

TENTAMEN: Dataanalys och statistik för I2, TMS135

Fredagen den 12 mars kl. 8:45 - 11:45 på V.

Jour: Jenny Andersson, ankn 8294 (mobil:070 3597858)

Hjälpmedel: Utdelad formelsamling med tabeller, BETA, på kursen använd ordlista och typgodkänd räknedosa.

Poängberäkning: Uppgifterna är av flervalstyp, där endast ett alternativ är rätt. Korrekt besvarad uppgift ger 2 poäng, obesvarad uppgift (vet inte) ger 0 poäng och felaktigt besvarad uppgift ger -0.5 poäng (flera ifyllda alternativ ger automatiskt -0.5 poäng). Inlämnade lösningar kommer ej tas hänsyn till vid rättningen. Fyll i och lämna in denna sida.

Svar: Lägg ut i studieportalen efter tentamens slut.

Uppgift	a	b	c	d	e	Vet inte	Poäng
1	☐	☐	☐	☐	☐	☐	
2	☐	☐	☐	☐	☐	☐	
3	☐	☐	☐	☐	☐	☐	
4	☐	☐	☐	☐	☐	☐	
5	☐	☐	☐	☐	☐	☐	
6	☐	☐	☐	☐	☐	☐	
7	☐	☐	☐	☐	☐	☐	
8	☐	☐	☐	☐	☐	☐	
9	☐	☐	☐	☐	☐	☐	
10	☐	☐	☐	☐	☐	☐	
11	☐	☐	☐	☐	☐	☐	
12	☐	☐	☐	☐	☐	☐	
13	☐	☐	☐	☐	☐	☐	
14	☐	☐	☐	☐	☐	☐	
15	☐	☐	☐	☐	☐	☐	

- 1 Under kursen har vi grafiskt undersökt stationäritet hos data. Ett av nedanstående påståenden om stationäritet är riktigt. Vilket?
- a ☐ Stationära data har en växande eller avtagande trend.
 - b ☐ Oberoende är ett krav för inferensmetoderna som ingår i kursen. Vi kontrollerar stationäritet eftersom det är samma sak som oberoende.
 - c ☐ Oberoende är ett krav för inferensmetoderna som ingår i kursen. Vi kan med våra kunskaper inte kontrollera oberoende utan undersöker istället det svagare kravet att data skall vara stationära.
 - d ☐ Stationaritet avgör vilken fördelning man skall arbeta med.
 - e ☐ Inget av påståendena a-d är sant.
 - f ☐ Vet ej.

2 För $X \sim N(\mu, \sigma)$ med $\mu = 2$ och $\sigma^2 = 4$, vilket svarsalternativ är korrekt?

- a ☐ $P[X = 3.6] = 0$
- b ☐ $P[X = 3.6] = 0.7881$
- c ☐ $P[X = 3.6] = 0.2119$
- d ☐ $P[X = 3.6] = 0.7422$
- e ☐ $P[X = 3.6] = 0.9998$
- f ☐ Vet ej.

3 Om man jämför t-fördelningen med normalfördelningen kan man säga följande.

- a ☐ t-fördelningen har tunnare svansar än normalfördelningen.
- b ☐ t-fördelningen är mer högerskev än normalfördelningen.
- c ☐ t-fördelningen är mer vänsterskev än normalfördelningen.
- d ☐ Normalfördelningen har tjockare svansar än t-fördelningen
- e ☐ Inget av påståendena a-d är sant.
- f ☐ Vet ej.

- 4 Inom hypotestest så är det viktigt att ha kännedom om testets styrka. Vilket av följande påståenden om testets styrka är rätt om vi betraktar ett test med fix signifikansnivå α av

$$H_0 : \mu = \mu_0$$

mot

$$H_a : \mu < \mu_0?$$

- a ☐ Styrkan av ett hypotestest är hur ofta man kan förkasta H_0 .
- b ☐ Styrkan är en funktion av μ .
- c ☐ Styrkan av testet är sannolikheten att förkasta H_0 då H_0 är sann.
- d ☐ Styrkan kan ökas genom att minska signifikansnivån α .
- e ☐ Styrkan kan ökas genom att minska stickprovsstorleken.
- f ☐ Vet ej.

5 Vilken/vilka av följande grafiska analysmetoder kan användas för att bestämma om en process är stationär; Histogram, boxplot, plot av rådata mot tid, plot av glidande medelvärde mot tid, qq-plot?

- a ☐ Histogram, boxplot, qq-plot.
- b ☐ Boxplot, plot av rådata mot tid, plot av glidande medelvärde mot tid.
- c ☐ Plot av glidande medelvärde mot tid, qq-plot.
- d ☐ Plot av rådata mot tid, plot av glidande medelvärde mot tid.
- e ☐ Histogram, qq-plot.
- f ☐ Vet ej.

- 6 Hur kommer det sig att Centrala gränsvärdessatsen kan användas för att approximera fördelningsfunktionen för EN binomialfördelad stokastisk variabel om n är tillräckligt stor?
- a ☐ Binomialfördelningen och normalfördelningen är båda symmetriska för alla parametervärden.
 - b ☐ En binomialfördelad stokastisk variabel är en summa av oberoende, Bernoullifördelade stokastiska variabler.
 - c ☐ En binomialfördelad stokastisk variabel definieras som en funktion av en standardnormalfördelad stokastisk variabel.
 - d ☐ Alla fördelningar med samma väntevärde har automatiskt samma fördelningsfunktion.
 - e ☐ Inget av påståendena a-d är sant.
 - f ☐ Vet ej.

- 7 Du har samlat in n observationer från en kontinuerlig process med okänd varians. Vilken av följande faktorer påverkar **inte** bredden på ett konfidensintervall för väntevärdet.
- a ☐ Addition av en konstant till var och en av observationerna.
 - b ☐ Storleken på stickprovet, n .
 - c ☐ Storleken på skattningen av standardavvikelsen.
 - d ☐ Konfidensgraden.
 - e ☐ Alla ovanstående alternativ påverkar längden av konfidensintervallet.
 - f ☐ Vet ej.

8 ”Morgonstund har guld i mund” heter det ju. Du har bestämt dig för att i fortsättningen alltid gå upp klockan 5 på söndagsmorgonarna om du kan visa att p , andelen söndagsmorgonar med guld i mund, är större än 0.5. Du testar genom att 5 söndagsmorgonar gå upp klockan 5 . Du tycker dock att ingen av dessa har haft guld i mund en enda gång. Du vill alltså testa

$$H_0 : p = 0.5$$

mot

$$H_a : p > 0.5.$$

Vilken av följande är den lägsta signifikansnivå α du kan förkasta nollhypotesen på?

a ☐ 10%

b ☐ 5%

c ☐ 1%

d ☐ 0.1%

e ☐ Kan inte förkasta H_0 på någon av dessa signifikansnivåer.

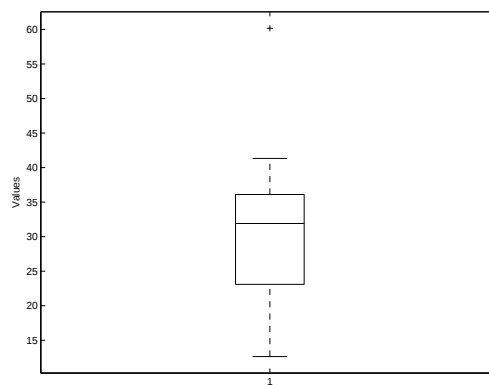
f ☐ Vet ej.

9 Eftersom detta är en flervalstenta kan man tänka sig att det går att chansa sig till att få godkänt (≥ 10 poäng). Vad är sannolikheten att få godkänt på tentan för någon som inte har en aning om rätt svar utan slumpmässigt väljer ett av alternativen a-e?

- a ☐ mindre än 0.0005
- b ☐ 0.018
- c ☐ 0.034
- d ☐ 0.048
- e ☐ 0.072
- f ☐ Vet ej.

10 Du har fått en boxplot av en viss datamängd enligt figuren nedan. Följande 4 påståenden vill du kunna göra om dina data:

- 1) Data är stationära.
- 2) Medelvärde är drygt 30 (ungefär 32).
- 3) Det finns en trolig outlier.
- 4) Data är approximativt normalfördelade.



Figur 1: Boxplot av 20 värden.

Kan verkligen alla dessa slutsatser dras genom att titta enbart på denna boxplot?

- a ☐ Ja.
- b ☐ Nej, påståenden 1 och 4 kan inte verifieras med en boxplot.
- c ☐ Nej, påståenden 1, 2 och 4 kan inte verifieras med en boxplot.
- d ☐ Nej, påståenden 1, 2 och 3 kan inte verifieras med en boxplot.
- e ☐ Nej, påståenden 2 och 3 kan inte verifieras med en boxplot.
- f ☐ Vet ej.

11 Enligt plikverket gäller följande för 17671 män inskrivna till värnplikt eller civilplikt 2003. Medelvärde av deras längder var 180.2 cm och standardavvikelsen 6.5 cm. Antag att de kan ses som ett stickprov på alla artonåringar i Sverige 2003. Mätningarna kan antas oberoende. Gör ett 99% prediktionsintervall för vikten hos nästa slumpvis utvalda artonåring som mäts.

a ☐ 180.2 ± 0.10 cm

b ☐ 180.2 ± 0.13 cm

c ☐ 180.2 ± 6.6 cm

d ☐ 180.2 ± 12.7 cm

e ☐ 180.2 ± 16.8 cm

f ☐ Vet ej.

- 12 Ett läkemedelsföretag har utvecklat en ny medicin för att lindra en viss sjukdom. Av 50 patienter med sjukdomen har symptomen hos 33 stycken lindrats efter en månads behandling med den nya medicinen. Ytterligare 50 sjuka patienter fick istället den mest använda medicinen på marknaden under en månad. Bland dessa lindrades symptomen hos 27 stycken. Låt p_1 vara sannolikheten att en slumpvis utvald patient får lindrigare symptom med den nya medicinen och p_2 motsvarande sannolikhet för patient som får den gamla medicinen. Vilken av nedanstående utsagor är sann vid ett test av

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2?$$

- a ☐ Testets p-värde ≤ 0.05 så nollhypotesen kan förkastas, d.v.s. den nya medicinen är bättre.
- b ☐ Testets p-värde ≥ 0.05 så nollhypotesen kan förkastas, d.v.s. den nya medicinen är bättre.
- c ☐ Testets p-värde ≤ 0.05 så nollhypotesen kan ej förkastas, d.v.s. det är ingen skillnad på medicinerna.
- d ☐ Testets p-värde ≥ 0.05 så nollhypotesen kan ej förkastas, d.v.s. det är ingen skillnad på medicinerna.
- e ☐ Inget av påståendena a-d är sant.
- f ☐ Vet ej.

13 Betrakta följande situationer:

- 1) Man tror att olja som transporteras i rör mellan de närliggande hamnarna A och B tar upp en viss förorening på vägen. Data samlas in från 10 rör och halten av föroreningen mäts i både hamn A och hamn B för varje rör.
- 2) För att testa om ett bedövningsmedel har olika effekt på pojkar och flickor, ges det till tvåäggstvillingpar med olika kön. Man ger medlet till både pojken och flickan bland 30 tvillingpar.
- 3) En steroids inverkan på vikten hos möss undersöks genom att 20 möss får ett foder med steroidtillskott och 20 andra möss får bara fodret utan steroidtillskott. Viktökningen hos mössen mäts efter en månad.
- 4) En författare vill undersöka om det går fortare att skriva med sin gamla skrivmaskin eller med datorn. Han väljer ut 15 texter och mäter tiden det tar att skriva av texterna på respektive maskin. Antag att inlärningseffekter kan försummas.

Följande av försöken analyseras bäst med hjälp av t-test för parade data.

- a ☐ 1 och 4 men inte 2 och 3.
- b ☐ 2 och 3 men inte 1 och 4.
- c ☐ 1, 2 och 4 men inte 3.
- d ☐ 2, 3 och 4 men inte 1.
- e ☐ 1, 2, 3 och 4.
- f ☐ Vet ej.

14 Under februari i förra året har NO_x halten mätts på Gibraltargatan. En plot av halt mot tid visar inte på något mönster som kan tyda på att mätningarna inte är stationära. Beräkna ett 99% konfidensintervall för NO_x haltens väntevärde. Resultatet av februarimätningarna var att medelhalten NO_x var 15 och att standardavvikelsen av NO_x -halten var 8.

a ☐ $15 \pm 2.473 \cdot 8 \sqrt{1 + 1/28}.$

b ☐ $15 \pm 2.771 \cdot 8 \sqrt{1 + 1/28}.$

c ☐ $15 \pm 2.467 \cdot 8 / \sqrt{28}.$

d ☐ $15 \pm 2.473 \cdot 8 / \sqrt{28}.$

e ☐ $15 \pm 2.771 \cdot 8 / \sqrt{28}.$

f ☐ Vet ej.

15 Man vet att inkomsterna hos ett visst företag under ett år följer en fördelning med väntevärde 1.5 miljoner kr och standardavvikelse 0.3 miljoner kr. Inkomsten kan betraktas som oberoende av företagets utgifter. Utgifterna under ett år följer i sin tur en fördelning med väntevärde 1.2 miljoner kr och standardavvikelse 0.1 miljoner kr. Vad är standardavvikelsen för företagets vinst (inkomster-utgifter) under ett år?

- a ☐ $\sqrt{0.08}$ miljoner kr.
- b ☐ $\sqrt{0.10}$ miljoner kr.
- c ☐ 0.10 miljoner kr.
- d ☐ 0.4 miljoner kr.
- e ☐ 0.2 miljoner kr.
- f ☐ Vet ej.