STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
TENTAMEN
16 februari 2022

Tentamen i Statistisk analys

16 februari 2022 kl. 8–13

Examinator: Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se
Tillåtna hjälpmedel: Formel- och tabellsamling (delas ut) och egen miniräknare (kan även lånas).
Återlämning: Måndag 21/2 kl 12.15 i rum 321 i hus 5 (obs kontorshuset!) kommer alla få se sina tentor, följt av en kortare genomgång av vanliga fel. (Tentor kan tas med hem först senare om de hämtas ut på expeditionen - de måste då kopieras).

Varje korrekt löst uppgift ger 10 poäng. Eventuella bonuspoäng adderas till tentamensresultatet. 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. Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

Lösningar till uppgifterna måste göras självständigt och 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) En punktskattning θ^* av parametern θ får innehålla (dvs vara funktion av) både data $(x_1, ..., x_n)$ och θ.

b) Antag att ett 95% konfidensintervall för θ har beräknats och blivit $(7.2,\ 10.4)$. Det betyder att $P(7.2 \leq \theta \leq 10.4) = 0.95$.

c) Ett av antagandena vid regressionsanalys är att $Y_1, \ldots Y_n$ är oberoende av varandra.

d) För ett stickprov från en fördelning för positiva talaxeln med tjock högersvans är det rimligt att anta att $\bar{x} > \tilde{x}$, där $\tilde{x}$ betecknar stickprovets median.

e) Ett konfidensintervall för väntevärdet μ i en normalfördelningen tenderar att ha kortare intervallbredd om σ är känd jämfört med om σ är okänd.

Uppgift 2

Följande stickprov från en (snäll) okänd fördelning erhölls: 10.7, 9.6, 11.2, 10.4, 9.9, 10.3, 10.5, 10.9, 9.5, 10.2. Flöjande gäller för dessa data: $\sum_i x_i = 103.2$ och $s = \sqrt{\sum_i (x_i - \bar{x})^2/(n-1)} = 0.545$.

a) Testa hypotesen att $H_0 : \mu = 10$ mot alternativet $H_1 : \mu \neq 10$ på 95%-nivån. (5 p)

b) Konstruera att två-sidigt 99% konfidensintervsll för μ. (5 p)

Uppgift 3

Antag att följande 6 observationer kommer från modellen för enkel linjär regression ($Y_i = \alpha + \beta x_i + \varepsilon_i$ med oberoende normalfördelade fel ($\varepsilon_i \sim N(0,\ \sigma^2)$). Datapunkterna ges av (x_i, y_i) = (0.0, -0.72), (2.4, 2.52), (4.0, 3.72), (5.1, 5.81), (6.7, 6.99) och (8.2, 7.68).

a) Skatta α och β samt ange 95%-konfidensintervall för respektive skattning (om du vill kan du utan poängavdrag anta att $\sigma = 1$). För ovan nämnda data gäller $\bar{x} = 4.4$, $\bar{y} = 4.33$, $\sum_i x_i^2 = 159.9$, $S_{xx} = 43.74$ och $S_{xy} = 45.97$. (5 p)

b) Härled MK-skattningen för β och visa vilken fördelning den har (Obs: du får utan bevis stoppa in skattningen för α: $\alpha^* = \bar{y} - \beta^*\bar{x}$). (5 p)

Uppgift 4

På tentamen i Statistisk analys den 12/1 2022 skrev 39 studenter. Själva tentaresultatet (x), dvs utan eventuella bonuspoäng, gav ett tal mellan 0 och 60. Dessutom kunde man erhålla mellan 0 och 6 bonuspoäng (y). Varje skrivande student gav således upphov till ett par (x_i, y_i). För att underlätta ges här bara följande relevanta summor (obs verkliga data!) $\sum_i x_i = 1447$, $\sum_i x_i^2 = 60933$, $\sum_i y_i = 106$, $\sum_i y_i^2 = 388$, $\sum_i x_i y_i = 4444$.

a) Beräkna datamaterialets (Pearson) korrelation ρ^*. (3 p)

b) Testa hypotesen $H_0 : \rho = 0$ mot den ensidiga hypotesen $\rho > 0$ (använd parametrisk inferens). (5 p)

c) (Obs: Kan göras även om du inte gjort a och b!). Förklara i ord vad det innebär *om* det gäller att $\rho > 0$. (2 p)

Uppgift 5

En pilkastare noterar hur många kast hon behöver för att kasta pilen i Bull's eye (=tavlans mittpunkt. Hon upprepar detta experiment 30 ggr. Första försöket krävdes 2 kast (inkl kastet som hamnade i Bull's eye), nästa försök 4 kast, därefter endast 1 kast, osv. Utfallen i dessa 30 omgångar sammanfattas enligt följande:

Antal kast	1	2	3	≥ 4
Antal omgångar	10	10	6	4

Det var alltså t ex 6 av de 30 omgångarna som resulterade i att pilkastaren träffade Bull's eye på tredje kastet. Vi ska nu testa om det verkar rimligt att pilkastaren träffar Bull's eye oberoende varje gång och med sannolikhet $p = 0.5$ varje gång. Under denna förutsättning (dvs H_0) gäller att antal kast som krävs följer för-första-gången fördelningen, dvs $P(X = k) = p(1-p)^{k-1}$, för $k = 1, 2, ...$ (med $p = 0.5$).

a) Bestäm under H_0: $E(N_i)$ = förväntat antal av de 30 omgångarna som krävde exakt i kast för $i = 1, ...3$ och $E(N_{4+})$ = förväntat antal som krävde minst 4. Om du inte kan räkna ut $E(N_{4+})$ så blir den 30 minus övriga 3 storheter. (2 p)

b) Om H_0 är sann så bör ju de observerade värdena n_i (i Tabellen ges dessa, t ex så är $n_1 = 10$) vara ganska lika dessa väntevärden. Beräkna därför följande kvadratsumma $Z = \sum_{i=1}^{4}(n_i - E(N_i))^2/E(N_i)$ där den sista termen ska vara "4+" (dvs minst 4). (3 p)

c) Under H_0 så gäller att Z är approximativ χ^2-fördelad med antal frihetsgrader svarande mot antal olika utfall i tabellen minus 1 (vi vet ju att summan är 30). Genomför detta test på 95%-nivån för att dra slutsats om data signifikant avviker från För-första-gången fördelningen eller ej. (5 p)

Uppgift 6

Två oberoende stickprov samlades in från två fördelningar med väntevärde μ_1 och standardavvikelse σ_1, respektive μ_2 och standardavvikelse σ_2. Det första stickprovet innehöll 10 observationer och är det som ges i uppgift 2 och som således har $\bar{x} = 10.320$ och en standardavvikelse $s_1 = 0.545$. Det andra stickprovet innehåller endast följande 8 observationer (y_j): 8.4, 7.3, 10.6, 9.5, 8.8, 10.2, 9.1, 11.0. Dessa observationer har medelvärde $\bar{y} = 9.362$ och standardavvikelse $s_2 = 1.225$.

a) Vi vill testa $H_0: \mu_1 = \mu_2$ mot två-sidig alternativhypotes, utan att anta att de två stickproven har samma standardavvikelse (vilket ju inte verkar gälla). Bilda en lämplig referensvariabel och beräkna dess värde. (5 p)

b) Testa H_0 på 99%-nivån. (5 p)

Lycka till!