

Statistikens grunder 1

2013 HT, dagtid

Statistiska institutionen

Orsak och verkan

N Kap 2 forts.

Annat ord: **kausalitet**

Något av det viktigaste för varje vetenskap. Varför?

Orsakssamband ger oss möjlighet att förklara **varför** något inträffar och möjlighet att **styra** åt ett gynnsamt håll.

Objektivitetskrav: en orsak till en händelse är ett **nödvändigt** krav för att det ska hända. **Problem?**

Orsak och verkan, forts.

Krav på verkliga orsaker:

Assymetri

- Om A orsakar B kan inte samtidigt B orsaka A
(återkopplande system?)

Kontrollerbarhet

- Man ska kunna ändra förutsättningarna och verifiera men även styra *(kan t.ex. kön vara en orsak?)*

Tidsfördröjning

- Det som sker idag kan inte påverka det som hände igår

Vetenskapens värderingar

Hur ska vi förhålla oss till våra metoder och våra resultat?

- **Objektivitet**

- Vi ska få samma resultat alldeles oavsett vilka vi röstar på, vem som har finansierat forskningen osv.

- **Transparens**

- Tydlighet i vad som gjorts, definitioner och antaganden

- **Etik**

- Vi ska inte våldföra oss på sanningen
- Vi ska inte heller störa omgivningen (mätningar)

Modeller

N Kap 3

I en generell mening är en **modell** något som på något sätt används för att **representera något annat**.

- **Fysiska objekt som modeller**
 - modelljärnväg, arkitektmodell
- **Konceptuella modeller**
 - finns bara i att sinnet

Konceptuella modeller används för att hjälpa oss förstå ämnet och den verklighet (?) de representerar.

Några viktiga begrepp

- **Population**

- En mängd av väldefinierade objekt som besitter egenskaper
- Kan vara ändlig eller oändlig

- **Urval, stickprov**

- Den delmängd av populationen som vi observerar
- Urvalet kan ske deterministiskt (inte så bra) eller slumpmässigt (bättre)

- **Variabler**

- De egenskaper som objekten i populationen besitter och som man avser att observera

Variabler

- **Kvantitativa variabler**
 - Antar numeriska värden
- **Kvalitativa variabler**
 - Antar icke-numeriska värden
- **Kontinuerliga variabler**
 - Kan anta *samtliga* värden inom ett intervall
 - Kan vara ändlig eller oändlig
- **Diskreta variabler**
 - Kan anta *endast vissa* värden
 - Uppräkneliga, listbara

Skalor

Värdena som en variabel kan anta anges på olika skaltyper:

- **Nominalskala**

- Icke-numeriskt, latin nomen = namn
- Ex. bilmärke, yrke m.m.

- **Ordinalskala**

- Icke-numeriskt men kan **ordnas**
- Ex. "bra, bättre, bäst", ledningshierarki i ett företag

- **Intervallskala**

- Numeriska värden där avstånden (differenser) är väldefinierade men inte kvoter
- Ex. Celsiusskala ("dubbelt så varmt"?), klockslag

- **Kvotskala**

- "20 kr är två ggr större än 10 kr"

Variabeltyp och skaltyp

Variabeltyp		Skaltyp	
Diskret	Kontinuerlig		
X	-	<i>Nominal</i>	Kvalitativ
X	-	<i>Ordinal</i>	
X	X	<i>Intervall</i>	Kvantitativ
X	X	<i>Kvot</i>	

Modeller, forts.

- En modell är en **förenklad** beskrivning av något verkligt
 - Vi tar bara med sådant som är väsentligt
 - Testa hållbarhet i ett material så spelar kanske inte färgen någon roll
- Vi kan ersätta de relevanta aspekterna av verkligheten med **symboler**
 - Vi gör det till "matematik"
 - Ex. $s = v \cdot t$
 - Sträcka, hastighet och tid är variablerna; s , v och t är symboler som representerar variablerna

Stokastiska modeller

I en **deterministisk modell** finns inget utrymme för undantag, allt är exakt beskrivet i modellen.

Ex. Ohms lag: Spänning/Strömstyrka = Resistans

Det som kännetecknar en **stokastisk modell** är att den innehåller en **slumpkomponent**. Vi vet inte exakt vad det kommer att bli men vi kan uttala oss om hur troligt det är.

Ex. X = antal dödsfall av åsnesparkar under ett år:

$$P(X = x) = \lambda^x e^{-\lambda} / x!$$

Utfallsrum

En uppräkning, listning eller beskrivning av **alla** tänkbara **utfall** av ett försök. Ett utfallsrum kan vara

- ändligt eller oändligt
- diskret eller kontinuerligt
- kvantitativt eller kvalitativt

Betecknas ibland med symbolen Ω

Ex.

- diskret mängd: {Krona, Klave}, {1,2,3}, {1,2,3,...}
- kontinuerligt intervall (0,100), [0,100], $(-\infty, \infty)$

Övning

Låt försöket vara "kast med två tärningar" såsom i Ex. 3.4

1. Definiera (X_1, X_2) = "prickar på tärning 1 resp. tärning 2"

Beskriv utfallsrummet för (X_1, X_2)

Är varje utfall lika sannolik?

2. Definiera Y = "summan av tärningarna"

Beskriv Ω_Y = utfallsrummet för Y

Är varje utfall lika sannolik?

3. Definiera Z = "åldern hos en här slumpvis vald person"

Sannolikhet

När vi har ett **utfallsrum** (vi vet vad som **kan** inträffa) så behöver vi också veta, för varje utfall, hur **troligt** det är att ett utfall inträffar.

- Låt e beteckna ett utfall.
- Vi vet att e finns i Ω och vi skriver $e \in \Omega$
- Vi låter $P(e)$ beteckna **sannolikheten** att utfallet blev e
- Sannolikheten $P(e)$ är ett tal

Sannolikhet och händelse

- Låt A beteckna en ***händelse***.
- A är en ***delmängd*** av Ω och vi skriver $A \subset \Omega$.
- Ex. $A = \{1,2\} \subset \{1,2,3,4,5,6\} = \Omega$
- Sannolikheten för A skrivs $P(A)$

En (så gott som) fullständig ***stokastisk modell*** kan nu sammanfattas enligt (se N Kap 3, sid 12):

1. Utfallsrummet Ω är definierat
2. För varje $A \subseteq \Omega$ kan $P(A)$ anges

Lite mängdlära

- Låt gemener e_1, e_2 , osv. beteckna **element**
- Låt versaler A, B, Ω beteckna **mängder** av element
 - Klamrar brukar användas $\{\cdot\}$
 - Ex. $A = \{1,2\}, B = \{4\}$
- Om e_i **tillhör** A , dvs. ligger i A , skriver vi $e_i \in A$
 - Ex. $1 \in \{1,2\}$ men $3 \notin \{1,2\}$
- Om A är en **delmängd** av B skriver vi $A \subset B$
 - Delmängd betecknas $\subseteq$
 - Strikt delmängd betecknas $\subset$
 - Ex. $A = \{1,2\} \subset \Omega = \{1,2,3,4,5,6\}$

Lite mängdlära, forts.

Antag att $\Omega = \{1,2,3,4,5,6\}$ och att $A = \{1,2\}$, $B = \{2,3,4\}$ och $C = \{3\}$

- **Komplementet** till en mängd är allt som inte ingår i mängden och betecknas med $\bar{A}$, A' , A^c , A^* , $\complement A$
 - Ex. $A^c = \{3,4,5,6\}$
- **Unionen** av mängder betecknas med $\cup$
 - Ex. $A \cup B = \{1,2,3,4\}$, $B \cup C = B$
- **Snittet** av mängder betecknas med $\cap$ (finns i både A och B)
 - Ex. $A \cap B = \{2\}$

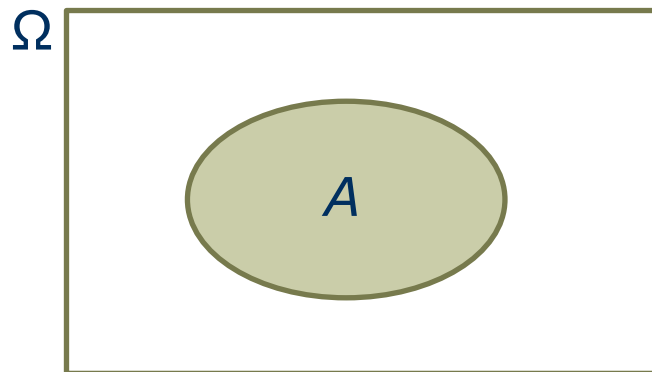
Lite mängdlära, forts.

- **Tomma mängden** är delmängden till Ω som inte innehåller några element alls. Betecknas med $\emptyset$.
- Två mängder är **disjunkta** (oförenliga) om snittet är tomt
 - Ex. $A = \{1,2\}$ och $C = \{3\}$
 $A \cap C = \emptyset$
- Vad är komplementet till Ω ?

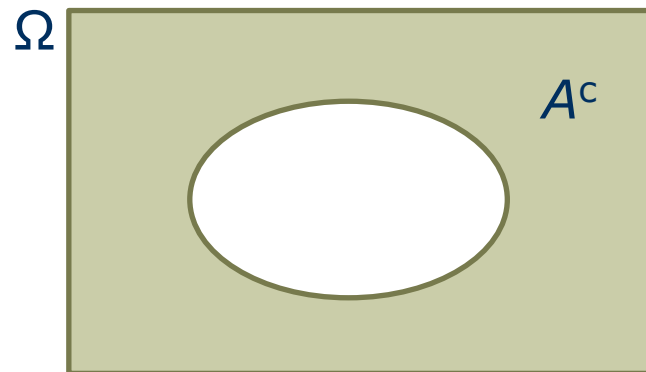
Venndiagram

- Ett sätt att illustrera mängder och deras relationer
- Låt det skuggade området beteckna mängden i fokus

$A \subset \Omega :$



$A^c \subset \Omega :$



Övning

Beskriv relationen mellan A och B i resp. Venndiagram och ange följande mängder i varje fall:

a) $A \cup B$

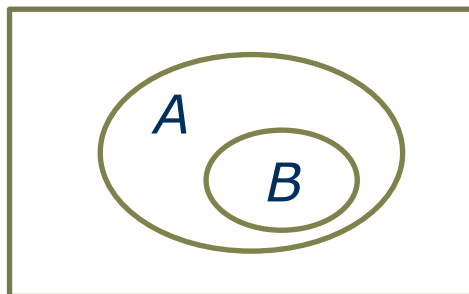
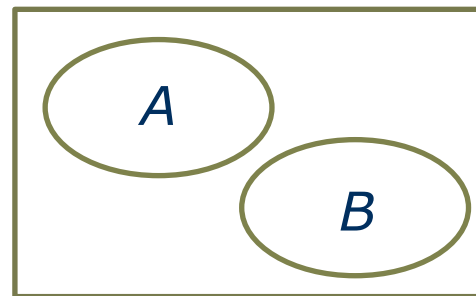
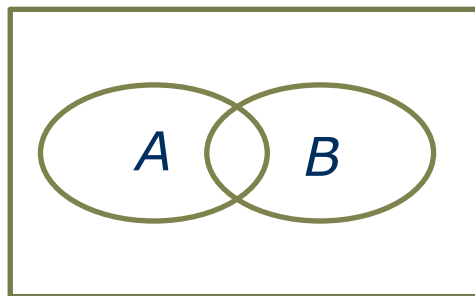
b) $A \cap B$

c) $A^c \cup B$

d) $A^c \cap B$

e) $(A \cup B)^c$

f) $(A \cap B)^c$



Vad är en sannolikhet?

- Sannolikheten $P(e)$ är ett tal
- Det ska i någon intuitiv mening säga hur troligt något är
- Intuitivt:
 - Om sannolikheten är noll kan det väl inte inträffa?
 - Om sannolikheten är 100 % måste det väl inträffa?
- Vad en sannolikhet *egentligen* är kan förklaras på lite olika sätt:
 - Frekventistiskt
 - Klassiskt
 - Subjektiv

Frekventistisk tolkning

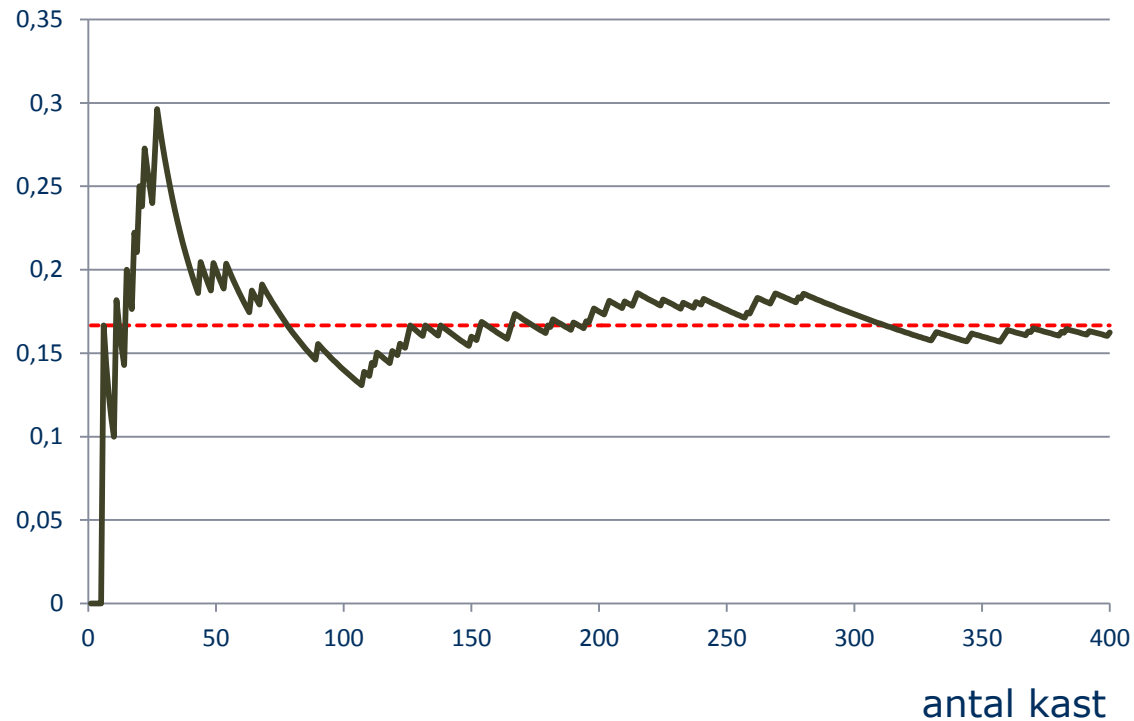
En intuitiv tolkning av begreppet sannolikhet är **hur ofta** vi tror att det ska inträffa (N 3.5.2):

- Vi utför experimentet upprepade gånger och räknar antalet gånger utfallet blev A .
- Efter n gånger noterar vi n_A lyckade utfall.
- Kvoten n_A/n är den relativa frekvensen för utfall A .
- Kvoten tenderar att stabiliseras när n ökar

$$n_A/n \rightarrow P(A) \text{ då } n \rightarrow \infty$$

Frekventistisk tolkning, forts.

$\frac{\text{antal 2:or}}{\text{antal kast}}$



Klassisk tolkning

Man kan också utgå ifrån (när så är möjligt) en jämförelse av "storleken" av delmängden A relativt "storleken" av Ω .

- Antag att man kan definiera Ω som en mängd av elementarhändelser, alla lika troliga.
- Räkna antal element som tillhör A .
- Jämför med antal element totalt.

$$\text{antal}(A) / \text{antal}(\Omega) = P(A)$$

Jämför med Ex 3.7 sid 13 i N

Klassisk tolkning

- Idé med Venndiagram

$A \subset \Omega :$



$$\text{storlek}(A) = 24$$

$$\text{storlek}(\Omega) = 60$$

$$P(A) = 24/60 = 0,4$$

Subjektiv sannolikhet

Sannolikhet kan också tolkas som grad av (personlig) tilltro.

Särskilt när de frekventistiska eller klassiska principerna inte fungerar (sannolikheten att Djurgården förlorar på lördag)

Kallas **subjektiva sannolikheter**.

- Sannolikheten bestäms av hur mycket du är villig att satsa och den vinst du kan kamma hem

$$\text{insats}/\text{total vinst} = P(A)$$

Övning 3.13 sid 18 i N

En axiomatisk teori

Kolmogorovs axiom: En sannolikhet är en funktion P som tilldelar varje möjlig händelse A i ett utfallsrum Ω , ett tal $P(A)$ så att följande villkor är uppfyllda:

- $P(A) \geq 0$ **(sannolikheter är aldrig negativa)**
- $P(\Omega) = 1$
- Om $A_1, A_2, \dots, A_k$, är parvis **disjunkta** händelser i Ω , då är

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k)$$

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$



En axiomatisk teori, forts.

Utifrån axiomen kan följande härledas, dvs. bevisas vara sanna:

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- Om $A \subset B$ så gäller $P(A) \leq P(B)$
- $P(A) \leq P(\Omega) = 1$ **(sannolikheter är aldrig >1)**
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

En axiomatisk teori, forts.

- Samtliga tre synsätt (frekventistisk, klassisk, subjektiv) på vad en sannolikhet egentligen är, är förenliga med Kolmogorovs axiom.
- Kom ihåg att vi har en formell definition på vad en sannolikhet är också.
 - Sannolikheten $P(\cdot)$ är ett tal
 - Utfallsrummet Ω är väldefinierat
 - För varje $A \subseteq \Omega$ kan $P(A)$ anges
 - Kolmogorovs axiom definierar resten

Har vi en teori nu?

- **Objekt**

- enstaka och grupper av elementarhändelser (ex. händelser och hela utfallsrummet Ω)
- sannolikheter; en funktion av händelser; $P(A)$

- **Relationer**

- hur händelserna förhåller sig till varandra via t.ex. mängdläran
- hur sannolikheterna förhåller sig till händelserna genom funktionen $P(A)$ och till varandra

- **Tolkningar**

- frekventistiska, klassiska, subjektiva tolkningar

Övning

Antag att $P(A) = 0,3$, $P(B) = 0,5$ och $P(C)=0,7$. Antag att A och B är disjunkta. Beräkna

1. $P(A^c) =$
2. $P(A \cap B) =$
3. $P[(A \cap B)^c] =$
4. $P(A \cup B) =$
5. $P(A \cup C) =$ [kuggfråga]
6. Kan $B \subset C$? Kan $C \subset B$?
7. Vad är det minsta värdet som $P(A \cup B \cup C)$ kan vara?
8. Vad är det största värdet som $P(A \cup B \cup C)$ kan vara?