Normalfördelning

Normalfördelning är ett specialfall av s.k. sannolikhetsfördelning
(som behandlas mer allmänt i matte 4)

Det innebär att en statistisk variabel med en viss förutsägbar sannolikhet kommer att hamna i ett visst intervall.

Vid normalfördelning kommer värdena med stor sannolikhet att hamna närmare medelvärdet för att med allt mindre sannolikhet hamna längre därifrån.

Konkretiserande exempel: Mätning av längd hos människor.

I) 250 slumpvisa människor mäts.
Tabellen till höger visar resultatet.

	171,6	157,1	174,5	158,8	176,0	171,5	177,5	173,6
158,5	165,6	174,5	158,0	161,3	183,7	172,6	170,3	175,7
164,0	168,8	168,1	177,8	162,4	186,5	174,3	148,6	165,2
172,2	188,9	180,2	194,3	185,8	157,9	174,1	168,1	154,3
174,7	173,7	173,9	159,1	169,6	175,8	156,4	172,0	170,1
186,3	173,7	175,7	170,3	176,3	175,5	158,5	172,8	151,1
161,3	182,2	173,5	172,3	162,3	178,0	178,7	168,8	167,5
178,3	165,8	163,6	165,9	155,5	174,8	188,3	178,8	165,5
167,5	184,2	181,5	178,9	164,8	164,3	167,5	173,9	174,7
175,0	172,2	183,4	180,3	158,5	174,6	177,6	179,1	174,9
175,9	160,2	162,6	180,9	159,0	168,5	178,7	165,2	171,2
167,7	177,0	165,3	169,9	175,8	178,3	163,5	174,5	
159,6	159,9	162,1	169,3	183,5	160,5	167,2	176,7	
183,6	183,4	175,8	159,3	163,8	188,5	157,7	171,8	
165,0	177,6	171,2	169,0	162,9	163,3	176,0	173,4	
153,5	157,1	175,1	172,8	176,1	182,0	172,5	166,5	
161,2	164,8	184,9	172,7	161,3	161,6	173,4	188,7	
172,3	165,5	153,4	168,5	185,2	179,9	190,8	174,3	
165,1	174,4	171,3	169,4	178,8	179,9	178,5	177,9	
167,7	182,3	165,2	167,9	167,4	171,4	171,5	188,2	
167,5	172,7	176,1	184,8	176,4	172,2	169,3	179,7	
173,2	181,0	166,1	169,3	186,9	170,1	172,1	166,8	
174,8	168,6	179,4	157,1	171,9	163,6	161,6	159,1	
180,1	180,4	178,7	158,5	172,2	179,1	175,7	167,8	
184,9	159,6	147,0	172,8	182,5	182,7	162,0	171,4	
161,4	160,3	166,4	172,0	171,6	164,2	169,3	168,3	
162,8	167,9	181,5	179,0	175,5	168,2	170,8	158,4	
178,7	158,7	170,4	171,9	169,7	180,2	177,3	171,3	
166,9	163,2	182,2	168,9	172,9	173,8	172,2	175,9	
169,4	179,4	168,6	177,1	172,5	175,3	180,1	175,9	

II) För att strukturera denna typ av data delas det in i intervall, dvs mindre grupper där antalet i varje grupp räknas.

exempelvis - det fanns 20 st värden mellan 165,9 cm och 168,3 cm

11 st mellan 180,1 och 182,5 cm

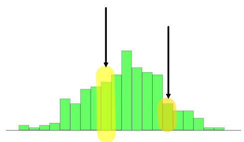
o s v

Intervall	Antal
147,0 - 149,4	2
149,4 - 151,7	1
151,7 - 154,1	2
154,1 - 156,5	3
156,5 - 158,8	13
158,8 - 161,2	11
161,2 - 163,6	17
163,6 - 165,9	18
165,9 - 168,3	20
168,3 - 170,6	23
170,6 - 173,0	33
173,0 - 175,4	26
175,4 - 177,7	24
177,7 - 180,1	23
180,1 - 182,5	11
182,5 - 184,8	8
184,8 - 187,2	8
187,2 - 189,5	5
189,5 - 191,9	1
191,9 - 194,3	1

III) Om dessa grupper ritas ut som staplar fås diagrammet till höger.

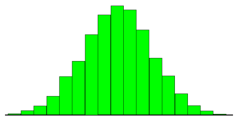
De två exempeln är markerade.

Även om det är relativt få värden är det tydligt att det är högre staplar i mitten, och att staplarnas längd glesnar mot kanterna

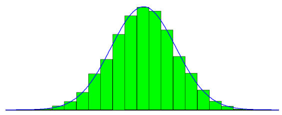


IV) Samlas ännu fler värden in blir detta mönster ännu tydligare.

När en variabel beter sig på detta sätt kallas variabeln normalfördelad



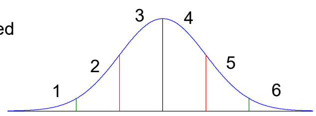
V) För normalfördelad data finns en matematisk modell, inritad som den blå grafen, som antyder hur hög respektive stapel "borde" varit enligt teorin.



Modellen säger följande:

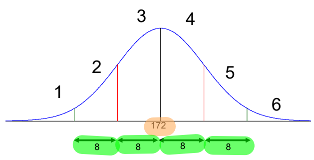
Högsta stapeln blir alltid talens medelvärde , och därefter minskar de enligt ett mönster som beror på talens standardavvikelse (se 4.2 vid oklarheter)

Datat delas in 6 stycken fack med hjälp av 5 streck
Mellan två streck är det alltid en standardavvikelse.

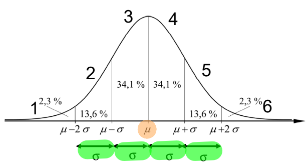


Räknar man på datat från längderna fås att

medelvärdet är 172 cm
standardavvikelsen är 8 cm



På formelbladet kan man utläsa den teoretiska fördelningen hur stor andel som ska finnas i respektive fack:

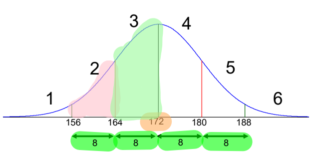


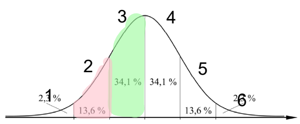
Notera att symbolerna längst ned är allmänna symboler för

medelvärde = μ och standardavvikelse = σ

I fallet med medellängden kan man t.ex. säga att 47,7 % av längderna är mellan 156 cm och 172 cm

(Det motsvarar fack 2 + 3 = 13,6 % + 34,1 % = 47,7 %)

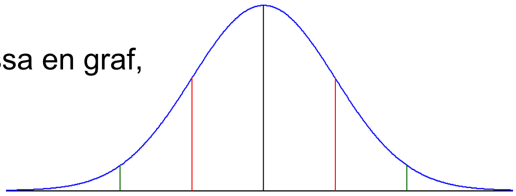




Exempel D1: Vikten hos en viss sorts knubbig (och påhittad) hamster s.k. tjocketroll är normalfördelad med medelvärdet 85 g, och standardavvikelsen 6 g.

Av 500 stycken tjocketroll, hur många väntas väga mer än 79 g?

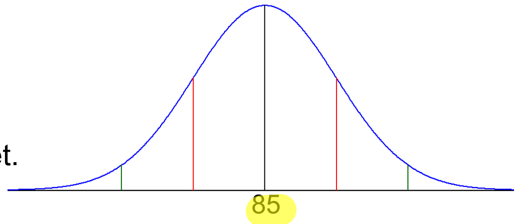
Lösning: Så fort det handlar om normalfördelning, börja med att skissa en graf, och dela in den i de 6 facken.



Fyll i siffrorna på axeln nedanför.

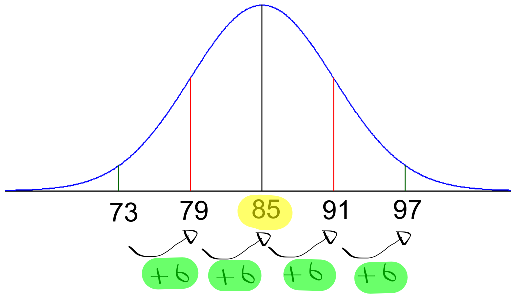
Börja med medelvärdet = 85 g.

Detta motsvarar alltid mittenstrecket.



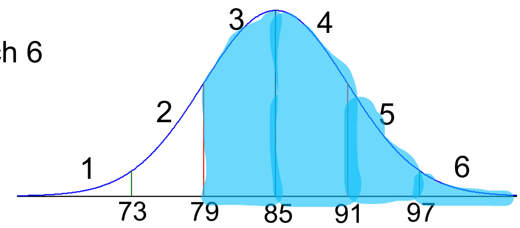
Använd informationen om standardavvikelse = 6 g till att fylla i resten

(det skiljer hela tiden 6 mellan två streck)



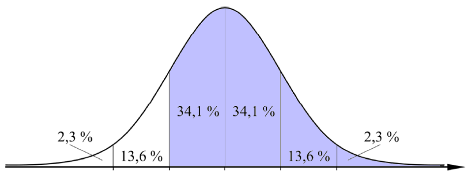
Markera de fack som frågan gäller

"Större än 79 g" ⟶ Fack 3, 4, 5 och 6



"Översätt" till procentsiffror med hjälp av formelbladet (eller Geogebra*)

34,1 + 34,1 + 13,6 + 2,3 = 84,1 %



* Se den skriftliga filen för detaljer om hur man gör i Geogebra

Med andelen känd kan frågan besvaras.

84,1 % av 500 = 0,841 · 500 ≈ 421 st

Svar: Av de 500 tjocketrollen väntas 421 stycken väga mer än 79 g

Exempel D2: Vikten på ett paket mjöl är normalfördelad med medelvärdet 2000 g och standardavvikelsen 50 g.

Anta att ett paket plockas ut slumpartat, och att detta paket har vikten V g

Vilket/Vilka av följande påståenden är då sanna om V?

A Det är 84 % sannolikhet att $V \geq 2050$

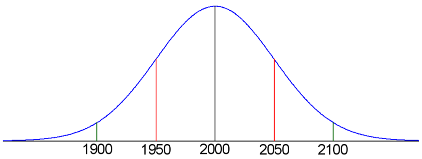
B Det är 50 % sannolikhet att $1950 \leq V \leq 2050$

C Det är 84 % sannolikhet att $V \leq 2050$

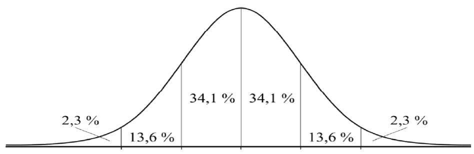
D Det är mindre än 3 % sannolikhet att $V \leq 2100$

E Det är 95 % sannolikhet att $1900 \leq V \leq 2100$

Lösning: Skissa grafen, och fyll i de 6 facken. Markera medelvärdet i mitten, och använd standardavvikelsen för att fylla i resten

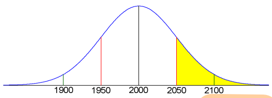


Använd formelbladet (eller Geogebra) för att jämföra siffrorna



Titta på varje påstående för sig.

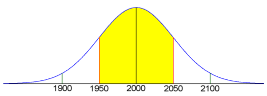
A Det är 84 % sannolikhet att $V \geq 2050$



13,6 + 2,3 = 15,9 %

nej

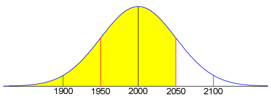
B Det är 50 % sannolikhet att $1950 \leq V \leq 2050$



34,1 + 34,1 = 68,2 %

nej

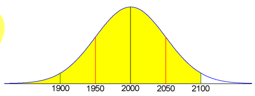
C Det är 84 % sannolikhet att $V \leq 2050$



50 + 34,1 = 84,1 %

ja

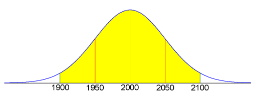
D Det är mindre än 3 % sannolikhet att $V \leq 2100$



100 - 2,3 = 97,7 %

nej

E Det är 95 % sannolikhet att $1900 \leq V \leq 2100$



13,6 + 13,6 + 34,1 + 34,1 = 95,4 %

ja

Svar: C och E