

## 16. Språkstatistik

I grundkursen lärde ni er hur olika krypton är beskaffade. Ni fick lära er att kryptera och dekryptera på olika sätt. I slutet av kursen fick ni nosa på kunskap hur man forcerar, knäcker krypton utan kunskap om hur de är gjorda. I denna del av kursen fortsätter ni att lära er forcering.

Ett nödvändigt hjälpmedel är språkstatistik. Man måste först av allt veta hur vanlig en viss bokstav är i klartexten. Därför börjar ni med att göra en enkel pinnstatistik. Skriv först alfabetets bokstäver lodrätt på ett rutat eller linjerat papper. Sedan tar ni fram en svensk text från en tidning eller en bok. Börja var som helst i texten, gå igenom texten bokstav för bokstav, hoppa över mellanslagen och sätt en "pinne" efter respektive bokstav. Sluta efter 80 bokstäver. Strunta i om ni slutar mitt i ett ord eller inte.

Antagligen kommer er statistiska uppställning att se ut ungefär som i tabellen till höger:

Det skall finnas 80 stycken "pinnar", en pinne för varje bokstav i den valda texten. I exemplet som vi visat här, ser ni också att det finns sju 'a', ett 'b' och så vidare och att det inte finns någon bokstav som har förekommit fler än 7 gånger.

Bokstäverna q, w, x, y och z är ovanliga i svenska språket. Därför är det nog tomt, "nollor", även i er språkstatistik. Vilka andra iakttagelser gör ni?

a	
b	
c	
d	
e	
f	
g	
h	
i	
j	
k	
l	
m	
n	
o	
p	
q	
r	
s	
t	
u	
v	
w	
x	
y	
z	
å	
ä	
ö	



## ÖVNING 16A

Gör fyra statistiker på avsnitt om 80 tecken svensk text och jämför resultaten.

Om man gör språkstatistik på en stor mängd svensk text finner man att bokstaven 'e' är vanligast, sedan kommer 'a' och därefter 'r' 'n' 't' 's' 'i' 'l' 'd' och 'o' i ordning. Ni såg i övning 16A att detta inte alltid stämmer för korta textavsnitt. I avsnitt 24 finns olika språkstatistiker för svenska och engelska. De är nödvändiga hjälpmedel när ni skall fördjupa er ännu mer.

Tänk på att använda små bokstäver för klartexter. För kryptotexter används stora bokstäver. Ni såg i avsnitten 10 och 11 i grundkursen att det är praktiskt att skilja klartext från kryptotext på detta sätt.

## ÖVNING 16B

Vad är det för märkligt med klartextstatistiken till höger här på sidan? Hur kan det ha det blivit så?

a	
b	
c	
d	
e	
f	
g	
h	
i	
j	
k	
l	
m	
n	
o	
p	
q	
r	
s	
t	
u	
v	
w	
x	
y	
z	
å	
ä	
ö	

