

Stockholms Universitet  
Statistiska institutionen  
Per Gösta Andersson

## Statistikens grunder 2

### SKRIFTLIG TENTAMEN

Måndagen den 28 oktober, 2019

Tillåtna hjälpmedel: miniräknare

Gräns för godkänt: 50 poäng av totalt 100.

För maximal poäng krävs på varje uppgift tydliga, utförliga och välmotiverade lösningar.

1. (20p)

992, 1011, 1020, 986, 997, 1010

är ett urval med  $n = 6$  på kapacitansen (i  $\mu\text{F}$ ) för en viss typ av kondensatorer. Antag att kapacitansen är  $N(\mu, \sigma^2)$ .

- Intervallskatta  $\mu$  med 99% konfidensgrad.
- Om vi fick information om att det skattade värdet  $s$  av  $\sigma$  i (a) faktiskt är det sanna värdet av  $\sigma$ , vilket konfidensintervall för  $\mu$  med 99% konfidensgrad får vi då?
- Om vi vill halvera längden på intervallet i (b), vilket värde på  $n$  behöver vi då?

2. (16p) I en undersökning vill man studera om elförbrukningen (i en viss typ av hushåll) under tider med hög belastning minskade om man fick rabatt på elpriser under tider med låg belastning. För 10 hushåll mättes elkonsumtionen i februari månad under hög belastning året innan rabatten infördes och året efter. Resultat: (enhet: kwh)

Hushåll	1	2	3	4	5	6	7	8	9	10
Före	200	180	240	424	120	333	418	380	340	516
Efter	160	185	210	370	130	298	368	250	305	477

Kan man påvisa en minskning av förbrukningen efter införande av rabatter? Använd lämpligt test under normalfördelningsantagande och signifikansnivå  $\alpha = 0.05$ . (Vi får utgå från att tex väderleksförhållandena vara desamma under dessa två år.)

3. (16p) Antag att det gjorts en undersökning av 1000 studenters rökvanor på riksnivå. För att undersöka om studenter vid Stockholms universitet har andra rökvanor än studenter i resten av landet, tillfrågades också 1000 studenter i Stockholm. Båda urvalen anses slumpmässiga. Resultat: (observerade frekvenser)

	Röker inte	Röker ca ett paket cigaretter per vecka	Röker ca ett paket cigaretter per dag
Stockholm	660	260	80
Riksnivå	700	240	60

Tyder dessa siffror på att studenterna vid Stockholms universitet har andra rökvanor än studenter i resten av landet? Utför lämplig hypotesprövning på approximativ signifikansnivå 10%.

4. (16p) Ett urval av 10 observationer har gett medelvärdet 20, medianen 18 och standardavvikelsen  $s = 5$ . Det största värdet är 39.27. Av misstag (felkodning) har detta värde ändrats till 392.7.

- (a) Vad blir medelvärdet efter ändringen?  
 (b) Vad blir medianen efter ändringen?  
 (c) Vad blir standardavvikelsen efter ändringen?

Tilläggsfråga (utanför poängbedömningen!) som är ett citat från annan lärobok:

*"Why doesn't anyone talk about the fourth quartile? Or do they?"*

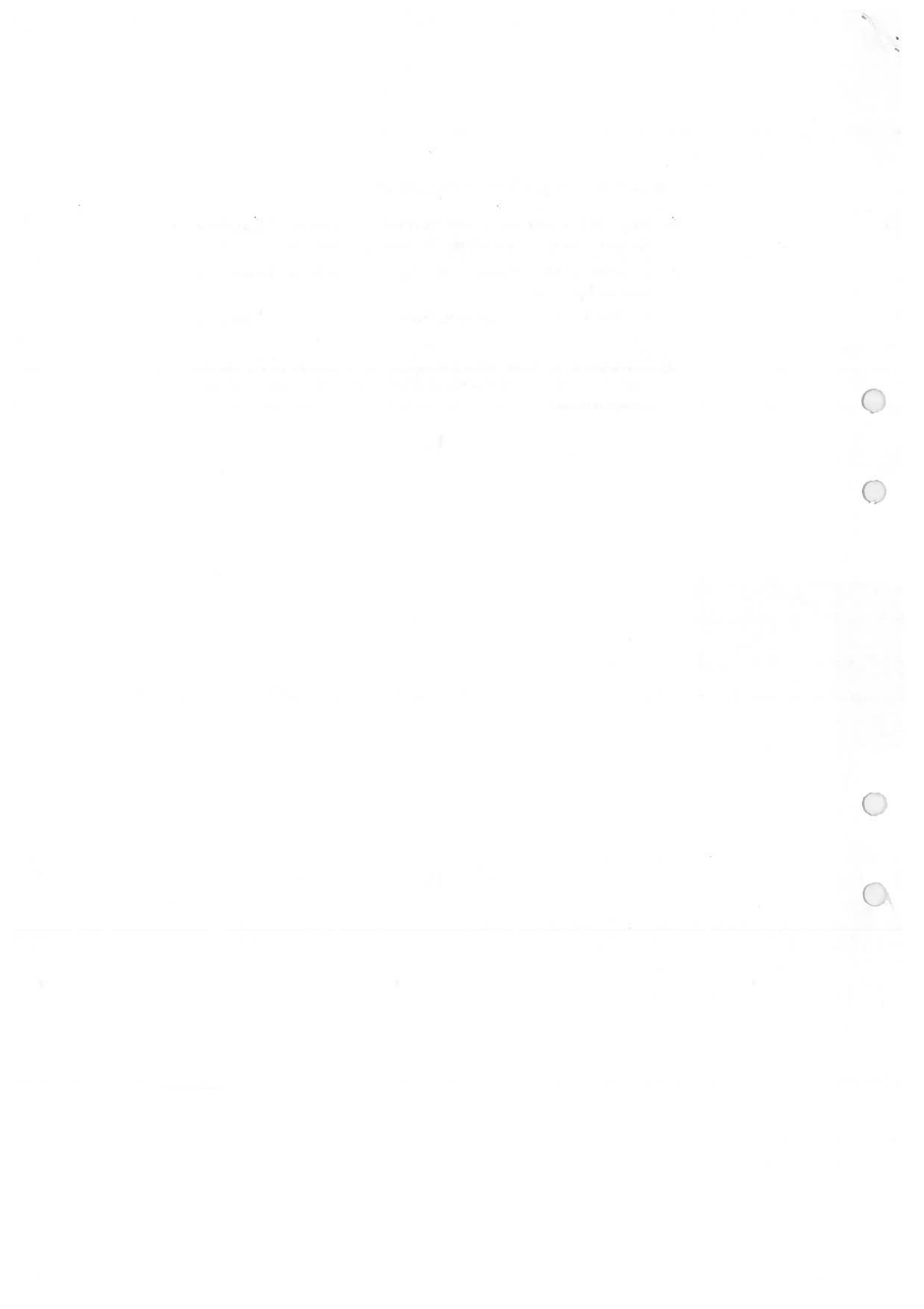
5. (16p) Favorit i repris: myntkast. Ett visst mynt har vid kast sannolikheten  $p$  att komma upp med "krona". Person  $A$  kastar myntet 100 gånger, får 42 st "krona" och skattar  $p$  med 0.42.  $B$  kastar 200 gånger, får 96 st "krona" och skattar  $p$  med 0.48. Vilken av skattningarna

$$\hat{p}_1 = \frac{0.42 + 0.48}{2} \quad \text{eller} \quad \hat{p}_2 = \frac{42 + 96}{100 + 200}$$

föredrar du? Motivera teoretiskt. Ledning: Jämför skattningarna genom motsvarande stokastiska variabler för antalet "krona" i de båda försöken och beräkna väntevärden och varianser.

6. (16p) Sant eller falskt? Motivering/kommentar krävs.

- (a) Maximin-kriteriet vid beslutsteori tar mer hänsyn till (garderar sig mot) "dåliga" nyttovärden än maximax-kriteriet.
- (b) P-värdet är sannolikheten vid hypotesprövning att förkasta en sann nollhypotes
- (c) För fixt  $\alpha$  vid hypotesprövning reducerar man  $\beta$  genom att minska urvalsstorleken.
- (d) Tillämpning av centrala gränsvärdessatsen för en  $\chi^2$ -fördelning innebär att ju fler frihetsgrader desto bättre anpassning till normalfördelning.





Stockholms  
universitet

Statistiska institutionen

## Rättningsblad

**Datum:** 28/10-2019

**Sal:** Ugglevikssalen

**Tenta:** Statistikens grunder 2

**Kurs:** Statistikens grunder

**ANONYMKOD:**

0035- FNZ

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	X				3
Lär.apt.	18	16	16	18	12				

POÄNG	96	BETYG	A	Lärarens sign.	PGW
-------	----	-------	---	----------------	-----



1) a) Let the variable of the kapacitansen be called  $X$ .

$$X \sim N(\mu, \sigma^2)$$

$$n = 6$$

$$\bar{X} = \frac{992 + 1011 + 1020 + 986 + 997 + 1010}{6}$$

$$= 1002.67$$

$$s^2 = \frac{1}{n-1} (\sum x_i^2 - (n \cdot \bar{x}^2))$$

$$= \frac{(6032890 - (6 \cdot 1002.67^2))}{5}$$

$$= 161.45 \quad (169.5)$$

$$\sigma_{\bar{x}} = \frac{\sqrt{161.45}}{\sqrt{6}}$$

$$= 5.19$$

$$99\% \text{ konfidensgrad} = \bar{X} \pm t_{0.005}(5-1) \cdot \sigma_{\bar{x}}$$

$$= 1002.67 \pm 4.032 \cdot 5.19$$

$$= 1002.67 \pm 20.91$$

$$(981.76 - 1023.58)$$

b) If  $s = 6$

Then, 99% konfidensintervall for  $\mu$

$$= \bar{X} \pm Z_{0.005} \cdot \sigma_{\bar{x}}$$

$$= 1002.67 \pm 2.58 \cdot 5.19$$

$$= 1002.67 \pm 13.39$$

$$(989.28, 1016.06)$$

$$c) \quad n \geq \frac{13.39}{2} = 6.695$$

$$n = 2.58^2 \cdot \frac{\sigma^2}{B}$$

$$n = 2.58^2 \cdot \frac{161.45}{6.695^2}$$

$$n = 23.96, \text{ At least 24 observations.}$$

/18

2)  $\mu_1 =$  Before discounts

$\mu_2 =$  After discounts

$$\mu_D = \mu_1 - \mu_2$$

$$H_0: \mu_D = 0$$

$$H_a: \mu_D > 0$$

Huzhall	1	2	3	4	5	6	7	8	9	10
Difference	40	-5	30	54	-10	35	50	130	35	39

$$\bar{d} = \frac{398}{10}$$

$$= 39.8$$

$$s_d^2 = \frac{1}{10-1} (28912 - 10 \cdot 39.8^2)$$

$$= 1452.4$$

$$\alpha = 0.05$$

$H_0$  accepted if  $t_{obs} < t_{0.05}(10-1)$

$H_0$  rejected if  $t_{obs} > 1.833$

$$t_{obs} = \frac{39.8}{\sqrt{1452.4}/\sqrt{10}}$$

$$= 13.3$$

$13.3 > 1.833$ ,  $H_0$  can be rejected on a 5% level.

From these results it can be seen that there has been a decrease after the discounts.

/16



3)  $H_0$ : Studenterna vid Stockholms universitet har samma rökvanor än studenterna i resten av landet.

$H_a$ : Studenterna vid Stockholms universitet har andra rökvanor än studenterna i resten av landet.

$\alpha = 0.05$

$H_0$  accepted if  $\chi^2_{obs} < \chi^2_{0.05}(3-1)(2-1)$

	Stockholm	Riksnivå	Total
Röker inte	660	700	1360
Röker ca ett paket cigaretter per vecka	260	240	500
Röker ca ett paket cigaretter per dag	80	60	140
Total	1000	1000	2000

$E(n_i) =$

Stockholm  
 $\frac{1360}{2000} \times 1000 = 680$

Riksnivå  
 $\frac{1360}{2000} \times 1000 = 680$

$\frac{500}{2000} \times 1000 = 250$

$\frac{500}{2000} \times 1000 = 250$

$\frac{140}{2000} \times 1000 = 70$

$\frac{140}{2000} \times 1000 = 70$

$$\chi^2_{obs} = \frac{(660-680)^2}{680} + \frac{(700-680)^2}{680} + \frac{(260-250)^2}{250} + \frac{(240-250)^2}{250} + \frac{(80-70)^2}{70} = 4.83$$

$H_0$  rejected if  $\chi^2_{obs} > \chi^2_{0.05}(2)$

$= \chi^2_{obs} > 5.991$

$4.83 < 5.991$ ,  $H_0$  can be accepted on a 5% nivel.

From the given results we can't say that students of Stockholms university have the same smoking habits than the rest of the country.

116

$$\begin{aligned}
 4) \quad n &= 10 \\
 \bar{x} &= 20 \\
 m &= 18 \\
 s &= 5
 \end{aligned}$$

Original biggest value = 39.27

New biggest value = 392.7

$$a) \text{ Before it was } - \frac{\sum_{i=1}^{10} x_i}{10} = 20$$

$$\sum_{i=1}^{10} x_i = 200$$

$$\sum_{i=1}^9 x_i + 39.27 = 200$$

$$\sum_{i=1}^9 x_i = 160.73$$

$$\text{Now it is then } - \frac{160.73 + 392.7}{10}$$

$$\bar{x} = 55.343$$

b) Median is not affected by the biggest or smallest numbers or the outliers.

It is dependent on the number of positions. Since the two values just around the median didn't change, the median is still 18.

$$c) \text{ Before it was } - \sqrt{\frac{\sum_{i=1}^{10} x_i^2 - (10 \times 20^2)}{9}} = 5$$

$$\sum_{i=1}^{10} x_i^2 = 225 + 4000$$

$$\sum_{i=1}^9 x_i^2 + (39.27)^2 = 4225$$

$$\sum_{i=1}^9 x_i^2 = 2682.87$$

$$\text{Now it is then } - \sqrt{\frac{2682.87 + (392.7)^2 - (10 \cdot (55.343)^2)}{9}}$$

$$s = 118.45$$

/16

We actually don't know the value of  $p$ .

$$5) X_1 \sim \text{bin}(n=100, p=0.5)$$

$$E(X_1) = 100 \times 0.5 = 50$$

$$V(X_1) = 100 \times 0.5 \times 0.5 = 25$$

$$\hat{p}_1 = \frac{X_1}{n_1}$$

$$\begin{aligned} E(\hat{p}_1) &= \frac{1}{n_1} E(X_1) \\ &= \frac{1}{100} \times 50 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} V(\hat{p}_1) &= \frac{1}{n_1^2} V(X_1) \\ &= \frac{1}{100^2} \times 25 \\ &= 0.00025 \end{aligned}$$

$$X_2 \sim \text{bin}(n=200, p=0.5)$$

$$E(X_2) = 200 \times 0.5 = 100$$

$$V(X_2) = 200 \times 0.5 \times 0.5 = 50$$

$$\hat{p}_2 = \frac{X_2}{n_2}$$

$$\begin{aligned} E(\hat{p}_2) &= \frac{1}{n_2} E(X_2) \\ &= \frac{1}{200} \times 100 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} V(\hat{p}_2) &= \frac{1}{n_2^2} V(X_2) \\ &= \frac{1}{200^2} \times 50 \\ &= 0.00125 \end{aligned}$$

It is best to choose  $\hat{p}_2$  because both of them have the same expected value but  $\hat{p}_2$  has a smaller variance meaning the estimation is more centered towards the true value.

otherwise good!

/18

6) a) True

Maximin-kriterium looks at the least profit for each case and choose the minimum. Or in other words it tries to minimize the loss. While maximin-kriterium only looks at how to maximize the profits, without taking the losses into consideration.

OK

b) False

Type I error  $\alpha$  is the probability to reject  $H_0$  given that  $H_0$  is true. While p-value is the probability to get the observed value we got or to get a even more unusual value given that  $H_0$  is true.

OK

c) False

For a fixed  $\alpha$ , to reduce  $\beta$  we need to increase the sample size. This is because the greater the sample size the less probability there is for type I and type II error to occur. This happens because variance becomes smaller and the curves become 'thinner', so both  $\alpha$  and  $\beta$  is reduced when  $n$  gets bigger.

OK

d) False

According to central limit theorem for t-distribution, the bigger the  $n$  gets, the better it can be approximated with normal distribution.  $\chi^2$ -distribution on the other hand is not even symmetrical and so for even bigger  $n$  values, it can't be approximated with normal distribution, at least not according to central limit theorem.

/12