



Blandning av prov

$$\begin{matrix} n_x & \bar{x} & s_x & \text{prov 1} \\ n_y & \bar{y} & s_y & \text{prov 2} \end{matrix}$$

$$\bar{x}_B = \frac{n_x \cdot \bar{x} + n_y \cdot \bar{y}}{n_x + n_y}$$

$$s_B^2 = [(n_x - 1) \cdot s_x^2 + n_x \cdot \bar{x}^2 + (n_y - 1) \cdot s_y^2 + n_y \cdot \bar{y}^2 - n \cdot \bar{x}_B^2]$$

OBS! Det är inte detsamma som "pooled variance"



Lektion 1

Grundbegrepp och beskrivande statistik



Rangkorrelation

Elev	A	B	C	D	E	F	G	H	I	J
Matte	7	4	3	10	6	2	9	8	1	5
Musik	5	7	3	10	1	9	6	2	8	4

$$r = \frac{S_{xy}}{\sqrt{S_{yy} \cdot S_{xx}}} \quad -1 \leq r \leq 1$$

$$r = \frac{c_{xy}}{s_x \cdot s_y} \quad c_{xy} = \frac{1}{n-1} S_{xy} \quad (\text{kovarians})$$

som förut !



Innehåll

- Blandning av prov
- Rangkorrelation
- Beskrivande statistik - uppgifter

Minitab !



Klassindelade data

I en undersökning av inkomsterna för viss grupp av löntagare visade sig månadsinkomsterna för 200 personer vara fördelade enligt nedanstående klassindelade material:

Klassmitt (kr)	Frekvens
17500	5
22500	19
27500	79
32500	33
37500	18
42500	15
47500	10
52500	5
57500	4
62500	5
67500	3
72500	4

- Beräkna medelvärde och standardavvikelse.
- Rita upp ett histogram och kumulativt histogram samt uppskatta med hjälp av det senare medianinkomsten.
- För att ange "medelinkomst" används ibland aritmetiska medelvärdet, ibland medianen. Blir dessa värden lika stora (ungefär)?
- Verkar det rimligt att anta att materialet kommer från en normalfördelning? Om inte, föreslå någon fördelning som verkar lämpligare?



Uppgifter

- 10.1 För 22 personer beräknades kariesindex, dvs antalet kariesangripna ytor bland de 100 tandytor man får om man bortser från visdomständerna samt den linguala ytan (= ytan mot tungan) på alla tänder. Resultat:

41	47	66	73	48	52	49	54	61	62	47
52	65	61	69	31	54	53	50	47	36	69

Beräkna medelvärde, median, varians, standardavvikelse samt variationskoefficient i detta material.

- 10.2 I en hamn räknade man för varje dygn under en månad ankommande lastfartyg. Man erhöll följande antal:

3	3	2	4	2	5	2	7	2	1	2	0	3	3	4
1	3	9	3	2	3	4	1	2	1	4	7	6	3	1

- Gör en frekvenstabell och rita upp ett stolpdiagram.
- Beräkna medelvärde och standardavvikelse.



Korrelation



I en kommun undersöktes sambandet mellan taxeringsvärde (x_i) och försäljningspris (y_i) av villor. För 10 villor hade man följande siffror (tusental kronor):

Taxeringsvärde	675	735	1200	560	550	980	780	600	750	840
Försäljningspris	910	990	1400	710	680	1300	950	800	910	1000

Beräkna korrelationskoefficienten mellan taxeringsvärde och försäljningspris.

- 10.3 Vid en industri har man för 50 batterier noterat livslängderna (livslängd = den tid ett batteri ger erforderlig spänning). Följande tabell ger livslängderna i dagar ordnade i storleksordning.

87	105	108	120	142	151	155	155	161	162
169	173	174	183	185	186	186	192	193	196
199	205	207	211	215	217	217	222	224	226
227	230	231	231	237	242	244	246	251	258
263	267	278	286	294	312	338	341	362	390

- Klassindela materialet med lämplig klassbredd och rita upp histogram. (Dessa blir skattningar av täthetsfunktionen respektive fördelningsfunktionen för den stokastiska variabeln X = livslängden för ett batteri).
- Beräkna stickprovets medelvärde och standardavvikelse. (Dessa blir skattningar av $E(X)$ respektive $D(X)$.)

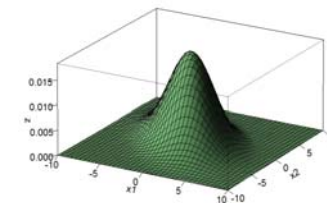
- 10.4 I två material x_1, x_2, \dots, x_{10} respektive y_1, y_2, \dots, y_5 beräknade man medelvärdena och standardavvikelserna varvid man fick

$$\bar{x} = 5313, \quad s_x = 5.2 \quad \text{respektive} \quad \bar{y} = 5309, \quad s_y = 3.0.$$

Om de femton talen hade betraktats som ett enda material, vilket medelvärde och vilken standardavvikelse hade man då fått?

- 10.5 Man beräknade med en dator medelvärdet och standardavvikelsen i ett material med 800 observationer. Man fick därvid $\bar{x} = 9.496$, $s = 0.345$. Vid kontroll av databasen visade det sig att en observation som skulle varit 9.56 hade lästs in som 1.56. Vilket medelvärde och vilken standardavvikelse hade datorn fått om det nämnda felet ej hade funnits?

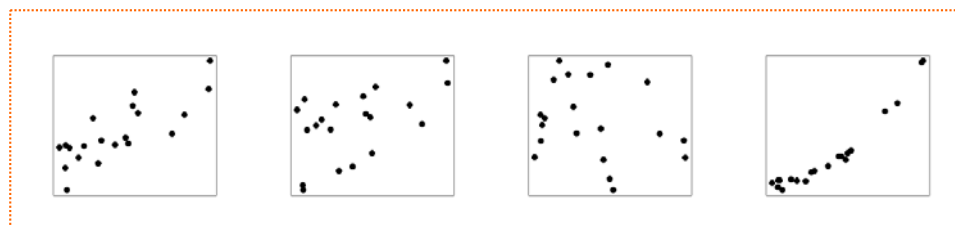




Normalfördelningen

Simulera 1000 observationer från en normalfördelning med väntevärde $\mu = 3$ och standardavvikelse $\sigma = 1$. Beräkna medelvärde och stickprovsstandardavvikelse av erhållna observationer och jämför med fördelningens väntevärde och stickprovsstandardavvikelse. Rita ett histogram med 30 klasser. Rita in normalfördelningens täthetsfunktion i histogrammet. För att skalan skall bli jämförbar måste täthetsfunktionen multipliceras med histogrammets totala area, som i detta fall är 1000 gånger klassbredden om alla klasser har samma bredd.

Korrelationskoefficient



$$r_1 = -0.23 \quad r_2 = 0.536 \quad r_3 = 0.798 \quad r_4 = 0.954$$

Vilken korrelationskoefficient tillhör vilken bild ?