

Tentamen i Statistisk analys

13 januari 2020 kl. 9–14

Examinator: Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se

Tillåtna hjälpmedel: Formel- och tabellsamling, samt miniräknare. Allt delas ut.

Återlämning: Tisdag 14/1 kl. 16.00 i sal 321, hus 6(!). Den som vill ha tentamensbesked per e-post ombeds skriva ett e-brev till Tom Britton (tomb@math.su.se) senast 14/1 kl 16.00.

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. Eventuella bonuspoäng adderas till tentamensresultatet. Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

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).

- Vid regression är observationerna (responsvariablerna) Y_1, \dots, Y_n likefördela men inte oberoende.
- Ett hypotestest som inte förkastar H_0 visar *inte* att H_0 är statistiskt bevisad.
- Vid Bayesiansk statistisk analys blir slutsaten i termer av apriorifördelningen olika beroende på vilken apriorifördelning man initialt väljer.
- Om insamlade data är normalfördelade så ger icke-parametriska metoder felaktiga resultat.
- Om data är normalfördelat visas i en normalfördelningsplot om punkterna ligger längs en horisontell linje, medan ju mer kurvan lutar ju mindre normalfördelade är data.

Uppgift 2

Ett stickprov av storlek 11 samlas in från en okänd men "snäll" fördelning (så normalfördelningsmetoder kan användas), där man vill skatta väntevärdet $\mu = E(X)$. Observationerna blev som följer: 56,7, 66,6, 57,8, 63,1, 59,2, 59,3, 57,4, 60,4, 57,6, 63,5, 55,7. Det gäller att $\sum_i x_i = 657.3$ och $\sum_i x_i^2 = 39389.45$.

- a) Skatta μ samt ge ett 95% konfidensintervall för μ . (7 p)
- b) Om man inte vill använda metoder baserade på normalfördelningen/centrala gränsvärdesatsen, men är villig att anta att fördelningen är symmetrisk runt sitt väntevärde (vilket data talar för). Hur skulle du då skatta μ med någon icke-parametrisk metod, och vad blir skattningen för dessa data? (3 p)

Uppgift 3

För att undersöka hur medelvikten på kvinnor påverkas av kvinnornas längd samlades 10 slumpvis valda kvinnor in och deras längd (x) och vikt (y) uppmättes. Resultatet blev som följer (x_i, y_i) : (162, 54), (168, 61), (166, 59), (173, 65), (169, 66), (167, 60), (177, 67), (171, 63), (164, 58), (170, 60). Mätvärdena ger följande summor som kan användas: $\sum_i x_i = 1687$, $\sum_i x_i^2 = 284769$, $\sum_i y_i = 613$, $\sum_i y_i^2 = 37721$, $\sum_i x_i y_i = 103553$.

- a) Ansätt lämplig modell och ange dess antaganden, och skatta parametererna. (3 p)
- b) Skapa ett 99% konfidensintervall för hur mycket vikten ökar i genomsnitt vid 1 cm längdökning. (4 p)
- c) Ytterligare en slumpvis vald kvinna väljs ut, och det visar sig att hon är 165 cm lång. Skapa ett 95% prediktionsintervall för hennes vikt. (3 p)

Uppgift 4

Frågan om hur det går för distansstudenter i jämförelse med vanliga kursstudenter har kommit upp. Antag att man jämförde resultaten för ett antal skrivande för att se om det föreligger någon skillnad på väntevärdet för respektive grupps resultat. Följande data erhöles (9 skrivande vanliga studenter och 8 distansstudenter):

Vanliga studenter: 25, 47, 33, 38, 58, 43, 20, 51, 34

Distansstudenter: 10, 31, 48, 32, 5, 36, 56, 37.

Testa på 95%-nivån om det föreligger någon statistisk säkerställd skillnad mellan de två studentkategorierna, under antagandet att data kan antas vara någorlunda normalfördelade (antag även att de två stickproven har samma varians trots att det inte riktigt tycks stämma). Ange vilka förutsättningar som antas och vilka hypoteser som testas mot varandra. (10 p)

Uppgift 5

Testa om de två studentgrupperna i uppgift 4 skiljer sig signifikant åt utan att anta något om att data liknar normalfördelningen, dvs genom att testa om de två studentgrupperna skiljer sig signifikant med hjälp av lämplig icke-parametrisk metod. (10 p)

Uppgift 6

I en medicinsk studie användes två olika behandlingsmetoder på allrvarligt sjuka patienter, och dessa följdes upp efter ett år. Dessa kunde efter ett år vara friskförklarade, fortfarande sjuka, eller avlidna. 30 patienter lottades till varje behandling och efter ett år blev utfallet som följer.

Metod	Frisk	Sjuk	Avliden	Total
Metod A	7	15	8	30
Metod B	13	5	12	30
Summa	20	20	20	60

Kan man baserat på detta hävda att de två behandlingsmetoderna skiljer sig signifikant åt på 5%-nivån?

(10 p)

Lycka till!