

Projekt 8.8 - Mendel og Hardy-Weinberg

Vores definition af sandsynligheder er enten empirisk begrundet eller eksperimentelt begrundet.

- Den empiriske begrundelse kræver et datamateriale
- Den eksperimentelle begrundelse kræver, at der gennemføres et eksperiment, fx en simulering, der genererer et datamateriale. Det kan evt. udbygges med et tankeeksperiment.

1. Sandsynlighedsbegrebet

1. Forklar hvad vi mener med en empirisk begrundet sandsynlighed, fx ud fra følgende oversigt over, hvordan blodtypesystemet ABO fordeler sig på forskellige befolkninger:

Land	O	A	B	AB
Danmark	41%	44%	11%	4%
Norge	39%	50%	8%	4%
Oprindelige Australier aboriginer	39%	61%	0%	0%
Peru indianere	100%	0%	0%	0%

Se evt i HEM 1, kapitel 9.

2. Sandsynligheden for at trække et billedkort ud af et almindeligt spil kort med 52 blade siger vi er $\frac{12}{52}$. Vi siger også, at dette er eksperimentelt begrundet. Hvad menes hermed?

Se evt i HEM 1, kapitel 9 og inddrag de store tals lov i din forklaring

2. Sandsynlighedsfelt

1. Hvad er et sandsynlighedsfelt? Illustrer din forklaring med at opstille sandsynlighedstabeller for følgende:

- Udfaldet ved kast med tre mønter
- Blodtypen i ABO-systemet for en tilfældigt valgt dansker
- Blodtypen i ABO-systemet for en tilfældigt valgt aboriginer
- De to mulige udfald: *to ens*, eller *to forskellige*, ved kast med to terninger.

2. Vi har givet en sandsynlighedstabel, hvor noget er udfyldt:

Udfald u	u_1	u_2	u_3	u_4
Sandsynlighed p	0,1	0,3	0,25	?

Færdigudfyld tabellen.

3. Symmetrisk sandsynlighedsfelt

1. Hvad er et symmetrisk sandsynlighedsfelt? Giv eksempler.

2. Følgende eksperiment giver ikke et symmetrisk sandsynlighedsfelt: *Kast med to terninger og udregn summen af øjnene*. Forklar hvordan vi alligevel kan anvende et symmetrisk sandsynlighedsfelt til at udfylde sandsynlighedstabellen.

3. Forklar, hvordan der ligger et symmetrisk sandsynlighedsfelt bag vores opstilling af sandsynlighedstabeller i opgaverne om blodtyper i 2.1.

4. Forklar hvad der menes med formelen:

$$\text{Sandsynligheden for hændelsen A} = \frac{\text{Antal gunstige for A}}{\text{Antal mulige i alt}}$$

4. Multiplikationsprincippet (se evt HEM 1, kap 9)

Ved kast med 1 terning er $P(\text{øjne} > 3) = \frac{3}{6}$, $P(\text{øjne} > 4) = \frac{2}{6}$

1. Vi kaster en rød og en sort terning. Hvad er $P(\text{rød viser} > 3 \text{ og sort viser} > 4)$?

Begrund svaret ud fra formelen i 3.4

2. Indfør betegnelsen A for hændelsen $\emptysetjne > 3$ og B for hændelsen $\emptysetjne > 4$.

Hvis vi kalder sandsynlighederne for p og q, dvs: $P(A) = p$ og $P(B) = q$, vis så, at 1) kan skrives:

$$P(A \text{ og } B) = p \cdot q$$

Dvs vi ganger sandsynligheder sammen, når hændelserne er uafhængige af hinanden.

5. Mendels spaltningsskema og multiplikationsprincippet

Antag en bestemt arvelig egenskab findes genetisk i to lige dominerende varianter A og B. (Disse kaldes *co-dominante*). Antag at 40% af alle gener er A og at 60% er B, og at fordelingen er ens for mænd og kvinder. 1. Opstil Mendels spaltningsskema for, hvordan egenskaben nedarves.

2. Vi forestiller os nu, vi har et skema på 100 x 100, der viser alle muligheder af kombinationer for de to varianter A og B. Dvs, der er 40 A'er og 60 B'er i kvinderækken og tilsvarende 40 A og 60 B i mandesøjlen. (Du skal ikke lave et sådan skema, men forestille dig det!) Bestem nu:

- Det samlede antal muligheder
- Det samlede antal børn af typen AA og af typen BB
- Det samlede antal børn af typen: mor A og far B, samt typen mor B og far A

3. Argumenter for, at hvis vi sætter $P(A) = 0,40 = p$, og $P(B) = 0,60 = q$, så gælder:

$$P(AA) = p^2, P(BB) = q^2, P(AB) = 2 \cdot p \cdot q$$

(udnyt formlen i 3.4)

6. Arvelighedsteori på Darwins tid

Antag egenskaber nedarves som et gennemsnit (evt et vægtet gennemsnit) af forældrenes egenskaber. Betragt en egenskab som farve (øjnenfarve, pelsfarve, kronblades farve), som vi kan angive på en skala fra 1 til 100. Gennemfør et eksperiment, hvor vi simulerer, hvad der sker med afkommets egenskaber efter nogle generationer. Vælg som udgangspunkt 2 x 100 forældrepar, hvor egenskaberne er tilfældigt fordelt.

Illustrer med boksplot og udregning af spredning i form af kvartilbredde, hvordan udviklingen er for de næste generationer. Kommenter.

7. Mendels løsning – artsrigdom bevares

Mendel formulerede en hypotese om, at egenskaber ikke nedarves som et gennemsnit, men at egenskaberne findes hos den enkelte i to versioner, et dominerende A og et vigende a. Han formulerede yderligere den hypotese, at kønsceller kun har én af disse egenskaber, enten A eller a.

1. Argumenter ud fra Mendels spaltningsskema, at vi kan se *to forskellige typer*, og at forholdet mellem deres relative andel må være 1:3.

2. Opstil selv et spaltningsskema for *to egenskaber*, eller find et sådant, og forklar hvordan man skal forstå det.

Argumenter for, at vi kan se *fire forskellige typer*, og at disse fordeler sig i forholdet $\frac{1}{16} : \frac{3}{16} : \frac{3}{16} : \frac{9}{16}$.

8. Mendels forsøg med ærteplanter

Her er et skema over Mendels egne data fra hans forsøg med ærteplanter:

Struktur	Egenskab	Dominerende	Vigende
Frø	Form (RUND, rynket)	5474	1850
Oplagsnæring i Kimbladene	Farve (GUL, grøn)	6022	2001
Bælg	Form (OPPUSTET, indsnøret)	822	299
Blomster	Farve (Rødviolet, hvid)	705	224
Umodne bælg	Farve (Grøn, gul)	428	152
Blomster	Stilling (SPREDTE, endestillede)	651	207
Stængel	Længde (HØJ, lav)	787	277

1. Forklar, hvordan Mendel kunne finde belæg for sin teori, der blev udmøntet i hans spaltningsskema, ud fra disse empiriske data. Hvordan vil du karakterisere Mendels videnskabelige metode?
2. Formuler Mendels hypotese som en nulhypotese, vi kan teste, og udregn for hver af hans forsøg det forventede antal dominerende og vigende under antagelse af hypotesen
3. Gennemfør et *binomialtest* for hver af de syv egenskaber om den observerede fordeling er i overensstemmelse med den forventede fordeling ifølge Mendels love.
4. Der har efterfølgende været kritik af Mendels data. Mendel skriver selv at det kun er et udsnit af data han offentliggør. Det samlede datasæt blev brændt efter hans død så vi har ikke efterfølgende mulighed for at undersøge efter hvilke kriterier han udvalgte sine data. Er der noget mønster i de ovenstående binomialtest, der kunne tyde på en form for datafusk?
5. Der findes tilfælde, hvor empiriske data tilsyneladende ikke følger Mendels skema. Fx har man i et kontrolleret forsøg med mus krydset man mus med gul pels med mus med grå pels. Krydser man derefter afkommet fås følgende fordeling:

Pelsfarve	Gul	Grå
Antal	1511	767

Der er ingen blandingsfarve og pelsfarven gul synes at være dominerende. Hvad kan være forklaringen på dette resultat

Hardy Weinberg - Indledende bemærkninger

Hvis en egenskab udtrykkes gennem to gener, et dominerende A, og et vigende a, så er der:

- tre forskellige *genotyper*: AA, aa og Aa. Typer som AA og aa kaldes homozygote, typer som Aa kaldes heterozygote.
- to forskellige *fænotyper*: typerne A og a.

(Ordet *gen* bliver indført af den engelske genetiker Bateson, og ordene *genotype* og *fænotype* bliver indført af den danske genetiker Wilhelm Johansen. De to er venner og arbejder tæt sammen. Begreberne formuleres først i 1900-tallet)

Den samlede %-andel af A-gener kaldes *gen-frekvensen* for A. Tilsvarende for a.

(Strengt taget er *gen* et ord for egenskaben, og A og a, som er den måde genet udtrykker sig på, kaldes for *alleler*. Her giver det god mening, at tale om gener.)

Vi antager i det følgende, at

- der er samme genfrekvenser blandt mænd og kvinder.
- man finder sin partner tilfældigt, dvs uden hensyn til om partneren er AA, Aa eller aa.
- genotyperne har samme fertilitet og samme overlevelseshastighed

Vi ser også bort fra mutationer, der ikke var et kendt begreb på Mendels, Hardys og Weinbergs tid

9. Indledende eksempel-opgave til Hardy Weinberg

Et af blodtypesystemerne hedder MN-systemet. Det er et eksempel på, at ikke alle par af gener er dominante og vigende, men at der findes tilfælde hvor begge er dominante (de kaldes co-dominante). Der findes her

- tre genotyper: MM (kaldes blodtype M), NN (kaldes blodtype N) og MN (kaldes blodtype MN)
- tre fænotyper, nemlig de samme

Antag vi har en population på 214, hvor der er:

- 60 med blodtype M
- 101 med blodtype MN
- 53 med blodtype N

1. Vis, at genfrekvensen for M er 0,52 og for N er 0,48.

2. Vis, at genotype-frekvensen for MM er 0,28, for MN er 0,47 og for NN er 0,25.

3. Opstil et diagram over **hvor stor en %-andel** de forskellige genotyper udgør i 2. generation:

	M	N
M		
N		

(Forekommer regning med sandsynligheder abstrakt, så gør som i opgave XX og antag, der er i alt 100 gener fra kvinder i øverste række, heraf 52 M og 48 N, og der er tilsvarende 100 gener fra mænd i venstre kolonne, også med 52 M og 48 N. Regn så på det).

Kontroller summen er 1.

4. Sammenlign frekvenserne af blodtyperne (genotyperne) i 2. generation med 1. generation. Fx finder vi, at frekvensen af NN i 2. generation er 0,2394. Opstil på baggrund heraf en prognose for, hvordan det vil udvikle sig i de kommende generationer.

5. Udregn genfrekvenserne for M og N i 2. generation. Kommenter resultatet!

10. Hardy-Weinberg ligevægt

Vi regner nu videre med genfrekvenserne i 2. generation.

1. Opstil et diagram som ovenfor over, **hvor stor en %-andel** de forskellige genotyper udgør i 3. generation:

	M	N
M		
N		

2. Sammenlign med diagrammet over 2. generation og giv en ny prognose for hvordan det vil udvikle sig i de kommende generationer.

3. Udregn genfrekvenserne for M og N i 3. generation.

4. Prøv at formulere dit resultat som en lov. Hvis alt er lykkedes for dig, så har du formuleret Hardy Weinbergs lov.

11. Udregning af Hardy-Weinberg ligevægt på symbolsk form

Vi betragter en egenskab der udtrykkes gennem det dominante A og det recessive a. Vi forestiller os, at vi har givet et tilfældigt udsnit af populationen, hvor antallet med genotype AA er x, antallet med genotype Aa er y og antallet med genotype aa er z. Dvs det samlede antal i vores gruppe er $N=x+y+z$.

1. Opstil formler for andelen af de tre genotyper. Summen giver naturligvis 1.

2. Betragt den samlede genpulje dannet af denne gruppe. Opstil formler for genfrekvensen p af A og genfrekvensen q af a.

Udfør kontrol på dine udregninger ved at indsætte talværdierne fra opgave 9: $x=60$, $y=101$ og $z=53$, og tjek at du fx får, at andelen af A er 51,64%

Kontroller at summen af de to giver 1.

Vi betragter dette som 1. generation.

3. Opstil et diagram der viser, **hvor stor en %-andel** de forskellige genotyper udgør i 2. generation, givet de genfrekvenser, du udregnede i punkt 2:

	A	a
A		
a		

Sammenlign med de andele du udregnede i punkt 1.

Udfør kontrol på dine udregninger ved at indsætte talværdierne fra opgave 9: $x=60$, $y=101$ og $z=53$, og tjek at du fx får, at andelen af AA er 26,66%

Kontroller, at summen af andelene giver 1 (100%)

4. Opstil en formel for genfrekvensen af A og genfrekvensen af a i 2. generation.

indsæt talværdierne fra opgave 9: $x=60$, $y=101$ og $z=53$, og sammenlign med de andele du udregnede i punkt 2. Hvad er din konklusion?

5. Begrund nu, at du umiddelbart, uden yderligere beregninger kan opskrive resultatet man får i 3. generation.

5. Formuler Hardy Weinbergs ligevægtslov.

Bemærk, at ligevægten ikke havde noget at gøre med det specifikke talmateriale vi anvendte som kontrol. Dvs Ligevægten indtræffer i ethvert udsnit af en befolkning vi vælger – og sætter ud på en øde ø!

Bemærk også, at ligevægten først indtræffer efter 2. generation.

12. Hardys oprindelige artikel om ligevægten

Læs Hardys artikel og svar på følgende:

1. Artiklen er et indlæg i en debat. Hvad drejede debatten sig om?
2. Hvordan argumenterer Hardy. Redegør for det taleksempel han anvender.
3. I løbet af artiklen får Hardy formuleret det vi i dag kalder Hardy Weinbergs lov. Hvor er det præcis han formulerer den. Sammenlign med den formulering du nåede frem til i opgave 11.
4. Hardy tager en række forbehold. Redegør for disse

13. Anvendelse af Hardy Weinberg

1. Øjenfarven blå eller brun bestemmes af et gen, der optræder med varianten B der koder for brun og varianten b, der koder for blå. Brun er dominant, blå er recessivt. Der er 64% af den danske befolkning, der har blå øjne.

- Hvad er genfrekvensen for b (blå)?
- Hvad er genfrekvensen for B (brun)?
- Hvor mange homozygote (BB) og hvor mange heterozygote (Bb) brunøjede er der?

2. I USA lider 1 ud af 3700 af cystisk fibrose. Sygdommen skyldes et recessivt gen. Vis, at over 3% af befolkningen er bærer af sygdomsgenet uden at mærke det.

3. Løs artiklen fra encyklopædien om eugenik. I slutningen står der følgende:

Recessive gener for alvorlige sygdomme, som fx Føllings sygdom (fenyketonuri, PKU) og Tay-Sachs' sygdom, udgør under 0,5% af genpuljen. Det betyder at antallet af homozygoter i en befolkning på 5 mio. vil være under 125, mens der vil være 400 gange så mange raske heterozygoter, altså bærere af sygdomsgenet. Det er derfor naivt at forestille sig, at man kan udrydde disse gener i befolkningen ved at forhindre homozygoterne i at forplante sig (negativ eugenik).

Argumenter for de beregningsmæssige påstande i afsnittet.

14. Forening af Darwin, Mendel og Hardy Weinberg: Matematisk modellering af 'survival of the fittest'

Antag vi har en egenskab, hvor A repræsenterer det dominante og a det vigende gen. Antag at genfrekvensen af A er 70%.

1. Opstil Mendels spaltningsskema og udregn genfrekvensen i de følgende generationer.
2. Ikke alle arter og individer er lige *fit to survive*, der er en af forudsætningerne for Hardy Weinberg. Antag nu, at der er en fitness faktor w , så $w(AA)$ er 1, $w(Aa)$ er 0.98 og $w(aa)$ er 0.92. Udregn genfrekvensen i den første og anden generation.
3. Anvend et regneark og udregn udviklingen gennem fx 50 generationer. Giv en grafisk fremstilling af udviklingen.