STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT 4001
TENTAMEN
11 februari 2020

Tentamen i Statistisk analys

11 februari 2020 kl. 9–14

Examinator: Tom Britton, tel. 08-164534, tom.britton@math.su.se
Tillåtna hjälpmedel: Formel- och tabellsamling, samt miniräknare. Allt delas ut.
Återlämning: Torsdag 13/2 kl. 12.30 i sal 321, hus 6(!). Den som vill ha tentamensbesked per e-post ombeds skriva ett e-brev till Tom Britton (tom.britton@math.su.se) senast 13/2 kl 12.30.

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).

a) Median är mer lämpligt som lägesmått än medelvärde utifall data är tjocksvansat (dvs innehåller extrema observationer).

b) Bortfall vid statistiska undersökningar är problematiskt eftersom det leder till färre observationer. Men om man samlar in fler observationer så att antalet erhållna observationer blir enligt planerat blir slutsatserna korrekta så enda problemet är att arbetsinsats/kostnad blivit större.

c) Bland 100 hypotesprövningar där nollhypotesen faktiskt är sann (vilket de som utför testet förstås inte vet) kan man förvänta sig att ca 5 av dem kommer resultera i att nollhypotesen förkastas på 5%-nivån.

d) Vid test av oberoende i en kontingenstabell med k rader och j kolumner blir det $k \cdot j - 1$ frihetsgrader i χ^2-testet.

e)

Uppgift 2

Vikten på 10 slumpvis valda trädgårdsjordsäckar av ett visst fabrikat uppmättes: 19.7, 20.3, 20.1, 20.5, 19.9, 20.0, 20.3, 20.6, 20.1, 20.2. Summan av vikterna och deras kvadratsumma ges av: $\sum_{i=1}^{10} x_i = 201.7$ och $\sum_{i=1}^{10} x_i^2 = 4068.95$.

a) Gör ett 90% konfidensintervall för genomsnittliga vikten på dessa jordsäckar. (6 p)

b) Testa på 95%-nivån om medelvikten överstiger 20 kg eller inte. (4 p)

Uppgift 3

För att bilda sig en uppfattning om hur årsmedeltemperaturen sjunker norrut i Sverige används enkel linjär regression. Årsmedeltemperaturen har samlats in från $n = 11$ orter i Sverige med känd lattitud: Jokkmokk: latt: 66.6, medeltemp: -0.6, Umeå: 63.5, 4.0, Östersund: 63.1, 4.2, Gävle: 60.4, 5.8, Karlstad: 59.2, 7.0, Stockholm: 59.3, 7.6, Jönköping: 57.4, 6.0, Visby: 57.6, 7.6, Göteborg, 57.8, 7.7, Kalmar: 56.7, 7.5, Lund: 55.7, 8.5. Följande sammanfattande storheter beräknades från data (x är lattitud och y årsmedeltemperatur): $\sum_i x_i = 657.3$, $\sum_i x_i^2 = 39389.45$, $\sum_i y_i = 65.3$, $\sum_i y_i^2 = 455.95$ samt $\sum_i x_i y_i = 3820.38$.

a) Skatta och ange ett 99% konfidensintervall för hur mycket medeltemperaturen sjunker per lattitud (inom Sverige). Obs: endast konfidensintervall för efterfrågad storhet ska anges. (En korrekt lösning under antagandet att $\sigma = 1$ ger 3p.) (5 p)

b) Vilken förklaringsgrad R^2 har modellen? (2 p)

c) Skatta medeltemperaturen för Bollnäs med lattitud 61.34 (endast punktskattning). (3 p)

Uppgift 4

Vid en tentamen i matematisk statistik noterades anmälningsnumret (som kommer i tidsordning) och tentamensresultatet (maxpoäng=40). Följande data erhölls bland 10 skrivande.

Anmälningsnummer	1	2	3	4	5	6	7	8	9	10
Tentamensresultat	38	30	24	28	20	25	18	12	15	7

Om det kan vara till någon användning kommer här några kvadratsummor: $\sum_i (x_i - \bar{x})^2 = 85.5$, $\sum_i (y_i - \bar{y})^2 = 762.1$, och $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = -238.73$.

a) Ange hur stort det *linjära sambandet* är mellan anmälningsnummer och tentamensresultat, och testa på 95%-nivån att det inte råder något sådant samband. (5 p)

b) Ange hur stort *ordningssambandet* är mellan anmälningsnummer och tentamensresultat, och testa på 95%-nivån att det inte råder något sådant samband. (5 p)

Uppgift 5

Antalet bilolyckor med dödligt utfall var i Stockholms län 10 på varandra följande år: 4, 6, 9, 14, 8, 6, 12, 7, 10, 11. På goda grunder kan man anta att antalet olyckor med dödlig utgång är oberoende och likafördelade, och att fördelningen bör vara $Poi(\lambda)$.

a) Härled ML-skattningen för λ och ange dess numeriska värde. (5 p)

b) Bilda ett approximativt 95% konfidensintervall för λ. (5 p)

Uppgift 6

Vikten på fullvuxna gäddor kan antas vara normalfördelad med väntevärde μ och $\sigma = 0.5$ kg, där μ varierar mellan olika sjöar resp olika delar av östersjön. Baserat på tidigare mätningar i ett flertal vattenstån så vet man att μ varierar mellan vattenstånd enligt en normalfördelningen med väntevärde $\mu_0 = 3.0$ kg och stndardavvikelse $\sigma_0 = 1.0$ kg. En ny sjö skall mätas och antag att du väljer den senare fördelningen, dvs $N(3.0, 1.0)$ som apriorifördelningen för μ. 10 slumpvis infångade fiskar fångas därefter in med vikter $x_1, \ldots, x_{10}$, med ett medelvärde på $\bar{x}$=3.50 kg.

a)

Härled vilken aposteriorifördelning som μ får baserat på data och vald apriorifördelning. (7 p)

b) Vad blir din bästa punktskattning av μ (motivera ditt svar)? (3 p)

Lycka till!