| | |
|---|---|
| STOCKHOLMS UNIVERSITET | MT4001 |
| MATEMATISKA INSTITUTIONEN | TENTAMEN |
| Avd. Matematisk statistik, TB | 9 januari 2023 |

# Tentamen i Statistisk analys

## 9 januari 2023 kl. 14–19

*Examinator:* Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se
*Tillåtna hjälpmedel:* Formel- och tabellsamling och miniräknare.
*Återlämning:* Tentan kommer vara rättad senast 18/2 2023 och återfinns då vid matematik-expeditionen (Obs! Notera att institutionen flyttat till Albano!).

Varje korrekt löst uppgift ger 10 poäng. Gränsen för godkänt är preliminärt 30 poäng (inkl ev bonuspoäng). För D krävs 34 p, för C 40 p, för B 48 p och för A krävs 54 p. Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

---

Det skall tydligt framgå hur beräkningar gjorts. Kommunikation med andra personer är **ej** tillåtet och kommer anmälas vid uppdagande.

---

### Uppgift 1

Nedan följer 5 påståenden att svara sant eller falskt på (eller ingenting om man inte vet). Korrekt svar på respektive påstående ger 2p, fel svar ger -2p och inget svar ger 0p (om total-summan skulle bli negativ sätts poängen till 0).

**a)** Ett 95% konfidensintervall för en parameter $\theta$ har 95% av sannolikhetsfördelning för $\theta$ i intervallet.

**b)** Box-plot (sv: låda-gram) är lämpligt att använda när man ska jämföra olika stickprov.

**c)** Styrkan (eng: power) för att påvisa en avvikelse från nollhypotesen $H_0 : \theta = \theta_0$ avtar ju längre ifrån $\theta_0$ som det sanna värdet $\theta$ ligger.

**d)** Valet av apriorifördelning är viktigare ju fler observationer som data består av.

**e)** Ett icke-parametriskt test kan användas även när data är normalfördelat, men det tenderar att ha mindre styrka än motsvarande parametriska test.

### Uppgift 2

Vid en undersökning av två näringsextrakt uppmäts tid till att blommor vissnade. Det första extraktet testades på sex blommor och det andra extraktet på sex andra blommor (vilken blomma som fick vilken näring avgjordes med lottning). Resultatet, med dygn som tidsenhet, blev för blommorna med första extraktet: $x_1 = 3.1, x_2 = 2.2, x_3 = 4.5, x_4 = 3.9, x_5 = 4.9, x_6 = 3.6$, och med andra extraktet: $y_1 = 5.1, y_2 = 3.2, y_3 = 4.0, y_4 = 4.4, y_5 = 5.4, y_6 = 3.7$. Antag att observationerna kan betraktas som någorlunda normalfördelade och med samma standardavvikelse $\sigma$.

**a)** Beräkna $\bar{x}, \bar{y}, \sum_i x_i^2, \sum_i y_i^2$. (1 p)

**b)** Testa hypotesen att de två extrakten har samma väntevärde på 95%-nivån, under förutsättning att $\sigma = 1$. (3 p)

**c)** Gör motsvarande test där $\sigma$ skattas. (6 p)

## Uppgift 3

Temperaturen på julskinkan i ugnen kan antas öka hyfsat linjärt under dess ugnsstekning. Temperaturen uppmättes, genom stick in i skinkan på olika platser i mitten, varje halvtimme sedan den sattes in och gav följande temperaturer: 0.5h: 17, 1h: 26, 1.5h: 35, 2h: 47, 2.5h: 58. Vi antar att variationen beror på stickens placering och att det inte finns något beroende mellan mätvärden nära i tid.

Om det kan vara till någon nytta får följande användas (där $x$ är tid och $y$ temperatur): $\bar{x} = 1.5, \sum_i x_i^2 = 13.75, S_{xx} = 2.5$, respektive $\bar{y} = 36.6, \sum_i y_i^2 = 7763, S_{yy} = 1065.2$, samt $\sum_i x_i y_i = 326$ vilket ger $S_{xy} = 51.5$.

**a)** Hur mycket ökar temperaturen per timme, och vilken förväntad temperatur hade julskinkan när den sattes in i ugnen (ange skattningar och deras 95% konfidensintervall). (7 p)

**b)** Prediktera vilken temperatur skinkan kommer ha vid 3h, dvs ange ett 95% prediktionsintervall (under förutsättning att extrapolation fungerar). (3 p)

## Uppgift 4

Ett jaktlag jagar rådjur på två olika jaktmarker. Tiden (i timmar, med decimaltal där t ex 1.5 betyder en och en halv timme) mellan att rådjur observeras noterades 6 gånger på första jaktmarken och 9 gånger på andra jaktmarken. Tiderna på första jaktmarken var: 1.3, 6.1, 2.4, 13.6, 3.5, 1.9. På andra jaktmarken var tiderna 5.4, 8.3, 4.1, 7.4, 6.2, 2.9, 16.8, 6.5, 4.0. Man undrar om tiden tills man får syn på rådjur har ungefär samma fördelning på de två jaktmarkerna eller inte.

**a)** Varför verkar det inte rimligt att testa frågeställning med ett parametriskt test? (2 p)

**b)** Genomför lämpligt icke-parametriskt test för att avgöra om väntetiden har samma fördelning i de två jaktmarkerna eller inte. (8 p)

## Uppgift 5

En astronomistudie studerar tider mellan stjärnfall och dessa kan antas följa exponentialfördelningen med intensitetsparameter $\beta$, dvs med väntevärde $1/\beta$. 10 tider mellan stjärnfall uppmättes (tidsenhet timmar): 9.6, 2.4, 3.6, 4.7, 1.1, 0.5, 7.4, 2.1, 3.1, 0.9.

**a)** Härled en skattning av $\beta$ med momentmetoden och ange dess numeriska värde. (4 p)

**b)** Härled en skattning av $\beta$ med ML-metoden (maximum-likelihoodmetoden) och ange dess numeriska värde. (6 p)

## Uppgift 6

För att testa om artbeståndet på olika öar i Stockholms skärgård liknar varandra begav sig ornitologer ut till 3 olika öar i skärgården, en i södra, en i mellersta och en i norra. Vid samtliga öar artbestämde man de första 40 skådade exemplaren med följande utfall.

| Ö-lokalisering | Starar | Måsar (m fl) | Sångare | Övrigt | Total |
|---|---|---|---|---|---|
| Södra | 16 | 7 | 10 | 7 | 40 |
| Mellersta | 8 | 14 | 9 | 9 | 40 |
| Norra | 9 | 10 | 16 | 5 | 40 |
| Totalt | 33 | 31 | 35 | 21 | 120 |

**a)** Testa huruvida fågelbeståndet vad gäller artfördelning verkar skilja sig åt eller inte. Använd 95% signifikansnivå. (5 p)

**b)** Antag att de tre öarna tillsammans är representativa för fågelbestådet i hela skärgården. Vilken fördelning har det totala antalet exemplar av starar (som när data samlats in blev 33)? Använd detta för att skapa ett 99% konfidensintervall för andelen $p$ av skärgårdens fåglar som är starar. (5 p)

*Lycka till!*