STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

TENTAMEN I STATISTIKENS GRUNDER 1

2018-02-14

Skrivtid: 10.00-15.00
Godkända hjälpmedel: Miniräknare, språklexikon.
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar. Samtliga formler som används för beräkningar ska anges tydligt.

Uppgift 1. (20 poäng)
Vid användning av ett lögndetektortest kan resultatet ibland vara missvisande. För ett lögndetektortest vill man ha en så hög sensitivitet och specificitet som möjligt. Sensitivitet innebär sannolikheten att testet avger resultatet att personen ljuger givet att personen ljuger. Specificitet innebär sannolikheten att testet avger resultatet att personen talar sanning givet att personen talar sanning. Sensitiviteten är 0.85 och specificiteten är 0.70 på detta test. Vi antar att sannolikheten att en person ljuger är 0.35.

a) Vad är sannolikheten att testet visar att personen ljuger givet att personen talar sanning?

b) Vad är sannolikheten att personen ljuger givet att lögndetektorn visar att personen ljuger?

Uppgift 2. (20 poäng)
Längden på ett stort industriprojekt Y antas vara normalfördelad med väntevärde 156 veckor och standardavvikelse 26 veckor.

a) Ange sannolikheten att projektet tar mer än 200 veckor

b) Ange ett 90-procentigt centrerat intervall för Y, dvs ange den lägre gränsen (y_1) och övre gränsen (y_2) så att $P(Y < y_1) = 0.05$ och $P(Y > y_2) = 0.05$

X är en stokastisk variabel med okänd fördelning med väntevärde μ och varians σ^2. Vi bildar den standardiserade variabeln $Z = \frac{X-\mu}{\sigma}$.

c) Vad blir väntevärde och varians för variabeln Z? Visa hur du kommer fram till svaret.

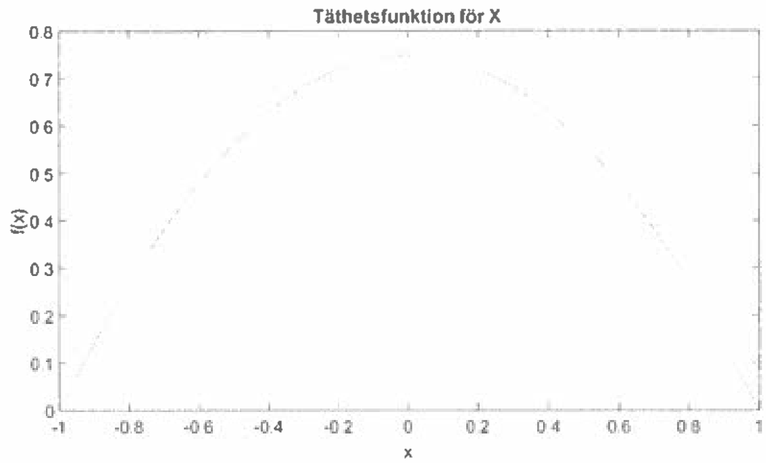
Uppgift 3 (20 poäng)
Den stokastiska variabeln X står för värdeförändringen i miljoner kronor för en specifik aktieportfölj ett år efter investering. Negativa värden på X innebär att aktieportföljen har tappat i värde sedan investeringen. Täthetsfunktionen för X antar värden mellan -1 och 1 och ges av

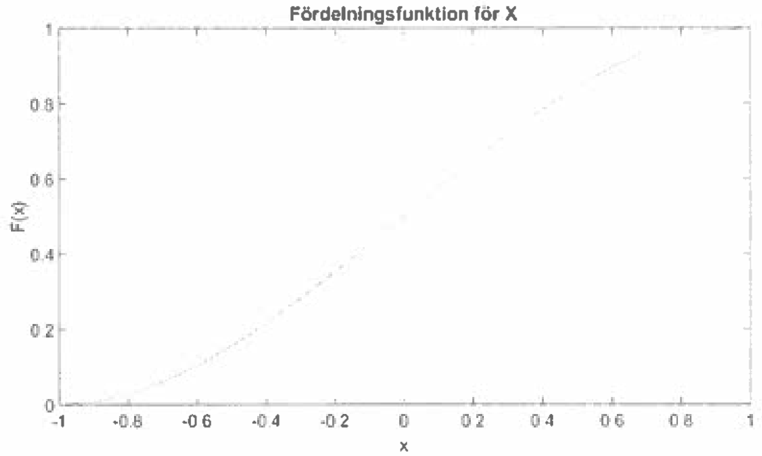
$$f(x) = 0.75\,(1 - x^2) \qquad -1 < x < 1$$

Fördelningsfuntionen ges då av

$$F(x) = 0.5 + 0.75x - 0.25x^3$$

De båda funktionerna återges i graferna nedan.





a) Beräkna sannolikheten att värdeförändringen ett år efter investeringen ligger mellan -200 000 och 400 000 kronor.

b) Kopiera de båda graferna ovan på ditt svarspapper och rita in vad sannolikheten uträknad i a) motsvarar i respektive graf. Kommentera vad det är du ritar in i graferna.

c) Vad är sannolikheten att värdeökningen blir över 800 000 kronor givet att vi vet att värdeökningen blir högre än 500 000. Räkna ut sannolikheten och rita in i (ytterligare en kopia av) täthetsfunktionen ovan, vad denna sannolikhet motsvarar i grafen. Kommentera vad det är du ritar in i grafen.

Uppgift 4 (20 poäng)
X är en stokastisk variabel som bara kan anta värdena 3, 5, 8 och 9. Man vet att $F(3) = 0.2$, $f(5) = 0.32$ och $f(9) = 0.07$

a) Ange frekvensfunktionen och fördelningsfunktionen för X.

Y är en stokastisk variabel som bara kan anta värdena 6 och 8. Man vet att $f_Y(6) = 0.6$, $f_{X,Y}(5,6) = 0.15$, $f_{X|Y}(9|6) = 0.08$ och $f_{Y|X}(8|8) = 0.39$

b) Beräkna korrelationen mellan X och Y och kommentera resultatet.

Uppgift 5. (20 poäng)
Ett företag förpackar russin i askar som fyllda ska väga 150 gram. I produktionsprocessen finns dock en viss osäkerhet och genom upprepade mätningar har man kunnat konstatera att russinaskarnas vikt är normalfördelad med en standardavvikelse om 10 gram. Företaget säljer russinaskarna för 2 kronor styck. Vid företagets kvalitetskontroll sorterar man ut alla askar som väger mindre än 150 gram och säljer dessa för 1.50 kronor. Tillverkningskostnaden för en ask russin är 50 öre.

a) Antag att vi sätter målvikten för maskinen som fyller askarna till 158 gram. Vad är sannolikheten att en slumpmässigt vald ask väger mindre än 150 gram?

b) Vi väljer slumpmässigt ut 100 askar ur produktionen. Vad är sannolikheten att fler än 15 av dessa askar väger mindre än 150 gram?

c) Vad är företagets förväntade vinst samt vad är variansen på vinsten för de 100 slumpmässigt utvalda askarna?

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 14/02/18

Sal: Ugglevikssalen

Tenta: Statistikens grunder

Kurs: Statistikens grunder 1

ANONYMKOD: 0057 RRC

☑ 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	X					10
Lär.ant. 20	20	20	20	18					

MW

POÄNG	BETYG	Lärarens sign.
98	A	JF

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0057 RRL Blad nr: 1

1.

Definierar följande händelser:

$\begin{cases} S: \text{testet visar att personen talar sanning} \\ L: \text{personen ljuger} \end{cases}$

$P(L) = 0{,}35$

$P(\bar{L}) = 1 - 0{,}35 = 0{,}65$

Sensitivitet: $P(\bar{S} \mid L) = \dfrac{P(\bar{S} \cap L)}{P(L)}$

$$0{,}85 = \frac{P(\bar{S} \cap L)}{0{,}35}$$

$$P(\bar{S} \cap L) = 0{,}85 \cdot 0{,}35 = 0{,}2975$$

Specificitet: $P(S \mid \bar{L}) = \dfrac{P(S \cap \bar{L})}{P(\bar{L})}$

$$0{,}70 = \frac{P(S \cap \bar{L})}{0{,}65}$$

$$P(S \cap \bar{L}) = 0{,}70 \cdot 0{,}65 = 0{,}455$$

NU HAR VI INFO FÖR EN FYRFÄLTSTABELL:

	S	$\bar{S}$	marginal
L	0,0525	0,2975	0,35
$\bar{L}$	0,455	0,195	0,65
marginal	0,5075	0,4925	

	S	$\bar{S}$	
L	$P(S \cap L)$	$P(\bar{S} \cap L)$	$P(L)$
$\bar{L}$	$P(S \cap \bar{L})$	$P(\bar{S} \cap \bar{L})$	$P(\bar{L})$
	$P(S)$	$P(\bar{S})$	

a) $P(\bar{S} \mid \bar{L}) = \dfrac{P(\bar{S} \cap \bar{L})}{P(\bar{L})} = \dfrac{0{,}195}{0{,}65} = \underline{\underline{0{,}3}}$

10

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0057RRC Blad nr: 2

1 b) $P(L \mid \bar{S}) = \dfrac{P(L \cap \bar{S})}{P(\bar{S})} = \dfrac{0,2975}{0,4925} = \underline{\underline{0,60}}$

10

2. Y: längd på projekt i veckor

$Y \sim N(\mu_Y = 156, \sigma_Y = 26)$

a) $P(Y > 200) = P\left(\dfrac{Y-156}{26} > \dfrac{200-156}{26}\right) =$

$= P(Z > 1,692307) = 1 - P(Z \leq 1,692307)$

$= \left\{\text{tabell över standardiserad normalfördelning}\right\} \approx 1 - 0,95449 = 0,04551 = \underline{\underline{0,046}}$

6

b) $P(Y < y_1) = 0,05$

$P\left(\dfrac{Y-156}{26} < \dfrac{y_1 - 156}{26}\right) = P\left(Z < \dfrac{y_1 - 156}{26}\right) =$

$1 - P\left(Z < -\dfrac{y_1 - 156}{26}\right) = 0,05$

Alltså måste $P\left(Z < -\dfrac{y_1 - 156}{26}\right) = 0,95$

↗

SU, STATISTIK

Skrivsal: Ugglevik   Anonymkod: 0057 RRC   Blad nr: 3

Enligt tabell måste då

$$-\frac{y_1 - 156}{26} \approx 1,645$$

$$y_1 = \underline{\underline{113,23}}$$

Så endast 5% av projekten ligger under 113 veckor.

Avstånd från medelvärdet: $156 - 113,23 = 42,77$

Eftersom normalfördelningen är symmetrisk kring sitt väntevärde kommer den övre gränsen ligga lika långt ovanför medelvärdet

$y_2 = 156 + 42,77 = 198,77$

$$P(113,23 \leq Y \leq 198,77) = 0,10$$

8

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0057 RRC Blad nr: 4

2 c)

$$Z = \frac{X-\mu}{\sigma} = \frac{X}{\sigma} - \frac{\mu}{\sigma}$$

väntevärdet på en konstant är själva konstanten

$$E(Z) = E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma}$$

$$= \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0 \qquad E(Z) = 0$$

$$V(Z) = V\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\cdot V(X) - 0 = \frac{1}{\sigma^2}\cdot\sigma^2 = 1 \qquad V(Z) = 1$$

en konstant varierar inte kring sitt väntevärde

6

Bra!

3.

X: Värdeförändring efter ett år

$$f(x) = 0{,}75\cdot(1-x^2) \qquad -1<x<1$$

$$F(x) = 0{,}5 + 0{,}75x - 0{,}25x^3$$

a)

$$P(-0{,}2 \le X \le 0{,}4) = F(0{,}4) - F(-0{,}2)$$

$$= 0{,}5 + 0{,}75\cdot 0{,}4 - 0{,}25\cdot 0{,}4^3 - \left(0{,}5 + 0{,}75(-0{,}2) - 0{,}25\cdot(-0{,}2)^3\right)$$

$$= 0{,}784 - 0{,}352$$

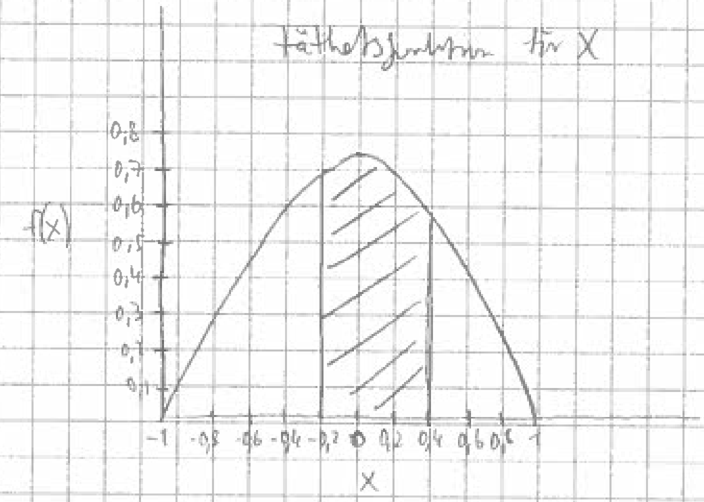
$$= 0{,}432$$

6

SU, STATISTIK

Skrivsal: Ugglevik   Anonymkod: 0057 RRC   Blad nr: 5

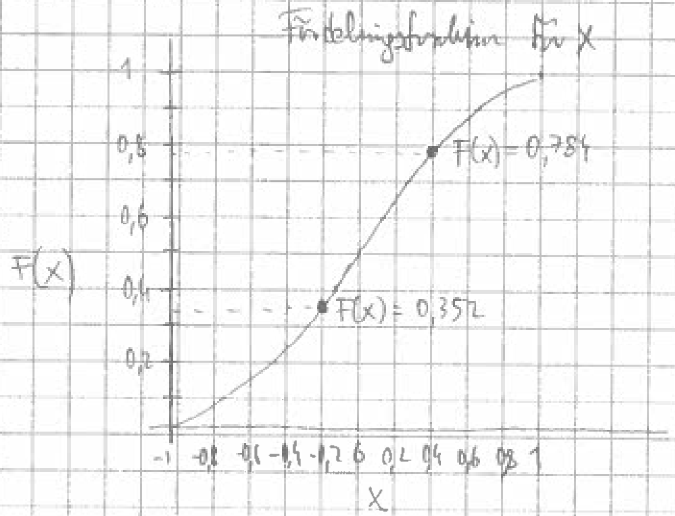
3 b)



totala arean under grafen är 1

Den streckade arean är 0,432, vilket är sannolikheten att

$-0{,}2 \leq X \leq 0{,}4$



Sannolikheten att x är mindre eller lika med 0,4 är 0,784

Sannolikheten att x är mindre eller lika med −0,2 är 0,352

Sannolikheten att ligga emellan är $0{,}784 - 0{,}352 = 0{,}432$

6

SU, STATISTIK

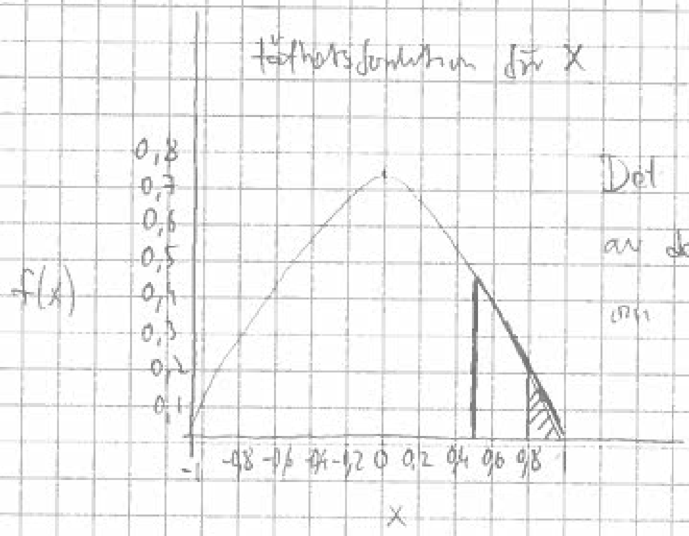
Skrivsal: Uggleviken   Anonymkod: 0057-RKC   Blad nr: 6

3 c)

$$P(X > 0{,}8 \mid X > 0{,}5) = \frac{P(X > 0{,}8 \cap X > 0{,}5)}{P(X > 0{,}5)} =$$

$$= \frac{P(X > 0{,}8)}{P(X > 0{,}5)} = \frac{1 - F(0{,}8)}{1 - F(0{,}5)} = \frac{1 - (0{,}5 + 0{,}75 \cdot 0{,}8 - 0{,}25 \cdot 0{,}8^3)}{1 - (0{,}5 + 0{,}75 \cdot 0{,}5 - 0{,}25 \cdot 0{,}5^3)}$$

$$= \frac{1 - 0{,}972}{1 - 0{,}84375} = \underline{\underline{0{,}1792}}$$



Det streckade området utgör 17,92% av det markerade området till höger om x = 0,5

8

SU, STATISTIK

Skrivsal: Uppleckh Anonymkod: 0057-RRC Blad nr: 7

4.

a)

X	f(x)	F(x)
3	0,20	$\boxed{0,20}$
5	$\boxed{0,32}$	0,52
8	0,41	0,93
9	$\boxed{0,07}$	1

F(3) = 0,20 ⟹ f(3) = 0,2 eftersom X inte kan vara mindre än 3

F(5) = f(3) + f(5) = 0,2 + 0,32 = 0,52

f(9) = 0,07 ⟹ F(8) = 0,93

F(9) = 1 per definition

f(8) = 1 − 0,20 − 0,32 − 0,07 = 0,41

6

b)

$f_Y(6) = 0,6$

$Corr(X,Y) = ?$

$f_{X,Y}(5,6) = 0,15$

$f_{X|Y}(9|6) = 0,08$

$f_{Y|X}(8|8) = 0,39$

Y \ X	3	5	8	9	
6	0,1519	0,15	0,2501	0,048	0,6 ← från uppgift
8	0,0481	0,17	0,1599	0,027	0,4
	0,2	0,32	0,41	0,07	← från a)

K

$f_{X,Y}(9,6) = F_{X|Y}(9|6) \cdot F_Y(6) = 0,08 \cdot 0,6 = \underline{\underline{0,048}}$

$f_{Y,X}(8,8) = F_{Y|X}(8|8) \cdot F_X(8) = 0,39 \cdot 0,41 = \underline{\underline{0,1599}}$

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0057-RRC Blad nr: 8

4b) Fortsättning

$E(X) = 0{,}2 \cdot 3 + 0{,}32 \cdot 5 + 0{,}41 \cdot 8 + 0{,}07 \cdot 9 = 6{,}11$ R

$E(Y) = 0{,}6 \cdot 6 + 0{,}4 \cdot 8 = 6{,}8$ R

$V(X) = \sum_x x^2 f(x) - E(X)^2 = 3^2 \cdot 0{,}2 + 5^2 \cdot 0{,}32 + 8^2 \cdot 0{,}41 + 9^2 \cdot 0{,}07 - 6{,}11^2$

$= 41{,}71 - 6{,}11^2 = \underline{\underline{4{,}3779}}$ R

$V(Y) = \sum_y y^2 f(y) - E(Y)^2 = 6^2 \cdot 0{,}6 + 8^2 \cdot 0{,}4 - 6{,}8^2 = 47{,}2 - 6{,}8^2 = \underline{\underline{0{,}96}}$ R

$Cov(X,Y) = \sum_x \sum_y xy f_{X,Y}(x,y) - E(X)E(Y) =$

$= 6 \cdot 3 \cdot 0{,}1519 + 6 \cdot 5 \cdot 0{,}15 + 6 \cdot 8 \cdot 0{,}2501 + 6 \cdot 9 \cdot 0{,}48 +$

$+ 8 \cdot 3 \cdot 0{,}0481 + 8 \cdot 5 \cdot 0{,}17 + 8 \cdot 8 \cdot 0{,}1599 + 8 \cdot 9 \cdot 0{,}02 - 6{,}11 \cdot 6{,}8$

$= 41{,}603 - 6{,}11 \cdot 6{,}8 = \underline{\underline{0{,}055}}$ R

$Corr(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \dfrac{0{,}055}{\sqrt{4{,}3779 \cdot 0{,}96}} = \underline{\underline{0{,}0268}}$ R

Om två variabler är oberoende av varandra är corr ati en noll. En korrelation på noll behöver dock inte alltid betyda att variablerna är oberoende av varandra

Korrelationen här ligger nära noll vilket skulle kunna tyda på att sambandet mellan variablerna [illegible] nyckel

14

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0057 RRC Blad nr: 9

(5.)

X: vikt på en russinask

$\mu_x = 158$

$\sigma_x = 10$

a) $P(X < 150) = P\left(\frac{X-158}{10} < \frac{150-158}{10}\right)$

$= P(Z < -0{,}8) = P(Z > 0{,}8) = 1 - P(Z \leq 0{,}8)$

$1 - 0{,}78814 = 0{,}21186 \approx 0{,}21$

4

b) Y: antal askar mindre än 150g

$Y \sim Bin(100, 0{,}212)$

$\mu_y = n \cdot p = 21{,}2$

$\sigma_y^2 = n \cdot p \cdot (1-p) = 16{,}7056$

$n \cdot p = 100 \cdot 0{,}212 = 21{,}2 > 5$

Vi kan approximera binomialfördelningen med en normalfördelning! K

$P(Y > 15) = P(Y \geq \cancel{14{,}5}\ 16) = P\left(\frac{Y - 21{,}2}{\sqrt{16{,}7056}} \geq \frac{\cancel{14{,}5}\ 15{,}5 - 21{,}2}{\sqrt{16{,}7056}}\right)$

kompensera för att normalfördelningen är kontinuerlig (vilket binomialfördelningen inte är)

$= P(Z \geq -1{,}639) = P(Z \leq 1{,}639) = 0{,}94950$ ✓

$\approx 0{,}95$

7

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0857 RRC Blad nr: 10

5 c)

Förväntat antal askar under 150g:

$\mu_Y = n \cdot p = 100 \cdot 0{,}21186 = 21{,}186$

Förväntat antal askar över eller lika med 150g:

$100 - 21{,}186 = 78{,}814$

kostnad per ask

Förväntad vinst: $21{,}186 \cdot 1{,}50 + 78{,}814 \cdot 2 - 100 \cdot 0{,}5 = 139{,}407$

≈ 139 kr K

Bättre räkna per ask inför nästa uppgift.

Y: vinst per ask

kostnad per ask

$E(Y) = \sum_y y f(y) = (2 - 0{,}5) \cdot 0{,}78814 + (1{,}5 - 0{,}5) \cdot 0{,}21186 = 1{,}39407$

Vinst för 100 askar: $E(100Y) = 100 \cdot E(Y) = 139{,}407 \approx 139$ kr K

Varians för vinsten per ask:

$V(Y) = \sum_y y^2 f(y) - E(Y)^2 = (2-0{,}5)^2 \cdot 0{,}78814 + (1{,}5-0{,}5)^2 \cdot 0{,}21186 - 1{,}39407^2$

$= 0{,}0417$ K

?

Det är skillnad på $V(100X)$ och $V(X_1 + X_2 + \ldots + X_{100})$

$V(100X) = 100^2 \cdot 0{,}0417 = \cancel{417}$

Du ska ta

$V(X_1 + X_2 + \ldots + X_{100}) = V(X_1) + V(X_2) + \ldots + V(X_{100}) = 100 \cdot V(X) =$

$= 100 \cdot 0{,}0417 = 4{,}17$