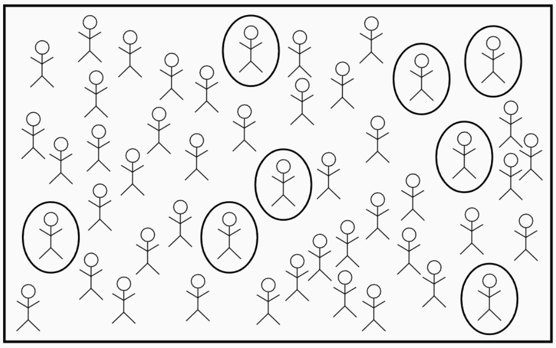
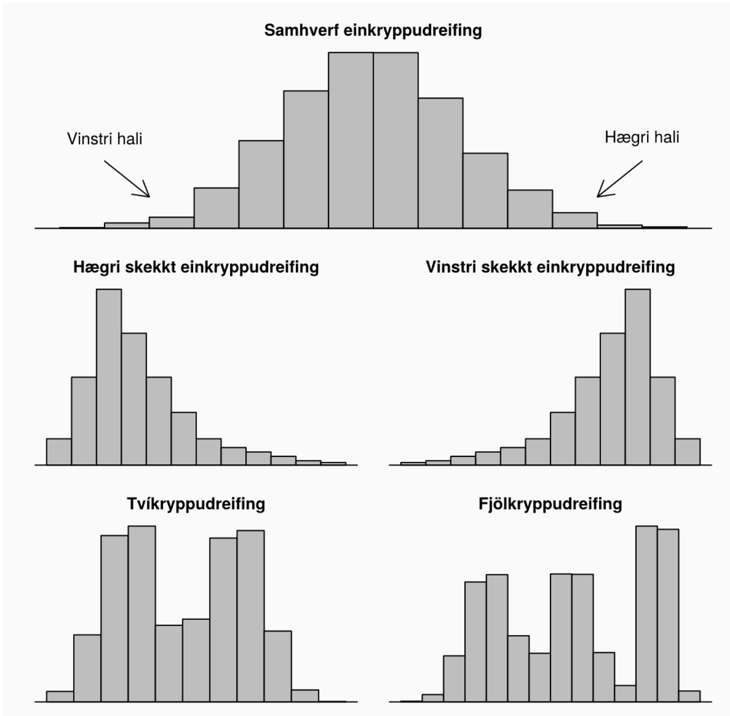
# Stærðfræði glósur



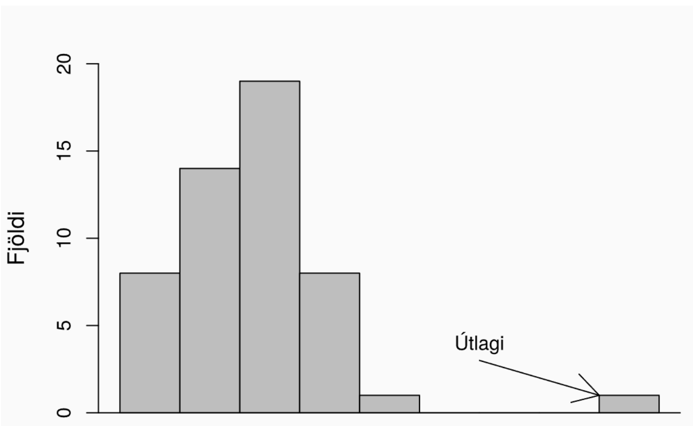
Þýði og úrtak

**Flokkabreytur**: geta verið raðaðar eða óraðaðar, taka óteljanleg gildi eins og kyn, og hárlitur. Allar flokkabreytur eru strjálar.
**Talnabreytur**: taka mælanleg gildi eins og hæð í cm. Skiptast í samfelldar og strjálar. Samfelldar breytur geta “technically” tekið öll gildi, en strjálar geta bara verið ákveðnar heilar tölur, eins og og fjöldi eggja í hreiðri. Sumar strjálar breytur geta tekið mjög mörg gildi, og eru oft álitnar sem samfelldar breytur.
**Svar og skýribreytur**: Fyrir sérhvert viðfangsefni mun gildi **skýribreytu** þess hafa áhrif á það hvaða gildi **svarbreytan** mun taka. Til einnar svarbreytu geta svarað margar skýribreytur sem hafa áhrif á hana. gildi skýribreytu hafa áhrif á gildi svar breytunnar, en ekki öfugt.




Allar flokkabreytur eru strjálar.
Talnabreytur geta verið samfelldar eða strjálar.
Talnabreytur sem eru strjálar en taka mörg gildi meðhöndlum við eins og samfelldar breytur.

**Ein eða tvær strjálar breytur** -Stöplarit
**Ein samfelld breyta-** Stuðlarit - fjöldi flokka=$5 \cdot \log$(fjöldi mælinga)
**Ein samfelld, ein strjál** -Kassarit
**Tvær samfelldar** -Punktarit



**Lýsisstærðir fyrir miðju**
Miðja spannar : (xmin + xmax ) / 2
Tiðasta gildi: þau gildi sem koma fyrir oftast
Miðgildi (median) : það gildi nákvæmlega í miðjunni, eða meðaltal þeirra tveggja í miðjunni (M)
Meðaltal (mean) : summa allra gilda deilt með heildarfjölda gilda ( $\overline{x}$ )
Vegið meðaltal: vægi sinnum gildi, deilt með fjölda ( $\overline{x_w}$)

**Lýsistærðir fyrir breytileika**
Spönn (range) : xmax-xmin
Fjórðungamörk: Q1, Q2, Q3. 25%, 50% og 75% mörk. Gildið í sæti nr 25%, etc.
Dreifni (variance) : fjarlægð mælinga frá meðaltali. ($s^2$)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Staðalfrávik (standard deviation): $s = \sqrt{s^2}$
Frávikshlutfall (coefficient of variation): $CV = \frac{s}{x}$. Eftir því sem CV er hærra, því breytilegri eru gögnin.
Fylgnistuðull úrtaks: r, er alltaf á bilinu -1, 1. Því nær 0, þeimur minni fylgni er á milli breytana. Basically hallatala punktarits.

| | | flokkabreyta | | talnabreyta | |
|---|---|---|---|---|---|
| | | óröðuð | röðuð | strjál | samfelld |
| Lýsistærð miðju | Miðja spannar | | | x | x |
| | Tíðasta gildi | x | x | x | |
| | Miðgildi | | x | x | x |
| | Meðaltal | | | x | x |
| | Vegið meðaltal | | | x | x |
| Lýsistærð breytileika | Spönn/dreifisvið | | x | x | x |
| | Fjórðungamörk | | x | x | x |
| | Fimm tölu samantekt | | x | x | x |
| | Fjórðungaspönn | | | x | x |
| | Prósentumörk | | x | x | x |
| | Dreifni/fervik | | | x | x |
| | Staðalfrávik | | | x | x |
| | Frávikshlutfall | | | x | x |

$P(X \leq a)$: Táknar líkur þess að útkoma slembistærðarinnar $X$ verði **minni eða jöfn** gildinu $a$.

$P(X \geq a)$: Táknar líkur þess að útkoma slembistærðarinnar $X$ verði **stærri eða jöfn** gildinu $a$.

$P(a \leq X \leq b)$: Táknar líkur þess að útkoma slembistærðarinnar $X$ verði **á milli** $a$ og $b$, bæði gildin meðtalin.

$P(X = a)$: Táknar líkur þess að útkoma slembistærðarinnar $X$ verði **nákvæmlega** gildið $a$.

Væntigildi slembistærðar = raunverulegt meðaltal slembistærðar = meðaltal þýðis, táknað með $\mu$ *eða* $E[X]$

Meðaltal mælinga: $\overline{x}$

Raunveruleg dreifni = dreifni þýðis: $\sigma^2$, $Var[X]$

Tvíkostadreifing

Bernoulli tilraun: tilraun hefur aðeins 2 útkomur (boolean), líkurnar á TRUE eru allar þær sömu, mælingarnar eru óháðar, fyrri útkoma hefur ekki áhrif á þá seinni.

X ~ B(n, p) þar sem n = fjöldi tilrauna, p = líkur á að hún heppnist, k = gildið sem við erum að skoða líkurnar á, hversu oft fáum við k.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \ldots n$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

Ef $X$ fylgir tvíkostadreifingu, $X \sim B(n, p)$ þá gildir

(10)

$$\mathrm{E}[X] = np$$

(11)

$$\mathrm{Var}[X] = np(1-p)$$

Poisson dreifing: fjöldi slembina atvika sem eiga sér stað *á ákveðinni einingu,* en mögulegar útkomur eiga sér engin eftir mörk.

X ~ Pois(**λ**)

Lambda = hvað við væntum margar jákvæðar útkomur að meðaltali.

Látum $X$ fylgja Poisson dreifingu með stikann $\lambda$, táknað $X \sim \mathrm{Pois}(\lambda)$. Ef $k$ er eitthvert jákvætt heiltölugildi (núll meðtalið) má reikna líkurnar á að slembistærðin $X$ taki gildið $k$ með massafalli Poisson dreifingarinnar :

(12)

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Ef $X$ fylgir Poisson dreifingu, $X \in \mathrm{Pois}(\lambda)$ þá gildir

(13)

$$\mathrm{E}[X] = \lambda$$

(14)

$$\mathrm{Var}[X] = \lambda$$

**Tvíkostadreifinguna** notum við þegar við höfum **endanlegan fjölda tilrauna** og við vitum líkurnar á því að hver og ein tilraun heppnist.
**Poisson dreifinguna** notum við þegar við höfum **engin efri mörk á fjölda tilrauna** og við vitum meðalfjölda jákvæðra útkoma á tiltekna einingu.

Normaldreifing: $\mu$ = meðaltal hennar og $\sigma$^2 = dreifni

Gerum ráð fyrir að slembistærðin $X$ fylgi normaldreifingu með meðaltal $\mu$ og dreifni $\sigma^2$, táknað $X \sim N(\mu, \sigma^2)$. Þá er þéttifall hennar, táknað með $\phi(x)$, gefið með

(18)

$$f(x) = \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Dreififall normaldreifingarinnar er táknað með $\Phi(x)$.

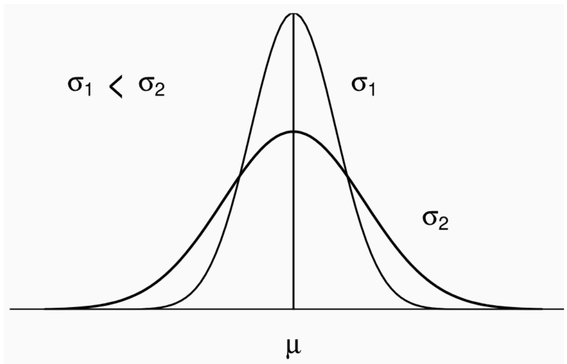
Ef $X$ og $Y$ eru tvær slembistærðir, þá er

(1)

$$E[X+Y] = E[X] + E[Y]$$

(2)

$$E[X-Y] = E[X] - E[Y]$$



*Tvær normaldreifingar með sama meðaltal en ólíka dreifni*

Ef $X$ og $Y$ eru tvær óháðar slembistærðir er

(3)

$$Var[X+Y] = Var[X] + Var[Y]$$

(4)

$$Var[X-Y] = Var[X] + Var[Y]$$

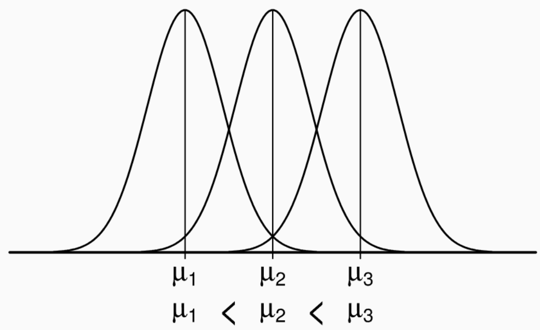
Ef $X_1, \ldots, X_n$ eru óháðar og einsdreifðar slembistærðir með væntigildi $E[X_i] = \mu$ og dreifni $Var[X_i] = \sigma^2$ þá gildir um meðaltal þeirra, $\bar{X}$ að:

(5)

$$E[\bar{X}] = \mu$$

(6)

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$



*Þrjár normaldreifingar með sömu dreifni en ólík meðaltöl*

**Athugið**

Ef $\bar{X}$ er meðaltal $X_1, \ldots, X_n$, óháðra og einsdreifðra slembistærða með dreifni $Var[X_i] = \sigma^2$, þá er *staðalskekkja* þeirra

$$\sigma/\sqrt{n}$$

Hún er staðalfrávik meðaltals mælinganna.

**Athugasemd**

Takið eftir að

$$\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{Var[\bar{X}]}$$

svo staðalskekkjan er kvaðratrótin af dreifni meðaltalsins, eins og við er að búast.

$\hat{p} = x/n =$ hlutfall slembistærðar, n = heildarfjöldi tilrauna, x = heppnaðar tilraunir

1-$\alpha$ öryggisbil, $\alpha$ = prósent ekki inn á bili.

## Tilgátupróf

Núlltilgáta: fullyrðing sem getur bara verið afsönnuð, ekki sönnuð. Táknuð H_0
H_0 : p = p0

Gagntilgáta: Sú fullyrðing sem við viljum staðfesta með rannsókninni. Táknuð H_1 eða H_a
H_1 : p1 = p2

Tvíhliða tilgátupróf: Ef gögnin leyfa þá fullyrðir tvíhliða tilgátupróf að einn stiki gagnanna sé annað hvort **stærri eða minni** en annar stiki eða eitthvað ákveðið gildi, ef gögnin leyfa.

Einhliða tilgátupróf:Til eru tvær gerðir einhliða tilgátuprófa:
Þau sem fullyrða að einn stiki gagnanna sé **stærri** en annar stiki eða eitthvað ákveðið gildi, ef gögnin leyfa.
Þau sem fullyrða að einn stiki gagnanna sé **minni** en annar stiki eða eitthvað ákveðið gildi, ef gögnin leyfa.

Framkvæmd tilgátuprófa:

1. Ákveða hvaða tilgátupróf er viðeigandi fyrir gögnin okkar.

2. Ákveða hæstu ásættanlegu villulíkur.
3. Setja fram núlltilgátu og ákveða um leið áttun prófsins (einhliða/tvíhliða).
4. Reikna prófstærðina sem svarar til tilgátuprófsins.
5. Kanna hvort prófstærðin falli á höfnunarsvæði tilgátuprófsins. Höfnunarsvæðið er $\alpha$ fyrir einhliða, og $\alpha$/2 fyrir tvíhliða próf báðum megin
6. Kanna p-gildi tilgátuprófsins. Hafna H_0 ef p-gildi er < $\alpha$, annars ekki.
7. Draga ályktun. Fullyrðum gagntilgátunni eða ekki?

Tafla mældrar tíðni, o:
Lxd tafla, lína sinnum dálkar, með summu gilda lína og dálka

Tafla væntanlegrar tíðni, e:
(Summa línu x summa dálks )/ summu töflu

Tafla prófstærðar: (o-e)^2 ) /e
Reiknað með gildunum úr fyrri töflum á sama stað.
Prófstærð fæst svo með því að leggja saman allar tölurnar úr töflunni.

Z-próf má nota þegar:
- Þýðin fylgja normaldreifingu og við gerum ráð fyrir að við þekkjum dreifni þýðanna.
- Úrtökin eru stór, óháð því hver dreifing þýðanna er.

t-próf má nota þegar:
- Þýðin fylgja normaldreifingu, sama hversu lítil úrtökin eru.
- Úrtökin eru stór, óháð því hver dreifing þýðanna er.

Fervikagreining: ANOVA

Aðferðin gengur út á að bera saman breytileika á gildum mælinga milli hópa annars vegar og innan hópa hins vegar. Út frá því er ályktað hvort meðaltölin séu ólík eða ekki.
Fervikagreining gerir ráð fyrir að úrtökin séu slembiúrtök, að þau séu valin úr þýðum sem fylgja normaldreifingu og að dreifnin sé sú sama í öllum þýðum.

Aðhvarfsgreining:

Aðhvarfsgreining er gífurlega almenn aðferð sem er notuð til að kanna samband háðrar breytu (e. dependent variable) og einnar eða fleiri óháðra breyta (e. independent variables). Sé óháða breytan aðeins ein er talað um einfalt línulegt aðhvarf (e. simple linear regression)

$y_{ij}$ : Við notum vísinn $i$ til að tákna númer hóps og vísinn $j$ til að tákna númer mælingu
innan hóps. $y_{ij}$ er því mæling númer $j$ úr hópi $i$.

$a$ : Við notum $a$ til að tákna fjölda hópa.

$n_i$ : Við notum $n_i$ til að tákna fjölda mælinga í hópi $i$.

$N$ : Við notum $N$ til að tákna heildarfjölda mælinga

$$N = n_1 + n_2 + \ldots + n_a$$

$\bar{y}_{i.}$ : Við notum $\bar{y}_{i.}$ til að tákna meðaltal fyrir hóp $i$

(1)

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$\bar{y}_{..}$ : Við notum $\bar{y}_{..}$ til að tákna meðaltal allra mælinga (úr öllum hópum)

(2)

$$\bar{y}_{..} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}}{N}$$