Afköst aukast eftir því sem meðaltalið er ólíkara því sem núlltilgátan segir. Það fer því eftir því hvað við viljum finna mikil frávik hvað afköstin verða mikil.

Við höfum áhrif á afköstin með stærð úrtaksins og með þeirri áhættu sem við tökum þegar núlltilgátan er rétt.

Við höfum lítil áhrif á staðalfrávikinu en þó getur mæliaðferðin og nákvæmni hennar haft áhrif á stærð þess.



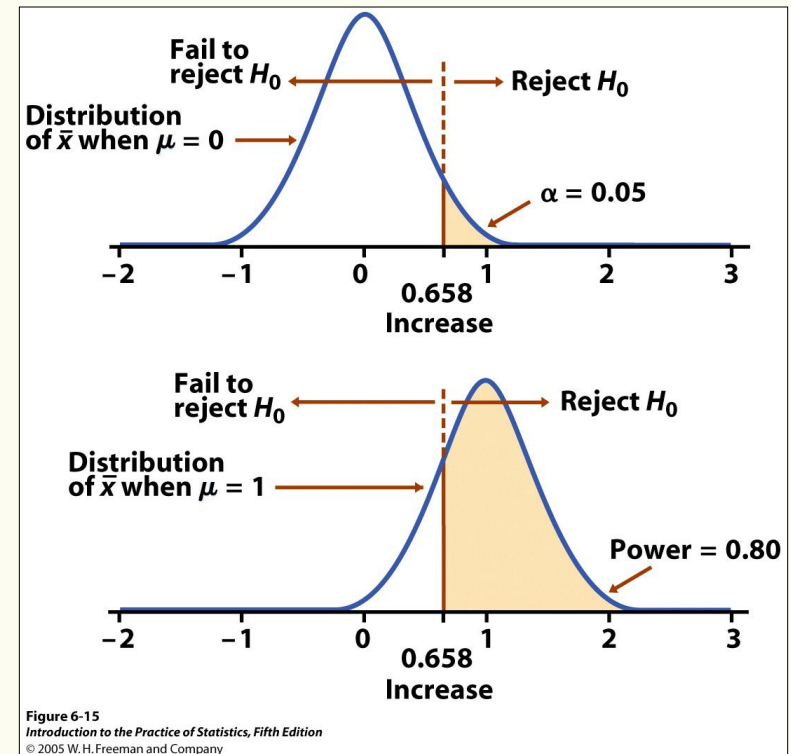
http://bcs.whfreeman.com/ips5e/content/cat_010/applets/power.html

Samspil afkasta og marktektar

Við vitum ekki hvort núlltilgátan er rétt eða ekki. Þótt hún væri röng, myndum við ekki vita hvert frávikið frá henni væri.

Efri myndin sýnir þekkta stærð, þ.e. svona virkar þetta þegar H_0 er rétt. Ég mun taka ranga ákvörðun í α tilvika.

Ef neðri normalkúrfan er rétt staðsett lýsir hún aðstæðum þegar H_a er rétt. Ég myndi komast að réttri niðurstöðu í 80% tilvika þegar frávikið væri það sem er sýnt.



Rökfræði tilgátuprófunar

Ef núlltilgátan er rétt getum við gert þau mistök að hafna henni. Tíðni þess ákvörðum við með α . Við vitum því hverjar líkurnar eru á því að hafna núlltilgátu þegar hún er rétt. Þetta er *ekki* það sama og að vita líkurnar á því að hafna réttri núlltilgátu—það fer einnig eftir líkum þess að H_0 sé rétt.

Ef núlltilgátan er röng getum við gert þau mistök að hafna henni ekki. Líkindi þess getum við reiknað miðað við eitthvert *tiltekið frávík* frá H_0 .

Ákvörðun	Núlltilgáta	
	Rétt	Röng
Ekki hafna	$1 - \alpha$	β
Hafna	α	$1 - \beta$
Samtals	1,00	1,00

Höfnunarmistök
Mistök af tegund I

Fastheldnimistök
Mistök af tegund II

Afköst

Val á afköstum rannsóknar

Oft er miðað við að 80% afköst séu æskilegt. Flestar rannsóknir ná því þó ekki og eru of afkastalítlar.

Afköst minnka eftir því sem α lækkar (t.d. úr 0,05 niður í 0,01). Ef við veljum minni áhættu þegar H_0 er rétt, minnka líkur á því að finna frávikið þegar H_0 er röng.

Afköst aukast með stækkandi úrtaki. En þar ráða fjárhagur, tími og umfang því hve stórt úrtakið getur verið.

Æskileg afköst eru því matsatriði hverju sinni.

Oft er litið svo á að höfnunarmistök séu alvarlegri en fastheldnimistök. Vísindi miðast við uppsöfnun þekkingar, uppsöfnun rangrar þekkingar er þá verri heldur en að missa *tímabundið* af upplýsingum.

Núlltilgáta sem *ekki* er hafnað getur samt orðið til þess að eitthvað er ekki rannsakað—sem er slæmt.

Stundum er það einnig alvarlegt að uppgötva ekki eitthvað, svo sem alvarlegar aukaverkanir lyfja.

Hvernig eru afköst reiknuð?

Hægt er að reikna afköst í höndunum með aðferðum kennslubókarinnar.

Auðveldast og hentugast er þó að nota forrit eða ýmsa forritlinga á vefnum.

Fyrst þarf að ákvarða það frávík sem við viljum finna. Þetta er minnsta frávikið sem er það stórt að við myndum ekki vilja missa af því. Allir afkastareikningarnir miðast við þetta frávík.

Velja α , meta staðalfrávík og velja þá úrtaksstærð sem miða á við.

Oft er erfitt að ákvarða það frávík sem miða á við. Þá eru nokkrir möguleikar.

Athuga hvað aðrar rannsóknir hafa komist að. Þannig má miða við að finna a.m.k. þau frávík frá H_0 sem eru svipuð og koma fram í öðrum sams konar rannsóknum.

Nota staðlaðar áhrifastærðir, þ.e. miða við Cohens d og viðmið hans. Þetta er ekki besti kosturinn en mjög oft sú leið sem þarf að fara.

Cohens d

Cohens d er staðlaður mælikvarði á frávikið frá núlltilgátunni. Stöðlunin fæst með því að mæla frávikið í staðalfrávikum. Þannig fæst mælikvarði sem leiðréttir fyrir breytileika hverrar rannsóknar.

Cohen skilgreindi lítil, miðlungs og mikil frávik.

Lítill frávik eru ekki greinanleg án rannsókna. Miðlungsfrávik eru greinileg fyrir þá sem hafa reynslu af viðfangsefninu. Mikil áhrif eru greinanleg án sérstakrar reynslu, þekkingar eða rannsókna.

$$d = \frac{\mu - \mu_0}{\sigma}$$

Áhrifastærð	d	Lýsing
Lítill	0,2	Erfitt að greina. Hæðarmunur 15 og 16 ára stúlkna.
Miðlungs	0,5	Sjáanlegur munur. Munur á hæð 14 og 18 ára stúlkna
Mikil	0,8	Augsýnilegur munur. Munur á hæð 13 og 18 ára stúlkna.

Cohens d er notað við útreikning afkasta með því einfaldlega að láta staðalfrávikið vera 1,0 og setja Cohens d inn fyrir frávikið.

Takmarkanir tilgátuprófunar

Takmarkanir

Núlltilgátu er hægt að hafna en aldrei samþykkja. Ef próf er ómarktækt er (a) núlltilgátan rétt, (b) óverulegt frávík eða (c) of lítil afköst (*power*).

Ef núlltilgátu er hafnað ber að líta svo á að aðaltilgátan sé rétt. Það þýðir ekki að hún sé raunverulega rétt né að mikil frávík séu frá H_0 .

Flestar núlltilgátur eru rangar strangt til tekið. Það er mjög ósennilegt að þýðistala sé nákvæmlega jöfn viðmiðsgildinu sem H_0 gefur upp.

Möguleg viðbrögð

Gæta að því að afköst séu nægjanleg svo áhugaverð frávík séu líkleg til að uppgötvast

Vanda alla þætti rannsókna svo utanaðkomandi þættir auki ekki dreifingu úrtaksins

Reikna öryggisbil fyrir þýðistöluna til að átta sig á raunverulegri stærð hennar.

Skoða niðurstöður margra rannsókna í senn; aldrei treysta niðurstöðum einstakra rannsókna.

Takmarkanir stakra rannsókna

Myndin sýnir mat stakra rannsókna á hlutfallslegri tíðni hjartaáfalla samfara notkun verkjalyfsins Vioxx.

Vioxx var tekið af markaði haustið 2004. Það er talið hafa valdið alvarlegum hjartasjúkdómum hjá tugum þúsunda Bandaríkjamanna.

Brotalínan sýnir neikvæð áhrif lyfsins miðað við að niðurstöður allra rannsókna séu teknar saman. Stakar rannsóknir hafa greinilega ekki haft nægileg afköst til að greina þetta.

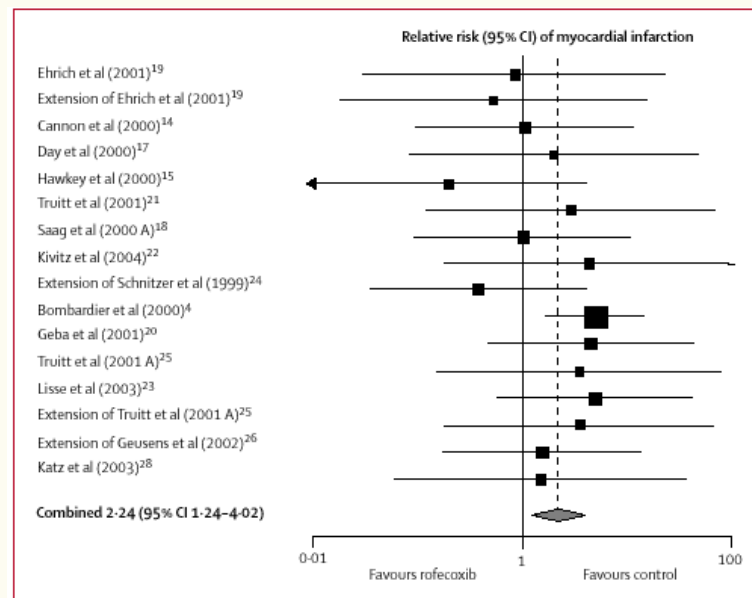


Figure 2: Meta-analysis of randomised trials comparing rofecoxib with control

Jüni, P., Nartey, L., Reichenbach, S, Sterchi, R., Dieppe, P.A., & Egger, M. (2004). Risk of cardiovascular events and rofecoxib: Cumulative meta-analysis. *Lancet*, 364, 2021–2029.

Uppsöfnun þekkingar

Sérstakar aðferðir (*metaanalysis*) má nota til að safna upp niðurstöðum rannsókna og meta þær í heild.

Fyrir Vioxx er greinilegt að fyrstu þrjú árin bendir ekkert til skaðlegra áhrifa á hjartað. Taktu samt eftir því að öryggisbilin útiloka alls ekki skaðleg áhrif.

Frá og með árinu 2000 hefði þó átt að vera ljóst að lyfið hefur þessar alvarlegu aukaverkanir. Samt var ekkert gert næstu fjögur árin.

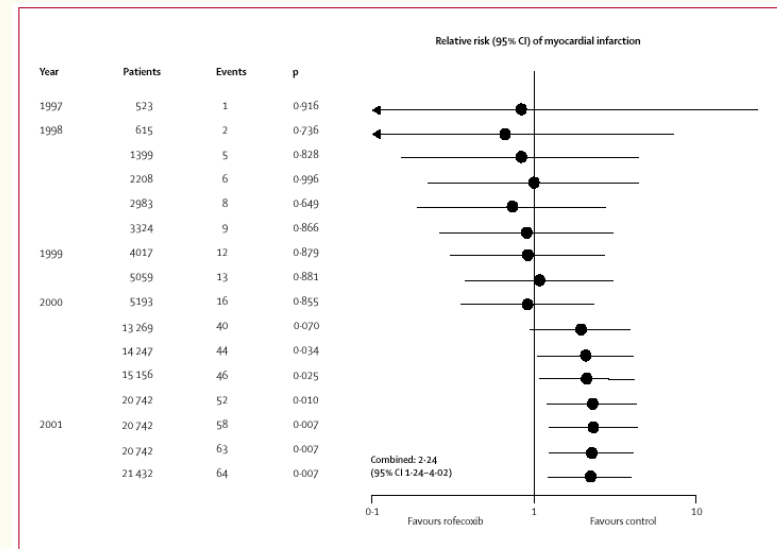


Figure 3: Cumulative meta-analysis of randomised trials comparing rofecoxib with control
See figure 2 for sequence of trials.

Jüni, P., Nartey, L., Reichenbach, S, Sterchi, R., Dieppe, P.A., & Egger, M. (2004). Risk of cardiovascular events and rofecoxib: Cumulative meta-analysis. *Lancet*, 364, 2021–2029.