

LÖSNINGSFÖRSLAG 2005-11-11

UPPGIFT 1

- a) X är hypergeometriskt fördelad med $N=20$, $n=10$, $S=1$ (antal vinstnummer) och $p=1/20=0,05$

$$\begin{aligned} \text{Sökt: } \Pr(X=1) &= \frac{\binom{1}{1} \binom{19}{9}}{\binom{20}{10}} = \frac{1 \cdot \frac{19!}{9!10!}}{\frac{20!}{10!10!}} = \frac{19 \cdot 18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11} = \\ &= \frac{1}{20} = \frac{1}{2} = 0,5 \\ &10 \end{aligned}$$

Med äran i behåll kan man också lösa uppgiften med ett sannolikhetsteoretiskt resonemang utifrån den klassiska sannolikhetsdefinitionen:

$A = \text{"Emma vinner"}$

$$\Pr(A) = \frac{\text{AntalGynnsammaUtfall}}{\text{AntalMöjligaUtfall}} = \frac{10}{20} = 0,5$$

- b) X är binomialfördelad med $n=10$ och $p=0,05$

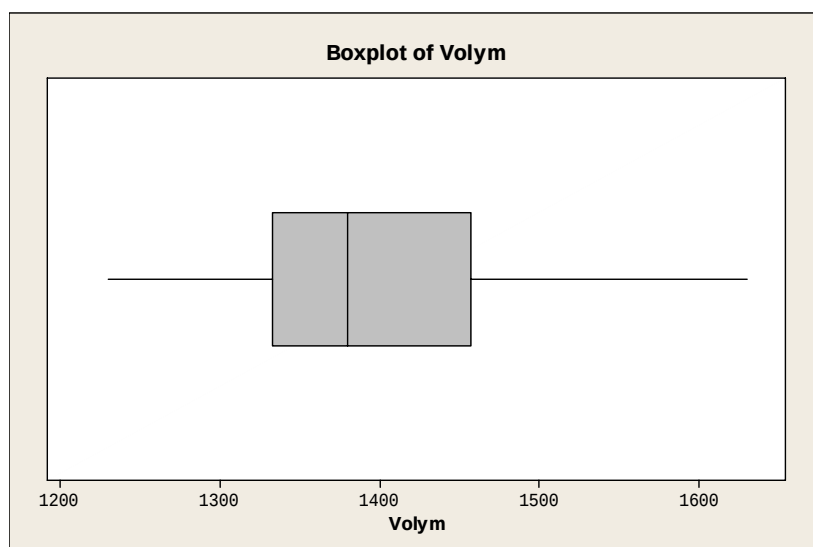
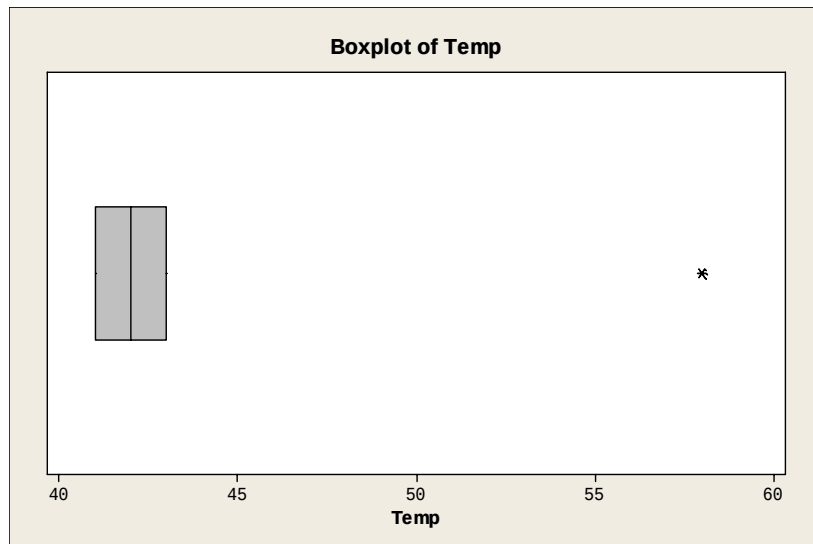
x	$\Pr(X=x)$
0	0,599
1	0,315
2	0,075
3	0,011
4	0,001
5	<0,001
6	<0,001
7	<0,001
8	<0,001
9	<0,001
10	<0,001

- c) Sannolikheten att inte vinna alls är större i alternativ b än a – det talar för att satsa alla pengar vid ett och samma spel. Förväntat antal vinster är dessutom detsamma ($n \cdot p = 0,5$), så om Emma vill ha största möjliga sannolikhet för att vinna en nalle ska hon välja det första alternativet. Dock kan hon då bara vinna en nalle – med det andra alternativet har hon hyfsad sannolikhet att åtminstone vinna två nallar.

UPPGIFT 2

Descriptive Statistics: Temp; Volym

Variable	Mean	StDev	Minimum	Q1	Median	Q3
Maximum						
Temp	44,73	6,62	41,00	41,00	42,00	43,00
58,00						
Volym	1394,6	108,4	1230,0	1333,0	1380,0	1457,5
1630,0						



Uteliggare (2 värden som är 58) för temperatur men ej volym om vi bedömer utifrån lådagrammen.

Om man istället väljer att diskutera i termer av medelvärde och standardavvikelse kan man konstatera att maxvärdena för temperatur är ganska exakt 2 standardavvikelser över medelvärdet medan maxvärdet för skallvolym är 2,2 standardavvikelser över det medelvärdet. Om man väljer att dra gränsen för extremvärde vid 2 standardavvikelser så är det alltså extremvärden även för skall-variabeln.

UPPGIFT 3

- Starkt linjärt samband mellan samtliga variabler. Såväl befolkning som annonsvolym är mycket starkt korrelerade med försäljningsvolymen, och därtill finns det ett nästan lika starkt samband mellan befolkning och annonsvolym.
- * Om man inte annonserade alls skulle försäljningen vara c:a 0,436 miljoner. Dock är denna skattning inte signifikant skild ifrån noll!
* När annonsvolymen ökar med 1 enhet, d v s 10 000 kr, ökar försäljningen i genomsnitt med 1,05 miljoner kr.
- Modellen förbättras eftersom korregerad förklaringsgrad ökar och koefficientskattningarna för båda förklarande variabler är signifikant skilda ifrån noll.

UPPGIFT 4

- 95%-igt konfidensintervall för p:

$$\hat{p} \pm 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Förutsättningar: OSU, oberoende mätningar, $np(1-p)$ minst 9 $\Rightarrow \hat{p}$ approximativt NF

$$n\hat{p}(1-\hat{p}) = 178 * 0,4 * 0,6 = 42,7 > 9$$

$$0,4 \pm 1,96 \sqrt{\frac{0,4(1-0,4)}{178}} \approx 0,4 \pm 1,96 \sqrt{0,00135} \approx 0,4 \pm 1,96 * 0,0367 \approx 0,4 \pm 0,072$$

Nedre gräns: $0,4 - 0,072 = 0,328$

Övre gräns: $0,4 + 0,072 = 0,472$

Med 95% säkerhet finns andelen ungdomar som känt sig hotade i intervallet 32,8-47,2%

$$b) n > \left(\frac{z}{m} \right)^2 p^*(1-p^*)$$

95% konfidens $\rightarrow z=1,96$

m =tillåten/önskvärd felmarginal=0,05

p^* ="sämsta" värde på p i den givna situationen, d v s det värde som ger största felmarginalen av de värden som är tänkbara=0,472

$$n > \left(\frac{1,96}{0,05} \right)^2 0,472 * 0,528 \approx 382,955$$

Stickprovet bör alltså bestå av minst 383 personer... kanske rentav kan avrundas till 400 eller så!

UPPGIFT 5

Lämpligt statistiskt test: Homogenitetstest

H_0 : Rökning har inget samband med plötslig spädbarnsdöd

H_1 : Rökning har samband med plötslig spädbarnsdöd

$$\text{Testfunktion: } X^2 = \sum \sum \frac{O - E}{E}$$

$$Df=(c-1)(r-1)=(2-1)(2-1)=1$$

Förutsättningar: OSU, oberoende mätningar, alla E minst 5

Rökning under graviditet	Cases	Controls		
Ja	79	168	247	0,400
Nej	46	324	370	0,600
Totalt	125	492	617	

Förväntade frekvenser	Cases	Controls	
Ja	50,0	197,0	247
Nej	75,0	295,0	370
Totalt	125	492	617

	Rökning under graviditeten	O	E	O-E	(O-E) ²	(O-E) ² /E
Cases	Ja	79	50	29	841	16,82
	Nej	46	75	-29	841	11,21333
Controls	Ja	168	197	-29	841	4,269036
	Nej	324	295	29	841	2,850847
		617	617			35,15322

Resultat: $X_{obs}^2 = 35,1 > X_{0,001}^2(1) = 10,83$; signifikant resultat => H_0 förkastas och H_1 antas

Slutsats: Resultatet tyder på att det finns samband mellan rökning och plötslig spädbarnsdöd och rökning kan därmed vara en bidragande orsak.

UPPGIFT 6

$$a) \Pr(X > 101) = \Pr(z > \frac{101 - 100}{1,25}) = \Pr(z > 0,8) = 1 - 0,788 = 0,212 = 21,2\%$$

b) Sannolikheten att en kartong inte behöver extra försegling = 0,788
Sökt: sannolikheten att ingen av 64 behöver extra försegling = $0,788^{64} = 0,0000002$

$$c) K = \text{Totalvikt för 64 kartonger} \\ \mu_K = \mu_T = 64 * 100 = 6400 \\ \sigma_K = \sqrt{64} * 1,25 = 10$$

P = Pallvikten

$$T = \text{Totalvikt för 64 kartonger och 1 pall} = K + P \\ \mu_T = 6400 + 150 = 6550 \\ \sigma_T = \sqrt{10^2 + 3^2} = 10,44$$

Sökt: Gränsvärdet för de 5% lägsta vikterna när $\mu = 6550$ och $\sigma = 10,44$ och vikten är NF = $6550 - 1,645 * 10,44 = 6532,8$