

TENTAMEN I STATISTIKENS GRUNDER 2
2019-03-21

Skrivtid: 09.00-14.00

Godkända hjälpmedel: Miniräknare, språklexikon.

Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

Vänligen använd inte röd penna på tentamen.

Lycka till!

Uppgift 1. (20 poäng) Man ville uppskatta genomsnittligt antal arbetsoperationer per arbetare och dag vid en viss typ av monteringsarbete. Sammanlagt studerades 100 personer (som kan betraktas som ett slumpmässigt stickprov ur en stor population) under en dag vardera. Man fann att de i genomsnitt hann med 140 arbetsoperationer per dag. Standardavvikelse i stickprovet blev 15.

- Beräkna ett 95% konfidensintervall för genomsnittliga antalet operationer per arbetare och dag.
- Hur påverkas konfidensintervallslängden (det kommer att vara kortare/samma/längre) av ändlighetskorrektur om det behövs introduceras? Motivera svaret.
- Hur stor ska vara urval om du vill att din punktskattning av μ inte ska ha en större statistisk felmarginal (dvs hälften av konfidensintervallets bredd) än 1.5?
- Ange median, \bar{x} , s och stam-bladdiagram för observerade antalet operationer per arbetare och dag: 60, 74, 150, 138, 160, 155, 139, 142, 89.

Uppgift 2. (20 poäng) På ett stor företag hade en attitydundersökning genomförts med hjälp av ett slumpmässigt urval bland de anställda. Resultat redovisas så här:

Kön	Attityd			Total
	Positiv	Negativ	Likgiltig	
Män	99	19	2	120
Kvinnor	328	92	10	430

- Analysera om det finns skillnader mellan män och kvinnor vad gäller attityd med χ^2 -metoden på signifikansnivå 5%.
- Analysera om proportion av person med positiv attityd överstiger 65% med hypotestest på signifikansnivå 5%.

Uppgift 3. (20 poäng) Finns det någon skillnad i koffeininnehåll mellan coladrycker av två olika märken? 36 flaskor av vardera märket undersöktes. Resultaten anges i μg (miligram) per volymenhet:

	Medelvärde	Urvalsstandardavvikelse
Märke A	20.1	2.35
Märke B	19.2	2.25

- Svara fråga med konfidensgrad 99% under antagande att $\sigma_1 \neq \sigma_2$ och population är mycket stor jämför med urvalsstorlek.
- Svara fråga med konfidensgrad 99% under antagande att $\sigma_1 = \sigma_2$ och populationsstorlek av vardera märket är $N = 100$ st.
- Beräkna p -värde för teststorhet i 3a).

Uppgift 4. (20 poäng) Elnätspriser för lägenheterna (pris per kWh, öre) ges i tabell samt med KPI:

År	2013	2014	2015	2016	2017	2018
Medelelnätspris (löpande priser)	64.1	66.7	68.0	70.8	73.8	76.3
KPI	314.06	313.49	313.35	316.43	322.11	328.4

- Beräkna indexserie för utveckling av elnätspriser för lägenheterna med basår 2013 och ange hur mycket medelelnätspris ökat från år 2014 till 2017.
- Beräkna elnätspriserna till prisnivån som gällde år 2018 (fastpriserna).
- Rita ett kurvdiagram över löpande och fasta elnätspriserna och kommentera kort om eventuell skillnad i prisutveckling.
- Använd KPI för att beräkna inflationen mellan 2013 och 2018. Beräkna genomsnitt inflationsnivån per 4 månader mellan 2013 och 2018.

Uppgift 5. (20 poäng) I en affär säljs dagsfärsk morotssaft. Denna levereras i kartonger med 64 burkar morotssaft i varje kartong. Varje kartong kostar 200 sek och säljs för 500 sek. Osålda kartonger måste kastas.

Man vet av erfarenhet att försäljningen ligger mellan en och tre kartonger per dag.

- Fyll in beslut matris med vinsten (i kronor) som nyttan. Strategierna S_1 , S_2 och S_3 betecknar att 1, 2 respektive 3 kartonger köps in. Händelserna E_1 , E_2 och E_3 betecknar att 1, 2 respektive 3 kartonger efterfrågas.

	Naturtillstånd		
	E_1	E_2	E_3
S_1
S_2
S_3

- Hur många kartonger ska inköpas enligt Maximax-kriteriet, Maximin-kriteriet, och hur många enligt MiniMax-regret-kriteriet?



Stockholms
universitet

Statistiska institutionen

Rättningsblad

Datum: 21/3-2019

Sal: Ugglevikssalen

Tenta: Statistikens grunder 2

Kurs: Statistikens grunder

ANONYMKOD:

0014-UN#

Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
x	x	x	x	^					8
Lär.ant.	20	18	20	20					

POÄNG

98

BETYG

A

Lärarens sign.

Prelasik

7. $n=100$ $\alpha=0,05$

$\bar{x}=140$

$s=15$

$n > 30$ är ojämnt men
 $n > 30$ så jag kan
 använda CGS.

a) $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

konfidensintervall

$140 \pm z_{0,025} \frac{15}{\sqrt{100}}$

$z_{\alpha/2} = z_{0,05/2} = z_{0,025}$

$140 \pm 1,96 \frac{15}{\sqrt{100}}$

$140 \pm 2,94$

R

$[137,06; 142,94]$

14

SVAR: konfidensintervallet omfattas av
 $[137,06; 142,94]$ på $\alpha=0,05$.

b) om ändlighet korrekturen och öns introduceras
 alltså om $\frac{N-n}{N-1} \geq 0,95$ vid en
 ändlig population utan återläggning,
 kommer längden på konfidensintervallet
 se ut som:

$2 \cdot \left(z_{\alpha/2} \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \right)$

Längden på
 $KI = 2 \cdot$ standard
 felmarginal

$2 \cdot \left(1,96 \frac{15}{\sqrt{100}} \cdot \sqrt{\frac{N-n}{N-1}} \right)$

med exempel från
 uppgiften

$> 0,95$



om jag antar att ändlighetskorrelationen är 0,75 blir exempelet så ut som följande:

$$2 \left(1,96 \frac{15}{\sqrt{100}} \cdot \sqrt{0,75} \right) \quad \text{med ändlighetskorrelation}$$

$$2(2,94 \cdot \sqrt{0,75})$$

$$2(2,55)$$

$$2(\bar{x} \pm 2,55)$$

och utan ändlighetskorrelation:

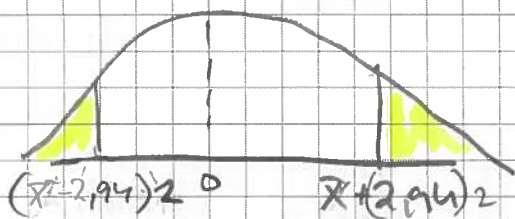
$$2 \left(1,96 \frac{15}{\sqrt{10}} \right)$$

$$2(2,94)$$

$$2(\bar{x} \pm 2,94)$$

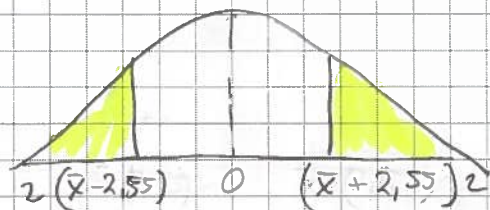
SVAR: om man adderar ändlighetskorrelation blir konfidensintervallet kortare.

Både undre och övre gränser av konfidensintervallet är nu kortare avstånd från 0, med ändlighetskorrelation.



långt avstånd

utan ändlighetskorrelation



kortare avstånd

med ändlighetskorrelation

är det så för alla N?

/2p

$$\left(Z_{\alpha/2} \frac{s}{\sqrt{n}} \right) \leq 1,5$$

har antag samma ~~beta~~
 α som i a), dussä
 $\alpha = 0,05$

c)

$$\left(1,96 \frac{15}{\sqrt{n}} \right) \leq 1,5$$

och $Z_{\alpha/2} = Z_{0,025} = 1,96$

$$1,96 \cdot 15 \leq 1,5 \sqrt{n}$$

$$29,4 \leq 1,5 \sqrt{n}$$

$$29,4 / 1,5 = 19,6$$

$$2 (19,6) \leq (\sqrt{n})^2$$

$$384,16 \leq n$$

$$n = 385$$

SVAR: För att den statistiska felmarginalen inte ska vara mindre än 1,5 måste urvalet vara minst 385, då $\alpha = 0,05$.

R 15p.

d) 60 74 89 138 139 142 150 155 160

$$n = 9$$

Stamblad-diagram

R median = 139 12p

R 12p. $\bar{x} = \frac{1107}{9} = 123$

$$s^2 = \frac{1}{8} (147651 - 9(123)^2)$$

$$\sum x^2 = 147651$$

$$s^2 = 1436,25$$

R $s = 37,9 = \sqrt{s^2} = \sqrt{1436,25}$

15p

Frekvens

Stam Blad

1	6	0
1	7	4
1	8	9
0	9	.
0	10	.
0	11	.
0	12	.
2	13	2,9
1	14	2
2	15	0,5
1	16	0

12p.

2.

Kön	Aktivitet		Likgiltigt	Totalt
	Positiv	Negativ		
Män	99	21 = 19	2	120
Kvinnor	328	102 = 92	10	430
Totalt	427	111	12	550

a) H_0 : Det finns ingen skillnad i aktivitet hos män och kvinnor

H_A : Det finns skillnad i aktivitet hos män och kvinnor

$$\alpha = 0,05$$

$$\chi^2_{0,05} (2-1) (2-1) = \chi^2_{0,05} (1) = 3,841$$

Test om homogenitet: $E(n)$

Kön	Aktivitet		Likgiltigt	Totalt
	Positiv	Negativ		
Män	93,16	26,84 = 24,22	2,62	120
Kvinnor	333,84	96,16 = 86,78	9,38	430
Totalt	427	111	12	550

Som följd av att $E(n) \stackrel{\text{mätte vara}}{\geq} 5$, och män, likgiltigt

finn 2,62, vilket lag slä ihop kategorierna "negativ" och "Liggiltigt".

H_A

Attityd

Kön	positiv	negativ / utslag	Totalt
Män	93,16	26,84	120
Kvinnor	333,84	96,16	430
Totalt	427	123	550

Efter att ha slagit ihop kategorierna ser E(n) ut som ovan.

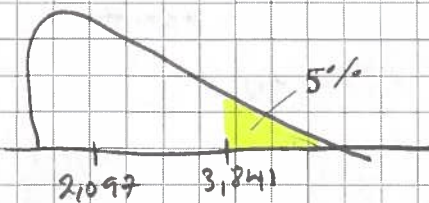
$$\chi^2_{OBS} = \sum \frac{(n - E(n))^2}{E(n)}$$

$$\chi^2_{OBS} = \frac{(99 - 93,16)^2}{93,16} + \frac{(21 - 26,84)^2}{26,84}$$

$$+ \frac{(328 - 333,84)^2}{333,84} + \frac{(102 - 96,16)^2}{96,16}$$

$$= 0,37 + 1,27 + 0,102 + 0,355 = 2,097 \quad R$$

$$\chi^2_{0,05} (2-1)(2-1) = \chi^2_{0,05} (1) = 3,841 \quad R$$



$$\chi^2_{OBS} < \chi^2_{0,05}(1)$$

$$2,097 < 3,841 \quad R$$

SVAR: H0 kan inte förkastas på $\alpha = 0,05$. Man kan också utreda några slutsatser om att det finns skillnad i attityd mellan män och kvinnor

110p

b)

	aktivitet
	positiv
kvinnor	
män	99
kvinnor	328
totalt	427

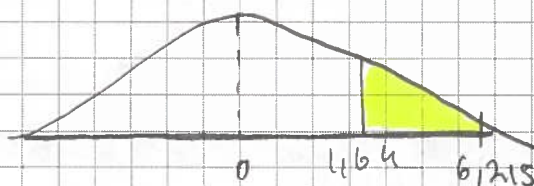
totalt i urvalet
120
430
550

$$n = 550$$

$$H_0: \pi_0 = 0,65$$

$$H_A: \pi > 0,65$$

$$\alpha = 0,05$$



$$p = \frac{\text{alla med positiv aktivitet}}{\text{totalt urval}} = \frac{427}{550} = 0,7764$$

$$z_{\text{obs}} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0,7764 - 0,65}{\sqrt{\frac{0,65 \cdot 0,35}{550}}} = \frac{0,1264}{0,0203} = 6,215$$

$$z_{0,05} = 1,64$$

$$z_{\text{obs}} > z_{0,05}$$

$$6,125 > 1,64$$

svår: H_0 kan förkastas på $\alpha = 0,05$ och man kan säga att personer med positiv aktivitet överstiger 65% i urvalet.

Man ska också kolla om det går använda testet.

$$\text{dvs om } n\hat{\pi} \geq 5$$

$$n(1-\hat{\pi}) \geq 5$$

//8p.

3.

H_0 : det finns ingen skillnad i koffeininnehåll mellan coladryckerna $\mu_1 = \mu_2$

$\sigma_0 = 0$

H_A : Det finns en skillnad i koffeininnehåll mellan coladryckerna

$\mu_1 \neq \mu_2$

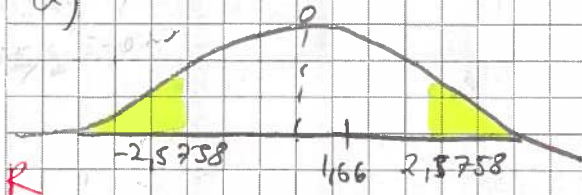
\bar{x} = Medelvärde

Urvalsstandardavvikelse = s

Märke A	$20,7 = \bar{x}$	$s_x = 2,35$	$n_x = 36$
Märke B	$19,2 = \bar{y}$	$s_y = 2,25$	$n_y = 36$

a)

Tränsvarvintervall



$\alpha = 0,01$

$\sigma^2 \neq \sigma^2$ gör att jag använder t-testmetod

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

som följd av stor population behövs inte ändrighetskonvention.

$$Z_{obs} = \frac{20,7 - 19,2 - 0}{\sqrt{\frac{2,35^2}{36} + \frac{2,25^2}{36}}} \approx \frac{0,9}{0,5422} \approx 1,66$$



$$Z_{0,01/2} = Z_{0,005} = 2,5758$$

Slut:

1,66 = $Z_{0,05}$ ligger mellan -2,5758 och 2,5758

vilket gör att vi inte kan förkasta

R

H_0 på $\alpha = 0,01$ och vi kan inte dra någon slutsats om skillnad i koffeininnehåll på coladrycker

17

b)

$\sigma_1 = \sigma_2$ vilket gör att vi antar

samma populations varians och jag använder då den pooled

variansen - S_p^2 .

		\bar{x}	S	n	N
X	Märke A	20,1	2,35	36	100
Y	Märke B	19,2	2,25	36	100

Som följd av $N_x = N_y = 100$ kan vi använda ändlighet korrektion:

$$\frac{N-n}{N-1} = \frac{100-36}{100-1} = \frac{64}{99} = 0,6465 > 0,95$$

R

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{35(2,35^2) + 35(2,25^2)}{36+36-2}$$

$$S_p^2 = \frac{370,475}{70} = 5,2925$$

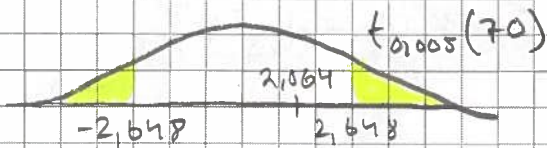
R

$$T = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{s_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)} \sqrt{\frac{N-n}{N-1}}}$$

$$R \quad T_{\text{Obs}} = \frac{20,1 - 19,2 - 0}{\sqrt{5,2925 \left(\frac{1}{36} + \frac{1}{36} \right)} \cdot \sqrt{\frac{100-36}{100-1}}}$$

$$R \quad T_{\text{Obs}} = \frac{0,9}{0,43598} = 2,064$$

$$R \quad T_{0,01/2} (n_1 + n_2 - 2) = T_{0,005} (70) = 2,648$$



SVAR: $T_{\text{Obs}} = 2,064$ hamnar mellan $2,648$ och

$-2,648$ vilket gör att den inte kan

förkastas på $\alpha = 0,01$ och vi kan ej

dra slutsatser om skillnad i koffeininnehåll i colaflaskor.

c) P-värde i 3a):

$$2 \cdot P(\bar{X} > 1,66)$$

$$2 \cdot (1 - P(Z \leq 1,66))$$

$$2(1 - \Phi(1,66)) = 2 \cdot (0,04846) = 0,09692$$

Ans svar: p-värde är 0,09692

$\approx 0,1 \approx 10\%$

13

4.

År	2013	2014	2015	2016	2017	2018
Medelvärd pris löpande	64,1	66,7	68	70,8	73,8	76,3
KPI	314,06	313,49	313,35	316,43	322,11	328,4

a)

Index 100 = 2013	2013	2014	2015	2016	2017	2018
	100	104,06	106,08	110,45	115,13	119,03

$$I_t = \frac{X_{2013}}{X_{2013}} = \frac{64,1}{64,1} \cdot 100 = 100$$

$$\frac{X_{2017}}{X_{2013}} = \frac{73,8}{64,1} \cdot 100 = 115,13$$

$$\frac{X_{2014}}{X_{2013}} = \frac{66,7}{64,1} \cdot 100 = 104,1$$

$$\frac{X_{2018}}{X_{2013}} = \frac{76,3}{64,1} \cdot 100 = 119,03$$

$$\frac{X_{2015}}{X_{2013}} = \frac{68}{64,1} \cdot 100 = 106,1$$

$$\frac{X_{2016}}{X_{2013}} = \frac{70,8}{64,1} \cdot 100 = 110,45$$

14

$$\frac{x_{2017}}{x_{2014}} = \frac{115,13}{104,06} \cdot 100 = 110,64 = 10,64\%$$

R

priset har gått upp med 10,64% /2p

vilket innebär att $66,7 \cdot 1,1064 = 73,8$

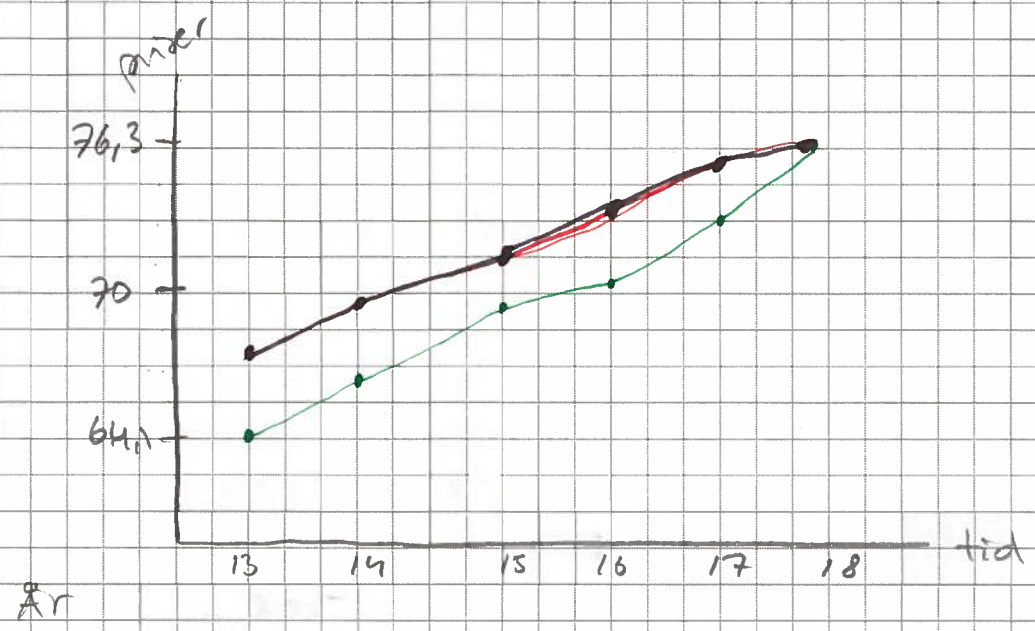
SVAR: medelvärdet har ökat med 10,64%
eller 7,1 öre/kWh.

b)	År	Pris	KPI = 2018 = 100	pris i 2018 års priser
	2013	64,1	104,57	67,03
	2014	66,7	104,76	69,87
	2015	68	104,8	71,264
	2016	70,8	103,78	73,47
	2017	73,8	101,95	75,24
	2018	76,3	100	76,3

$$\frac{KPI(2018) \cdot 100}{KPI(\text{År})}$$

/6p

c)



År	2013	2014	2015	2016	2017	2018
• Löpande priser	64,1	66,7	68	70,8	73,8	76,3
		+2,6	+1,3	+2,8	+3	+2,5
• Fakta priser	67,03	69,89	71,264	73,47	75,24	76,3
		+2,86	+1,4	+2,21	+1,77	+1,06

SVAR: Fakta om löpande priser förändras (ökar) till i början av halvperioden, korrekter går priserna mot varandra, då prismedelningen i fakta priser ~~är~~ stannar av medlen löpande priser har en fortsatt prismedelning.

14

d)

$$314,06 \cdot p^x = 328,4$$

Jag räknar på tre indelningar av varje år, där 328,4 är KPI i början av 2018, alltså är inte 2018 års trimester månad

R

$$x = 3 \cdot 5 = 15$$

trimester är 2013, 2014, 2015, 2016, 2017

$$314,06 \cdot p^{15} = 328,4$$

$$\sqrt[15]{p^{15}} = (1,0456)^{1/15}$$

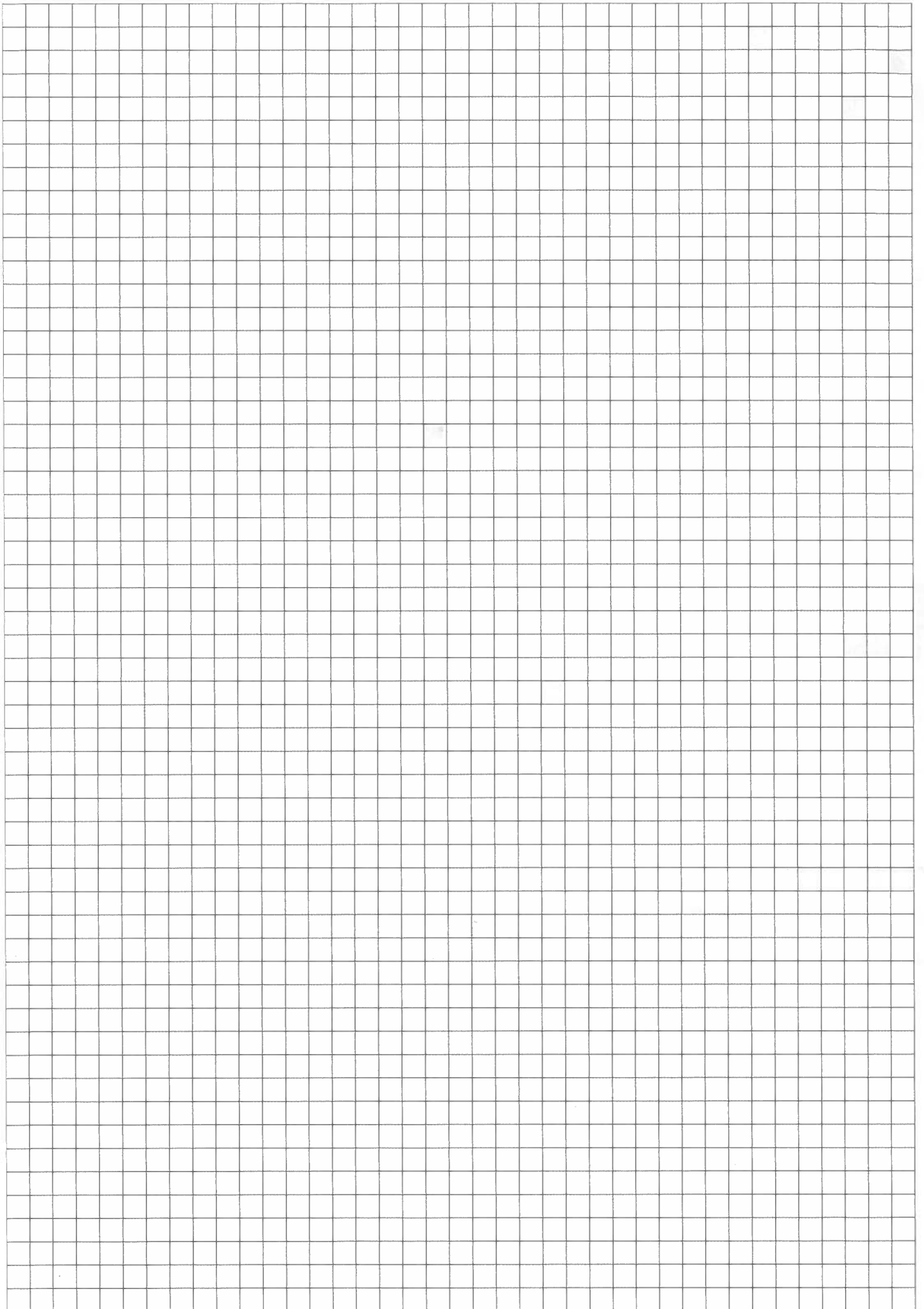
R

$$p = 1,00298 = 0,298\%$$

SVAR: varje 4 månad mellan 2013-2018 hade inflationen med 0,298%. Totalt

under perioden hade inflationen med 4,56% = $\frac{328,4}{314,06} = \frac{2018}{2013}$

/6



5.

64 blubber i varje kartong

kostnader = 200 / kartong

intäkter = 500 / kartong

Kartong	Ett förpackning	Ett förpackning			Maxi max	Maxi min	Minimax regret
		E_1 1	E_2 2	E_3 3			
S_1	1	300 ⁰	300 ³⁰⁰	300 ⁶⁰⁰	300	300	600
S_2	2	100 ²⁰⁰	600 ⁰	600 ³⁰⁰	600	100	300
S_3	3	-100 ⁴⁰⁰	400 ²⁰⁰	900 ⁰	900	-100	400
Maxi min		$s^* = 300$	$s^* = 600$	$s^* = 900$	S_3	S_1	S_2

$$S_1 E_1 = 500 \cdot 1 - 200 \cdot 1 = 300$$

$$S_2 E_1 = 500 \cdot 1 - 200 \cdot 2 = 100$$

$$S_3 E_1 = 500 \cdot 1 - 200 \cdot 3 = -100$$

$$S_1 E_2 = 500 \cdot 2 - 200 \cdot 1 = 800$$

$$S_2 E_2 = 500 \cdot 2 - 200 \cdot 2 = 600$$

$$S_3 E_2 = 500 \cdot 2 - 200 \cdot 3 = 400$$

$$S_1 E_3 = 500 \cdot 3 - 200 \cdot 1 = 1300$$

$$S_2 E_3 = 500 \cdot 3 - 200 \cdot 2 = 1100$$

$$S_3 E_3 = 500 \cdot 3 - 200 \cdot 3 = 900$$

SVAR: Enligt Maximax ska 3 kartonger köpas

Enligt Maximin ska 1 kartong köpas

Enligt Minimax-regret ska 2 kartonger köpas