

SDAI (ST1101), Tentamen 2, 6 hp

Stockholms universitet, statistiska institutionen

Kurs: Statistik och dataanalys I, 15 hp

Tentamensdatum: 2024-03-17

Skrivtid: kl. 09–14 (5 timmar).

Godkända hjälpmedel: Miniräknare utan lagrade formler och text.

Bifogade hjälpmedel: Formel- och tabellsamling för statistik och dataanalys I, 15 hp.

Tentamen består av 5 uppgifter uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.

Svar med fullständiga redovisningar ska lämnas för fulla poäng.

- Använd endast skrivpapper som tillhandahålls i skrivsalen.
- För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
- Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan ge poängavdrag.
- Om du inte lyckas lösa en deluppgift och behöver det svaret för en senare deluppgift så kan du hitta på värdet för att kunna göra beräkningar i de efterföljande uppgifterna.
- I beräkningar från R-utskrifter får du utgå från det som är givet.

Tentamen kan maximalt ge 100 poäng. För godkänt resultat krävs minst 50 poäng.

Betygsgränser

A	90–100
B	80–89
C	70–79
D	60–69
E	50–59
Fx	40–49
F	0–40

Obs! Fx och F är underkända betyg som kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.

Lösningförslag läggs ut på athena efter att tentamenstiden är över.

Lycka till!

Uppgift 1 (16 poäng)

Låt A och B vara två oberoende händelser med $P(A) = 0.4$ och $P(B) = 0.5$.

- a) Vad är sannolikheten att både A och B inträffar? (4p)
- b) Vad är sannolikheten att åtminstone en av A och B inträffar? (4p)
- c) Vad är sannolikheten att varken A eller B inträffar? (4p)
- d) Illustrera händelserna A och B tillsammans med utfallsrummet S med ett Venndiagram. Markera även ut följande händelser i diagrammet (4p)
 - Både A och B inträffar
 - Varken A eller B inträffar

Uppgift 2 (22 poäng)

Ett lotteri har tryckt upp en miljard lotter. Varje lott kostar 30 Kronor. 75 % av lotterna är nitlotter, och 25 % är vinstlotter. Det är ett enkelt lotteri där varje vinstlott ger 100 Kronor i vinst.

- a) Du börjar köpa lotter och fortsätter tills du får en vinstlott. Vad är sannolikheten att du får din första vinst på femte lotten? (3p)
- b) Om du köper lotter tills du får en vinstlott, hur många lotter förväntar du dig att du behöver köpa? (2p)
- c) Vad är den förväntade förlusten (i Kronor) för varje lott du köper? (4p)
- d) Vad är sannolikheten att du får färre än två vinstlotter, om du köper tio lotter? (6p)
- d) Din kusin köper ett slumpmässigt antal lotter varje fredag, där antalet lotter kan antas följa en Poissonfördelning med $\lambda = 2$. Vad är sannolikheten att din kusin köper minst två lotter en given fredag? (7p)

Uppgift 3 (21 poäng)

Fitnessgurun Demona driver ett gym. Hon får ofta klagomål om att det spelas för mycket dunka dunka-musik, vilket leder till att folk spenderar mindre tid på gymmet. Eftersom medlemmarna i gymmet betalar en fast månadsavgift så inser Demona att det bara är bra om träningspassen är korta, och börjar därför spela ännu mer irriterande dunka-dunka för att se om hon kan korta ner tiden kunderna spenderar i gymmet ännu mer. Innan förändringen genomfördes var den genomsnittliga träningstiden 58.3 minuter per gympass.

- a) Bland 6 slumpmässigt valda kunder visar det sig att den genomsnittliga längden på deras gympass efter att musiken förändrats är $\bar{x} = 47.1$ minuter, med en standardavvikelse på $s = 18.4$. Utför ett hypotestest på 1% signifikansnivå för att undersöka om den genomsnittliga träningstiden har ändrats, och alltså skiljer sig från 58.3. Ställ upp hypoteser, utför testet och dra korrekta slutsaser. Vilket antaganden om populationsfördelningen krävs för att testet ska vara giltigt? (9p)
- b) Det visar sig att den genomsnittliga tiden innan förändringen är en skattning baserad på 7 observationer, med standardavvikelse $s = 15.6$. Genomför ett nytt hypotestest som undersöker om den genomsnittliga tiden skiljer sig före och efter förändringen som tar hänsyn till osäkerheten i träningstiden innan. Ställ upp hypoteser, utför testet och dra korrekta slutsaser. (12p)

Uppgift 4 (22 p)

Ett företag som genomför opinionsundersökningar har frågat ett slumpmässigt urval på 400 personer i Sverige om hur dom ställer sig till höjda bensinskatter. Av dom tillfrågade var det 180 som svarade att dom var positiva.

- a) Låt p vara andelen i populationen som är positiva till höjda bensinskatter. Beräkna ett approximativt 99% konfidensintervall för p . Tolka intervallet på ett sätt som visar att du *förstått* dess innebörd. Vilka antaganden behöver vara uppfyllda för att intervallet ska vara giltigt? (10p)
- b) Bland dom som tillfrågades bodde 327 i tätort och 73 på landsbygden (se tabell nedan). Använd ett chi2-test för att avgöra om inställningen till höjda bensinskatter är oberoende av var personen bor. Ställ upp hypoteser, utför testet och dra korrekta slutsatser. Använd signifikansnivå $\alpha = 0.05$. (9p)
- c) Nämn minst två av dom tre antagandena som krävs för att testet i 4b ska vara giltigt. Är antagandena uppfyllda? (3p)

	Positiv	Negativ	Total
Tätort	222	105	327
Landsbygd	18	55	73
Total	240	160	400

Uppgift 5 (20 poäng)

En skogsvaktare har mätt upp omkretsen på stammen (*circumference*, mätt i centimeter) på 35 apelsinträd av olika ålder (*age*, mätt i dagar) för att se hur denna utvecklas över tid. Följande populationsmodell används

$$\text{circumference} = \beta_0 + \beta_1 \text{age} + \varepsilon, \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Skattningen från regressionen hittar du i tabellen nedan.

- Tolka det skattade värdet *age*. Vad säger det om hur snabbt träden växer? Konstruera ett 90% konfidensintervall för β_1 . (8p)
- Skogsvaktaren inser att hon glömt att mäta ett av träden. Trädet hon glömde är 405 dagar gammalt, och den genomsnittliga åldern på träden är $\bar{x} = 674$ dagar. Konstruera ett 95% prediktionsintervall för omkretsen på trädet hon glömde att mäta. (8p)
- Ange de fyra grundantagandena för populationsmodellen. (4p)

Parameter estimates

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.39965	8.6226598	2.0179	5.1793e-02
age	0.10677	0.0082766	12.9002	1.9306e-14