



Stockholms
universitet

Statistiska institutionen, HT 2024

STATISTIK OCH DATAANALYS III

Tentamen 1, 8.5 hp, ST2101

Formalia

Datum: 9 oktober 2024

Skrivtid: 14.00 - 19.00

Resultat meddelas: senast 30 oktober 2024

Tillåtna hjälpmedel

- R/R studio
- Miniräknare
- Språklexikon

Tentamen består av fem uppgifter på maximalt 20 poäng per uppgift. För maximal poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar. Lösningarna skrivs och lämnas in på papper. Ingen digital inlämning ska göras. R får användas som hjälpmedel under tentamen men alla steg i lösningarna ska tydligt redovisas på papper.

Tentan ska göras anonymt, kontrollera därför innan du lämnar in din tentamen att du bara angett din anonymkod och INTE skrivit ditt namn någonstans i tentan.

Uppgift 1 (20 poäng)

Ett försäkringsbolag har analyserat ett stort antal skadeståndsansökningar av en viss kategori. Bolaget har funnit att denna typ av ansökningar resulterar i en utbetalning med sannolikheten 0.2.

- Vad är sannolikheten att en utbetalning sker efter att företaget har hanterat exakt 4 skadeståndsansökningar?
- Vad är sannolikheten att den 5:e utbetalningen sker efter att företaget har hanterat exakt 10 skadeståndsansökningar?
- Beräkna väntevärdet och variansen för antalet skadeståndsansökningar som behöver hanteras för att bolaget ska genomföra 5 utbetalningar.
- Försäkringsbolaget har också tittat närmare på den administrativa kostnaden för att genomföra utbetalningar. Man har dels en kostnad på 700 kronor per utbetalning (oavsett hur många ansökningar som hanterats) och dels en kostnad för varje skadeståndsansökan som hanteras (oavsett om den leder till utbetalning eller inte) på 1600 kronor. Beräkna väntevärdet och standardavvikelsen för den totala administrativa kostnaden för att genomföra 5 utbetalningar.

Uppgift 2 (20 poäng)

Ett spelbolag vill analysera följande två speltyper:

Spel 1: I detta spel drar spelaren ett antal lotter där varje lott kan vara en vinstlott eller en nitlott. Antalet vinster X är en diskret slumpvariabel som representerar antalet vinstlotter som dras. Sannolikheten för antal vinstlotter ges av följande sannolikhetsfördelning:

$$P(X = 0) = 0.4, \quad P(X = 1) = 0.3, \quad P(X = 2) = 0.2, \quad P(X = 3) = 0.1.$$

Spel 2: Summan vinstpengar, Y , som spelaren vinner i detta spel representeras av en kontinuerlig slumpvariabel som är likformigt fördelad över intervallet $(0, 100)$ kronor.

- Bestäm fördelningsfunktionen för antal vinster (X) för *Spel 1*.
- Bestäm fördelningsfunktionen för summan vinstpengar (Y) för *Spel 2*.

- c) För att ge spelet en annan dynamik och öka spänningen har spelbolaget beslutat att kvadrera vinsterna i *Spel 2*. Den nya vinsten för varje spelomgång, U , ges av transformationen

$$U = Y^2.$$

Bestäm täthets- och fördelningsfunktion för U .

Uppgift 3 (20 poäng)

Gurra Gambler är en duktig nätpokerspelare. Hans pokerspel genererar alltid en vinst på Y_2 kronor (i tiotusental) i månaden. Gurra tycker även om att spekulera i aktier. Varje månad köper han aktier för Y_1 (i tiotusental) kronor. Antag att Y_1 och Y_2 har simultan täthetsfunktion

$$f(y_1, y_2) = \begin{cases} ky_1 y_2, & 0 < y_1 \leq 1, \quad 0 < y_2 \leq 2 \\ 0, & \text{annars.} \end{cases}$$

- Bestäm konstanten k .
- Bestäm marginalfördelningarna för Y_1 och Y_2 . Hitta på ett värde på k om du inte har löst a)-uppgiften.
- Beräkna sannolikheten att Gurra köper aktier för mer än 5000 kr givet att hans pokervinst är mindre än 12000 kr.

Uppgift 4 (10 poäng)

- Förklara och ge exempel på begreppet konjugerande prior.
- Förklara skillnaden mellan ett klassiskt konfidensintervall och ett "Bayesian credible interval".

Uppgift 5 (10 poäng)

Kreditupplysningsföretaget F^3 håller på att utveckla en ny metod för att identifiera risker. Om metoden fungerar i mindre än 50 % av fallen läger man ner utvecklingen av metoden. Man genomför därför ett test av hypotesen

$$\begin{aligned}H_0 &: p = 0.5 \\H_a &: p < 0.5\end{aligned}$$

genom att tillämpa metoden på 12 fall som man vet är risker för att se hur många av dessa man kan identifiera. Nollhypotesen förkastas, d.v.s. utvecklingen läggs ner, om tre eller färre risker kan identifieras.

- Bestäm testets signifikansnivå.
- Bestäm testets styrka då $p = 0.3$.
- Vad händer med signifikansnivån och styrkan (då $p = 0.3$) om förkastelseområdet ändras så att den kritiska gränsen blir lägre?

Uppgift 6 (20 poäng)

Ett forskningsföretag inom läkemedelsindustrin studerar effektiviteten av en ny medicin som påverkar patienters reaktionstid i ett test. Forskarna har antagit att reaktionstiderna, som normaliserats till intervallet $(0, 1)$, följer en beta-fördelning med täthetsfunktion

$$f(y | \alpha) = \alpha y^{\alpha-1}, \quad 0 < y < 1$$

- Härled maximum likelihood-skattningen $\hat{\alpha}_{ML}$.
- I tidigare studier har man funnit att $\alpha = 2$ och forskarna vill nu testa om detta fortfarande stämmer genom att använda ett likelihoodkvot-test baserat på nya data som samlas in. Antag att ett slumpmässigt urval av n reaktionstider görs, där n är stort. Bestäm teststatistikan och ange kritiskt område (RR) för signifikansnivån 0.01.