Lýsandi tölfræði

Tölfræði frá grunni

Anna Helga Jónsdóttir og Sigrún Helga Lund

Háskóli Íslands

Yfirlit

Yfirlit

Lýsistærðir

- Við notum lýsandi tölfræði til að lýsa tilteknum eiginleikum mælinganna okkar.
- Góð aðferð til þess er að nota **lýsistærðir** en þær eru tölur sem lýsa tilteknum eiginleikum mælinga.

Lýsistærð (statistic)

Lýsistærð er tala sem er reiknuð með einhverjum ákveðnum hætti út frá mælingunum okkar.

Lýsistærðir

- Til eru margar gerðir af lýsistærðum, sem lýsa ólíkum eiginleikum mælinga
- Skoðum nú tvær tegundir lýsistærða:
  - Lýsistærðir sem lýsa **miðju** (center) gagna
  - Lýsistærðir sem lýsa **breytileika** (spread) gagna
- Algengastar eru meðaltal og staðalfrávik

Vísir

Vísir (index)

Gerum ráð fyrir að við höfum n mælingar á tiltekinni breytu og köllum mælingarnar $x_1, \ldots, x_n$. Tölurnar $1, \ldots, n$, sem mælingarnar eru tölusettar eftir kallast **vísar** (index) mælinganna.

Þannig hefur fyrsta mælingin, x_1, vísinn einn, mæling x_2 hefur vísinn tvo og svo framvegis. Síðasta mælingin x_n hefur vísinn n en við getum líka talað um ótilgreind x_i þar sem það er ekki ákvarðað hvert gildið á vísinum i er. Þá stendur x_i fyrir hvaða mælingu sem er.

Yfirlit

Lýsistærðir fyrir miðju

Skoðum fimm mismunandi lýsistærðir sem allar lýsa miðju mælinga

1. Miðja spannar (mid range)
2. Tíðasta gildi (mode)
3. Miðgildi (median)
4. Meðaltal (mean, arithmetic mean)
5. Vegið meðaltal (weighted mean)

Miðja spannar

Miðja spannar (mid range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Látum x_{min} tákna þá minnstu og x_{max} þá stærstu. **Miðja spannar** er reiknuð með

$$\text{Miðja spannar} = \frac{x_{\mathsf{min}} + x_{\mathsf{max}}}{2}.$$

Dæmi

Hver er miðja spannar talnasafnins 10,2,5,3,8,5,6?

Miðja spannar

Miðja spannar (mid range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Látum x_{min} tákna þá minnstu og x_{max} þá stærstu. **Miðja spannar** er reiknuð með

$$\text{Miðja spannar} = \frac{x_{\text{min}} + x_{\text{max}}}{2}.$$

Dæmi

Hver er miðja spannar talnasafnins 10,2,5,3,8,5,6?
Svar: miðja spannar er 6.

Tíðasta gildi

Tíðasta gildi (mode)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. **Tíðasta gildið** er sú útkoma sem oftast kemur fyrir í mælingunum okkar. Það er sú eina af lýsistærðunum fyrir miðju sem við fjöllum um sem er hægt er að nota til að lýsa flokkabreytum. Hins vegar er ekki við hæfi að reikna tíðasta gildið þegar mældar eru samfelldar talnabreytur.

Dæmi

Hvert er tíðasta gildi talnasafnins `10,2,5,3,8,5,6`?

Tíðasta gildi

Tíðasta gildi (mode)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. **Tíðasta gildið** er sú útkoma sem oftast kemur fyrir í mælingunum okkar. Það er sú eina af lýsistærðunum fyrir miðju sem við fjöllum um sem er hægt er að nota til að lýsa flokkabreytum. Hins vegar er ekki við hæfi að reikna tíðasta gildið þegar mældar eru samfelldar talnabreytur.

Dæmi

Hvert er tíðasta gildi talnasafnins `10,2,5,3,8,5,6`?
Svar: tíðasta gildið er 5.

Miðgildi

Miðgildi (median)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Byrjum á að raða þessum mælingum upp í stærðarröð, frá minnst gildi upp í stærsta gildi. Reiknum svo

$$\text{Sæti í röð} = 0.5 \cdot (n+1).$$

Miðgildi er oft táknað með M. Það fer eftir því hvort n sé oddatala eða slétt tala hvernig við reiknum út miðgildið.

- Ef n er oddatala þá er miðgildið staðsett í sæti $0.5 \cdot (n+1)$ í röðinni.
- Ef n er slétt tala þá er miðgildið meðaltalið af þeim tveimur mælingum sem standa við sæti $0.5 \cdot (n+1)$ í röðinni.

VARÚÐ: $0.5 \cdot (n+1)$ er númerið á sætinu í röðinni, ekki miðgildið sjálft!

Dæmi

Hvert er miðgildi talnasafnins 10,2,5,3,8,5,6?

Miðgildi

Miðgildi (median)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Byrjum á að raða þessum mælingum upp í stærðarröð, frá minnst gildi upp í stærsta gildi. Reiknum svo

$$\text{Sæti í röð} = 0.5 \cdot (n+1).$$

Miðgildi er oft táknað með M. Það fer eftir því hvort n sé oddatala eða slétt tala hvernig við reiknum út miðgildið.

- Ef n er oddatala þá er miðgildið staðsett í sæti $0.5 \cdot (n+1)$ í röðinni.
- Ef n er slétt tala þá er miðgildið meðaltalið af þeim tveimur mælingum sem standa við sæti $0.5 \cdot (n+1)$ í röðinni.

VARÚÐ: $0.5 \cdot (n+1)$ er númerið á sætinu í röðinni, ekki miðgildið sjálft!

Dæmi

Hvert er miðgildi talnasafnins 10,2,5,3,8,5,6?
Svar: miðgigldið er 5

Meðaltal

Meðaltal (mean, arithmetic mean)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. **Meðaltalið** fæst með að leggja mælingarnar saman og deila með fjölda mælinga.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Dæmi

Hvert er meðaltal talnasafnins `10,2,5,3,8,5,6`?

Meðaltal

Meðaltal (mean, arithmetic mean)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. **Meðaltalið** fæst með að leggja mælingarnar saman og deila með fjölda mælinga.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Dæmi

Hvert er meðaltal talnasafnins 10,2,5,3,8,5,6?
Svar: meðaltalið er 5.57.

Vegið meðaltal

Þegar meðaltal er reiknað eins og við gerðum hér að framan fá allar mælingarnar sama vægi. Í sumum tilfellum viljum við gefa mælingunum misjafnt vægi, þá er talað um **vegið meðaltal**.

Vegið meðaltal (weighted mean)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$ og vægi þeira $w_1, w_2, ...w_n$. Vegið meðaltal er reiknað sem

$$\bar{x}_w = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + ...w_n \cdot x_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i}$$

Vegið meðaltal - dæmi

Dæmi

Í töflunni hér að neðan eru meðaltekjur á hvern vinnandi íbúa í þremur bæjarfélögum A, B og C.

Bæjarfélag	Fjöldi vinnandi	Meðaltekjur
A	50	220000
B	10	240000
C	20	300000

Hverjar eru meðaltekjur vinnandi íbúa bæjarfélaganna þriggja?

Vegið meðaltal - dæmi

Dæmi

Í töflunni hér að neðan eru meðaltekjur á hvern vinnandi íbúa í þremur bæjarfélögum A, B og C.

Bæjarfélag	Fjöldi vinnandi	Meðaltekjur
A	50	220000
B	10	240000
C	20	300000

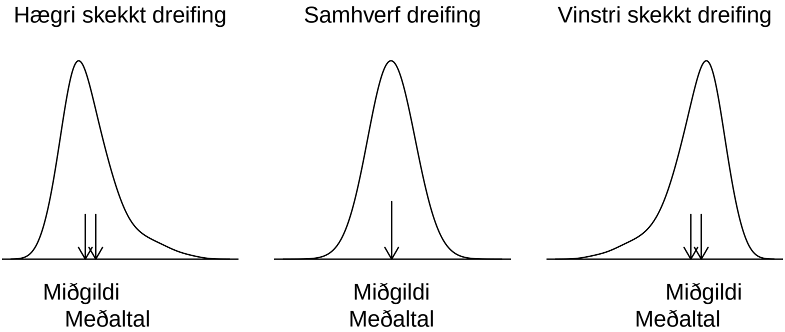
Hverjar eru meðaltekjur vinnandi íbúa bæjarfélaganna þriggja?
Svar: meðaltekjurnar eru 242500 krónur.

Samanburður á lýsistærðum fyrir miðju

- Allar þær lýsistærðir sem við höfum skoðað eru áhugaverðar og gott að reikna út þegar við fáum ný gögn í hendurnar
- Áður en ákvörðun er tekin um hvaða lýsistærð skal nota til að lýsa miðju gagnanna er gott að skoða gögnin myndrænt til að átta sig á dreifingu gagnanna
- Sé dreifingin skekkt, tvíkryppu- eða fjölkryppu dreifing skal nota miðgildið fram yfir meðaltalið
- Miðgildi er betri mælikvarði á miðju gagna ef útlagar eru í gagnasafninu

Samanburður á meðaltali og miðgildi

- Ef dreifingin er skekkt til hægri er meðaltalið hærra en miðgildið.
- Ef dreifingin er samhverf er meðaltalið og miðgildið það sama.
- Ef dreifingin er skekkt til vinstri er meðaltalið lægra en miðgildið.



Yfirlit

Lýsistærðir fyrir breytileika

Breytileiki mælinga (spread)

Breytileiki mælinga er aðferð sem lýsir því hversu dreift mælingarnar liggja.

Við munum fjalla um sjö lýsistærðir sem allar lýsa breytileika mælinga

1. Spönn/dreifisvið (range)
2. Fjórðungamörk (quartiles)
3. Fjórðungaspönn (interquartile range)
4. Prósentumörk (percentiles)
5. Dreifni/fervik (variance)
6. Staðalfrávik (standard deviation)
7. Frávikshlutfall (coefficient of variation)

Spönn/dreifisvið

Spönn (range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$ og látum x_{min} tákna þá minnstu og x_{max} þá stærstu. **Spönn** gagnanna er reiknuð með

$$\text{Spönn} = x_{\text{max}} - x_{\text{min}}$$

Dæmi

Hver er spönn talnasafnins `10,2,5,3,8,5,6`?

Spönn/dreifisvið

Spönn (range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$ og látum x_{min} tákna þá minnstu og x_{max} þá stærstu. **Spönn** gagnanna er reiknuð með

$$\text{Spönn} = x_{\text{max}} - x_{\text{min}}$$

Dæmi

Hver er spönn talnasafnins `10,2,5,3,8,5,6`?

Svar: spön talnasafnsins er 8.

Fjórðungamörk

Fjórðungamörkin eru þrjú og er algengt að kalla þau, Q_1, Q_2 og Q_3. Í sumum kennslubókum og ritum eru fjórðungamörkin kölluð $Q_{25\%}, Q_{50\%}$ og $Q_{75\%}$. Við munum halda okkur við fyrri ritháttinn í þessari bók.

Q_1: Um fyrsta fjórðungamarkið gildir að 25% af mælingunum eru lægri en Q_1. Q_1 er því miðgildi neðri helmings mælinganna, að undanskildu miðgildinu.

Q_2: Um annað fjórðungamarkið gildir að 50% af mælingunum eru lægri en Q_2. Q_2 er því miðgildið, $Q_2 = M$.

Q_3: Um þriðja fjórðungamarkið gildir að 75% af mælingunum eru lægri en Q_3. Q_3 er því miðgildi efri helmings mælinganna, að undanskildu miðgildinu.

Fjórðungamörk

Fjórðungamörk (quartiles)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Byrjum á að raða mælingum upp í stærðarröð, frá lægsta gildinu upp í hæsta gildið. Reiknum svo

$$Q_1 \text{ - sæti í röð: } = 0.25 \cdot (n+1)$$
$$Q_2 \text{ - sæti í röð: } = 0.50 \cdot (n+1)$$
$$Q_3 \text{ - sæti í röð: } = 0.75 \cdot (n+1)$$

- Q_1 er mælingin sem stendur í sæti $0.25 \cdot (n+1)$ eða meðaltalið af þeim tveimur mælingum sem standa við sæti $0.25 \cdot (n+1)$ í röðinni.
- Q_2 er mælingin sem stendur í sæti $0.50 \cdot (n+1)$ eða meðaltalið af þeim tveimur mælingum sem standa við sæti $0.50 \cdot (n+1)$ í röðinni.
- Q_3 er mælingin sem stendur í sæti $0.75 \cdot (n+1)$ eða meðaltalið af þeim tveimur mælingum sem standa við sæti $0.75 \cdot (n+1)$ í röðinni.

Fjórðungaspönn

Fjórðungaspönn (interquartile range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$ og látum Q_1 tákna fyrsta fjórðungamark og Q_3 þriðja fjórðungamark.
Fjórðungaspönn gagnanna er táknuð með IQR og reiknuð með

$$IQR = Q_3 - Q_1.$$

Dæmi

Hver eru fjórðungamörk og fjórðungaspönn talnasafnins `10,2,5,3,8,5,6`?

Fjórðungaspönn

Fjórðungaspönn (interquartile range)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$ og látum Q_1 tákna fyrsta fjórðungamark og Q_3 þriðja fjórðungamark. **Fjórðungaspönn** gagnanna er táknuð með IQR og reiknuð með

$$IQR = Q_3 - Q_1.$$

Dæmi

Hver eru fjórðungamörk og fjórðungaspönn talnasafnins `10,2,5,3,8,5,6`?
Svar: fjórðunamörkin eru $Q_1 = 3$, $Q_2 = 5$ og $Q_3 = 8$.
Fjórðungaspönnin er 5.

Prósentumörk

Hugmyndin að baki **prósentumörkum** (percentiles) er svipuð og sú að baki fjórðungamörkum nema í stað þess að skoða eingöngu mörkin við 25 %, 50 % eða 75 % mælinganna getum við leyft hvaða hlutfall sem er.

Prósentumörk (percentiles)

Með $a\%$ prósentumörkum er átt við þá tölu sem er þannig að $a\%$ mælinganna hafa lægra gildi en sú tala.

Líkt og með fjórðungamörkin eru nokkrar ólíkar leiðir til þess að reikna prósentumörk og er það nær aldrei gert „í höndunum" heldur er notast við tölfræðihugbúnað.

Dreifni

Dreifni (variance)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, \ldots x_n$. Dreifni mælinga er táknuð s^2 og er reiknuð með

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

$s^2 = 0$ þá og því aðeins að allar mælingarnar séu jafnar, annars er s^2 ávallt stærra en 0. Því lengra sem mælingarnar liggja frá meðaltalinu því hærra verður s^2.

Dæmi

Hver er dreifni talnasafnins `10,2,5,3,8,5,6`?

Dreifni

Dreifni (variance)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Dreifni mælinga er táknuð s^2 og er reiknuð með

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

$s^2 = 0$ þá og því aðeins að allar mælingarnar séu jafnar, annars er s^2 ávallt stærra en 0. Því lengra sem mælingarnar liggja frá meðaltalinu því hærra verður s^2.

Dæmi

Hver er dreifni talnasafnins `10,2,5,3,8,5,6`?

Svar: dreifni talnasafnsins er 7.62.

Staðalfrávik

Staðalfrávik (standard deviation)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Staðalfrávik mælinga er táknað með s og er reiknað með

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

$s = 0$ þá og því aðeins að allar mælingarnar eru jafnar, annars er s ávallt stærra en 0. Því lengra sem mælingarnar liggja frá meðaltalinu því hærra verður s.

Dæmi

Hvert er staðalfrávik talnasafnins `10,2,5,3,8,5,6`?

Staðalfrávik

Staðalfrávik (standard deviation)

Gerum ráð fyrir að við höfum n mælingar $x_1, x_2, ...x_n$. Staðalfrávik mælinga er táknað með s og er reiknað með

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

$s = 0$ þá og því aðeins að allar mælingarnar eru jafnar, annars er s ávallt stærra en 0. Því lengra sem mælingarnar liggja frá meðaltalinu því hærra verður s.

Dæmi

Hvert er staðalfrávik talnasafnins `10,2,5,3,8,5,6`?
Svar: staðalfrávik talnasafnsins er 2.76.

Frávikshlutfall

- Varast skal að bera saman staðalfrávik gagna þegar mælingarnar eru í misjöfnum mælieiningum eða meðaltal þeirra mjög frábrugðið.
- Í þeim tilvikum reiknum við frávikshlutfall til að bera saman breytileika tveggja eða fleiri hópa. Það er táknað með CV.

Frávikshlutfall (coefficient of variation)

Frávikshlutfall reiknum við með

$$CV = \frac{s}{\bar{x}}$$

Eftir því sem CV er hærra því dreifðari eru gögnin.

Samanburður á lýsistærðum fyrir breytileika

- Allar þær lýsistærðir sem við höfum skoðað sem lýsa breytileika eru áhugaverðar og gott að reikna út þegar við fáum ný gögn í hendurnar
- Dreifni og staðalfrávik eru notuð til að lýsa breytileika mælinga umhverfis meðaltalið og á aðeins að nota þegar meðaltal er notað sem mælikvarði á miðju
- Staðalfrávik er yfirleitt notað fram yfir dreifni þar sem mælieiningin á staðalfrávikinu er sú sama og á mælingunum
- Staðalfrávik er viðkvæmt fyrir skekkingu og útlögum. Aðeins fáir útlagar geta gert staðalfrávikið mjög hátt
- Séu mælingarnar skekktar eða ef útlagar eru til staðar er fimm tölu samantekt besti mælikvarðinn á breytileika gagnanna

Fimm tölu samantekt

Fimm tölu samantekt (five-number summary)

Fimm-tölu samantekt samanstendur af minnsta gildi (min), fjórðungamörkunum og stærsta gildi (max), þ.e.a.s

$$\text{min},\ Q_1,\ Q_2,\ Q_3,\ \text{max}$$

Yfirlit

Jafna beinnar línu

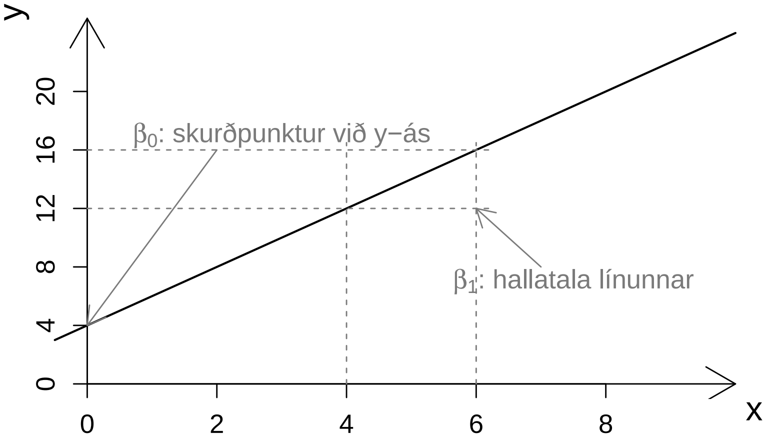
Jafna beinnar línu

Jafna beinnar línu lýsir línulegu sambandi tveggja breyta, y og x. Jöfnuna má skrifa sem

$$y = \beta_0 + \beta_1 x$$

þar sem β_0 er **skurðpunktur** (intercept) línunnar við y-ás og β_1 er **hallatala** (slope) línunnar.

Jafna beinnar línu



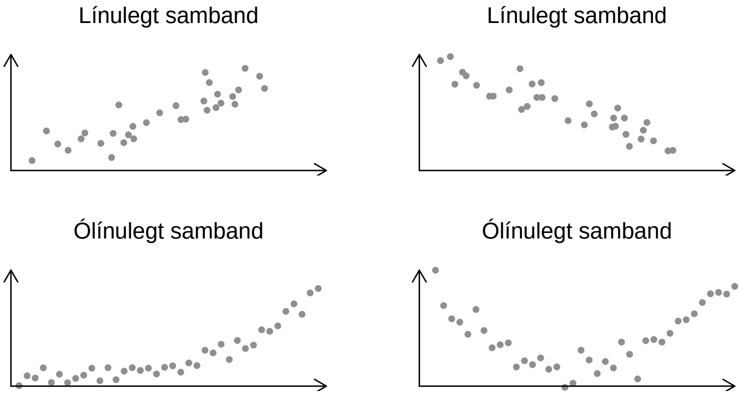
Línulegt samband

Línulegt samband

Við segjum að samband tveggja breyta sé **línulegt** (linear) ef nota má jöfnu beinnar línu til spá fyrir um gildi háðu breytunnar út frá gildi óháðu breytunnar.

Athugið að það geta verið margs konar aðrar gerðir af samböndum milli tveggja breyta. Til dæmis ef lýsa má sambandinu með fleygboga, veldisvísisfalli og svo framvegis. Þau sambönd eru einu orði nefnd ólínuleg og eru utan efni þessa námskeiðs.

Línulegt og ólínulegt samband



Fylgnistuðull úrtaks

Fylgnistuðull úrtaks

Gerum ráð fyrir að við höfum n mælingar á tveimur breytum x og y.

Táknum meðaltal og staðalfrávik x breytunnar með $\bar{x}$ og s_x og meðaltal og staðalfrávik y breytunnar með $\bar{y}$ og s_y.

Fylgnistuðul úrtaksins (sample coefficient of correlation) reiknum við með

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right).$$

Gætið ykkar að við notum fylgnisstuðul eingöngu til að meta **línulegt** samband!

Stefna og styrkleiki línulegs sambands

Stefna línulegs sambands

Formerki fylgnistuðulsins segir til um það hver **stefna** línulegs sambands er. Hún er annað hvort jákvæð eða neikvæð.

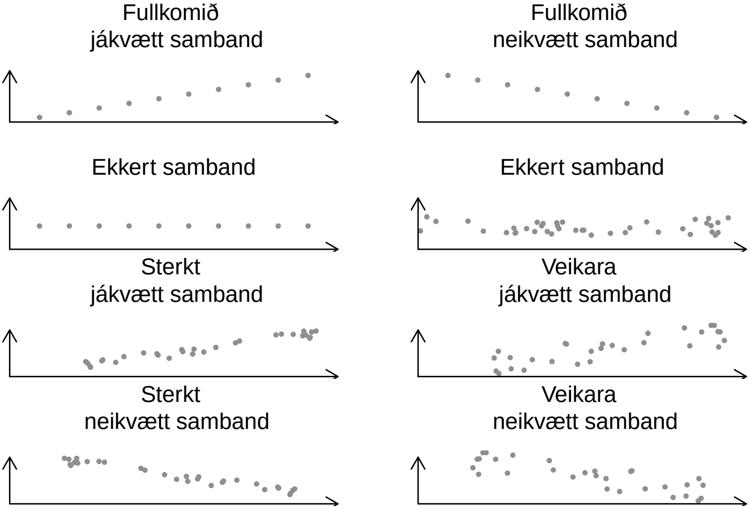
- Ef fylgnistuðull tveggja breyta er jákvæður, þá segjum við að fylgni þeirra sé **jákvæð**.
- Ef fylgnistuðull tveggja breyta er neikvæður, þá segjum við að fylgni þeirra sé **neikvæð**.

Styrkleiki línulegs sambands

Algildi (absolute value) fylgnistuðuls lýsir **styrkleika** línulega sambandsins sem gildir milli breytanna.

Hann segir okkur hversu vel við getum ákvarðað gildi svarbreytunnar út frá gildi skýribreytunnar.

Stefna og styrkleiki línulegs sambands



Fylgni og orsakasamband

- **Orsakasamband** (causation) er til staðar þegar breyting á annarri breytunni **veldur** breytingu í hinni breytunni.
- Oft má finna sterka fylgni á milli breyta þó svo að orsakasamband sé ekki til staðar.
- Í mörgum tilfellum eru breyturnar þá undir áhrifum þriðju breytunnar sem þá er kölluð **dulin breyta** (lurking variable).
- Því dugar há fylgni aldrei ein og sér til að fullyrða að orsakasamband sé á milli tveggja breyta.

Yfirlit

Áhættuhlutfall

Áhættuhlutfall er lýsistærð sem er mikið notuð til að bera saman tíðni í tveimur þýðum.

Áhættuhlutfall

Áhættuhlutfall, táknað RR, er hlutfall útkomu í tilteknu þýði táknað p_1, deilt með hlutfalli sömu útkomu í viðmiðunarþýði, táknað p_2.

$$RR = \frac{p_1}{p_2} \qquad (1)$$

Túlkun áhættuhlutfalls

Túlkun áhættuhlutfalls

- Sé RR $= 1$ er hlutfallið það sama í þýðunum tveimur.
- Sé RR <1 er hlutfallið minna hjá tiltekna þýðinu en í viðmiðunarþýðinu.
- Sé RR >1 er hlutfallið meira hjá tiltekna þýðinu en í viðmiðunarþýðinu.

Áhættuhlutfall - Dæmi

Árið 2003 fæddust 18,2% allra barna með keisaraskurði á Íslandi.
Árið 1989 fæddust 11,6% allra barna með keisaraskurði á Íslandi.
Hvert er áhættuhlutfallið á því að fæðast með keisaraskurði fyrir börn fædd 2003 á móti börnum fæddum 1989 á Íslandi og hvernig túlkum við það?

Áhættuhlutfall - Dæmi

Árið 2003 fæddust 18,2% allra barna með keisaraskurði á Íslandi.
Árið 1989 fæddust 11,6% allra barna með keisaraskurði á Íslandi.
Hvert er áhættuhlutfallið á því að fæðast með keisaraskurði fyrir börn fædd 2003 á móti börnum fæddum 1989 á Íslandi og hvernig túlkum við það?

Í þessu dæmi er $p_1 = 0.182$ og $p_2 = 0.116$ svo áhættuhlutfallið er

$$RR = \frac{p_1}{p_2} = \frac{0.182}{0.116} = 1.569.$$

Það segir okkur að börn fædd 2003 eru í 1.569 meiri hættu á að vera tekin með keisaraskurði heldur en börn fædd 1989.

Gagnlíkindi

- Áhættuhlutfall ber saman tíðni í tveimur þýðum.
- Oft viljum við bera saman hlutföll tveggja breyta án þess að við vitum eiginlega tíðni í þýðum.
- Þá notum við gagnlíkindahlutföll í staðinn.

Gagnlíkindi

Ef líkurnar á að tiltekinn atburður eigi sér stað eru p, þá eru **gagnlíkindi** atburðarins, táknaðar með o reiknaðar með jöfnunni

$$o = \frac{p}{1-p} \tag{2}$$

Gagnlíkindahlutfall

Gagnlíkindahlutfall

Gagnlíkindahlutfall, táknað OR, eru gagnlíkindi breytu í tilteknum hópi, táknað o_1, deildar með gagnlíkindum sömu breytu í öðrum hópi, táknaðar o_2.

$$OR = \frac{o_1}{o_2} \tag{3}$$

Gagnlíkindahlutfall er einfalt að reikna með því að „margfalda í kross". Við sjáum dæmi um það á eftir.

Túlkun gagnlíkindahlutfalls

Túlkun gagnlíkindahlutfalls

- Sé OR $=1$ eru gagnlíkindin þær sömu í hópunum tveimur.
- Sé OR <1 eru gagnlíkindin minni hjá tiltekna hópnum en í viðmiðunarhópnum.
- Sé OR >1 eru gagnlíkindin meiri hjá tiltekna hópnum en í viðmiðunarhópnum.

Gagnlíkindahlutfall - Dæmi

Á árunum 2003 til 2010 gengust 378 sjúklingar undir gallblöðrutöku á HVE á Akranesi. Níu fengu alvarlega fylgikvilla, 6 þeirra höfðu áður fengið gallblöðrubólgu. Að auki höfðu 84 áður fengið gallblöðrubólgu en fengu ekki alvarlega fylgikvilla. Eykur saga um gallblöðrubólgu hættu á alvarlegum fylgikvillum?

Gagnlíkindahlutfall - Dæmi

Á árunum 2003 til 2010 gengust 378 sjúklingar undir gallblöðrutöku á HVE á Akranesi. Níu fengu alvarlega fylgikvilla, 6 þeirra höfðu áður fengið gallblöðrubólgu. Að auki höfðu 84 áður fengið gallblöðrubólgu en fengu ekki alvarlega fylgikvilla. Eykur saga um gallblöðrubólgu hættu á alvarlegum fylgikvillum?

	Alvarlegur fylgikvilli	Ekki alvarl. fylgikv.
Saga um gallblöðrubólgu	6	84
Ekki áður gallblöðrubólga	3	285

$$OR = \frac{6 \cdot 285}{3 \cdot 84} = 6.79$$

Sjúklingar sem hafa áður fengið gallblöðrubólgu hafa 6.79 sinnum meiri gagnlíkindi á að fá alvarlega fylgikvilla við gallblöðrubrottnám.

Yfirlit

Sannar og falskar, jákvæðar og neikvæðar mælingar

Sannar og falskar, jákvæðar og neikvæðar mælingar

- Sannar jákvæðar (SJ): Mælingar sem eru **flokkaðar með** eiginleikann og hafa hann **í raun**.
- Falskar jákvæðar (FJ): Mælingar sem eru **flokkaðar með** eiginleikann en hafa hann **ekki í raun**.
- Sannar neikvæðar (SN): Mælingar sem eru **flokkaðar án** eiginleikans og hafa hann **ekki í raun**.
- Falskar neikvæðar (FN): Mælingar sem eru **flokkaðar án** eiginleikans en hafa hann **í raun**.

Sannar og falskar, jákvæðar og neikvæðar mælingar

Sannar og falskar, jákvæðar og neikvæðar mælingar

Mörgum finnst hugtökin skýrari þegar þau eru sett upp í litla töflu. Þá teljum við fjölda mælinga sem falla undir hvern og einn flokk.

	Með eiginleikann	Án eiginleikans
Flokkuð með eiginleikann	SJ	FJ
Ekki flokkuð með eiginleikann	FN	SN

Tafla: Sannar og falskar, jákvæðar og neikvæðar mælingar.

Næmi

Næmi

Næmi flokkunaraðferðar er fjöldi sannra jákvæðra mælinga deildur með samanlögðum fjölda sannra jákvæðra og falskra neikvæðra mælinga.

$$\frac{SJ}{SJ + FN} \tag{4}$$

Hversu hátt hlutfall af þeim sem hafa eiginleikann greinum við?

Sértæki

Sértæki

Sértæki flokkunaraðferðar er fjöldi sannra neikvæðra mælinga deildur með samanlögðum fjölda sannra neikvæðra og falskra jákvæðra mælinga.

$$\frac{SN}{SN + FJ} \tag{5}$$

Hversu hátt hlutfall af þeim sem hafa ekki eiginleikann greinum við ekki?

Næmi og sértæki - dæmi

Klínískrar greiningar þunglyndis og kvíða skv. viðtali hjá þeim sem komu til hjartaendurhæfingar á Reykjalundi voru bornar saman við viðurkenndan þunglyndis- og kvíðakvarða, Hospital Anxiety and Depression Scale (HAD).

	Þunglynd s. HAD	Ekki þunglynd s. HAD
Þunglynd s. viðtali	14	23
Ekki þunglynd s. viðtali	5	159

Hvert er næmi klínískrar greiningar þunglyndis borið saman við niðurstöður HAD?

Næmi og sértæki - dæmi

Klínískrar greiningar þunglyndis og kvíða skv. viðtali hjá þeim sem komu til hjartaendurhæfingar á Reykjalundi voru bornar saman við viðurkenndan þunglyndis- og kvíðakvarða, Hospital Anxiety and Depression Scale (HAD).

	Þunglynd s. HAD	Ekki þunglynd s. HAD
Þunglynd s. viðtali	14	23
Ekki þunglynd s. viðtali	5	159

Hvert er næmi klínískrar greiningar þunglyndis borið saman við niðurstöður HAD?
Næmið er $\frac{SJ}{SJ+FN} = \frac{14}{14+5} = \frac{14}{19} = 0.737 = 73.3\%$

Jákvætt forspárgildi

Jákvætt forspárgildi

Jákvætt forspárgildi flokkunaraðferðar er fjöldi sannra jákvæðra mælinga deildur með samanlögðum fjölda sannra jákvæðra og falskra jákvæðra mælinga.

$$\frac{SJ}{SJ+FJ} \tag{6}$$

Hversu hátt hlutfall af þeim sem við greinum með eiginleikann hefur hann í raun?

Neikvætt forspárgildi

Neikvætt forspárgildi

Neikvætt forspárgildi flokkunaraðferðar er fjöldi sannra neikvæðra mælinga deildur með samanlögðum fjölda sannra neikvæðra og falskra neikvæðra mælinga.

$$\frac{SN}{SN + FN} \tag{7}$$

Hversu hátt hlutfall af þeim sem við greinum ekki með eiginleikann hefur hann í raun ekki?

Jákvætt og neikvætt forspárgildi - dæmi

Skoðum sama dæmi og áðan.

	Þunglynd s. HAD	Ekki þunglynd s. HAD
Þunglynd s. viðtali	14	23
Ekki þunglynd s. viðtali	5	159

Hvert er forspárgildi jákvæðar klínískrar greiningar þunglyndis skv. viðtali?

Jákvætt og neikvætt forspárgildi - dæmi

Skoðum sama dæmi og áðan.

	Þunglynd s. HAD	Ekki þunglynd s. HAD
Þunglynd s. viðtali	14	23
Ekki þunglynd s. viðtali	5	159

Hvert er forspárgildi jákvæðar klínískrar greiningar þunglyndis skv. viðtali?

Jákvætt forspárgildi er

$$\frac{SJ}{SJ+FJ} = \frac{14}{14+23} = \frac{14}{37} = 37.8\%$$

Yfirlit

Samantekt um lýsistærðir fyrir miðju

	flokkabreyta		talnabreyta	
	óröðuð	röðuð	strjál	samfelld
Miðja spannar			x	x
Tíðasta gildi	x	x	x	
Miðgildi		x	x	x
Meðaltal			x	x
Vegið meðaltal			x	x

Tafla: Tafla sem sýnir hvaða lýsistærð er viðeigandi að nota fyrir hverja gerð af breytu.

Samantekt um lýsistærðir fyrir breytileika

	flokkabreyta		talnabreyta	
	óröðuð	röðuð	strjál	samfelld
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

Tafla: Tafla sem sýnir hvaða lýsistærð er viðeigandi að nota fyrir hverja gerð af breytu.

Samantekt um lýsistærðir

	meðaltal staðalfrávik	miðgildi fjórðungamörk
Samhverf dreifing	x	
Skekkt dreifing		x
Útlagi		x

Tafla: Tafla sem sýnir hvort meðaltal/staðalfrávik eða miðgildi/fjórðungamörk eru meira viðeigandi til að lýsa talnabreytum með tilsvarandi dreifingu.