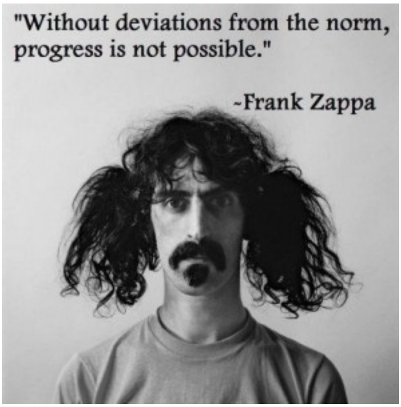
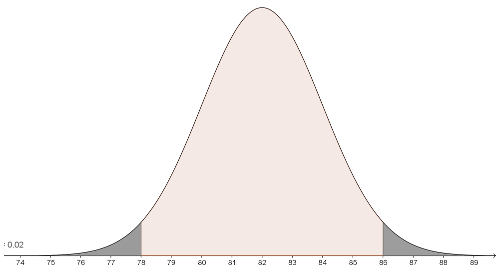
Lota 3

Dreifing



Í fyrsta kafla skoðuðum við nokkrar leiðir til að mæla miðju dreifingar eða miðsækni. Í þessum kafla ætlum við að skoða annan mikilvægan eiginleika gagnasafna, nefnilega dreifingu.
Í síðari hluta kaflans skoðum við svo ýmsar líkindadreifingar.



3.A Meðalfrávik

Frávik táknar hversu langt ákveðið úrtak er frá meðaltalinu í gagnasefni. Þ.e. fyrir gildi x í gagnasafni er frávikið

$$|x - \overline{x}|.$$

Táknið $|\ldots|$ merkir **algildi** eða tölugildi. Það gerir ekki annað en að losa okkur við neikvæðar tölur. Enda, ef við hugsum okkur að meðalskóstærð hóps sé 41 og Guðrún noti skóstærð 38, þá ætti frávikið að vera 3 en ekki -3. $|38 - 41| = |-3| = 3$.
Á ensku kallast algildi **absolute value** svo í tölvum og reiknivélum er þetta kallað `abs`. Þ.e. `abs(38 - 41) = abs(-3) = 3`.

Verkefni 3.A.1
Í kaflaprófi fengu nemendur einkunirnar; 7,8,5,8,6,6,6,9,7. Bergþóra fékk 9 en Njáll fékk 7. Finndu frávikið fyrir hvort þeirra:

Þegar við skiljum hvað frávik merkir ætti hugtakið **meðalfrávik** að vera nokkuð gegnsætt. Meðalfrávik er meðaltal frávikanna. Þetta er því táknað svona:

$$\frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

En hér er átt við að frávikin séu lögð saman fyrir sérhvert úrtak. Ef við vitum tíðni hvers gildis, eins og í tíðnitöflu, þá getum við frekar notað eftirfarandi stæðu:

$$\frac{\sum |x - \overline{x}| f}{n}$$

Verkefni 3.A.2
Finndu frávikjð fyrir alla nemendurna í verkefni 3.A.1 og finndu svo meðalfrávikið.

Svona lagað er þægilegt að setja upp í tíðnitöflu. Í síðasta kafla sáum við tíðnitöflu fyrir aldur í ákveðnum nemendahóp. Bætum við frávikinu:

Gildi (x)	Tíðni (f)	xf	Frávik ($\lvert x-\overline{x}\rvert$)	$\lvert x-\overline{x}\rvert f$
16	4	64	2,3	9,2
17	6	102	1,3	7,8
18	5	90	0,3	1,5
19	9	171	0,7	6,3
20	4	80	1,7	6,8
21	2	42	2,7	5,4
Samtals	30	549	9	37

Verkefni 3.A.3
Notaðu töfluna hér að ofan til þess að finna:

a) Meðalaldur nemendanna.

b) Frávik þess sem er 17 ára.

c) Meðalfrávik aldursins.

3.B Fervik

Í stað þess að nota algildi til þess að losna við neikvæðar tölur er hægt að nota annað veldi. Önnur mæling á dreifingu er því **fervik** (táknað með s^2) og er ekki annað en meðaltal frávikanna í öðru veldi, eða:

$$s^2 = \frac{\sum(x-\overline{x})^2 f}{n}$$

Með tiltölulega einföldum rökstuðningi má finna aðra jöfnu fyrir fervikið. Algengast er að þessi jafna sé notuð í reikningum:

$$s^2 = \frac{\sum x^2 f}{n} - \overline{x}^2$$

Verkefni 3.B.1
Hér eru tíðnitafla úr sömu gögnum og sú á fyrri blaðsíðunni:

Gildi (x)	Tíðni (f)	xf	$\lvert x-\overline{x}\rvert$	$(x-\overline{x})^2 f$	$x^2 f$
16	4	64	2,3		
17	6	102	1,3		
18	5	90	0,3		
19	9	171	0,7		
20	4	80	1,7		
21	2	42	2,7		
Samtals	30	549	9		

a) Fylltu inn þá dálka sem vantar.

b) Finndu fervikið með fyrri aðferðinni og sýndu útreikninga.

c) Finndu fervikið með seinni aðferðinni og sýndu útreikninga.

3.C Staðalfrávik

Eitt sem hægt væri að setja út á varðandi fervikið er að útkoman eru ekki í sömu mælieiningu og gögnin. Þ.e. ef gögnin eru mæld í metrum (m) myndi fervikið vera í metrum í öðru veldi (m^2). Til þess að laga þetta tökum við rót af fervikinu og fáum mælingu á dreifingu sem hefur sömu mælieiningu og gögnin:

$$s = \sqrt{s^2}$$

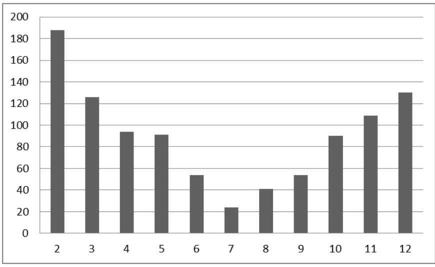
Hér er komin ástæða fyrir því að fervik er táknað s^2. s stendur fyrir **staðalfrávik** og er sú mæling á dreifingu sem mest er notuð og því er mikilvægt að kunna góð skil á henni.

Verkefni 3.C.1
Finndu staðalfrávik gagnanna í verkefni 3.B.1.

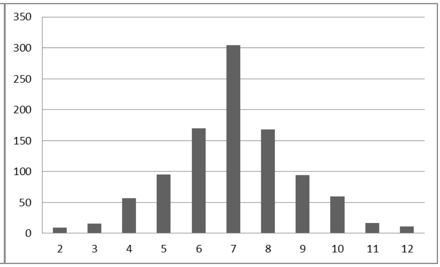
Verkefni 3.C.2

Fjórir teningaframleiðendur bera saman afurð sína með því að kasta teningapari 1000 sinnum. Myndirnar hér að neðan sýna hversu oft hver útkoma kom upp.

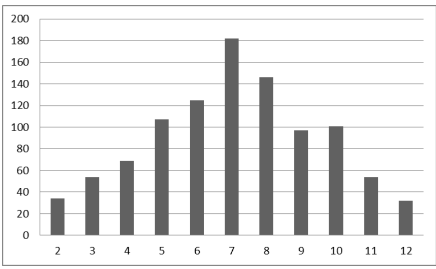
A



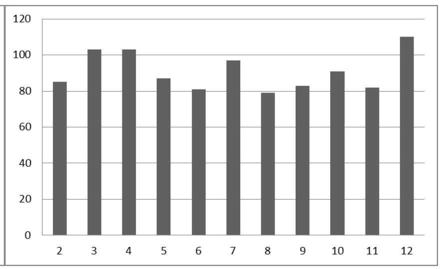
B



C



D



a) Hver var útkoman að meðaltali hjá framleiðendunum fjórum?

b) Raðaðu framleiðendunum eftir staðalfráviki. Þ.e. hvar er staðalfrávikið mest, næst mest, o.s.frv.?

c) Gefðu stutta skýringu á því hvernig má sjá staðalfrávikið af myndum sem þessum.

d) Hvaða teningaframleiðandi heldur þú að framleiði eðlilegustu teningana?

Verkefni 3.C.3
Gefin er þessi tíðnitafla:

x	f	xf	x^2f	$\lvert x-\overline{x}\rvert f$
37	2	74	2738	9,2
38	7	266	10108	25,2
39	6	234	9126	15,6
40	5	200	8000	8
41	8	328	13448	4,8
42	10	420	17640	4
43	6	258	11094	8,4
44	7	308	13552	16,8
45	6	270	12150	20,4
Samtals	60	2496	104204	125,6

Finndu eftirfarandi:

a) Fjöldi úrtaka.

b) Meðaltal.

c) Meðalfrávik.

d) Fervik.

e) Staðalfrávik.

3.D Líkindadreifingar

Verkefni 3.D.1 *Systkinahópar*
Veltum aðeins fyrir okkur systkinahópum. Segjum að þið ætlið að eignast 3 börn.

a) Hvað er hægt að eiga marga mismunandi barnahópa með tilliti til fjölda stráka og í hvaða aldursröð strákarnir og stelpurnar eru (t.d. strákur, stelpa, strákur)?

b) Hvað er þá margir möguleikar á því að eignast 3 stráka, 2 stráka og eina stelpu, o.s.frv.? Skoðið liðunina $(a+b)^3$ og takið saman alla liði (finnið stuðlana).

c) Hvað segir þetta okkur um líkurnar á því að það séu 3 strákar, 2 strákar o.s.frv. í þriggja manna fjölskyldu?

Verkefni 3.D.2 *Tvíliðustuðlar* a) Finnið stuðlana í $(a+b)^4$?

b) Setjið stuðlana í liðunum úr þessu dæmi og dæminu á síðustu blaðsíðu í tvær línur beint fyrir ofan/neðan hverja aðra. Athugið hvort þið getið svo fyllt inn nýjar raðir fyrir ofan og neðan.

3.E Tvíkostadreifing

Dreifingin á útkomunum sem við sáum í dæmi **2.2** hér á undan er kölluð tvíkostadreifing. Ef þið hafið leyst dæmið samviskulega og raðað stuðlunum upp í röð þá hafið þið ef til vill þekkt mynstrið sem sést. Ef þið hafið ekki nú þegar gert það skulu þið reyna að rita stuðlana í línunum fyrir ofan og eina til tvær línur fyrir neðan. Þríhyrningurinn sem myndast er þekktur sem **Pascalþríhyrningurinn**.

Í raun sýnir tvíkostadreifing allar mögulegar útkomur tilrauna þar sem hver tilraun hefur tvær mögulegar útkomur. Hún annað hvort heppnast eða heppnast ekki (gerist eða gerist ekki). Ekki þarf að margfalda uppúr svigum eða teikna þríhyrning til að reikna út hvað hægt er að láta r tilraunir heppnast af n fjölda tilrauna. Þetta er stundum táknað $\binom{n}{r}$ sem lesist n yfir r eða (fjöldi r staka hlutmengja úr n staka mengi.

$$\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!r!}$$

Geogebra getur reiknað þetta beint með skipuninni:

<Tvíliðustuðull[<Tala n>, <Tala r>]>

Á mörgum reiknivélum er líka til takki sem reiknar þetta ${}_nC_r$ eða ${}_nC_k$

Verkefni 3.E.1

a) Hvað er hægt að velja tvo nemendur úr hópi 20 á marga mismunandi vegu?

b) Á uppáhalds kínverska veitingastaðnum hans Njáls eru 13 réttir á matseðlinum. Hann kaupir alltaf tilboð sem eru þrír réttir að eigin vali. Hvað getur hann farið oft á veitingastaðinn án þess að koma með nákvæmlega sömu þrjá réttina heim?

c) Hvað eru til margar mismunandi hendur í póker) (hvað er hægt að velja 5 spil úr venjulegum spilastokki á marga mismunandi vegu?)

3.E.1 Bernoulliferli/tvíliðuformúlan

Bernoulliferli er mjög mikilvægt tilraunaferli þar sem við framkvæmum sömu tilraunina aftur og aftur. **Bernoulliferli** er runa af tilraunum þar sem:

- Hver tilraun hefur tvær mögulegar útkomur heppnast/heppnast ekki.

- Tilraunirnar eru óháðar innbyrðis. (t.d. líkurnar á að fá sexu í næsta kasti er óháð því hvort upp kom sexa í síðasta kasti)

- Líkurnar á því að tilraun heppnist er sú sama í hverri tilraun. Ef við köllum líkurnar á því að tilraun heppnist, p, þá eru líkurnar á því að hún heppnist ekki $q = 1 - p$.

Verkefni 3.E.2
Sýnidæmi Teningi er kastað sex sinnum. Reiknum líkurnar á því að við fáum yfir fjóra í að minnsta kosti fimm köstum?
Líkurnar á fá yfir fjóra (tilraun heppnast) er $p = 2/6 = 1/3$ þannig að $q = 1 - 1/3 = 2/3$. Við reiknum þá fyrst líkurnar að fá yfir 4 í 5 köstum:

$$\binom{6}{5} \cdot (\frac{1}{3})^5 \cdot (\frac{2}{3}) = 0.0165$$

og líkurnar á að fá yfir 4 í 6 köstum:

$$\binom{6}{6} \cdot (\frac{1}{3})^6 = 0.0014$$

og líkurnar eru því samanlagt

$$0.0165 + 0.0014 = 0.0179$$

Verkefni 3.E.3
Skilgreinum nú þetta formlega. Líkindi þess að tilraun heppnist í r skipti af n í Bernoulliferli eru:

$$\binom{n}{r} p^r q^{n-r}$$

Segðu með þínum orðum hvað allir liðir í þessari formúlu merkja.

Verkefni 3.E.4

a) Á krossaprófi eru 20 spurningar og við hverja spurningu eru 3 valmöguleikar. Ef nemandi gískar á allar spurningar, hver eru líkindi þess að hann fái fimm í prófinu?

b) Líkindi þess að Skarphéðinn sofi yfir sig og mæti seint í skólann eru 0.2. Hverjar eru líkurnar á því að í næstu viku sofi hann yfir sig...

   i) Nákvæmlega einu sinni?

   ii) Oftar en einu sinni?

iii) Að minsta kosti einu sinni?

Til að geta svarað 2.6a hér að framan kemur tvennt til greina. Annars vegar reiknar þú út tvíliðaformúluna fyrir mörg tilvik í „höndunum”. Þá þarf að reikna líkurnar á því að svara 10,11,12... spurningum rétt og það er tímafrekt. Hins vegar er hægt að láta Geogebra reikna mörg tilvik í runu með smá forritun.

Verkefni 3.E.5
Runuskipunin - upprifjun: Hægt er að láta Geogebru reikna út mörg dæmi í einu og láta hana skila útkomunum í lista eins og við sáum í fyrsta kafla.
Skipunin <Runa[<Stæða>, <Breyta>, <frá>, <til>]> býr til runu af útkomum þar sem hún reiknar einhverja formúlu (Stæða) sem hefur einhverja breytu (Breyta) sem byrjar á að hafa gildið (frá) og hækkar síðan um einn heilan í hvert skipti þar til við erum komin upp í topp (til).

a) Hvernig er runuskipunin sem reiknar fyrstu 20 tölurnar í þrisvarsinnum töflunni?

b) Legðu nú saman tölurnar í listunum hér að ofan. <Summa[<Listi>]>.

Verkefni 3.E.6
Runuskipunin - framhald: Reiknaðu nú svarið í dæmi 3.E.4 með runuskipuninni en reyndu nú að reikna líkurnar á því að fá 5 eða hærra í prófinu. Athugaðu að nú er stæðan (Stæða) eins og í dæmi 2.5 þar sem
$\binom{n}{r}$ er í raun <Tvíliðustuðull[<Tala n>, <Tala r>]>.
n í rununni okkar er 20 þar sem við ætlum að gera 20 tilraunir (gíska á 20 spurningar) en r er breytilegt þar sem við ætlum að reikna líkur á að hafa mismunandi fjölda réttra svara. r er því breytan í rununni.

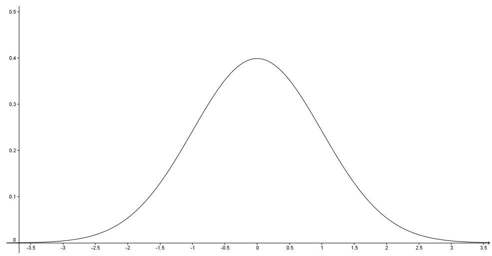
a) Ef við ætlum að reikna líkurnar á að ná hvað eigum við þá að láta r ganga (frá) og (til)?

b) Hverjar eru líkurnar á því að fá yfir 5 í prófinu?

3.F Normaldreifing

Normaldreifingin er þekktust allra líkindadreifinga. Hún hefur bjöllulaga feril. Margar náttúrulegar dreifingar eru normaldreifðar s.s. hæð og þyngd í einsleitum samfélögum, greind og fleira. Normaldreifing er mikið notuð í ályktunartölfræði sem við könnum í næsta kafla. Normaldreifing er samfelld sem þýðir að öfugt við tvíkostadreifingu (eins og bernoulli) þá getum við ekki reiknað líkurnar fá ákveðið gildi heldur notum við normaldreifinguna til að meta líkurnar á að gildið lendi á ákveðnu bili. T.d. ef við ákveðum að velja karlmann af handahófi úr þekktum hópi (t.d. íslenskur framhaldskólanemi). Þá getum við ekki notað normaldreifingu til að meta líkurnar á því að hann sé nákvæmlega 180 sentimetar á hæð. Við getum hins vegar metið líkurnar á því hann sé á bilinu 175-185 eða líkurnar á því að hann sé hærri en 190 sentimetrar.
Normaldreifing hefur nokkur mikilvæg einkenni:

- Normaldreifing er bjöllulaga og algerlega samhverf um meðaltalið.
- Meðaltalið er undir hæsta punkti ferilsins.
- Ferillinn nær endalaust í báðar áttir.
- Normaldreifingin ákvarðast af meðaltali og staðalfráviki þannig að ef tvær normaldreifingar hafa sama meðaltal og staðalfrávik þá eru þær nákvæmlega eins útlítandi (eru sama dreifingin)



3.G Poissondreifing

Poissondreifing er líkindadreifing sem vert er að minnast á þó við notum hana ekki meira í þessum áfanga. Hún er gjarnan notuð til að skrá fjölda atburða sem gerist á tilteknu tímabili. T.d. Fjöldi innhringinga hjá þjónustuveri, eða fjöldi viðskiptavina á hverjum tíma hjá skyndibitastað. Hún er einnig mjög mikilvæg í líffræði, t.d. ef skoðaður er fjöldi dýra á einhverju ákveðnu svæði.