**STOCKHOLMS UNIVERSITET**
Statistiska institutionen
Jolanta Pielaszkiewicz

# TENTAMEN I STATISTIKENS GRUNDER 2
## 2018-10-29

---

**Skrivtid:** 15.00-20.00
**Godkända hjälpmedel:** Miniräknare, språklexikon.
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

---

**Uppgift 1.** (20 poäng) Det brukar att ta några minuter innan Nisse svarar på sms. Vi skriver ner väntetid till svaret. Resultat (i minuter): 10, 15, 22, 2, 4, 15, 28, 16.

a) Beräkna typvärde, median, urvalsmedelvärde, urvalsvarians.

b) Konstruera ett histogram med klassbredd 6 över väntetiderna.

c) Förklara begrepp:
- Centrala gränsvärdessatsen
- Signifikansnivå

d) Anta att väntetiden är normalfördelad med $\sigma = 9$ och att vi ska förkasta $H_0 : \mu = 10$ mot $H_A : \mu > 10$ om urvalsmedelvärdet är större än 15. Beräkna signifikansnivå för testet.

**Uppgift 2.** (20 poäng) Vi fick ett datamaterial angående betygsättning vid online och traditionell klass-undervisning. Två oberoende slumpmässiga grupper av 50 studenter (som vid slutet av terminen tilldelades en betyg i skala: A, B, C, D-E, F) ska analysera. Datamaterialet:

| Undervisningsmetod | Betyg | | | | |
|---|---|---|---|---|---|
| | A | B | C | D-E | F |
| Online | 10 | 13 | 16 | 9 | 2 |
| Traditionell | 4 | 12 | 15 | 9 | 10 |

a) Hypotespröva om undervisningsmetod och betygsättning är oberoende av varandra på 2.5 % signifikansnivå. Kom ihåg att skriva $H_0$, $H_1$, teststorhet, slutsatser (med motivation).

b) Hypotespröva på $\alpha = 5\%$ om betygsfördelning vid online-undervisning följer:
P(Betyg=A)=0.14, P(Betyg=B)=0.16, P(Betyg=C)=0.26,
P(Betyg=D eller E)=0.32, P(Betyg=F)=0.12.

**Uppgift 3.** (20 poäng) En viss sorts konstfiberväv, som produceras i mycket stora mängder, genomgår en kemisk behandling för att få bättre hållfasthet. Två olika kemikalier provades på 9 stycker väv och draghållfastheten bestämdes. Resultat:

| | mätningar av draghållfastheten | | | | |
|---|---|---|---|---|---|
| KEM1 | 1.3 | 1.6 | 0.5 | 1.2 | 1.4 |
| KEM2 | 2.2 | 2.4 | 0.4 | 2.2 | |

a) Anta normalfödelningen av de två stickproven med samma varians $\sigma^2$. Kan man visa att det finns någon signifikant skillnad när det gäller draghållfastheten mellan de två kemikalier? Svara med lämpligt test på $\alpha = 5\%$.

b) Ytterligare mätningar gjordes för den första av de studerade kemikalierna (KEM1). Urvalsmedelvärdet $\bar{x} = 1.3$ och urvalsvarians $s_1^2 = 0.2$ erhölls för urval av storlek 150. Histogram visar att datamaterialet inte är normalfördelad. Ange ett 99% kondifensintervall för draghållfasthetensmedelvärde vid kemisk behandling KEM1.

**Uppgift 4.** (20 poäng) Från ett företags försäljningsstatistik har följande uppgifter hämtats:

| Plagg | Pris | | *Kvantitet* | |
|---|---|---|---|---|
| | 1993 | 1994 | *1993* | *1994* |
| Blus | 117 | 125 | 27 | *33* |
| Skjorta | 106 | 112 | 13 | *16* |
| T-shirt | 95 | 107 | 43 | *64* |

Basår=1993

a) Beräkna Laspeyers och Paasches index för prisutvecklingen från 1993 till 1994.

b) Använda Laspeyers index från a) för att svara fråga: Med hur många procent per halv år har prisnivån i genomsnitt stigit från år 1993 till 1994?

c) Företagsägare undrar om proportion av "rea produkter" inom försäljningen. En urvalsproportion på 25% (dvs. $p = 0.25$) erhölls för ett slumpmässigt urval av storlek 1000. Kan man visa att populationsproportion av "rea produkter" är mindre än 30%? Om möjligt använda $p$-värdet för att dra slutsatser. $\alpha = 5\%$.

OBS. Populationen av företagetsprodukter är mycket stor i jämförelse med urvalsstorlek.

**Uppgift 5.** (20 poäng) Vi förväntar oss att vår nya innovativa bilhantering påverkar nyttan man har från en bil (metoden är anpassad för Nissan, Volvo, BMW och Audi).

| | Naturtillstånd | | | |
|---|---|---|---|---|
| | Nissan | Volvo | BMW | Audi |
| nya bilhantering | 117 | 125 | 100 | 122 |
| vanlig bilhantering | 110 | 112 | 90 | 120 |

a) Vilken bilhanteringsmetod (den nya eller den vanliga) ska vi välja för att få den största nyttan enligt Maximax-kriteriet och vilken enligt Laplace-kriteriet?

b) Nu antar vi att:
- nyttan i tabellen ovan är normalfördelade
och att
- nyttan var bestämt vid den nya och den vanliga bilhanteringen för samma individ (från population av alla Nissan, Audi, BMW och Volvo bilar).
Hypotespröva om det finns en signifikant skillnad i nyttan mellan de två bilhanteringsmetoderna. Använda $\alpha = 1\%$.

Statistiska institutionen

Stockholms universitet

# Rättningsblad

**Datum:** 29/10/18

**Sal:** Ugglevikssalen

**Tenta:** Statistikens grunder

**Kurs:** Statistikens grunder 2

**ANONYMKOD:** 0023-0HW

☒ 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 | | | | | 4 |
| Lär.ant. 20 | 19 | 20 | 20 | 19 | | | | | |

PG

| POÄNG | BETYG | Lärarens sign. |
|---|---|---|
| 98 | A | YPrelaskra |

**SU, DEPARTMENT OF STATISTICS**

Room: Vöglerkissalen Anonymous code: 0023-OHW Sheet number: 1

1. a. typvärde: 15 R /2p

2, 4, 10, 15, 15, 16, 22, 28

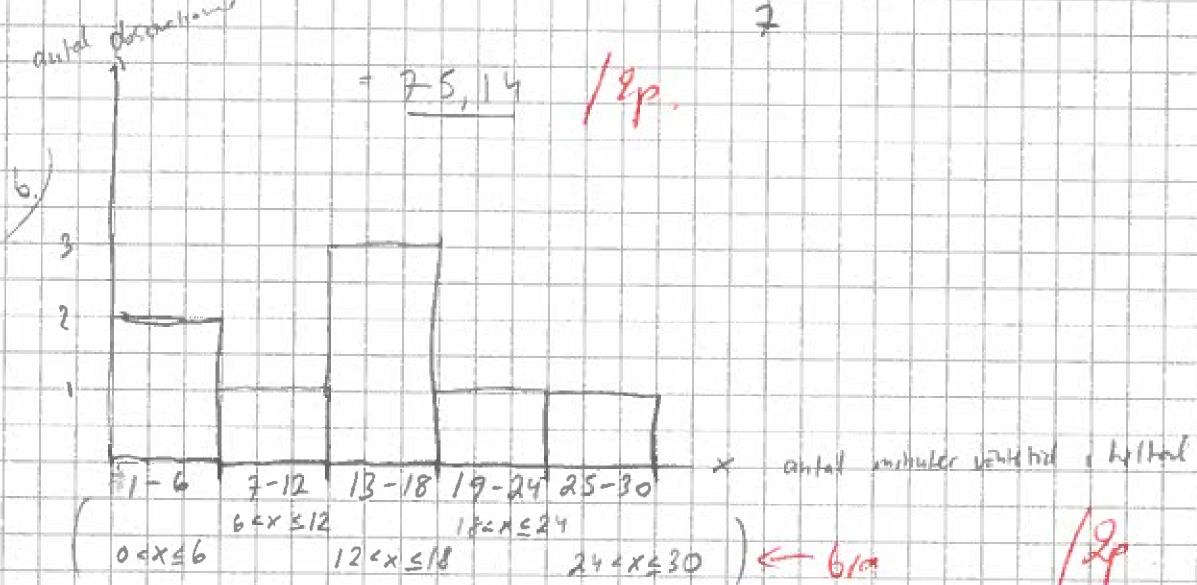
median: $\frac{15+15}{2} = \underline{15}$ /2p

urvalsmedelvärde: $\frac{2+4+10+15+15+16+22+28}{8} = \underline{14}$ /2p

urvalsvarians: $\frac{2^2+4^2+10^2+15^2+15^2+16^2+22^2+28^2-8\cdot 14^2}{7}$

$= \underline{75,14}$ /2p

b.)




← bra /2p

c. CGS: Om urvalsstorleken är stor så kan samplingfördelningen av approximeras med normalfördelningen. Allmänt gäller vid $n > 30$:

$$n > 30 \begin{cases} \bar{X} \overset{approx}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \\ Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \end{cases}$$

/3⁻

signifikansnivå = $\alpha$ = sannolikheten att nollhypotesen förkastas givet att den är sann. (T.ex. om $\alpha = 0{,}05$ så förkastas 5% av de mest extrema möjliga värdena i dess fördelning trots att $H_0$ är sann.)

/3

d. $H_0: \mu_0 = 10$

$H_A: \mu_A > 10$ } ensidigt test

$X$ = väntetiden

om $H_0$: $X \sim N(10, 9^2)$.

Känd varians och normalfördelning:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

enl. text förkastas $H_0$ vid $\bar{x}_{obs} > 15$

$$P(\bar{X} > 15) = 1 - P(\bar{X} \le 15) =$$

$$1 - P\left(Z \le \frac{15 - 10}{9/\sqrt{8}}\right) = 1 - \phi(1,57) = 1 - 0,94179 \approx 0,0582$$

Svar: testets signifikansnivå är 0,0582 /4

2. a. Oberoende test:

$H_0$: Betyg & undervisningsmetod är oberoende av varandra

/1 $H_A$: — " — beroende av varandra.

obs. värden

| | A | B | C | D-E | F | |
|---|---|---|---|---|---|---|
| Online | 10 | 13 | 16 | 9 | 2 | 50 |
| Tradit | 4 | 12 | 15 | 9 | 10 | 50 |
| | 14 | 25 | 31 | 18 | 12 | 100 |

förväntade värden

| | A | B | C | D-E | F | |
|---|---|---|---|---|---|---|
| | 50/100·14 = | 50/100·25 = | 50/100·31 = | 50/100·18 = | 5/100·12 = | |
| Online | 7 | 12,5 | 15,5 | 9 | 6 | 50 |
| Tradition | 7 | 12,5 | 15,5 | 9 | 6 | 50 |
| | 14 | 25 | 31 | 18 | 12 | 100 |

R /3

bra Obs alla förväntade värden är $\ge 5$, således behövs ingen ihopslagning göras. Antal frihetsgrader = (antal rader − 1)·(antal kolumner − 1) $= 1 \times 4 = 4$

Förkasta $H_0$ om $\chi^2_{obs} > \chi^2_{0,025}(4) = 11,143$ R /2

$$\chi^2_{obs} = \frac{(10-7)^2}{7} + \frac{(13-12,5)^2}{12,5} + \frac{(16-15,5)^2}{15,5} + \frac{(9-9)^2}{9} + \frac{(2-6)^2}{6}$$

$$+ \frac{(4-7)^2}{7} + \frac{(12-12,5)^2}{12,5} + \frac{(15-15,5)^2}{15,5} + \frac{(9-9)^2}{9} + \frac{(10-6)^2}{6} = \text{forts.}$$

**SU, DEPARTMENT OF STATISTICS**

Room: Vegaluhssalen Anonymous code: 0023-OHW Sheet number: 2

$1,286 + 0,02 + 0,016 + 0 + 2,667 + 1,286 + 0,02 + 0,016 + 0 +$
$2,667 = 7,978$ /2

$7,978 < 11,143$ och vi kan således ej förkasta $H_0$ och vi drar därför slutsatsen att betyg och undervisningsmetod är oberoende av varandra vid $\alpha = 0,025$. /2

b. Goodness-of-fit-test

$H_0$: observerade betyg i onlinegruppen följer fördelningen $0,14 : 0,16 : 0,26 : 0,32 : 0,12$ /1

$H_A$: betygen i onlinegruppen följer ej ovan nämnd fördelning.

?

Förkasta $H_0$ om $\chi^2_{obs} > \chi^2_{0,05}(k-1-m) = \chi^2_{0,05}(3) = 7,815$ /1

förväntade värden om $H_0$: (Obs alla är $\geq 5$!)

| | A | B | C | D-E | F | tot |
|---|---|---|---|---|---|---|
| | $0,14 \cdot 50 = 7$ | $0,16 \cdot 50 = 8$ | $0,26 \cdot 50 = 13$ | $0,32 \cdot 50 = 16$ | $0,12 \cdot 50 = 6$ | 50 |

/3

$$\chi^2_{obs} = \frac{(10-7)^2}{7} + \frac{(13-8)^2}{8} + \frac{(16-13)^2}{13} + \frac{(9-16)^2}{16} + \frac{(2-6)^2}{6}$$

$$= 1,286 + 5 + 0,692 + 3,063 + 2,667 = 12,708$$ /2

Svar: Förkasta $H_0$ på $\alpha = 0,05$, dvs de observerade betygen följer ej sannolikhetsfördelningen som beskrivs i uppgiften vid $\alpha = 0,05$. /2

3. a.

$H_0: \mu_X = \mu_Y$ (tvåsidigt test) $\mu_X$ = medelvärde för KEM1

/1 $H_1: \mu_X \neq \mu_Y$ $\mu_Y$ = " KEM2

Till uppg antas normalfördelning och samma $\sigma^2$ i båda populationerna.

/1
$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \quad \text{där} \quad S_p^2 = \frac{(n_x - 1) \cdot S_x^2 + (n_y - 1) S_y^2}{n_x + n_y - 2}$$

förkasta $H_0$ vid $|t_{obs}| > t_{\frac{0,05}{2}}(7) = 2,365$ /2

$$\bar{x} = \frac{1,3 + 1,6 + 0,5 + 1,2 + 1,4}{5} = \frac{6}{5} = 1,2$$ /1

$$S_x^2 = \frac{1,3^2 + 1,6^2 + 0,5^2 + 1,2^2 + 1,4^2 - 5 \cdot 1,2^2}{4} = 0,175$$ /1

$$\bar{y} = \frac{2,2 + 2,4 + 0,4 + 2,2}{4} = 1,8$$ /1

$$S_y^2 = \frac{2,2^2 + 2,4^2 + 0,4^2 + 2,2^2 - 4 \cdot 1,8^2}{3} = 0,88$$ /1

$$S_p^2 = \frac{4 \cdot 0,175 + 3 \cdot 0,88}{7} = 0,4771$$ /2

$$t_{obs} = \frac{1,2 - 1,8}{\sqrt{0,4771 \cdot \left(\frac{1}{5} + \frac{1}{4}\right)}} = -1,295$$ /1

Svar: Vi kan ej förkasta $H_0$ på $\alpha = 0,05$, dvs det finns inte statistiskt signifikanta belägg för att det finns en skillnad i draghållfasthet mellan Kem1 & Kem2 vid $\alpha = 0,05$.

/2

**SU, DEPARTMENT OF STATISTICS**

**Room:** Ugglevikssalen **Anonymous code:** 0023-OHW **Sheet number:** 3

b. Enl. CGS kan vi approximera med normalfördelningen vid stort n (ex x = 150 som i detta exempel)

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$1,3 \pm 2,5758 \cdot \sqrt{\frac{0,2}{150}}$$

ok

$$1,3 \pm 0,094$$

$$1,3 \; [1,206 \; ; 1,394]$$

/7

4. a. $$L_{1997}^{1993} = \frac{125 \cdot 27 + 112 \times 13 + 107 \times 45}{117 \cdot 27 + 106 \cdot 13 + 95 \cdot 45} \cdot 100$$

$$= \frac{9646}{8812} \cdot 100 \approx 109,46$$

/4

$$P_{1997}^{1993} = \frac{125 \cdot 33 + 112 \cdot 16 + 107 \cdot 64}{117 \cdot 33 + 106 \cdot 16 + 95 \cdot 64} \cdot 100$$

$$= \frac{12\,765}{11\,637} \cdot 100 \approx 109,69$$

/4

Svar: + 9,46 % (Laspeyres) resp + 9,69 % (Paasches).

b.

$$100 \cdot x^2 = 109,46$$

$$x = \sqrt{\frac{109,46}{100}} \approx 1,0462$$

Svar: Ökning med ca 4,62 % per halvår.

/4

c. $H_0: \pi_0 = 0,3$ } enkelsidigt $\alpha = 0,05$

/1.5 $H_A: \pi_0 < 0,3$

/1 $np = 0,3 \cdot 1000 = 300 > 5$ normalfördelningsapprox. ok!

$n(1-p) = 0,7 \cdot 1000 = 700 > 5$

$p \overset{approx}{\sim} N(\pi_0, \sigma^2)$

om $H_0$: $Z = \dfrac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$ förkasta $H_0$ om $Z_{obs} < -1,6449$

om $H_0$:

$Z = \dfrac{0,25 - 0,3}{\sqrt{\dfrac{0,3 \cdot 0,7}{1000}}} = -3,45$ → således förkasta $H_0$!

/2

Vårat p-värde är $P(P \leq 0,25) = P(Z \leq -3,45) = \phi(-3,45)$

$= 1 - \phi(3,45) = 1 - 0,99972 = 0,00028$ /2

Svar: Man kan dra slutsatsen att andelen kaprodukter är mindre än 30% vid signifikansnivå 5%. Det är först vid orimligt låga signifikansnivåer (under 0,028%) som man ej kan förkasta $H_0$. /1.5

**SU, DEPARTMENT OF STATISTICS**

Room: Ynglingasalen. Anonymous code: (T)23-0HW Sheet number: 4

Beslut under osäkerhet.

| S.E. | $S_1$ | $S_2$ | $S_3$ | $S_4$ | Maximax nytta | LaPlace |
|---|---|---|---|---|---|---|
| $A_1$ | 117 | 125 | 100 | 122 | 125 | 116 |
| $A_2$ | 110 | 112 | 90 | 120 | 120 | 108 |

förväntad nytta enl Laplace: $A_1 = \frac{117 + 125 + 100 + 122}{4} = 116$ R

$A_2 = \frac{110 + 112 + 90 + 120}{4} = 108$ R

Svar: enl. både Maximax- & LaPlace kriterierna bör man välja $A_1$ (ny tillverkning).

/10

5.

relaterade urval

$H_0: \bar{D} = 0$ dvs det finns ingen skillnad i nytta mellan tillverk.

$H_A: \bar{D} \neq 0$ dvs det finns en skillnad i nytta.

dubbelsidigt test

$$T = \frac{\bar{D}}{S_D / \sqrt{n}}$$

Förkasta $H_0$ om $|t_{obs}| > t_{\frac{0.01}{2}}(3) = 5{,}841$ R

$$\bar{D} = \frac{(117-110)+(125-112)+(100-90)+(122-120)}{4} = \frac{7+13+10+2}{4}$$

R $= 8$

$$S_D^2 = \frac{(7-8)^2+(13-8)^2+(10-8)^2+(2-8)^2}{3} = \frac{1^2+5^2+2^2+6^2}{3}$$

R $= 22$

$t_{obs} = \dfrac{8}{\sqrt{22/4}} = 3{,}411$ R

Kan förkasta ej $H_0$ på $\alpha = 0{,}01$, på denna signifikansnivå
R kan man ej ~~x~~ [illegible] signifikanta skillnader i
vikten mellan hög eller vanlig behandling.

/9