

Tentamen i Statistisk analys

22 augusti 2024 kl. 14–19

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

Tillåtna hjälpmedel: Formel- och tabellsamling (delas ut) och miniräknare.

Återlämning: Tentan kommer vara rättad senast två veckor efter tentamensdagen. När tentan är rättad meddelas detta via kursforumet och därefter kan tentan hämtas ut 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.

Tillägg till formelsamlingen: Om observationer sker oberoende med k antal möjliga utfall och sannolikheten för ett utfall i under H_0 är p_i , så kan H_0 testas genom att bilda $Q = \sum_i (n_i - e_i)^2 / e_i$, där n_i är antalet observationer som resulterade i utfall i och e_i motsvarande förväntat antal observationer under H_0 (summan kan i vissa fall skrivas som en dubbelsumma).

Under H_0 är $Q \sim \chi^2$ vilket gäller approximativt om alla $e_i \geq 5$. Antalet frihetsgrader är antalet möjliga olika utfall subtraherat med antal icke-fria sådana (beroende på antal skattade parameterar och att totalt antal observationer är givet).

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) Under förutsättning att X har ändligt väntevärde ($E(X) = \mu$ med $|\mu| < \infty$) så är medelvärdet $\bar{x} = (x_1 + \dots + x_n)/n$ av ett stickprov från X en väntevärdesriktig skattning av μ .

b) En medicinsk studie med mätresultaten frisk, lätt symptom, allvarlig symptom och livshotande symptom, är på *nominalskala*.

- c) Oberoende upprepade 95% konfidensintervall av en parameter θ täcker in det sanna värdet θ i ca 95% av fallen.
- d) En viktig förutsättning för regressionsanalys är att Y -observationerna är likafördelade, men inte nödvändigtvis oberoende av varandra.
- e) Minstakvadratskattningen θ_{MK}^* av en parameter θ är en punktskattning av θ .

Uppgift 2

Ett stickprov av storlek åtta från en okänd kontinuerlig fördelning med täthet f samlades in. Följande värden erhöles: 13.4, 14.1, 15.1, 14.3, 14.9, 13.9, 14.4, 14.0. Vi kan anta att fördelningen f är tillräckligt snäll så att vi kan normalapproximera medelvärden från stickprov. Till er hjälp har följande beräknats: $\sum_i x_i = 114.1$, samt $\sum_i x_i^2 = 1629.45$. Låt μ beteckna fördelningens väntevärde och σ dess standardavvikelse.

- a) Konstruera ett 95% konfidensintervall för μ under förutsättning att $\sigma = 0.5$ är känt på förhand. (5 p)
- b) Konstruera ett 95% konfidensintervall för μ under förutsättning att σ inte är känt på förhand. (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

I ett seniorbasketlag (dvs 50+) bestod de 12 spelarna av 6 australienskor och 6 svenskor. Månadslönerna för australienskorna översatt till tkr var (i växande ordning) 23.4, 25.1, 29.4, 42.8, 49.3 och 95.0 tkr. För svenskorna vara lönerna 35.1, 47.3, 50.2, 53.0, 61.8 och 80.4. Frågan man vill ha besvarad är om lönerna skiljer sig åt mellan spelarna från de två länderna (under antagandet att de är representativa).

- a) Varför är det olämpligt att jämföra medellönerna mellan de två grupperna? (2 p)
- b) Genomför lämpligt icke-parametrisk test för att avgöra om lönerna är någorlunda lika varandra eller skiljer sig åt signifikant. Välj signifikansnivå själv. (8 p)

Uppgift 5

Man vill undersöka om preferensen för olika inriktningar på studierna vid Matematiska institutionen skiljer sig åt mellan manliga och kvinnliga studenter. Utfallet ett visst år blev enligt nedan.

Inriktning	Matematik	Matstat	Beräkn mat
Kvinnor	8	12	9
Män	15	10	12

Testa med lämplig metod om preferenserna skiljer sig åt mellan manliga och kvinnliga studenter under förutsättning att det givna året var representativt. Använd 95% signifikansnivå. Kommentera utfallet och om det tycks rimligt givet observationerna. (10 p)

Uppgift 6

Antalet diagnosticerade patienter av en ovanlig cancerform de senaste 6 åren var 14, 19, 16, 12, 18, 23. Dessa observationer kan på goda grunder antas vara ett stickprov från Poissonfördelningen med okänd parameter λ .

- Härled ML-skattningen för λ och ange dess numeriska värde för dessa data. (5 p)
- Använd normalapproximation för skattningen i a), tillsammans med Poissonfördelningens standardavvikelse, för att beräkna ett 95% konfidensintervall för λ . (5 p)

Lycka till!