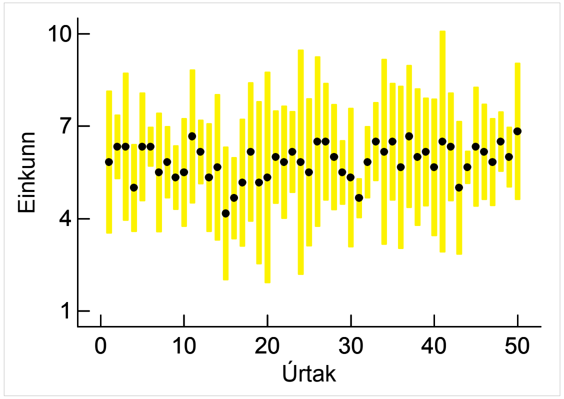


Öryggisbil

Fyrirlestur í Aðferðafræði II

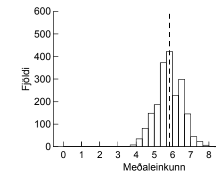
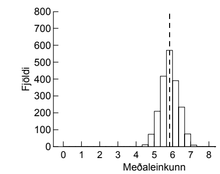
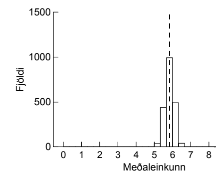
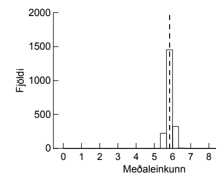
Úrtakadreifing

- Dreifing úrtaka (*sampling distribution*) er dreifing mælitölu milli úrtaka
  – Mismunandi einstaklingar veljast í úrtökin
  – Þetta veldur breytileika milli úrtaka
  – Breytileikinn er háður stærð úrtaksins
    - Því minni úrtök, því ólíkari verða þau hvert frá öðru

A:4.3, bls. 94–99

Breytileiki eftir stærð

- Hver mynd sýnir 2.000 úrtök
  - Meðaltal meðaltalanna er svipað óháð úrtaksstærð
  - Breytileiki (staðalfrávik) minnkar með stærra úrtaki

Úrtaks stærð	Meðal tal	Staðal frávik	Dreifing
N= 5	5,83	0,696	
N= 10	5,85	0,483	
N= 40	5,84	0,246	
N= 100	5,85	0,152	

Þýðistölur
$\mu = 5{,}849$
$\sigma = 1{,}531$

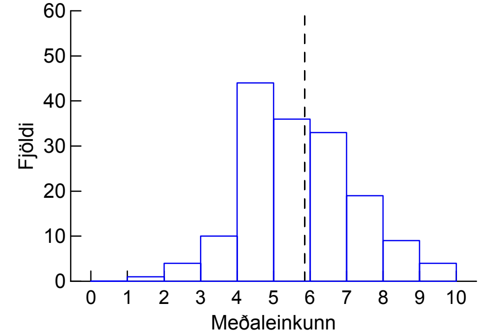
A:4.3, bls. 94–99

Markgildissetningin

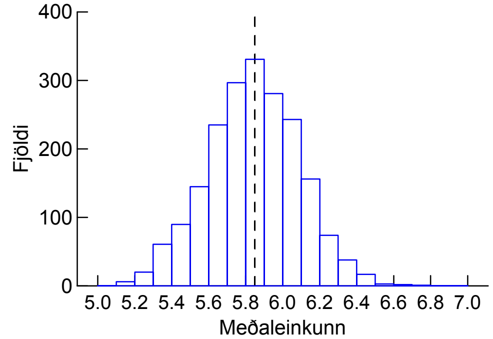
Ef mjög mörg tilviljunarúrtök af stærðinni N eru dregin úr þýði, verður úrtakadreifing (*sampling distribution*) meðaltalanna (dreifing úrtaksmeðaltalanna) mjög nálægt því að vera *normal*, svo fremi að N sé stórt.

Úrtakadreifingin líkist normaldreifingu jafnvel þótt úrtaksstærðin sé aðeins 10. Ástæðan er sú að þýðisdreifingin víkur ekki stórkostlega frá normaldreifingu.

Þýðis dreifing



Úrtaka dreifing N= 10



A:4.4, bls. 103–106; 4.5, bls. 107–110

Fræðidreifing meðaltala

- Úrtaksmeðaltöl …
  - eru normaldreifð
  - með miðju á μ
  - og staðalvillu (*staðalfrávik meðaltala*) sem reikna má með þessum formúlum

Ef staðalfrávik þýðis er óþekkt má nota staðalfrávik úrtaks sem besta spágildi þess.

Staðalfrávik þekkt :

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}$$

Staðalfrávik óþekkt :

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}_X}{\sqrt{N}}$$

A:4.4, bls. 99–102

Útreikningur staðalvillu

Úrtak	Fjöldi	Staðalvilla
1	5	$\sigma_{\overline{X}} = \frac{1{,}531}{\sqrt{5}} = \frac{1{,}531}{2{,}236} = 0{,}685 \approx 0{,}69$
2	10	$\sigma_{\overline{X}} = \frac{1{,}531}{\sqrt{10}} = \frac{1{,}531}{3{,}162} = 0{,}484 \approx 0{,}48$
3	40	$\sigma_{\overline{X}} = \frac{1{,}531}{\sqrt{40}} = \frac{1{,}531}{6{,}325} = 0{,}242 \approx 0{,}24$
4	100	$\sigma_{\overline{X}} = \frac{1{,}531}{\sqrt{100}} = \frac{1{,}531}{10} = 0{,}153 \approx 0{,}15$

Athugasemd. Miðað er við staðalfrávikið 1,531 í þýðinu

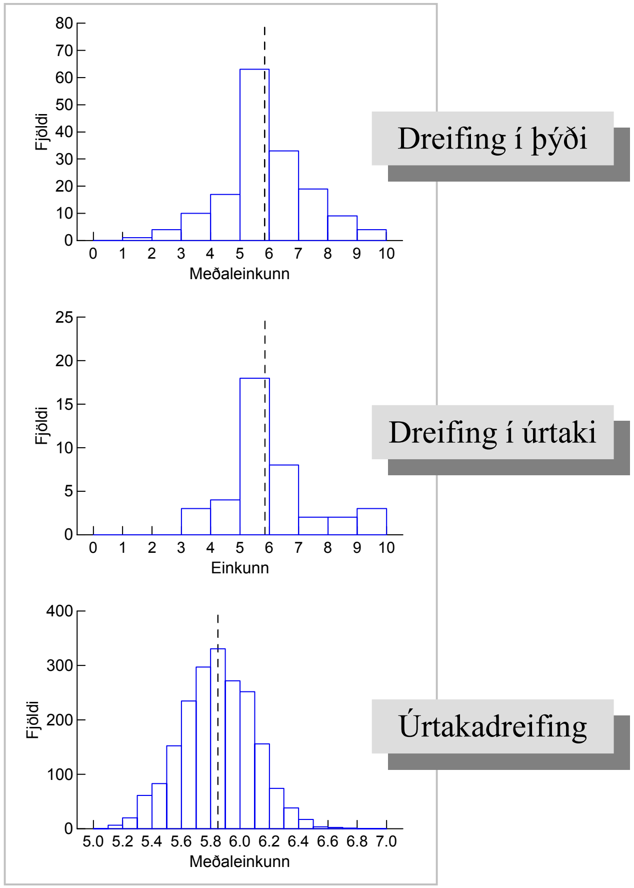
Þýðis- úrtaks- og úrtakadreifing

Við erum að vinna með þrjár ólíkar tegundir dreifinga.

Þýðisdreifingin sýnir hvernig mæligildin dreifast í þýði.

Dreifing úrtaksins sýnir hvernig þau stök sem veljast í úrtakið dreifast. Hún þarf ekki að vera dæmigerð fyrir dreifingu þýðisins þar sem tilviljun ræður því hverjir veljast í úrtakið.

Úrtakadreifing sýnir hvernig mælitölur breytast frá einu úrtaki til annars. Meðaltöl normaldreifast ef úrtakið er nægilega stórt.



A:4.5, bls. 107–109

Öryggisbil úrtaks- og þýðistalna

- Öryggisbil úrtakstölu
  - Ég þekki eiginleika þýðis og vil segja fyrir um dreifingu úrtaka úr þessu þýði
  - T.d. öryggisbil úrtaksmeðaltala, þ.e. á hvaða bil þau eru líkleg til að dreifast
- Öryggisbil þýðistölu
  - Ég þekki eiginleika úrtaks og vil álykta um eiginleika þýðisins
  - T.d. öryggisbil þýðismeðaltals—hversu sennilegt er að bilið innihaldi þýðismeðaltalið

Öryggisbil úrtaksmeðaltals

Algengast að miða við 95% öryggi

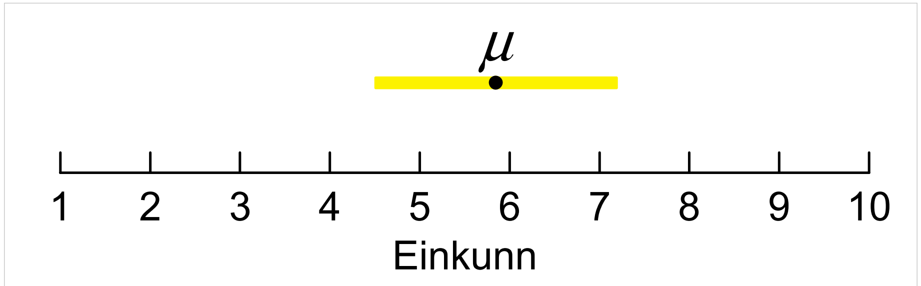
- Þá falla 5% úrtaksmeðaltalanna utan bilsins. Þetta er táknað með α, þ.e. α= 0,05.
- Þessi 5% sem lenda utan bilsins skiptast til helminga; 2,5% lenda fyrir ofan bilið og 2,5% fyrir neðan það.
- Við flettum því upp í töflu A miðað við ½α ; 0,025 ef 95% öryggi er valið en 0,005 fyrir 99% öryggi.

z	Annar aukastafur z-gildisins									
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,460	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,421	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,382	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,345	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,309	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,274	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,242	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,212	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,184	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,159	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,136	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,115	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,097	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,081	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,067	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,055	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,045	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,036	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,029	0,0281	0,0274	0,0268	0,0262	0,0256	$\boxed{0,0250}$	0,0244	0,0239	0,0233
2,0	0,023	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,018	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,014	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,011	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,008	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,006	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,005	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,003	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,003	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,002	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,00135									
3,5	0,000233									
4,0	0,0000317									
4,5	0,00000340									
5,0	0,000000287									

Öryggisbil fyrir úrtaksmeðaltöl

$$\mu - z_{\frac{1}{2}\alpha} \cdot \sigma_{\overline{X}} < \overline{X} < \mu + z_{\frac{1}{2}\alpha} \cdot \sigma_{\overline{X}}$$

Útreikningur öryggisbils



Við þekkjum eiginleika þýðis og viljum vita hvernig eiginleikar úrtaka muni dreifast.

Við vorum búin að sjá að meðaltöl úrtakanna normaldreifast þrátt fyrir smæð úrtakanna. Við vorum einnig búin að reikna út staðalvillu meðaltalanna.

95% öryggisbil úrtaksmeðaltala

$\mu = 5{,}849$

$\sigma = 1{,}531$

$N = 5$

$\sigma_{\overline{X}} = 0{,}685$

$\alpha = 0{,}05$

$z_{\frac{1}{2}\alpha} = 1{,}96$

$\mu - z_{\frac{1}{2}\alpha} \cdot \sigma_{\overline{X}} < \overline{X} < \mu + z_{\frac{1}{2}\alpha} \cdot \sigma_{\overline{X}}$

$5{,}849 - 1{,}96 \cdot 0{,}685 < \overline{X} < 5{,}849 + 1{,}96 \cdot 0{,}685$

$5{,}849 - 1{,}342 < \overline{X} < 5{,}849 + 1{,}342$

$4{,}507 < \overline{X} < 7{,}191$

Heimaverkefni

Úrtak	Stærð	95% öryggisbil
1	5	$4{,}5 < \overline{X} < 7{,}2$
2	10	
3	40	
4	100	$5{,}5 < \overline{X} < 6{,}1$

Reiknaðu öryggisbilin sem vantar út frá eftirfarandi upplýsingum og túlkaðu niðurstöður.

μ: 5,849

σ:1,531

Þetta er aðeins eitt af nokkrum heimaverkefnum sem verða birt innan skamms.

Öryggisbil þýðismeðaltals

Við ályktum um þýði á grundvelli úrtaks.

Staðalfrávik þýðis er ekki þekkt svo við notum spágildi staðalvilluna.

Við notum *t* dreifingu í stað normaldreifingar. *t* dreifing fer eftir frígráðunum (*df*) og breytist því með stærð úrtaksins.

Öryggisbil fyrir þýðismeðaltal

$$\overline{X} - t_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\overline{X}} < \mu < \overline{X} + t_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\overline{X}}$$

$$df = N - 1$$

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{N}}$$

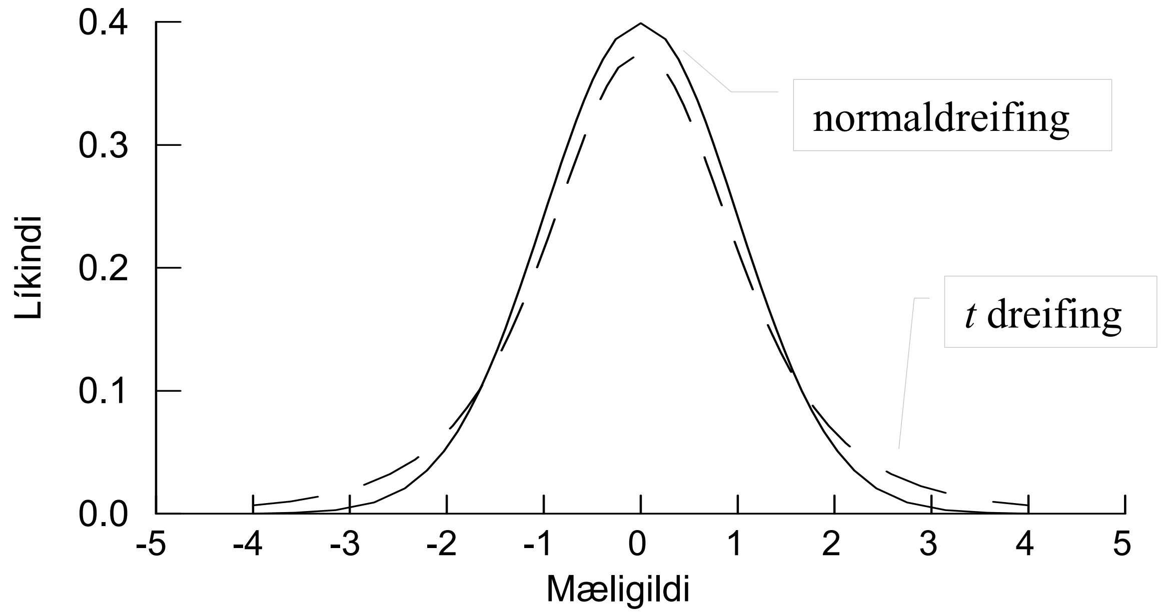
Agresti notar normaldreifingu (z) í stað t dreifingar sem er í lagi ef $N \geq 30$.

A:5.2, bls. 125–128; 6.5, bls. 183

t dreifing

Svipuð og normaldreifing en með meiri dreifingu til endanna.

Dreifingarnar verða nánast eins við $N = 30$.



A:6.5, bls. 180–183

Nokkur úrtök

Minnsta úrtakið kemst næst meðaltali þýðis.

Á grundvelli þessara upplýsinga má reikna öryggisbil fyrir meðaltal þýðis.

Úrtak	Stærð	Meðaltal	Staðalfrávik
1	5	5,83	1,559
2	10	5,92	1,732
3	40	6,08	1,870
4	100	5,48	1,778

Þótt þýðistölur séu yfirleitt óþekktar, vitum við um þær hér
$\mu = 5{,}849$
$\sigma = 1{,}531$

Við höfum yfirleitt aðeins aðgang að einu úrtaki úr þýði. Hvert þessara fjögurra gæti verið okka eina úrtak. Þar sem meðaltöl og staðalfrávik eru ólík milli úrtaka verða öryggisbil þýðismeðaltala einnig ólík eftir því hvert úrtakið er.

Útreikningur fyrir þýðismeðaltal

Grunnupplýsingar

$N = 100$

$\overline{X} = 5{,}48$

$\hat{\sigma} = s = 1{,}778$

$t_{0,025}(99) = 1{,}960$ (sbr. töflu B)

Að jafnaði munu 95% öryggisbila byggð á sams konar óháðum úrtökum innihalda meðaltal þýðisins.

Öryggisbil fyrir þýðismeðaltal

$$\overline{X} - t_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\overline{X}} < \mu < \overline{X} + t_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\overline{X}}$$

$$df = 99$$

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{N}} = \frac{1{,}778}{\sqrt{100}} = 0{,}178$$

$$5{,}48 - 1{,}96 \cdot 0{,}178 < \mu < 5{,}48 + 1{,}96 \cdot 0{,}178$$

$$5{,}48 - 0{,}349 < \mu < 5{,}48 + 0{,}349$$

$$5{,}13 < \mu < 5{,}83$$

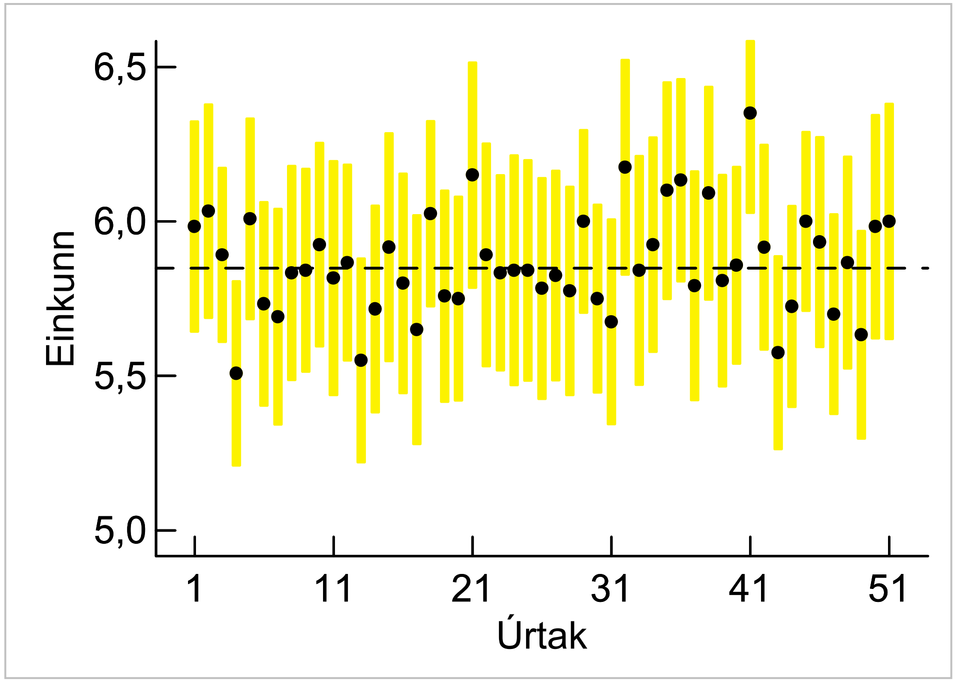
Túlkun öryggisbila

Hér hafa verið dregin 51 úrtak úr þýði með þýðismeðaltal 5,849 og staðalfrávik 1,531.

Punktar sýna úrtaksmeðaltal og línur 95% öryggisbil í hverju úrtaki fyrir sig. Lárétta brotalínan sýnir þýðismeðaltalið.

Að jafnaði ættu 5% úrtakanna að vera með öryggisbil sem inniheldur ekki þýðismeðaltalið. Í reynd hafa 2 úrtök (nr. 4 og 41) öryggisbil sem innihalda ekki þýðismeðaltalið eða 3,9% úrtakanna.

Að jafnaði ættu að vera 2–3 öryggisbil sem ekki innihalda þýðismeðaltalið.



A:5.2, bls. 129–131

Kóðun tvíkostabreyta

Tvíkostabreytur má kóða þannig að hægt sé að nota þær sem megindlegar breytur.

Best er að kóða tvíkostabreytur — t.d. nafnbreytuna kyn—sem 0 og 1.

Þá má lesa beint út úr meðaltalinu hlutfall þeirra sem hafa gildið 1, sbr. töfluna.

Það má nota aðrar kóðanir, þ.e. hvaða tvö gildi sem er—gildin 1 og 2 valda t.d. engum sérstökum vandræðum — en þá getur reynst erfitt að túlka meðaltal tvíkostabreytunnar.

	Meðaltal	
Hlutfall	Kóðað 0 og 1	Kóðað 1 og 2
10%	0,10	1,10
20%	0,20	1,20
30%	0,30	1,30
40%	0,40	1,40
...	...	...
80%	0,80	1,80
90%	0,90	1,90
100%	1,00	2,00

A:5.3, bls. 132

Margkostabreytur

Margkostabreytur eru í öllum skilningi eigindlegar. Því verður ekki breytt með kóðun.

Meðaltal margkostabreytu nafnbreytu hefur því enga merkingu. Þótt við kóðuðum breytuna Aðalgrein, yrði meðaltal hennar merkingalaust.

Aðalgrein	Fjöldi	%
Bókasafnsfr.	26	9,8
Félagsfr.	54	20,4
Mannfræði	12	4,5
Sálarfræði	121	45,7
Stjórnmálafr.	42	15,8
Uppeldisfr.	2	0,8
Annað	8	3,0
Samtals	135	100,0

Normaldreifing hlutfalla

Tvíkostabreytur taka aðeins tvö gildi. Ef ákveðnum forsendum er fullnægt normaldreifast meðaltöl þeirra (hlutföllin).

Meginreglan er að a.m.k. 10 stök (einstaklingar) séu í þeim flokki sem færra er í.

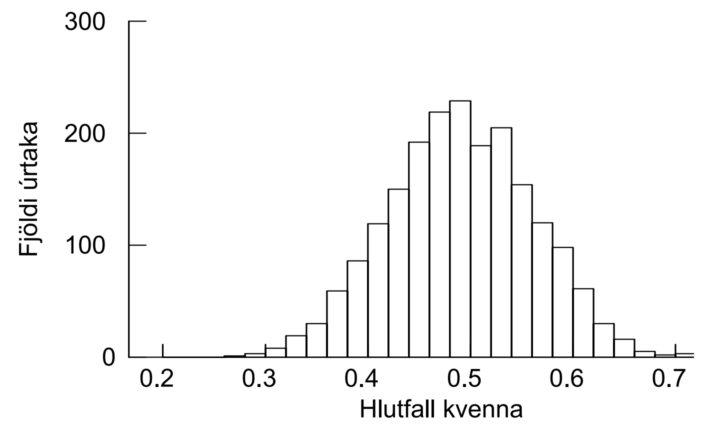
Umfjöllun Agresti á bls. 134 er óskýr en betri í textareit efst á bls. 135.

Meðaltöl tvíkostabreytu normaldreifast ef :

$$N \cdot \hat{\pi} \geq 10$$

og

$$N \cdot (1 - \hat{\pi}) \geq 10$$



A:5.3, bls. 135

Staðalfrávik tvíkostabreytu

Staðalfrávik tvíkostabreytu

$$\sigma = \sqrt{\pi \cdot (1 - \pi)}$$

$$\hat{\sigma} = \sqrt{\hat{\pi} \cdot (1 - \hat{\pi})}$$

Staðalfrávik tvíkostabreyta er tiltölulega stöðugt fyrir hlutföll á bilinu 0,3–0,7.

Oftast nægja því fremur ónákvæmar upplýsingar um þýðishlutfallið til að meta staðalfrávik tvíkostabreytu.

π	$1-\pi$	σ
0,0	1,0	0,00
0,1	0,9	0,30
0,2	0,8	0,40
0,3	0,7	0,46
0,4	0,6	0,49
0,5	0,5	0,50
0,6	0,4	0,49
0,7	0,3	0,46
0,8	0,2	0,40
0,9	0,1	0,30
1,0	0,0	0,00

Staðalvilla hlutfalla

Þessar formúlur samsvara sams konar formúlum fyrir staðalvillu meðaltala.

Þú getur reiknað staðalfrávik tvíkostabreytunnar sérstaklega eða reiknað staðalvilluna beint.

Ef forsendur fyrir normaldreifingu hlutfalla eru fyrir hendi má reikna staðalvillu hlutfalla og ákvarða þannig úrtakadreifingu (*sampling distribution*) hlutfallsins.

$$\sigma_{\hat{\pi}} = \frac{\sigma}{\sqrt{N}}$$

$$\hat{\sigma}_{\hat{\pi}} = \frac{\hat{\sigma}}{\sqrt{N}}$$

eða

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi} \cdot (1 - \hat{\pi})}{N}}$$

A:5.3, bls. 131–135

Útreikningur staðalvillu hlutfalls

Útreikningur á staðalvillu hlutfalls

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi \cdot (1-\pi)}{N}} = \sqrt{\frac{0{,}5 \cdot 0{,}5}{50}} = \sqrt{\frac{0{,}25}{50}} = \sqrt{0{,}005} = 0{,}07071$$

Útreikningar okkar gefa nánast sömu niðurstöðu og þegar við reiknum staðalfrávik hlutfalla (staðalvillu) yfir 2.000 úrtök.

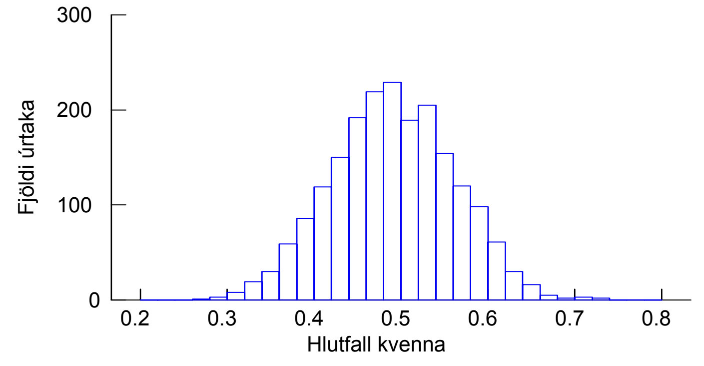
2.000 úrtök

$N = 50$

$\pi = 0{,}5$

$\sigma = 0{,}5$

$s_{\hat{\pi}} = 0{,}0707$



Öryggisbil hlutfalla

Agresti gefur aðeins upp formúlu fyrir öryggisbil hlutfalla í *þýði*.

Hér miðast formúlurnar við að fyrst sé reiknað staðalfrávik tvíkostabreytunnar. Einnig má reikna staðalvilluna beint og setja inn í viðeigandi formúlu eins og Agresti gerir.

Öryggisbil fyrir hlutföll í úrtaki

$$\pi - z_{\frac{1}{2}\alpha} \cdot \sigma_{\hat{\pi}} < p < \pi + z_{\frac{1}{2}\alpha} \cdot \sigma_{\hat{\pi}}$$

Öryggisbil fyrir hlutföll í þýði

$$\hat{\pi} - z_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\hat{\pi}} < \pi < \hat{\pi} + z_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\hat{\pi}}$$

A:5.3, bls. 133, 135

Útreikningur öryggisbils hlutfalls

95% öryggisbil fyrir þýðishlutfall

$$\hat{\pi} = 0,64$$

$$z_{0,025} = 1,96$$

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{0,64 \cdot 0,36}{50}} = \sqrt{\frac{0,2304}{50}} = \sqrt{0,004608} = 0,06788$$

$$\hat{\pi} - z_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\hat{\pi}} < \pi < \hat{\pi} + z_{\frac{1}{2}\alpha} \cdot \hat{\sigma}_{\hat{\pi}}$$

$$0,64 - 1,96 \cdot 0,06788 < \pi < 0,64 + 1,96 \cdot 0,06788$$

$$0,64 - 0,133 < \pi < 0,64 + 0,133$$

$$0,507 < \pi < 0,773$$

Nauðsynleg úrtaksstærð

Til að ákvarða úrtaksstærð þurfum við að ákveða fyrirfram æskilega nákvæmni niðurstöðunnar (B) og hversu miklu öryggi við óskum eftir ($z_{½\alpha}$).

Oftast þekkjum við ekki π. Ef það er á bilinu 0,4–0,6 þarf ekki að ákvarða það nákvæmlega. Ef π er 0,3–0,7, þá er úrtaksstærðin aðeins lítillega ofmetin með því að gera ráð fyrir að $\pi = 0,5$. Ef π er utan þessara marka þarf að giska á rétta gildið eða sætta sig við ofmat á úrtaksstærðinni.

$$N = \pi \cdot (1-\pi) \cdot \left(\frac{z_{\frac{1}{2}\alpha}}{B} \right)^2$$

B : Nákvæmni niðurstöðunnar

A:5.4, bls. 135–138

Útreikningur úrtaksstærðar

Ég nota úrtakshlutfallið sem bestu ágiskun á þýðishlutfallinu og vel 5 prósentustig sem æskilega nákvæmni.

Samkvæmt niðurstöðunni þarf um 350 manna úrtak til að ákvarða þýðishlutfall þannig að 95% öryggisbil verði 10 prósentustiga (2·5) breitt.

Ég þarf greinilega að vega og meta annars vegar æskilega nákvæmni niðurstöðunnar og hins vegar þá úrtaksstærð sem ég hef völ á.

$$N = \pi \cdot (1-\pi) \cdot \left(\frac{z_{\frac{1}{2}\alpha}}{B} \right)^2$$

$$B = 0{,}05$$

$$z_{\frac{1}{2}\alpha} = 1{,}96$$

$\hat{\pi} = 0{,}64$ ég nota þetta sem ágiskun á π

$$N = 0{,}64 \cdot (1-0{,}64) \cdot \left(\frac{1{,}96}{0{,}05} \right)^2$$

$$= 0{,}2304 \cdot (39{,}2)^2 = 0{,}2304 \cdot 1536{,}64$$

$$= 354{,}042 \approx 354$$

Úrtaksstærð fyrir meðaltöl

Formúlan fyrir meðaltöl er nánast eins nema hvað við setjum staðalfrávik þýðis beint inn í formúluna.

Gættu að því að *B* gefur nákvæmni niðurstöðunnar sem frávik frá réttu gildi. Lengd öryggisbilsins er því 2·B.

$$N = \sigma^2 \cdot \left(\frac{z_{\frac{\alpha}{2}}}{B} \right)^2$$

B : Nákvæmni niðurstöðunnar

$z_{\frac{\alpha}{2}}$: z - gildi miðað við valið α

A:5.4, bls. 138–141