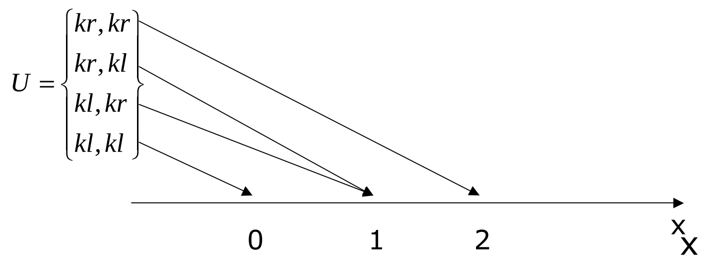
F10

Slumpvariabler
Väntevärden
Slutsatser från urval till population

Slumpvariabler (Stokastiska variabler)

- En **slumpvariabel** är en funktion från utfallsrummet till tallinjen
  - Ex kast med ett mynt 2 ggr
  - X=antalet krona
  - X={0,1,2}

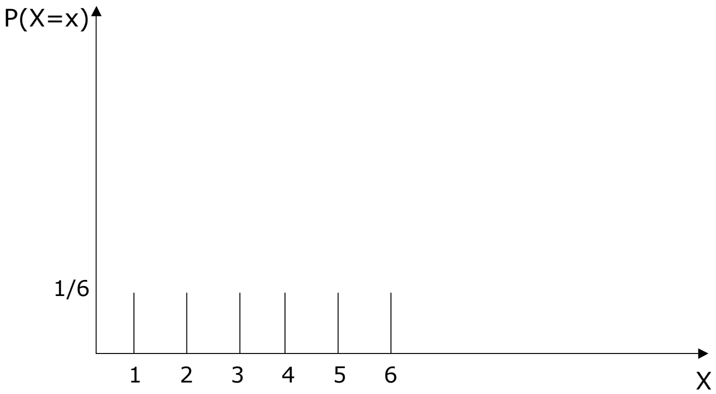


Sannolikhetsfördelning för en diskret variabel

- Ex X=antalet prickar vid kast med tärning
- X={1,2,3,4,5,6}
- X är en diskret slumpvariabel
- Sannolikhetsfördelningen för X

x	P(X=x)	$P(X \leq x) = F_x(x)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6=1

Sannolikhetsfördelning för en diskret variabel



Väntevärde

- **Väntevärdet** för X
  (Expectation, expected value)
  - Vid ”oändligt” många upprepningar av försöket
  - E(X)=μ_X

$$E(X) = \sum_{i=1}^{k} x_i \times P(X = x_i)$$

    - k=antalet variabelvärden
  - Jmfr medelvärdet i ett grupperat material

Exempel:kast med tärning

- Förväntat antal prickar vid kast med tärning:

$$E(X) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3{,}5$$

Varians

$$V(X)=\sigma_x^2=E\left[(X-\mu_x)^2\right]=\sum_{i=1}^{k}(x_i-\mu_x)^2\times P(X=x_i)$$

- Ex μ_X=3,5

$$V(X)=(1-3{,}5)^2\times\frac{1}{6}+(2-3{,}5)^2\times\frac{1}{6}+(3-3{,}5)^2\times\frac{1}{6}+(4-3{,}5)^2\times\frac{1}{6}+(5-3{,}5)^2\times\frac{1}{6}+$$

$$+(6-3{,}5)^2\times\frac{1}{6}=\frac{35}{12}$$

- σ_X=**standardavvikelsen** för X

$$\sigma_x=\sqrt{V(X)}=\sqrt{\sigma_x^2}$$

$$\sigma_x=\sqrt{\frac{35}{12}}\approx 1{,}7$$

Urval från en population

- μ_X i en population
- Ex μ_X=medellängd
  - Urval från en population
  - Bestäm en **skattning** av μ_X
  - Använd medelvärdet i urvalet som skattning

    $$\overline{X}=\frac{\sum_{i=1}^{n}x_i}{n}$$

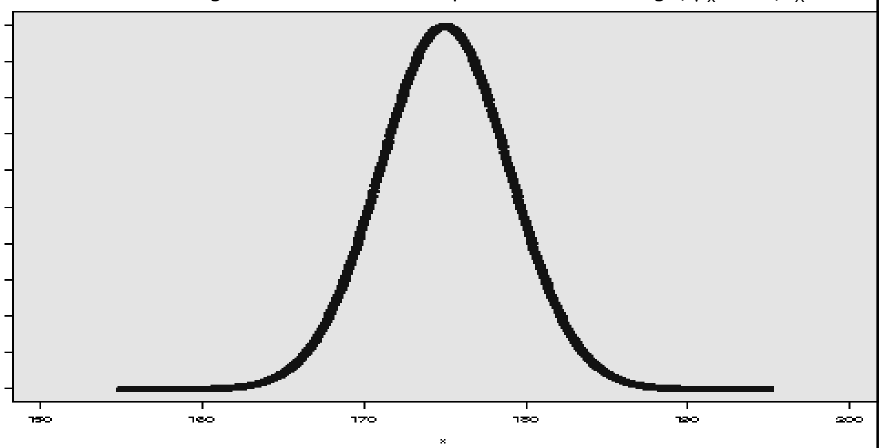
  - ***<u>Det skattade medelvärdet "X-bar" är en slumpvariabel</u>***

Kontinuerlig slumpvariabel

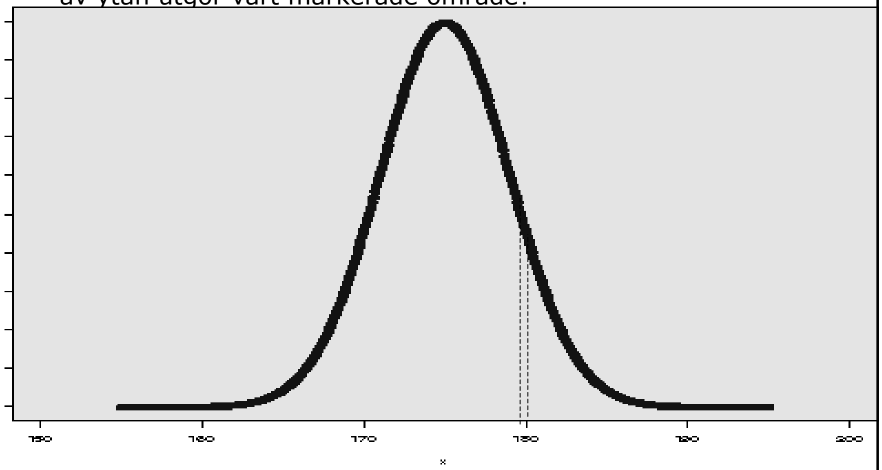
- X kan anta oändligt många värden i ett intervall
- Då X kan anta ett oändligt antal värden i ett intervall säger vi att sannolikheten att X antar ett specifikt värde är noll.
- Vi tittar istället på sannolikheten att X antar värden inom ett intervall
  - Ex sannolikheten att en slumpmässigt utvald person är 190 cm lång? Vi tittar inte på sannolikheten att en person är *exakt* 190 cm, utan på sannolikheten att personen är mellan 189,5 och 190,5.

Sannolikhetsfördelning för en kontinuerlig variabel

- En kontinuerlig normalfördelad slumpvariabel. Här: längd, μ_x=175,σ_X=4



- P(179,5 ≤X≤180,5)=? Ytan under kurvan=1. Hur stor del av ytan utgör vårt markerade område?



Normalfördelning

- Viktigaste sannolikhetsfördelningen för en kontinuerlig variabel
- Kurvan beräknas med formeln (behöver ej kunna denna formel!):

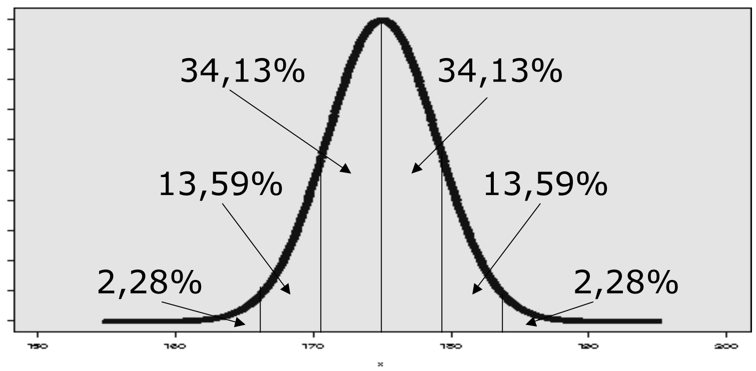
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-\left(\frac{x-\mu}{2\sigma^2}\right)^2}$$

- Kurvan bestäms av medelvärdet och standardavvikelsen – Det finns alltså ett oändligt antal normalfördelningar

$$X \in N(\mu_x, \sigma_x)$$

Varför behöver vi kunna något om normalfördelningen?

- Ex längd: $X \in N(175,4)$
  - Dvs X är normalfördelad med medelvärde 175 och standardavvikelse 4



Slumpmässigt urval

- Slumpmässigt urval
  - Population: N st elemet
  - Urval (stickprov): n st element
  - Varje element i populationen ska ha en känd sannlikhet >0 att bli utvald.
  - Populationen har ett okänt medelvärde μ_X
  - Urvalet:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Exempel

- En population består av N = 4 personer och variabeln X = vikt. Observationerna är 77, 83, 90 och 90 kg. Populationsmedelvärdet μ blir 85 kg.
- Vi drar nu ett obundet slumpmässigt urval (OSU) med återläggning om n = 2 personer. Det finns då 4 x 4 = 16 olika urval som alla har samma sannolikhet, 1/16. Vi antar nu att vi inte vet populationsmedelvärdet utan vi vill skatta det från den information vi får från urvalet. Vi väljer stickprovsmedelvärdet som en skattning för μ. Denna skattning är en slumpvariabel och vi kan nu bestämma alla tänkbara värden på denna variabel utifrån de olika möjliga urvalen.

Sannolikhetsfördelningen för X-bar

- Vad vet vi om sannolikhetsfördelningen för X-bar?

$$E(\bar{X}) = ?$$

$$V(\bar{X}) = ?$$

- Hur ser fördelningen för X-bar ut?

Väntevärdesriktig skattning

- Vi kan för varje urval också bestämma urvalsfelet $\overline{X} - \mu$ dvs avvikelsen mellan medelvärdet i stickprovet och populationen
- Om urvalsfelet i genomsnitt är 0 så är skattningen väntevärdesriktig, vi har inget systematiskt fel

Centrala gränsvärdesatsen, CGV

- Oavsett hur populationen ser ut gäller att fördelningen för X-bar närmar sig en normalfördelning om vi drar ett tillräckligt stort stickprov av n oberoende observationer
- ”Tillräckligt stort”. Tumregel: n>30
- Kravet på stickprovsstorleken är beroende av hur populationen ser ut:
  - ju mer symmetrisk, desto bättre kan fördelningen för X-bar approximeras med en normalfördelning för en given stickprovsstorlek.
  - Om populationen är normalfördelad är också fördelningen för X-bar normalfördelad för alla storlekar på n, även om n är mycket litet.