

Statistik och dataanalys I - F11

Statistiska institutionen

Jonas Bjermo

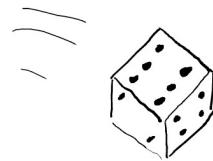
VT2026

Innehåll föreläsning 11

- Från slumpmässighet till sannolikhet
- Försök, utfall och händelser
- Sannolikheter och sannolikhetslära
- Sannolikhetsberäkningar
- Kombinatorik

Slumpmässighet

- Går det att förutsäga utfallet av ett tärningskast?

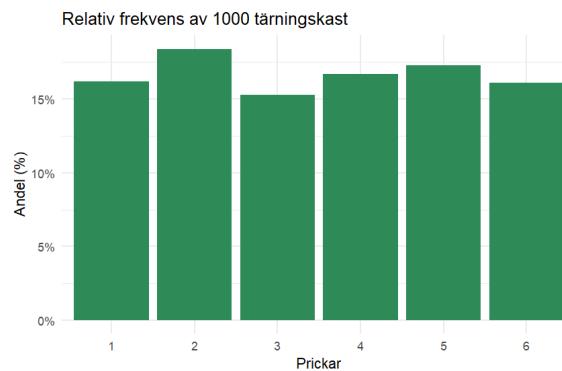
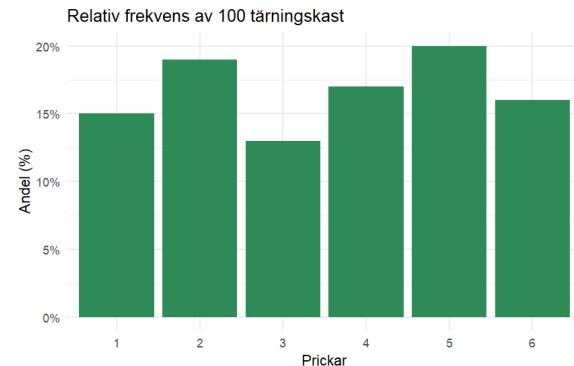
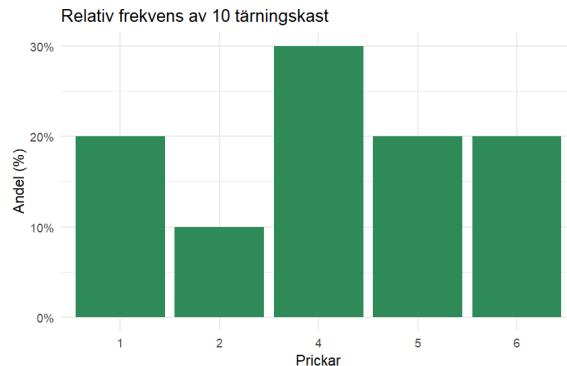


- En slantsingling?



Från slumpmässighet till sannolikhet

- **Svårt!** Eftersom det är **slumpmässiga** händelser.
- Men vi kan förutsäga **proportioner** av tärningskast, slantsingling osv. i det det **långa loppet**.



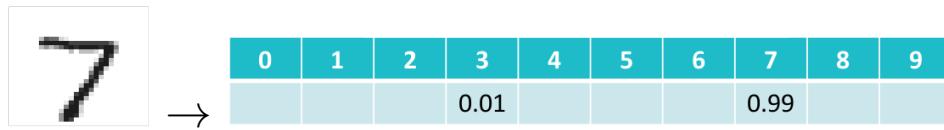
Sannolikheter

- Sannolikheter för en kärnkraftsolycka.
- Sannolikheten för en stor översvämning.
- Sannolikheten för att två personer har en matchande DNA-profil.

Exempel: Ett test för en ovanlig sjukdom har 99 % pricksäkerhet. Sjukdomen finns hos 0,1 % av befolkningen. Om du testar positivt, vad är sannolikheten att du faktiskt har sjukdomen?

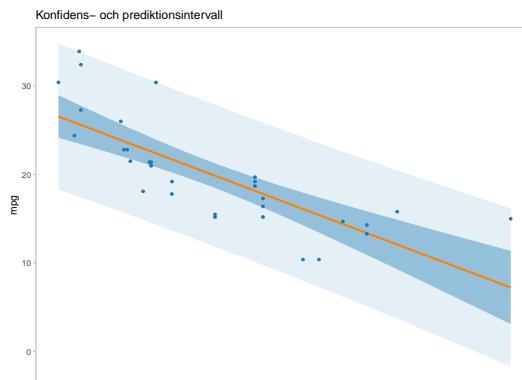
Sannolikhetslära

- Beskriver och kvantifierar **osäkerhet** och **slumpmässiga fenomen**.
 - ▶ Statistiska modeller bygger på **sannolikhetsmodeller**.
 - ▶ Att **kvantifiera osäkerheten** i en **prediktion**.
 - ▶ Hjälper till att fatta **optimala** beslut i en osäker värld.



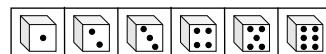
Sannolikheter för regression

- Hittills på kursen:
 - ▶ skattat **regressionslinjen**: $\hat{y} = b_0 + b_1 x$
 - ▶ **prediktion** för en ny observation: $\hat{y}_i = b_0 + b_1 x_i$
- Sannolikhetslära hjälper oss att göra ännu mer:
 - ▶ Om $b_1 \neq 0$, är x och y verkligen korrelerade? Från **stickprov** till **population**.
 - ▶ **Osäkerhetsintervall** för b_1 som troligen täcker det sanna värdet.
 - ▶ **Osäkerhetsintervall** för **prediktionen** \hat{y}_i



Försök, utfall och utfallsrum

- Vi utför ett **försök** (eng: trial): singlar slant.
- Observerar ett **utfall** (eng: outcome): Krona.
- **Utfallsrummet** är **alla möjliga** utfall: $S = \{\text{Krona}, \text{Klave}\}$
- Kast med tärning:



- Kast med två tärningar: Summan av antal prickar.

	•	•	•	•	•	•
•	2	3	4	5	6	7
•	3	4	5	6	7	8
•	4	5	6	7	8	9
•	5	6	7	8	9	10
•	6	7	8	9	10	11
•	7	8	9	10	11	12

Händelser

- En **händelse** består av en mängd av **utfall**.
- Händelse $A = \{\text{pricksumman är } 7 \text{ vid två slagna tärningar}\}$

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

	•	••	•••	••••	•••••	••••••
•	2	3	4	5	6	7
••	3	4	5	6	7	8
•••	4	5	6	7	8	9
••••	5	6	7	8	9	10
•••••	6	7	8	9	10	11
••••••	7	8	9	10	11	12

Händelser

- Händelse B = {samma antal prickar på två slagna tärningar}

$$B = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

		•	••	•••	••••	•••••	••••••
		2	3	4	5	6	7
•	2	3	4	5	6	7	8
••	3	4	5	6	7	8	9
•••	4	5	6	7	8	9	10
••••	5	6	7	8	9	10	11
•••••	6	7	8	9	10	11	12
••••••	7	8	9	10	11	12	

Sannolikhet

- Vi kan ju inte förutsäga ett enskild utfall från t.ex. ett tärningskast.
- Vi vill kvantifiera hur troligt ett utfall är med en **sannolikhet**.
- Vad är **sannolikheten** att få 6 prickar med en tärning?
 - ▶ Utfallsrum: $S = \{1, 2, 3, 4, 5, 6\}$
 - ▶ Händelse: $A = \{6\}$
 - ▶ Sannolikhet: $P(A)$. Måste uppfylla $0 \leq P(A) \leq 1$.

Tre sannolikhetsbegrepp

- **Logisk sannolikhet:** Fysiska egenskaper hos en tärning
→ samma sannolikhet för alla utfall

$$P(A) = \frac{\text{antal utfall i A}}{\text{totalt antal möjliga utfall}} = 1/6 \approx 1.667$$

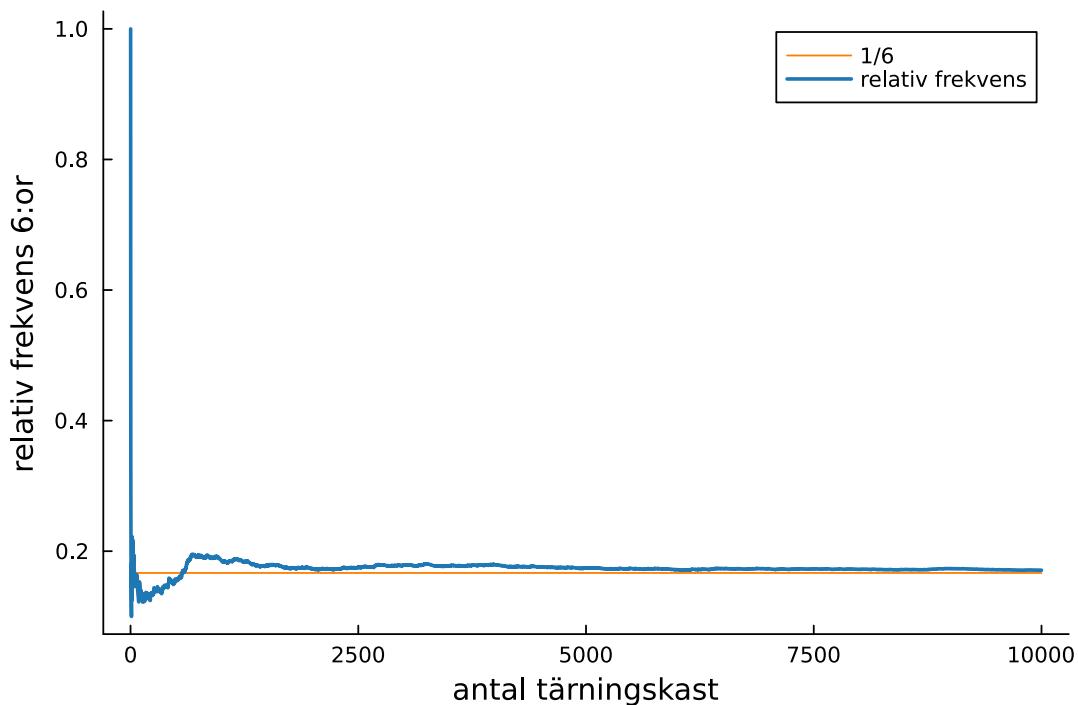
- **Empirisk sannolikhet:** Andelen 6:or om tärningen kastas ”oändligt” många gånger.

$$P(A) = \frac{\text{antal gånger som A inträffar}}{\text{totalt antal försök}}$$

- **Subjektiv sannolikhet:** Min tidigare erfarenhet av tärningskast och min uppfattning om en tärnings symmetri säger mig att min sannolikhet att få en 6:a är $1/6 \approx 0.1667$.

Stora talens lag

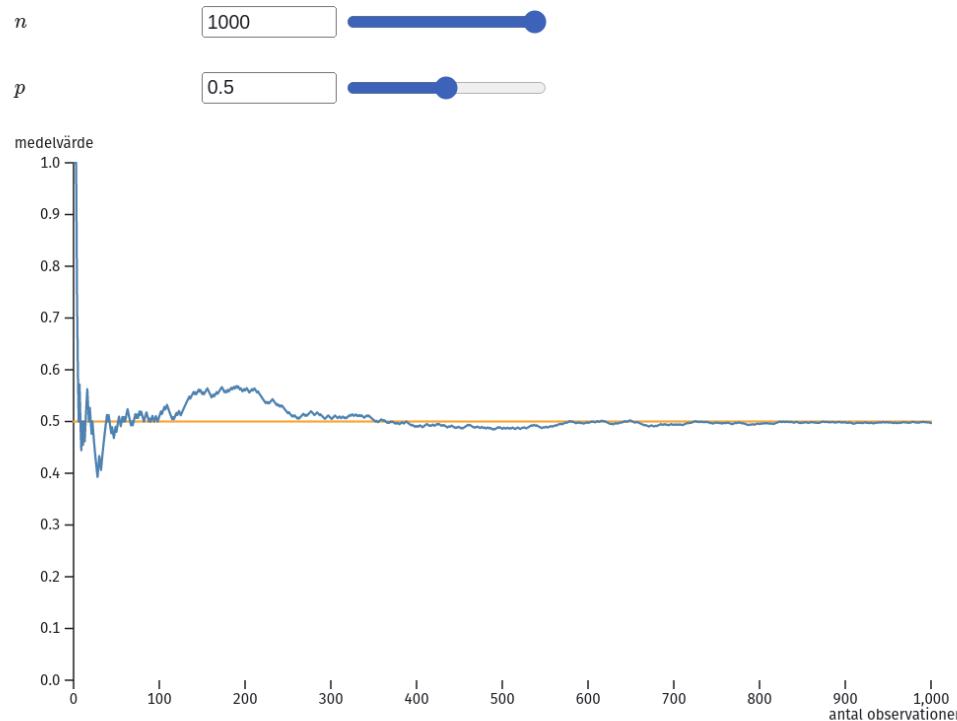
När vi upprepar ett slumpmässig försök om och om igen, stabiliseras andelen gånger ett visst utfall inträffar → ett tal.



Samma sannolikhet vid varje försök och **oberoende** försök.

Stora talens lag - slantsingling

Stora talens lag - slantsingling



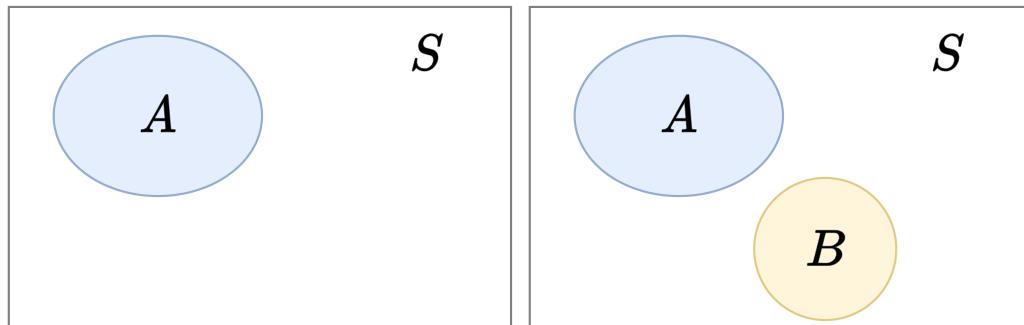
Mattias Villani Stora talens lag för slantsingling



Observable

Venndiagram

- Praktiskt att visualisera händelser i ett **Venndiagram**.
- **Utfallsrummet** (allt som kan inträffa) visas med **rektangel**.
- **Händelser** ritas som **cirklar, ellipser eller rektanglar**.



(a) En händelse A

(b) Två händelser A och B

	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

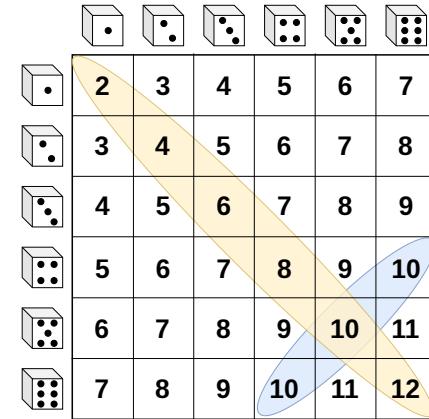
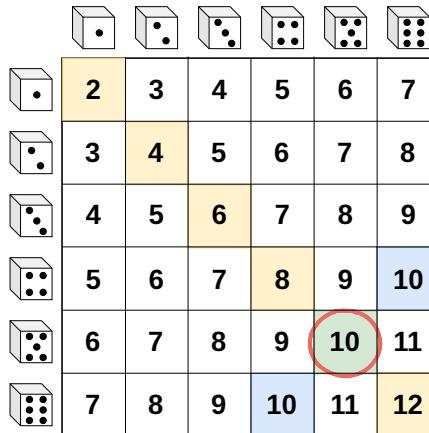
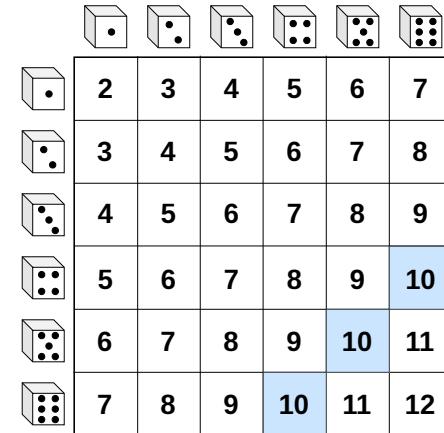
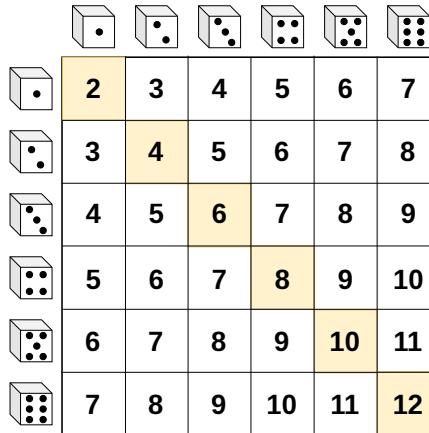
Venndiagram - kast med två tärningar

	2	3	4	5	6	7	
	3	4	5	6	7	8	
	4	5	6	7	8	9	
	5	6	7	8	9	10	
	6	7	8	9	10	11	
	7	8	9	10	11	12	

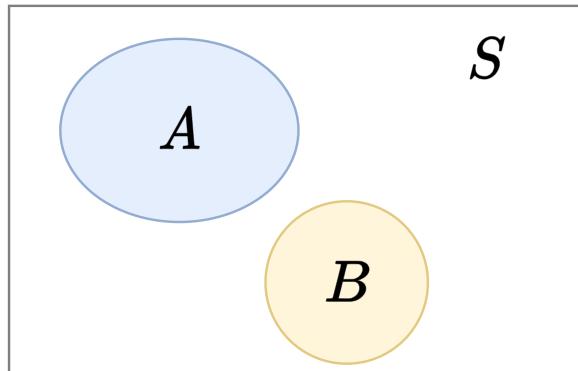
	2	3	4	5	6	7	
	3	4	5	6	7	8	
	4	5	6	7	8	9	
	5	6	7	8	9	10	
	6	7	8	9	10	11	
	7	8	9	10	11	12	

	2	3	4	5	6	7	
	3	4	5	6	7	8	
	4	5	6	7	8	9	
	5	6	7	8	9	10	
	6	7	8	9	10	11	
	7	8	9	10	11	12	

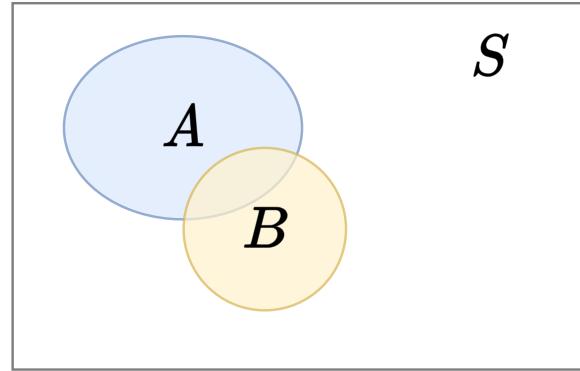
Venndiagram - kast med två tärningar



Disjunkta händelser



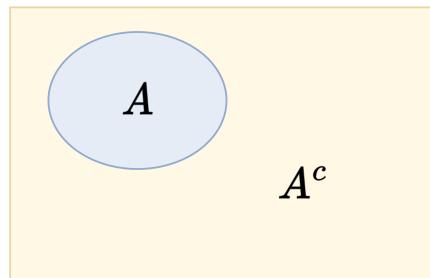
Disjunkta händelser inga gemensamma element.



Överlappande händelser med gemensamma element.

Komplementhändelse

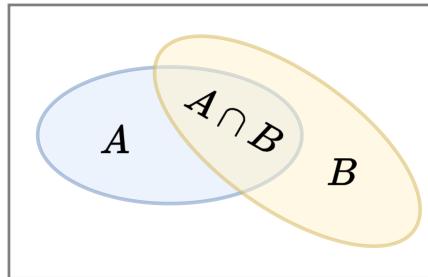
- **Komplementet** till A inträffar när A **inte** inträffar.
- Skrivs A^c , där c = "Complement".



- **Exempel:** Tärningskast
 - ▶ $A = \{\text{udda antal prickar}\} = \{1, 3, 5\}$
 - ▶ $A^c = \{\text{jämnt antal prickar}\} = \{2, 4, 6\}$
- **Exempel:** Inflation
 - ▶ $A = \{\text{inflationen nästa månad} \leq 2\}$
 - ▶ $A^c = \{\text{inflationen nästa månad} > 2\}$

Snitthändelse

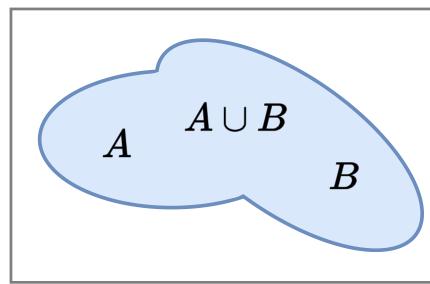
- **Snitthändelsen** är händelsen där **både** A och B inträffar.
- Skrivs A och B eller $A \cap B$.



- Två tärningskast:
 - ▶ $A = \{\text{samma antal prickar på båda tärningarna}\}$
 - ▶ $B = \{\text{summan av prickarna är tio}\}$
 - ▶ $A \cap B = \{5:\text{a på båda tärningarna}\}$
- Lågkonjunktur:
 - ▶ $A = \{\text{BNP-tillväxt kvartal } 1 < 0\}$
 - ▶ $B = \{\text{BNP-tillväxt kvartal } 2 < 0\}$
 - ▶ $A \cap B = \{\text{Negativ BNP-tillväxt två kvartal i rad}\}$
- Snitt av **disjunkta** mängder A och B kallas den **tomma mängden**, $A \cap B = \emptyset$.

Unionhändelse

- **Unionhändelsen** är händelsen där **A och/eller B** inträffar.
- **Minst** en av händelserna inträffar.



- **Exempel:** Universitetsstudier.
 - ▶ $A = \{\text{Kommer in på en kurs på betyg}\}$
 - ▶ $B = \{\text{Kommer in på en kurs på högskoleprov }\}$
 - ▶ $A \cup B = \{\text{Kommer in på kursen}\}$

Formell sannolikhet - med ord

- ① En sannolikhet är ett tal mellan 0 och 1.
- ② Sannolikheten för en händelse som **säkert** inträffar är 1.
- ③ Sannolikheten för att händelsen **inte** inträffar är 1 minus sannolikheten för att den inträffar.
- ④ Av två händelser som **inte kan inträffa** samtidigt är sannolikheten för att **åtminstone en** inträffar summan av båda händelsernas sannolikheter.
- ⑤ Sannolikheten att två oberoende händelser **båda** inträffar är lika med produkten av händelsernas sannolikheter.

Formell sannolikhet - matematisk

Sannolikheten $P(A)$, för händelse A på utfallsrummet S

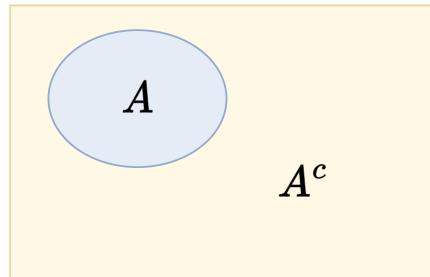
- ① $0 \leq P(A) \leq 1$
- ② $P(S) = 1$
- ③ $P(A^c) = 1 - P(A)$
- ④ $P(A \cup B) = P(A) + P(B)$ om A och B är **disjunkta**
- ⑤ $P(A \cap B) = P(A) \cdot P(B)$ om A och B är **oberoende**

Komplementsregeln

- Händelserna A och A^c kan **inte** inträffa samtidigt; de är ju disjunkta.
- Någon av A och A^c **måste** då inträffa $\Rightarrow P(A) + P(A^c) = 1$

Komplementsregeln

$$P(A^c) = 1 - P(A)$$



- **Exempel (kodning):** $A = \{\text{ingen bugg i koden}\}$,
 $A^c = \{\text{minst en bugg i koden}\} = \{\text{en bugg, två buggar, ...}\} \Rightarrow$
 $P(\{\text{minst en bugg i koden}\}) = 1 - P(\{\text{ingen bugg i koden}\})$

Allmänna additionsregeln

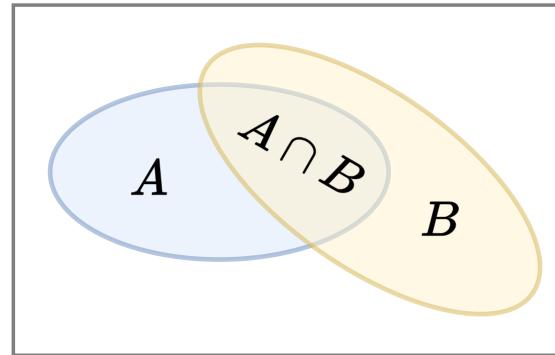
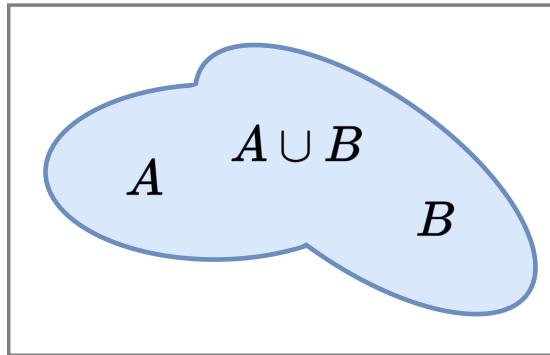
- **Additionsregeln:** A och B är disjunkta

$$\Rightarrow P(A \cup B) = P(A) + P(B)$$

Allmänna additionsregeln

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Om inte snittet $A \cap B$ subtraheras, kommer det att räknas två gånger.



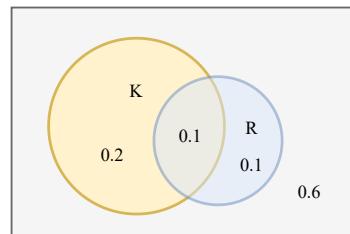
Exempel: Röst på socialdemokraterna (S)

- Låt $R = \{\text{En person röstar på S i riksdagsvalet}\}$, $P(R) = 0.2$.
- Låt $K = \{\text{En person röstar på S i kommunalvalet}\}$, $P(K) = 0.3$.
- Låt sannolikheten att personen röstar på S i båda valen vara $P(R \cap K) = 0.1$.
- Vad är sannolikheten att personen röstar på S i **åtminstone** ett av valen? Additionsregeln:

$$P(R \cup K) = P(R) + P(K) - P(R \cap K) = 0.2 + 0.3 - 0.1 = 0.4$$

- Vad är sannolikheten för att personen inte röstar på S i något av valen?

$$P(R^c \cap K^c) = P((R \cup K)^c) = 1 - P(R \cup K) = 1 - 0.4 = 0.6$$



Multiplikationssregeln för oberoende händelser

- Händelserna A och B är **oberoende** om vetskäpen att B har inträffat **inte påverkar** sannolikheten för A . Och vice versa.

Multiplikationsregeln

För **oberoende** händelser A och B gäller

$$P(A \cap B) = P(A) \cdot P(B)$$

- Hur beräknar man sannolikheten för snittet $A \cap B$ för händelser som inte är oberoende? Kommer i F12.

Multiplikationsregeln för oberoende händelser

- Vad är sannolikheten att få 2 st krona i rad vid slantsingling?

$$0.5 \cdot 0.5 = 0.5^2 = 0.25$$

- Vad är sannolikheten att få 5 st krona i rad vid slantsingling?

$$0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 = 0.5^5 = 0.03125$$

- 1 % risk att streaming laggar under en kväll. Oberoende kvällar.

$$P(\text{ingen lagg hela veckan}) = (1 - 0.01)^7 = 0.99^7 \approx 0.932.$$

Multiplikationsregeln för oberoende händelser

- Sannolikheten att dra två klöver ur en blandad kortlek?

$$P(1:\text{a kortet klöver}) = \frac{13}{52} = \frac{1}{4}$$

$$P(2:\text{a kortet klöver givet 1:a kortet klöver}) = \frac{12}{51}$$

$$P(2:\text{a kortet klöver givet 1:a kortet inte klöver}) = \frac{13}{51}$$

- $A = \{\text{klöver på 1:a}\}$ och $B = \{\text{klöver på 2:a}\}$ är inte oberoende.
- Svaret: $P(1:\text{a och 2:a kortet klöver}) = \frac{13}{52} \cdot \frac{12}{51}$, kommer i F12.
- Vad händer om vi lägger tillbaka första kortet innan vi drar det andra?

Röst på (S) - multiplikationsregeln

- Låt $\mathbf{R} = \{\text{En person röstar på S i riksdagsvalet}\}, P(\mathbf{R}) = 0.2.$
- Låt $\mathbf{K} = \{\text{En person röstar på S i kommunalvalet}\}, P(\mathbf{K}) = 0.3.$
- Låt sannolikheten att personen röstar på S i båda valen vara $P(\mathbf{R} \cap \mathbf{K}) = 0.1.$
- Är händelserna \mathbf{R} och \mathbf{K} **oberoende**? Vi måste undersöka om

$$P(\mathbf{R} \cap \mathbf{K}) \stackrel{?}{=} P(\mathbf{R}) \cdot P(\mathbf{K})$$

- Händelserna är inte oberoende:

$$P(\mathbf{R}) \cdot P(\mathbf{K}) = 0.2 \cdot 0.3 \neq 0.1 = P(\mathbf{R} \cap \mathbf{K})$$

Kombinatorik - på hur många olika sätt?

- **Exempel:** Vi väljer helt slumpmässigt fem kort från en kortlek med 52 kort (4 färger med 13 valörer). Vad är sannolikheten att få ”färg”?
- Kombinatorik underlättar för att lösa detta och liknande problem.
- Antag att n personer ska ställa sig i kö. På hur många olika sätt kan detta utföras?

Kombinatorik - permutationer

- Första personen kan väljas på n olika sätt.
- Andra personen kan väljas på $n - 1$ olika sätt, tredje personen på $n - 2$ olika sätt osv.
- **Multiplikationsprincipen** säger då att det totala antalet sätt är $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 = n!$
- $n!$ sägs vara antalet möjliga **permutationer**.

Kombinatorik - permutationer

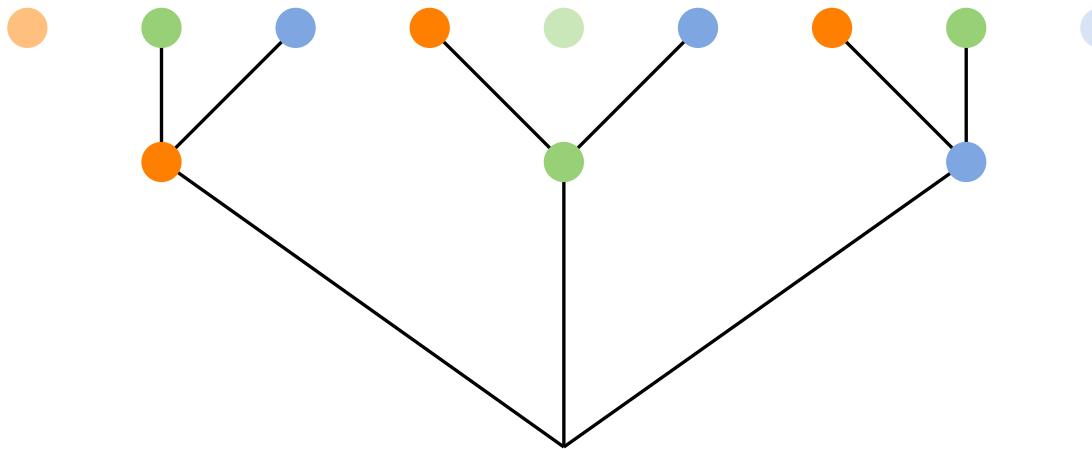
- **Exempel:** Nio löpare i ett lopp. Antalet möjliga sätt att bestämma prispallar? Multiplikationsprincipen $\rightarrow 9 \cdot 8 \cdot 7 = 504$ olika sätt. Antal permutationer av 3 objekt valda bland 9.
- Vi betecknar antal **permutationer** av k objekt valda bland n , ($k \leq n$), med $_nP_k$.
- $_nP_k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$
- $_nP_k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) \cdot (n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$
- **Exempel:** $n = 20, k = 10$
$$_{20}P_{10} = 20 \cdot 19 \cdot \dots \cdot 11 = \frac{20!}{(20-10)!} = \frac{20!}{10!} = 670442572800.$$

Utan återläggning, med hänsyn till ordning

välj $k = 2$ bland $n = 3$

utan återläggning - med hänsyn till ordning

$${}_nP_k = \frac{n!}{(n-k)!} = \frac{3!}{1!} = \frac{3 \cdot 2 \cdot 1}{1} = 6$$



Kombinatorik - kombinationer

- **Exempel:** Välj tre kolor glass från fem olika smaker (nougat, lakrits, jordgubb, vanilj, choklad).
- ${}_5P_3 = \frac{5!}{(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 60$
- Men för den som beställer en glass spelar kanske ordningen ingen roll, permutationen "jordgubb, vanilj, choklad" är likvärdig med "jordgubb, choklad, vanilj".
- Vi talar då om antal **kombinationer**.

Kombinatorik - kombinationer

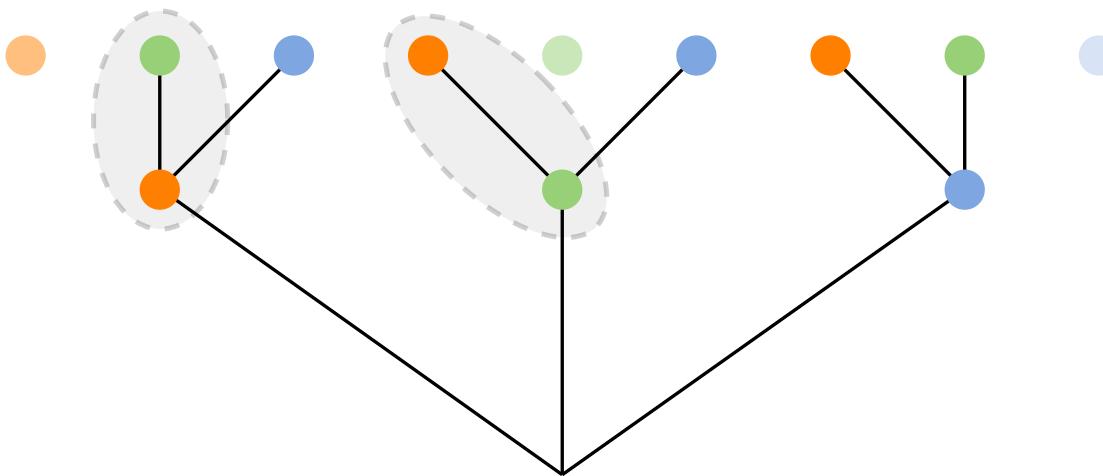
- Varje mängd av k objekt kan ju ordnas på $k!$ olika sätt.
- Därför får vi antal möjligheter till $\frac{n!}{(n-k)!}/k! = \frac{n!}{k!(n-k)!}$.
- Betecknas $_nC_k = \frac{n!}{k!(n-k)!}$ (även $(n)_k$).
- Antal **kombinationer** av k objekt som kan väljas bland n .
- Glass-exempel forts. $_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10$.

Utan återläggning, utan hänsyn till ordning

välj $k = 2$ bland $n = 3$

utan återläggning - utan hänsyn till ordning

$${}_nC_k = \frac{n!}{k!(n-k)!} = \frac{3!}{2!1!} = 3 \text{ sätt}$$



Kombinationer - med återläggning

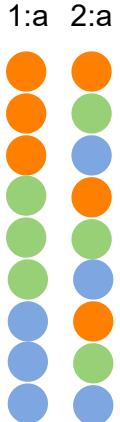
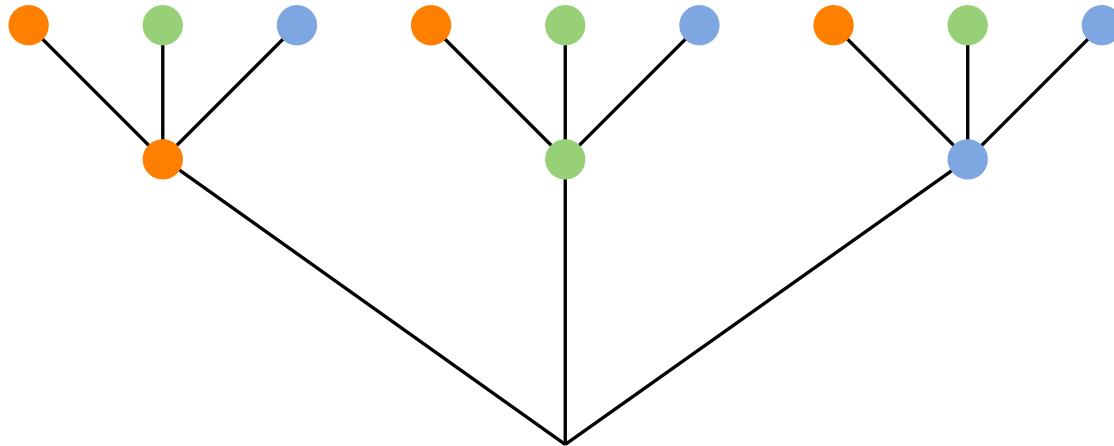
- **Exempel:** Hur många möjliga koder finns det på ett sifferlås med 4 ”platser” som kan ha siffrorna 0-9?
- Varje position i koden kan fyllas av 10 olika siffror (0, 1, 2, 3, 4, 5, 6, 7, 8, 9).
- Alltså kan första platsen ha 10 möjliga siffror, andra 10, tredje 10 och fjärde 10.
- $10 \cdot 10 \cdot 10 \cdot 10 = 10^4 = 10000$
- Allmänt gäller n^k

Med återläggning, med hänsyn till ordning

välj $k = 2$ bland $n = 3$

med återläggning - med hänsyn till ordning

$$n^k = 3^2 = 9$$



Kombinatorik

Hur många sätt att välja k element bland n element?

	med återläggning	utan återläggning
med ordning	n^k	$nP_k = \frac{n!}{(n - k)!}$
utan ordning	ej på kurs	$nC_k = \frac{n!}{k!(n - k)!}$

Credits

Dessa slides är gjorda utifrån slides som skapades för kursen statistik och dataanalys 1 av Mattias Villani HT 2023, och har modifierats av Oscar Oelrich VT 2024, Oskar Gustafsson för VT 2025, och Jonas Bjermo VT 2026.