STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
TENTAMEN
16 februari 2023

Tentamen i Statistisk analys

16 februari 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 2/3 2023 och återfinns då vid matematikexpeditionen.

Varje korrekt löst uppgift ger 10 poäng. Gränsen för godkänt är preliminärt 30 poäng. För D krävs 34 p, för C 40 p, för B 48 p och för A krävs 54 p (alla gränser gäller inkl ev bonuspoäng). 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 totalsumman skulle bli negativ sätts poängen till 0).

a) Ett p-värde anger sannolikheten att observera de data som föreligger.

b) Ett prediktionsintervall tar hänsyn till både osäkerhet i parameterskattningar *och* att framtida observationer i sig innehåller slump.

c) En normalfördelningsplot skapas genom att det storleksordnade stickprovet transfomeras med inversen till normalfördelningsfunktionen.

d) En maximimlikelhoodskattning är en punktskattning.

e) Styrkan i ett parametriskt test är (oftast) lägre jämfört med styrkan i icke-parametriska test.

Uppgift 2

Ett företags 10 telefonförsäljares försäljningsresultat månaden innan respektive efter säljarna genomgick en påkostad kurs blev enligt följande (före, efter): (10.8, 12.7), (14.1, 35.2), (21.1, 20.3), (15.7, 17.9), (12.5, 19.4), (14.7, 15.3), (12.9, 12.6), (16.4, 20.1), (13.5, 14.9), (19.6, 24.6). Frågan är om det finns någon signifikant skillnad på försäljningen före respektive efter kursen, eller om dessa har samma fördelningen.

a) Motivera varför en icke-parametrisk metod kan vara att föredra för dessa data. (2 p)

b) Testa med lämplig icke-parametrisk metod om det föreligger någon skillnad (använd 5% signifikansnivå). (8 p)

Uppgift 3

Istjockleken vid 5 småsjöar mättes och frågeställningen berör hur istjockleken påverkas av sjöarnas nord/sydläge mätt i latitud. Följande resultat erhölls per den 31/1 2023, (latitud, istjocklek i cm): (60, 12), (62, 23), (64, 40), (66, 49), (68, 66). (Påhittade data)

a) Ansätt lämplig förklaringsmodell för hur istjockleken beror av latituden. (2 p)

b) Skatta hur mycket tjockare isen i genomsnitt blir per latitud (4 p)

c) En ny sjö belägen på latitud 63 skall mätas. Skapa ett 95% prediktionsintervall för hur tjock isen kommer vara där (samma datum). (4 p)

Uppgift 4

I syfte att undersöka skärmtiden hos svenskar genomfördes en undersökning bland ungdomar som ombads ange deras genomsnittlga skärmtiden (per dag) den senaste veckan. Resultatet bland 10 slumpvis valda ungdomar blev enligt följande (i timmar/dag): 1.7, 2.5, 3.1, 0.9, 4.0, 3.0, 2.7, 4.1, 1.9, 3.6. Om det kan vara till någon nytta ger dessa värden $\bar{x} = 2.75$ och $s_x = 1.03$

a) Ange på lämpligt sätt den information undersökningen gav om den genomsnittliga skärmtiden bland svenska ungdomar. (4 p)

b) Samtidigt gjordes en undersökning även bland unga vuxna angående skärmtid med följande resultat (8 slumpvis valda testpersoner): 2.1, 0.5, 0.9, 1.4, 0.8, 2.6, 2.7, 0.2, vilket ger följande medelvärde och standardavvikelse: $\bar{y} = 1.40$ och $s_y = 0.96$

Ger studien tecken på att ungdomar och vuxna har lika eller olika genomsnittlig skärmtid? Genomför lämpligt statistiskt test. (6 p)

Uppgift 5

Vid tentamen den 9/1 2023 i Statistisk analys blev den totala tentapoängen (x) (utan bonus) samt bonuspoäng (y) enligt följande för de 31 skrivande (obs: verkliga data!): (15, 0), (46, 5), (30, 0), (15, 0), (28, 0), (40, 2), (34, 1), (15, 0), (35, 2), (54, 2), (23, 1), (18, 0), (47, 4), (49, 2), (56, 5), (42, 4), (10, 1), (30, 2), (50, 3),(25, 5), (46, 0), (42, 3), (38, 1), (52, 5), (43, 3), (42, 3), (43, 5), (53, 5), (56, 6), (60, 6), (48, 1), (0, 0).

Dessa data ger följande summor: $\sum_i x_i = 1185$, $\sum_i y_i = 77$, $\sum_i x_i^2 = 51199$, $\sum_i y_i^2 = 315$ och $\sum_i x_i y_i = 3527$.

a) Vi ska använda Pearsons korrelation. Vilka antaganden/approximationer förutsätter vi därmed? (2 p)

b) Beräkna Pearsonkorrelationen för datamaterialet, samt formulera och testa hypotesen att tentamensresultatet (utan bonuspoäng) inte påverkas av hur mycket bonuspoäng man erhållit. (4 p)

c) Gör ett (approximativt) 95% konfidensintervall för den teoretiska korrelationen r baserat på insamlat data. (4 p)

Uppgift 6

Betrakta "sluttande-planet-fördelningen" med täthetsfunktion $f(x;\theta) = 1 + \theta - 2\theta x$, för $0 \leq x \leq 1$. Möjliga parametervärden är $-1 \leq \theta \leq 1$. Ett antal observationer $x_1, \ldots, x_n$ görs i syfte att skatta θ.

a) Rita upp tätheten för $\theta = -0.5$ respektive för $\theta = 0.8$. (2 p)

b) Härled $\tilde{\theta}_{MM}$, dvs momentmetodens skattning av θ. (6 p)

c) Vad blir skattningen för följande 6 x-observationer: 0.92, 0.34, 0.56, 0.83, 0.49, 0.79. (2 p)

Lycka till!