

Lösningsförslag till statistiktenta 061201

UPPGIFT 1

a)

	A	Icke A	
B	0,3	0,3	0,6
Icke B	0,1	0,3	0,4
	0,4	0,6	

$$\Pr(A | B) = \frac{\Pr(A \text{ och } B)}{\Pr(B)} = \frac{0,3}{0,6} = 0,5$$

b)

	A	Icke A	
B	0,4*0,6=0,24	0,36	0,6
Icke B	0,16	0,24	0,4
	0,4	0,6	

$$\Pr(A | B) = \frac{\Pr(A \text{ och } B)}{\Pr(B)} = \frac{0,24}{0,6} = 0,4 = \Pr(A)$$

c)

	A	Icke A	
B	0	0,3	0,6
Icke B	0,1	0,3	0,4
	0,4	0,6	

$$\Pr(A | B) = \frac{\Pr(A \text{ och } B)}{\Pr(B)} = \frac{0}{0,6} = 0$$

UPPGIFT 2

a) X är hypergeometriskt fördelad

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - \frac{\binom{5}{0} \binom{4}{3}}{\binom{9}{3}} = 1 - \frac{\left(\frac{5!}{0! \cdot 5!}\right) \left(\frac{4!}{3! \cdot 1!}\right)}{\left(\frac{9!}{3! \cdot 6!}\right)} = 1 - \frac{1 \cdot 4}{\frac{9 \cdot 8 \cdot 7 \cdot 6!}{3 \cdot 2 \cdot 1 \cdot 6!}} = 1 - \frac{4}{3 \cdot 4 \cdot 7} = 1 - \frac{1}{21} = 0,952$$

b) X är binomialfördelad med n=3 och p=5/9=0,556

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - \binom{3}{0} \cdot 0,556^0 \cdot 0,444^3 = 1 - 0,444^3 = 0,912$$

UPPGIFT 3

Homogenitetstest

H₀: Stickproven kommer från populationer med samma medianvärde

H₁: Stickproven kommer inte från populationer med samma medianvärde

Förutsättningar: OSU, oberoende mätningar, alla E minst 5

Obs frekv:	Stickprov 1	Stickprov 2	Stickprov 3	
Under 40,5	3	11	7	21
Över 40,5	11	3	7	21
	14	14	14	42

Förv frekv	Stickprov 1	Stickprov 2	Stickprov 3	
Under 40,5	7	7	7	21
Över 40,5	7	7	7	21
	14	14	14	42

Stickprov		O	E	O-E	(O-E) ²	(O-E) ² /E
1	Under 40,5	3	7	-4	16	2,285714
1	Över 40,5	11	7	4	16	2,285714
2	Under 40,5	11	7	4	16	2,285714
2	Över 40,5	3	7	-4	16	2,285714
3	Under 40,5	7	7	0	0	0
3	Över 40,5	7	7	0	0	0
		42	42			9,142857

$$\text{Frihetsgrader} = (c-1)(r-1) = (3-1)(2-1) = 2$$

$$X^2_{\text{obs}} = 9,14 > 7,815 = X^2_{3;0,05}$$

P < 5% => H₀ förkastas, H₁ antas.

Slutsats: Resultatet tyder på att stickproven inte kommer från populationer med samma medianvärde.

UPPGIFT 4

a)

Variabeln *antal rätta svar* är en kvantitativ variabel. Variabeln antar bara heltalsvärden mellan 0-122 och är därmed diskret. Skalnivån är kvotskala.

b)

Antal rätta svar	Klassmitt	Antal provdeltagare	Kumulerad frekvens	
x	m	f	F	m*f
00-32	16	699	699	11184
33-38	35,5	1228	1927	43594
39-41	40	939	2866	37560
42-44	43	1059	3925	45537
45-47	46	1294	5219	59524
48-50	49	1470	6689	72030
51-54	52,5	2042	8731	107205
55-58	56,5	2250	10981	127125
59-62	60,5	2369	13350	143324,5
63-66	64,5	2282	15632	147189
67-70	68,5	2220	17852	152070
71-74	72,5	2141	19993	155222,5
75-78	76,5	1944	21937	148716
79-82	80,5	1750	23687	140875
83-87	84,5	1878	25565	158691
88-91	89,5	1262	26827	112949
92-96	93,5	1255	28082	117342,5
97-100	98,5	713	28795	70230,5
101-104	102,5	521	29316	53402,5
105-108	106,5	305	29621	32482,5
109-122	115,5	166	29787	19173
		29787		1955427

$$\mu_x = \frac{\sum_{i=1}^K f_i m_i}{N} = \frac{1955427}{29787} = 65,65$$

Medianen är den 50:e percentilen. Det ska vara observation nr $0,5 \cdot 29787 = 14893,5$ i storleksordningen, vilken återfinns i klassen 63-66 antal rätta svar.

$$P_i = x_0 + w \frac{(i/100) \cdot n - kf_0}{f} = 63 + 3 \frac{0,5 \cdot 29787 - 13350}{2282} = 65,03$$

c)

Histogram, lådagram(boxplot) eller dotplot fungerar bra.

d)

Man presenterar materialet i ett histogram och studerar ifall figuren ser symmetrisk ut. Man kan se ifall medelvärde och median överensstämmer och man kan se ifall avstånden mellan respektive kvartil och medianen är ungefär lika stora.

UPPGIFT 5

a)

Förutsättningar:

Slumpmässigt stickprov, oberoende mätningar

Stickprovet består av $n=1903$ observationer och förutsättningarna för normalapproximation är uppfyllda

p = andel av populationen som röstar på socialdemokraterna

Vid valet är $p=0,3499$

Vid ett senare stickprov är stickprovsandelen $\hat{p} = 0,389$

Ensidig prövning. Vi testar följande hypoteser mot varandra:

$$H_0 : p=0,3499$$

$$H_1 : p>0,3499$$

$$\text{Ett lämpligt test för detta är: } Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0,389 - 0,3499}{\sqrt{\frac{0,3499 \cdot (1 - 0,3499)}{1903}}} = \frac{0,0391}{0,0109} = 3,576$$

Vi förkastar nollhypotesen på 5%-nivån då Z_{obs} är betydligt större än Z_{krit} ($3,576 > 1,645$)

Vi kan konstatera att ökningen är statistiskt säkerställd

b)

Om vi väljer signifikansnivån till 5% så kommer det kritiska värdet att vara 1,645
standardavvikelse större än den hypotetiska populationsandelen som är 0,3499

$$\text{dvs } 0,3499 + 1,645 \cdot 0,0109 = 0,3678.$$

För stickprovsandelar som är större än 36,8% kommer vi att förkasta nollhypotesen och tro att det skett en ökning.

c)

Styrkan i testet = Sannolikheten att förkasta nollhypotesen när den är falsk, dvs

$P(p > 0,3678 \text{ då } p = 0,35)$ är ungefär samma som signifikansnivån 5%

UPPGIFT 6

a)

Ja, modellen är bra. Modellen är signifikant och förklaringsgraden är relativt hög. De ingående parameterskattningarna är signifikanta. Bostadsytan påverkar priset.

b)

För en bostad utan bostadsyta ligger grundpriset på 20,5 tusen dollar

För varje kvadratfot som bostaden ökar med så ökar priset med 18,4 dollar

c)

Ett förväntat värde på denna bostad är:

$20,5 + 0,0184 \cdot 2420 = 65,028$ tusen dollar

d)

Korrelationskoefficienten $r = \sqrt{\text{föklaringsgraden}(R\text{-sq})}$, dvs roten ur $0,561 = 0,749$

Korrelationskoefficienten tyder på ett ganska starkt linjärt samband. Ju större bostadsyta desto högre pris

e)

Nja, egentligen inte. Såvida inte några tomma tomter ingår i materialet så kan man inte uppskatta detta. En bostadsyta på noll kvadratfot innebär en obebyggd tomt, alternativt en tomt med någon annan byggnad som inte avsetts för boende.