Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 24/11/17

Sal: Brunnsvikssalen

Tenta: Statistikens grunder

Kurs: Statistikens grunder 2

ANONYMKOD: SG2-XOH-HTC

☒ 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					6
Lär.ant. 14	16	18	14	18					

POÄNG	BETYG	Lärarens sign.
85	B	[illegible]

Jag ber om ursäkt om vad jag skrev ser konstigt ut: svenska är inte mitt modersmål. :)

SU, STATISTIK

Skrivsal: [illegible] Anonymkod: SG2-XOH-HTC Blad nr: 1

Uppgift 1.

a) Båda test är χ^2-test, som används när man vill jämföra en (eller fler) urvals observerade värde med förväntade värde.

– I ett goodness-of-fit test följer de förväntade värdena en given fördelning: man vill testa ~~att~~ om de observerade värde också följer den givna fördelningen.

Ex: Vi kastar en tärning 120 gånger. ~~och~~ Vi förväntar oss att varje utfall ~~vid varje kast~~ kan inträffas med sannolikhet $\frac{1}{6}$

→ H_0: Observationerna följer fördelningen $\frac{1}{6}$.

→ H_A: Observationerna följer inte fördelningen $\frac{1}{6}$

– Ett homogenitetstest genomförs när ~~vi~~ vi vill veta om två variabler / urval följer den samma fördelningen, även när fördelningen är okänd.

Ex: en politiker vill veta om fördelningen av individer som sympatiserar för henne är densamma i de fyra dela av hennes valdistrikt.

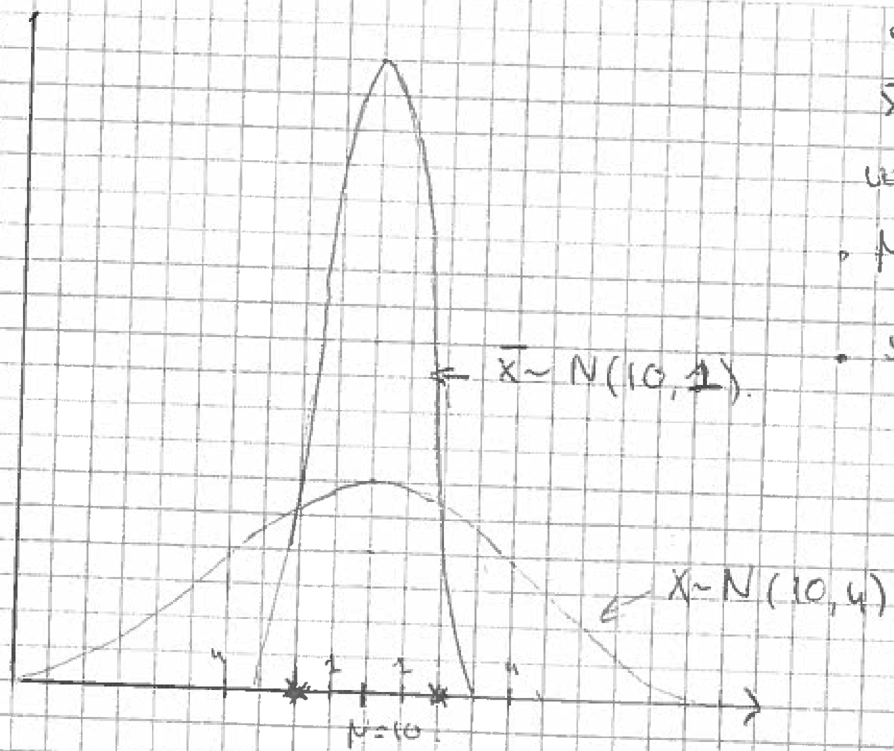
→ H_0: fördelningen av "sympatiserar" är den samma i alla delar.

→ H_A: fördelningen av "sympatiserar" är inte den samma i alla delar.

5

→

b/



$n = 16$.

$\bar{X}$ är en normalfördelad variabel med:

• $\mu_{\bar{x}} = \mu_x = 10$.

• $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{4}{4} = 1$

c/ – Tidsserie: vi tittar på utvecklingen av en variabels värde vid olika punkter i tiden.

– Indexserie: vi tittar på utvecklingen av en variabels värde vid olika punkter i tiden, men vi använder en enda bastidpunkt i varje beräkning så att vi kan jämföra resultaten vid varje tidpunkt med resultatet vid bastidpunkt (hur variabel utvecklar sig i tiden relativt en viss tidpunkt (ex år, månad, osv)).

Ex: en individ tittar på sin inkomst per månad varje år mellan 2010 och 2014, och sedan använder 2010 som basår för att se hur inkomsten har utvecklat.

år	Inkomst / mån	Index (basår 2010)
2010	20 000	100
2011	22 000	110
2012	23 000	115
2013	25 000	125
2014	28 000	[illegible]

SU, STATISTIK

Skrivsal: Kräftriket/Aulasalen Anonymkod: SGL-XOH-HTC Blad nr: 2.

För att beräkna indexen använder personen formeln:

$$I_t^{2010} = \frac{\text{Inkomst } t}{\text{Inkomst } 2010} \times 100.$$

d/ ~~D~~en centrala gränsvärdessatsen är en approximation som säger att när fördelningen av ett urval inte är känd, och när n är tillräckligt stor ($n > 30$), liknar fördelningen en normalfördelning. CGS är viktig i statistik inferens eftersom en förutsättning för att beräkna och testa samplingfördelningar av ~~…~~ ett urvals karaktäristiker är att urvalet är normalfördelad.

STATISTIK

Skrivsal: Brunnsvikssalen   SG2-XOH-HTC.   Anonymkod: ______   Blad nr: 3

Uppgift 2

X = pulstrycket.

n = 144 : vi antar att X är normalfördelad enligt den centrala gränsvärdessatsen ($n > 30$).

$\rightarrow X \sim N(\mu = 47, \sigma = 25)$

a/ X approximeras till en normalfördelade variabel Z.

$Z \sim N(0, 1)$ $Z = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}}$

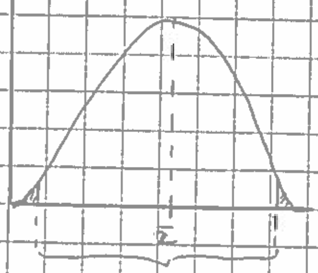
En 95%-igt konfidensintervall ges av

$$\bar{x} \pm Z_{0{,}05/2} \times \frac{\sigma}{\sqrt{n}} = 47 \pm 1{,}96 \times \frac{25}{\sqrt{144}}$$

$$= 47 \pm 4{,}083$$

$\rightarrow$ 95%-igt konfidensintervall: $[42{,}917 - 51{,}083]$

Intervallet ska tolkas som ~~[illegible]~~ värdena som ska observeras i 95% av fallen



95%-igt konfidensintervall

6

b/ Antagande: Observationerna är oberoende

Hypotes: H_0: $\mu_{obs} = \mu_0 = 45$

H_A: $\mu_{obs} > \mu_0 = 45$ (ensidigt)

Testvariabel: $Z = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

$$p = 1 - \phi\left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

Observationer: $z_{obs} = \dfrac{47 - 45}{25/12} = 0{,}96$

$$p = 1 - \phi(0{,}96)$$
$$= 1 - 0{,}83147$$
$$= 0{,}16853$$

Slutsats: H_0 förkastas vid 16,853 %-igt signifikansnivå och högre. Om vi skulle genomföra ett test vid 5 % eller 10 % signifikansnivå så skulle H_0 inte förkastas, vilket skulle tyda på ~~tyda~~ att patientens pulstryck inte överstiger 45 mm hg.

2

c/ ~~1,96 = 1,96x~~

$$1{,}96 \times \frac{25}{\sqrt{n}} = 4 \iff \frac{25}{\sqrt{n}} = \frac{4}{1{,}96}$$

?

$$\sqrt{n} = 25 \times \frac{1{,}96}{4}$$

$$n = \left(\frac{25 \times 1{,}96}{\cancel{4}\ 2}\right)^2$$

$$= 150{,}06$$

$$\Rightarrow \underline{n \geq 151}$$

4

Man måste genomföra minst 151 mätningar.

..., STATISTIK

Skrivsal: Brunnsvikssalen   Anonymkod: SG2-XOH-HTC   Blad nr: 4.

Uppgift 3.

Observationer: 11,7 ; 12,4 ; 12,8 ; 12,9 ; 13,3

X = koncentrationen giftämne.

$\bar{x} = \frac{\sum x_i}{n} = \underline{12,62}$

$s_x^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$

$= \frac{797,78 - \frac{63,1^2}{5}}{4}$

$= \underline{0,367}$

a) Hypotes: $H_0: \mu_x = \mu_0 = 12$ [Avtagande observerade obs]

$H_1: \mu_x > \mu_0 = 12$ ← ensidigt test

Signifikansnivå: $\alpha = 0,05$

Testvariabel: $T_{obs} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ $T \sim t_{0,95}(4)$ om H_0 är sann.

Beslutregel: H_0 förkastas om $t_{obs} > t_{0,05}(4) = 4,604$ 2,132

Observationer: $t_{obs} = \frac{12,62 - 12}{\sqrt{0,367/5}}$

$\approx \underline{2,29}$

Slutsats: $t_{obs} < t_{0,05}(4)$. H_0 förkastas inte på signifikansnivån $\alpha = 0,05$.

→

8

b) Vi antar att observationerna är oberoende och okorrelerade

Hypotes: $H_0: \mu_{öst} = \mu_{väst} \iff \mu_{öst} - \mu_{väst} = 0$

$H_A: \mu_{öst} > \mu_{väst}$

Signifikansnivå: $\alpha = 0,025$

Testvariabel $Z = \dfrac{\bar{x}_{öst} - \bar{x}_{väst}}{\sqrt{\frac{s^2_{öst}}{n_{öst}} + \frac{s^2_{väst}}{n_{väst}}}}$ $Z \sim$ approx $N(0,1)$ om H_0 är sann.

Beslutsregel H_0 förkastas om $Z_{obs} > Z_{0,025} = 1,96$ K

Observationer

$$Z = \frac{10,66 - 9,01}{\sqrt{\frac{12,1}{35} + \frac{11,6}{35}}}$$

$$= \frac{1,65}{0,82}$$

$$= \underline{2,00}$$ K

Slutsats: H_0 förkastas på 2,5%-igt signifikansnivån koncentrationen av giftämnen är högre öster om fabriken

10

, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: S62-XOH-HTC Blad nr: 5.

Uppgift 4.

a) När man beslutar under risk känner man sannolikheten att en viss naturtillstånd kommer att inträffa. Vi antar att myntet som kastas är "normalt" och att sannolikheten att "krona" eller "klave" kommer att inträffa är den samma: $\frac{1}{2}$.
Vi antar också att varje deltagande till spelet har 1000 kronor och att en deltagande ~~spelare~~ som har 1000 kronor och väljer att inte spela håller kvar sina 1000 kr oavsett resultatet (krona eller klave).

A_1 = spelar S_1 = krona → vinner.

A_2 = spelar ej S_2 = klave → förlorar.

	S_1 $(P(S_1) = 0{,}5)$	S_2 $(P(S_2) = 0{,}5)$	$E(U)$
A_1	$u_{1,1} = 2500$	$u_{1,2} = 0$	$2500/2 = 1225$
A_2	$u_{2,1} = 1000$	$u_{2,1} = 1000$	1000

→ Man bör spela.

b) Om jag vill minimera förlusten skulle jag inte spela:
- om jag inte spelar håller jag kvar mina 1000 sek
- om jag spelar finns det chans att jag förlorar allting, vilket går emot om jag vill minimera förlusten.

Om företaget vill gå "break-even" måste det förändra vinstsumman X så att $\frac{X+0}{2} = 1000$ i stället för 1225.

(Så att den förväntade nyttan ...

oavsett man spelar eller inte.

$$\Rightarrow \frac{x}{2} = 1000 \iff \underline{x = 2000}.$$

Företaget måste sänka vinstsumman till 2000 kr.

4

d/ Resonemanget är det samma. För att gå "break-even" måste företaget förändra sannolikheten att vinna, så att den förväntade summan E(V) blir detsamma oavsett man spelar eller inte.

$$\Rightarrow p_{vinna} \times 2500 = 1000 \iff \underline{p_{vinna} = 0,4}$$

4

SU, STATISTIK

Skrivsal: [illegible] Anonymkod: SGZ-XOH-HTC Blad nr: 6

Uppgift 5:

Antagande: Observationerna är oberoende.

Test: Vi söker en given fördelning hos observationerna.
↳ Vi gör därför ett goodness of fit test.

Hypotes: H_0: Observationerna följer en normalfördelningen med $\mu = 12$ och $\sigma = 3$.
H_A: Observationerna följer inte en normalfördelning med $\mu = 12$ och $\sigma = 3$.

Signifikansnivå: $\alpha = 0,05$.

Testvariabel: $\chi^2_{obs} = \sum \frac{(n_i - E(n_i))^2}{E(n_i)}$, $\chi^2_{obs} \sim \chi^2_{0,05}(5)$ Om H_0 är sann.

* Frihetgrader = 5: klasserna "under 6" och "6-8" samt klasserna "16-18" och "över 18" grupperas eftersom E(n "under 6") och E(n "över 18") < 5
↳ Det finns då 6 klasser i stället för 8.

Beslutregel: H_0 förkastas om $\chi^2_{obs} > \chi^2_{0,05}(5) = 11,07$.

→

Observationer och förväntade värde

Längd	Antal (n_i)	Förväntade Antal (E_i)	χ^2
under 8*	19	$P(X \le 8) \times 92 = 8{,}44$. Visa hur ihr räknat ut sh för de olika intervallen	$\frac{(19-8{,}44)^2}{8{,}44} = 13{,}2$
8 - 10	13	$P(8 \le X \le 10) \times 92 = 14{,}98$	$\frac{(13-14{,}98)^2}{14{,}98} = 0{,}2$
10 - 12	18	$P(10 \le X \le 12) \times 92 = 22{,}57$	$\frac{(18-22{,}57)^2}{22{,}57} = 0{,}9$
12 - 14	16	$P(12 \le X \le 14) \times 92 = 22{,}57$	$\frac{(16-22{,}57)^2}{22{,}57} = 1{,}9$
14 - 16	9	$P(14 \le X \le 16) \times 92 = 14{,}98$	$\frac{(9-14{,}98)^2}{14{,}98} = 2{,}38$
över 16*	17	$P(X \ge 16) \times 92 = 8{,}44$	$\frac{(17-8{,}44)^2}{8{,}44} = 8{,}68$
Summa	92	92	$\chi^2_{obs} = 27{,}50$

* Klasserna som grupperades: "under 6" och "6-8".
"16-18" och "över 18"

Slutsats: $\chi^2_{obs} > \chi^2_{0,05}(5) = 11{,}07$. H_0 förkastas vid $\alpha = 0{,}05$.
Observationerna följer inte en normalfördelning med $\mu = 12$ och $\sigma = 3$.

18

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

OMTENTAMEN I STATISTIKENS GRUNDER 2

2016-11-24

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
Förklara följande begrepp:
a) Centrala gränsvärdessatsen
b) Median
c) Homogenitets-test
d) Ändlighetskorrektion
e) Typ-I fel

Uppgift 2
Vid en fabrik tillverkas rep av nylon. Enligt en tidigare produktionsmetod har den förväntade hållfastheten på repen varit 8320 kg. Man vill undersöka om ett nytt produktionssätt ger ökad hållfasthet och kontrollerar därför hållfastheten för 300 repstumpar tillverkade på det nya sättet. Man antar oberoende mellan repstumparna. Medelvärdet och standardavvikelsen av hållfastheten för de 300 repstumparna blev $\bar{x} = 8450$ och $s = 1000$.

a) Utför ett lämpligt test för att testa om det nya tillverkningssättet ger ökad hållfasthet. Använd signifikansnivån 5%.

b) Beräkna och tolka p-värdet för testet.

c) Beräkna ett 90%-igt konfidensintervall. Hur ska konfidensintervallet tolkas?

Uppgift 3

Man studerar sambandet mellan användande av bilbälte (i en stat i USA där det inte krävs av lagen att man använder bilbälte) och rökning. En teori är att människor som röker är mindre bekymrade över sin hälsa och säkerhet och är därför mindre benägna att använda bilbälte. En undersökning av rökvanor och bältesanvänding bland 639 slumpmässigt valda vuxna, genererade följande resultat.

	Antal rökta cigaretter per dag			
	0	1-14	15-34	35 och över
Använder bilbälte	195	167	52	9
Använder inte bilbälte	129	50	31	6

a) Är användandet av bilbälte och rökning oberoende? Utför ett lämpligt test och använd signinfikansnivån 5%.
b) Man vill testa om andelen vuxna som använder bilbälte är under 70 procent. Utför ett lämpligt test och använd siginifikansnivån 5%.

Uppgift 4

Man vill testa om personers reaktionstid påverkas av ett visst sorts läkemedel. Ett reaktionstest utfördes bland ett urval av totalt 40 personer - 15 stycken som åt läkemedlet och 25 stycken som inte åt läkemedlet. Reaktionstiderna mättes för båda grupperna. Medelreaktionstiden för personer som åt läkemedlet var 0.87 sekunder med en standardavvikelse på 0.24 sekunder. Bland de som inte åt läkemedlet var medelreaktionstiden 1.13 sekunder med en standardavvikelse på 0.25 sekunder.

a) Verkar det vara en skillnad i reaktionstid mellan patienter som äter läkemedlet och de som inte äter läkemedlet? Utför ett lämpligt test och testa på signifikansnivån 5%. Redogör för nödvändiga antaganden.
b) Räkna ut ett 95%-igt konfidensinterall för reaktionstiden för respektive grupp dvs de som äter respektive de som inte äter läkemedlet.
c) Finns det något koppling i tolkningen av svaren i uppgift a) och b)?

Uppgift 5

Ett skogsområde skall säljas genom anbud, d.v.s. de som vill köpa området lägger ett bud som hålls hemligt fram till en viss tidpunkt då de inlämnade buden offentliggörs. Den som har det högsta budet vid offentliggörandet vinner budgivningen och får köpa området för det pris som anges i budet. Svenska Skogar AB bedömmer att skogsområdet kan ge en avkastning på 100 000. VD:n överväger att antingen ge budet 40 000 kr, budet 60 000 kr eller budet 80 000 kr. VD:n bedömer också att konkurrenten Fina skog AB kan ge ett bud på 50 000 kr, ett bud på 70 000 kr eller inte ge något bud alls. Ingen annan väntas ge något bud.

a) Hjälp VD:n för Svenska skogar att sätta upp beslutsmatrisen.

b) Vilket bud bör VD:n lägga om minimax-regretkriteriet används?

c) Vilket bud bör VD:n välja om sannolikheten att Fina skog AB lägger budet 50 000 är 0.50 och sannolikheten att de lägger budet 70 000 är 0.25?

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

OMTENTAMEN I STATISTIKENS GRUNDER 2

2017-11-24

Skrivtid: 10.00-15.00
Godkända hjälpmedel: Miniräknare.
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

Uppgift 1

a) Förklara skillnaden mellan ett Godness-of-fit test och ett Homogenitetsttest. Vad testar man med respektive test? Ge ett påhittat exempel på en korrekt formulerad noll- och mothypoten för respektive test. 5

b) Skissa upp i samma graf fördelningarna för $X \sim N(10, 4)$ och $\overline{X}$ beräknat på ett stickprov av storleken 16. 5

c) Beskriv vad en tidsserie respektive en indexserie är? Ge även ett eget påhittat exempel på en tidsserie angiven vid fem tidpunkter och skapa sedan en indexserie från denna tidsserie. 5

d) Förklara vad "Centrala gränsvärdessatsen" innebär. 5

Uppgift 2

Man vill skatta en patients genomsnittliga pulstryck (differensen mellan systoliskt och diastoliskt blodtryck) under en 24-timmars period. Detta görs genom att patienten under 24 timmar bär en maskin som var tionde minut mäter blodtrycket (systoliskt och diastoliskt blodtryck). För patienten resulterade de totalt 144 mätningarna under 24 timmar i ett medelvärde av pulstrycket på 47 mm Hg och en standardavvikelse på 25 mm Hg.

a) Beräkna ett 95-procentigt konfidensintervall för patientens genomsnittliga pulstryck under 24-timmars perioden. Hur ska konfidensintervallet tolkas? 7

b) Ett genomsnittligt pulstryck högre än 45 mm Hg är onormalt. Undersök om patientens genomsnittliga pulstryck överstiger 45 mm hg med hjälp av ett lämpligt test. Beräkna och tolka p-värdet. 7

c) Hur många mätningar måste man göra under 24-timmars perioden om man vill att längden på ett 95-procentigt konfidensintervall för det genomsnittliga pulstrycket inte ska överstiga 4 mm Hg? 6

Uppgift 3

Man granskar koncentrationen av ett giftämne i havsbotten alldeles utanför en fabrik. Miljöföreskrifterna säger att koncentrationen inte får överstiga 12 (mg/cm^3). För att kontrollera detta tas prover från havsbotten precis utanför fabriken, resultaten visas nedan. Antag att provvärdena Y är normalfördelade.

11.7 12.4 12.8 12.9 13.3

a) Kan man dra slutsatsen att giftkoncentrationen på havsbotten alldeles vid fabriken är över 12 (mg/cm^3)? Utför ett lämpligt test på signifikansnivån 0.05.

Man misstänker att gifterna sprider sig från fabriken mer i östlig riktning än i västlig riktning pga av strömmar i vattnet. Man gör 35 mätningar 1 km öster om fabriken och lika många mätningar 1 km väster om fabriken. Giftkoncentrationen framgår nedan.

$\bar{x}_{öst} = 10.66$ $\quad s^2_{öst} = 12.1$

$\bar{x}_{väst} = 9.01$ $\quad s^2_{väst} = 11.6$

b) Är koncentrationen av giftämnet högre öster om fabriken jämfört med väster om fabriken? Utför ett lämpligt test på signifikansnivån 2.5 %.

Uppgift 4

Ett spel som du erbjuds spela ser ut på följande sätt. Ett mynt kastas, om "krona" kommer upp vinner du 2500 kronor. Om "klave" kommer upp vinner du ingenting. Insatsen i spelet är 1000 kronor.

a) Bör du spela spelet eller inte om du utgår från beslutskriteriet "Beslut under risk"? 6

b) Bör du spela spelet om du utgår fran att du vill minimera förlusten som kan uppstå av att välja fel alternativ. 6

Spelbolaget som erbjuder spelet gar back pa att erbjuda detta spel. De funderar pa att korrigera antingen vinstsumman eller sannolikheten att vinna.

c) Spelbolaget bestämmer sig för att korrigera vinststumman. Vid vilken vinstsumma går spelbolaget "break-even" d.v.s. vid vilken vinstsumma är det förväntade resultatet för Spelbolaget 0? Insatsen är oförändrad d.v.s. 1000 kronor. 4

d) Om man istället korrigerar sannolikheten att vinna (genom att frångå myntkastningen), vid vilken sannolikhet för vinst i spelet går spelbolaget "break-even". Insatsen och vinstsumman är oförändrade d.v.s. 1000 kronor respektive 2500 kronor. 4

Uppgift 5 20

En biolog mäter längden på 90 stycken ödlor av en viss sort. Tabellen nedan visar antalet ödlor i olika längdintervall. Testa med lämpligt test om data följer en normalfördelning med väntevärde 12 cm och standardavvikelse 3 cm. Använd signifikansnivån 5%.

Längd (cm)	Antal
under 6	9
6-8	10
8-10	13
10-12	18
12-14	16
14-16	9
16-18	9
över 18	8
Totalt	90 92

Tentamen SG2 2017-11-24

Uppgift 1

a) Ett Godness-of-fit test tester om en variabel (på ordinal och nominalskalenivå) följer en viss specificerad fördelning. Ett homogenitetstest tester istället om två variabler (på ordinal och nominalskalenivå) eller fler följer samma fördelning. Vilken fördelning de båda ev. har är inte specificerad.

Godness-of-fit

H_0: Fördelningen för variabel X i 3 olika kategorier är 0,25, 0,5 och 0,25

H_A —"— —"— —"— är inte 0,25, 0,5 och 0,25

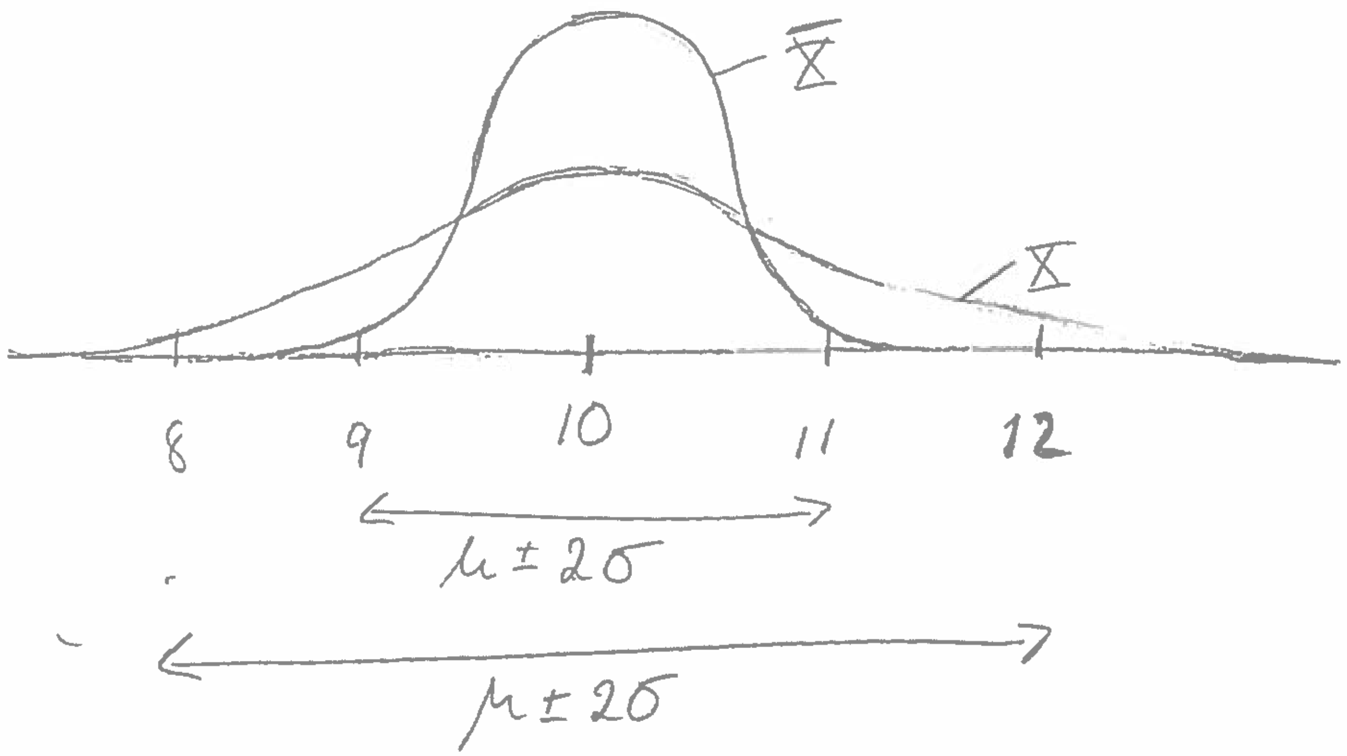
Homogenitetstest

H_0: Förd. för variabel X och Y i de olika kategorierna är desamma

b/ $X \sim N(10,4)$ $\sigma^2 = 4 \Rightarrow \sigma = 2$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\bar{X} \sim N(\mu, \frac{4}{16})$ $\sigma^2 = \frac{4}{16} \Rightarrow \sigma = \sqrt{\frac{4}{16}} = \frac{2}{4} = 0{,}5$



Ca 95% av alla observationer ligger inom 2 standardavvikelser från μ åt vardera håll i en normalfördelning.

9) En tidsserie är faktiska observationer av en variabel tex Pris vid olika tidpunkter. En indexserie visar procentuell förändring av tidsserien i förhållande till en vald bastidpunkt.

Tidsserie

t	Pris på vara	Index (bastid 2000) $I_t = \frac{X_t}{X_0} \cdot 100$
2000	45,3	100
2001	46,8	$\frac{46,8}{45,3} \cdot 100 = 103,31$
2002	49,5	$\frac{49,5}{45,3} \cdot 100 = 109,27$
2003	48,0	$\frac{48,0}{45,3} \cdot 100 = 105,96$
2004	50,2	$\frac{50,2}{45,3} \cdot 100 = 110,82$

d) <u>Centrala gränsvärdessatsen</u>

Samplingfördelningen för urvalsmedelvärdet närmar sig en normalfördelning då urvalsstorleken ökar. Detta gäller oavsett vilken fördelning urvalet är taget ifrån. Som direkt konsekvens av detta närmar sig samplingfördelningen för en Z-transformation av urvalsmedelvärdet ($Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$) en standardiserad normalfördelning då urvalsstorleken ökar.

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

Uppgift 2

X = pulstryck

$n = 144$

$\bar{x} = 47$

$s = 25$

Ett 95%-igt konfidensintervall för μ_X ges av:

$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ enligt CGS

$47 \pm 1{,}96 \frac{25}{\sqrt{144}}$

$47 \pm 4{,}0833$

$\boxed{[42{,}92; 51{,}08]}$

Av ett stort antal k.i. uträknade på samma sätt med samma stickprovs-storlek kommer 95% av dessa intervall täcka det sanna värdet på μ.

1) $H_0: \mu = 45$

$H_A: \mu > 45$

$Z = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim \text{approx } N(0,1)$ enligt CGS

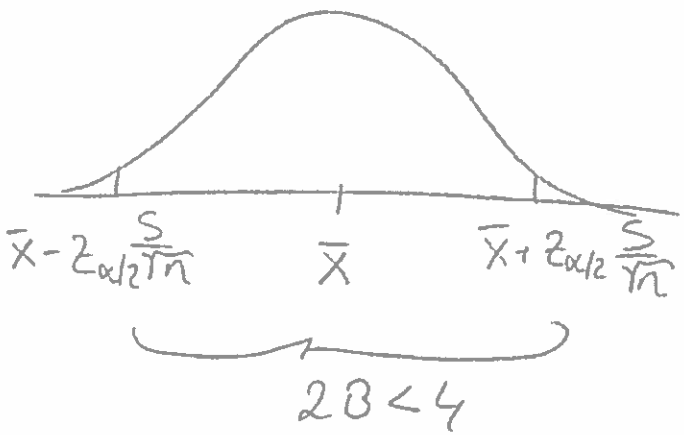
$Z_{obs} = \dfrac{47-45}{25/\sqrt{144}} = 0{,}96$

P-värdet $= P(Z > Z_{obs}) = P(Z > 0{,}96) = 1 - P(Z < 0{,}96)$

$= 1 - 0{,}83147 = 0{,}16853$

P-värdet är minsta signifikansnivå som H_0 kan förkastas på. H_0 förkastas på sign nivåer större än 16,9% men inte på nivåer lägre än 16,9%. Vi kan inte förkasta H_0 på några rimliga sign. nivåer så vi kan inte dra slutsatsen att det genomsnittliga pulsslaget överstiger 45.

c)



Totala längden $2B < 4 \Rightarrow B < 2$

$$B = z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$2 = 1{,}96\,\frac{25}{\sqrt{n}}$$

$$n = \left(\frac{1{,}96 \cdot 25}{2}\right)^2 = \underline{\underline{600{,}25}}$$

dvs 601 mätningar behövs.

Uppgift 3

$\underline{Y}$ = konc. av giftämne

a) $H_0: \mu = 12$

$H_A: \mu > 12$

$$\bar{X} = \frac{11,7 + 12,4 + 12,8 + 12,9 + 13,3}{5} = 12,62$$

$$S^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{11,7^2 + 12,4^2 + 12,8^2 + 12,9^2 + 13,3^2 - 5 \cdot 12,62^2}{4} = \frac{1,468}{4} = 0,367$$

Litet stickprov och normalfördelade data

$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$ om H_0 är sann.

$$t_{obs} = \frac{12{,}62 - 12}{\sqrt{0{,}367}/\sqrt{5}} = 2{,}29$$

H_0 förkastas om $t_{obs} > t_{0{,}05}(5-1) = 2{,}132$

$t_{obs} = 2{,}29 > 2{,}132$ så H_0 förkastas. Vi kan påvisa att giftkoncentrationen är större än 12.

, Oberoende observationer

$\bar{X}_{öst} = 10{,}66$ $S^2_{öst} = 12{,}1$ $n_{öst} = 35$

$\bar{X}_{väst} = 9{,}01$ $S^2_{väst} = 11{,}6$ $n_{väst} = 35$

$H_0 : \mu_{öst} = \mu_{väst}$

$H_A : \mu_{öst} > \mu_{väst}$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim \text{approx } N(0,1) \text{ enligt CGS}$$

$n > 30$ och $m > 30$

Beslutsregel: H_0 förkastas om $Z_{obs} > Z_{0,025} = 1{,}96$

$$Z_{obs} = \frac{10{,}66 - 9{,}01}{\sqrt{\frac{12{,}1}{35} + \frac{11{,}6}{35}}} = 2{,}005$$

(eller t-test)

$Z_{obs} = 2{,}005 > 1{,}96 = Z_{0,025}$

H_0 förkastas på sign nivån 2,5% och vi kan påvisa att koncentrationen är högre öster om fabriken jämfört med öster om fabriken

Uppgift 4

a) Nyttan = Vinstsumma − insats

Nyttomatris:

p	0,5	0,5	
	Krona	Klave	$E(U_i)$
Spela	2500−1000= 1500	0−1000= −1000	$E(U_1) = 1500 \cdot 0{,}5 - 1000 \cdot 0{,}5 = 250$
Spela inte	0	0	$E(U_2) = 0$

Beslut under risk utgår från högsta förväntade nyttan dvs Spela är det alternativ vi bör gå på.

b) Minimax-regel kriteriet

Maximala nyttan för de två tillstånden

$S_1^+ = 1500 \qquad S_2^+ = 0$

Alternativförlusten $r_{ij} = S_j^+ - U_{ij}$

Regretmatris:

	Krona	Klave	$a_i^r = \max(r_{i1}, r_{i2})$ Minimax regret Störst "regret"
Spela	1500 − 1500 = 0	0 − (−1000) = 1000	$a_1^r = 1000$
Spela inte	1500 − 0 = 1500	0 − 0 = 0	$a_2^r = 1500$

Vi väljer det alternativ som minimerar den maximala regreten, dvs alt. att spela.

c) X = Vinstsumma

	Slh	Spelbolagets resultat Intäkt (insats) − utgift (vinstsumma)
Krona (Vinst)	0,5	1000 − X
Klave (Ingen vinst)	0,5	1000 − 0 = 1000

Vid break-even är förväntat resultat = 0

$E(\text{resultat}) = (1000 - X) \cdot 0,5 + 1000 \cdot 0,5 = 0$

$500 - 0,5X + 500 = 0$

$0,5X = 1000$

$\underline{\underline{X = 2000}}$

— Svar: Vid en vinstsumma på 2000 kronor går spelbolaget break-even på spelet.

d) P = Slh för vinst

	Slh	Spelbolagets resultat
Vinst	P	$1000 - 2500 = -1500$
Ingen vinst	$1-P$	$1000 - 0 = 1000$

$E(\text{resultat}) = -1500P + 1000(1-P) = 0$

$-1500P + 1000 - 1000P = 0$

Svar: Vid en sannolikhet för vinst på 0,4 går spelbolaget break-even.

Uppgift 5

a) X = längd på ödla

Testa om $X \sim N(12, 3^2)$

Detta görs med ett Godness-of-fit test.

Om $X \sim N(12, 3^2)$ så är sannolikheten för respektive intervall:

$$P(X<6) = P\left(Z < \frac{6-\mu}{\sigma}\right) = P\left(Z < \frac{6-12}{3}\right) = P(Z<-2) = 1-P(Z<2) = 1-0{,}97725$$
$$= \underline{\underline{0{,}02275}}$$

$$P(6<X<8) = P(X<8) - P(X<6) = P\left(Z<\frac{8-12}{3}\right) - P(X<6) =$$

$$P(Z<-1{,}34) - 0{,}02275 = [1-P(Z<1{,}34)] - 0{,}02275 = (1-0{,}90988) - 0{,}0227\ldots$$
$$= \underline{0{,}06737}$$

$P(8 < X < 10) = P(X < 10) - P(X < 8) = P\left(Z < \frac{10-12}{3}\right) - P(X < 8) =$

$P(Z < -0{,}67) - (1 - 0{,}90988) = [1 - P(Z < 0{,}67)] - (1 - 0{,}90988) =$

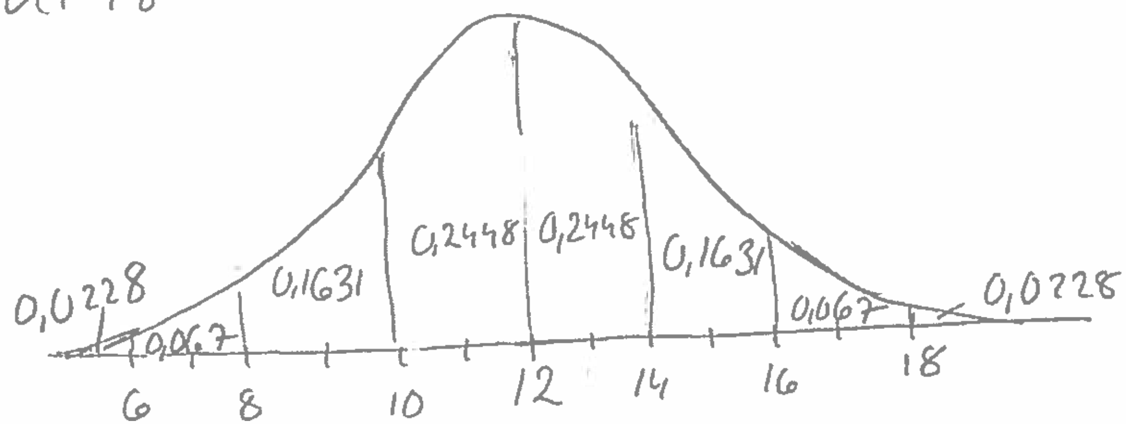
$= (1 - 0{,}74857) - (1 - 0{,}90988) = -0{,}74857 + 0{,}90988 = \underline{\underline{0{,}16131}}$

$P(10 < X < 12) = P(X < 12) - P(X < 10) = 0{,}5 - (1 - 0{,}74857) = \underline{\underline{0{,}2485}}$

Pga symmetri i normalfördelningen blir slh för intervallen till höger om $\mu = 12$ motsvarande som de till vänster om $\mu = 12$ dvs.

$\Rightarrow$

Längd	slh	Förväntad frekvens ($n=90$)	Obs. frekvens
Under 6	0,02275	0,02275 · 92 = 2,093	9
6 - 8	0,06737	0,06737 · 92 = 6,198	10
8 - 10	0,16131	0,16131 · 92 = 14,841	13
10 - 12	0,2448	0,2448 · 92 = 22,5216	18
12 - 14	0,2448	22,5216	16
14 - 16	0,16131	14,841	9
16 - 18	0,06737	6,198	9
över 18	0,02275	2,093	8
			92



Godness-of-fit test

H_0: Data är normalfördelad dvs enligt fördelningen på föregående sida

H_A: Data är ej fördelade enligt fördelningen på föregående sida. dvs n.f.

De förväntade frekvenserna i första och sista gruppen är mindre än 5 så vi slår ihop med intillliggande grupper.

	Upp till 8	8-10	10-12	12-14	14-16	16 och över
Obs. (n_i)	19	13	18	16	9	17
Förväntat $E(n_i)$	8,3	14,8	22,5	22,5	14,8	8,3

$E(n_i) \geq 5$

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E(n_i))^2}{E(n_i)} \sim \text{approx } \chi^2(k-1)$$

Beslutsregel: Förkasta om $\chi^2_{obs} > \chi^2_{0,05}(5) = 11{,}070$

$$\chi^2_{obs} = \frac{(19-8{,}3)^2}{8{,}3} + \frac{(13-14{,}8)^2}{14{,}8} + \frac{(18-22{,}5)^2}{22{,}5} + \frac{(16-22{,}5)^2}{22{,}5} + \frac{(9-14{,}8)^2}{14{,}8} + \frac{(17-8{,}3)^2}{8{,}3} = 28{,}18$$

$$\chi^2_{obs} = 28{,}18 > 11{,}070 = \chi^2_{0,05}(5)$$

H_0 förkastas. Vi kan på sign.nivån 5% påvisa att data inte är normalfördelat.