Kurskod: TAMS24
Provkod: TEN1

Statistisk teori, grundkurs

Tentamen måndagen den 18 augusti 2014 kl 8–12.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen och/eller formelsamling ”Formel- och tabellsamling i matematisk statistik TAMS24/TAMS65 - 2013/2014 (Martin Singull)” (formelsamlingen ska ha en vattenmarkering TAMS24/TAMS65 - 2013/2014 i bakgrunden på varje sida). Inga anteckningar i formelsamlingarna är tillåtet. Miniräknare med tömda minnen.
Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 ger betyg 4 och 15-18 poäng ger betyg 5.
Examinator: Martin Singull, 013–281447
Resultatet meddelas *normalt* via LADOK inom 12 arbetsdagar.
Tydliga svar och motiveringar krävs till varje uppgift.

1. Någon har ställt upp hypotesen att dubbelt så många bilolyckor med dödlig utgång inträffar under lördagar och söndagar som under övriga veckodagar. Det vill säga sannolikheten att en måfå dragen dödsolycka inträffar på en lördag är 2/9, på en söndag 2/9 och för varje annan veckodag är den 1/9. Från riksstatistiken dras 90 olyckor på måfå. Dessa ger följande fördelning av olyckor över veckans dagar:

Måndag	Tisdag	Onsdag	Torsdag	Fredag	Lördag	Söndag
6	8	11	7	10	18	30

   Motsäger dessa resultat den ställda hypotesen på nivån $\alpha = 5\%$? (2p)

2. Töjbarheten hos provbitar av stål som tillverkats på två olika sätt, A och B, har mätts:

A	4.45	4.54	3.87	5.76	4.26	4.65	4.23	5.03	4.85	5.45	4.54	3.95
B	4.02	3.45	4.21	4.65	4.32	3.76	3.54	4.08	4.31	3.87	4.07	3.59

   Modell: Vi ha oberoende observationer $x_1, ..., x_{12}$ från $N(\mu_1, \sigma_1)$ för A och $y_1, ..., y_{12}$ från $N(\mu_2, \sigma_2)$ för B. Vidare är medelvärdena och stickprovsstandardavvikelserna $\bar{x} = 4.6317$, $s_1 = 0.5676$ respektive $\bar{y} = 3.9892$ och $s_2 = 0.3608$.

   a) Har man utgående från detta datamaterial någon anledning att tro att de båda standardavvikelserna σ_1 och σ_2 är olika? Genomför ett lämpligt test på nivån 10% och redovisa din slutsats. (1p)

   b) Antag nu att $\sigma_1 = \sigma_2$. Redan innan man gjorde mätningarna misstänkte man att behandling B skulle ge mindre töjbarhet. Bekräftas denna hypotes? Konstruera ett lämpligt 95% konfidensintervall och redovisa din slutsats. (2p)

3. Tiderna mellan "successiva sammanbrott för ett datorsystem" noterades under en tidsperiod och man fick följande observerade tider i timmar

1	10	20	30	40	52	63	70	80	90	100
102	130	140	190	210	266	310	530	590	640	1340

Modell: Vi antar att vi har observationer av oberoende stokastiska variabler X_i, som är exponentialfördelade med väntevärde μ dvs med täthetsfunktionen

$$f(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu}, & \text{för } x \geq 0, \\ 0, & \text{annars.} \end{cases}$$

a) Härled maximum-likelihood-skattningen av μ och beräkna variansen för skattningsvariablen. (2p)

b) Punktskatta med hjälp av resultatet i a) sannolikheten $P(X_i > 500)$ (1p)

4. I datamaterialet nedan finns uppgifter för två olika tidskrifter om antalet helsidesannonser, x, och det totala antalet sidor, y, fö fyra olika utgåvor under åren 2009 ($t = 1$), 2011 ($t = 2$) och 2013 ($t = 3$). För att kunna skilja på de båda tidskrifterna har tidskrift A kodats med $u = 0$ och tidskrift B med $u = 1$.

	t	x	u	y
1	1	60	0	76
2	1	61	0	76
3	1	45	0	84
4	1	72	0	100
5	2	36	0	74
6	2	63	0	74
7	2	34	0	68
8	2	60	0	90
9	3	27	0	76
10	3	53	0	82
11	3	36	0	76
12	3	61	0	90
13	1	53	1	88
14	1	56	1	122
15	1	33	1	88
16	1	44	1	106
17	2	32	1	82
18	2	58	1	104
19	2	55	1	96
20	2	64	1	140
21	3	75	1	202
22	3	58	1	90
23	3	45	1	86
24	3	35	1	86

Vid en preliminär analys visade det sig att t inte gjorde någon större nytta som förklaringsvariabel, så modellen som används är

$$Y = \beta_0 + \beta_1 x + \beta_2 u + \varepsilon$$

där ε-variablerna är oberoende och normalfördelade med väntevärde noll och standardavvikelse σ.

Skattad regressionslinje: $y = 23.5 + 1.13x + 27.0u$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	23.45	16.78
1	1.1259	0.3107
2	27.000	8.233

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	9715.5
RES	21	8540.5
TOT	23	18256.0

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.692616 & -0.012025 & -0.083333 \\ -0.012025 & 0.000237 & 0.000000 \\ -0.083333 & 0.000000 & 0.0502 \end{pmatrix}$$

a) Visa hur förklaringsgraden R^2 beräknas och förklara kortfattat vad den mäter (0.5p)

b) Verkar det finnas en systematisk skillnad i antalet sidor mellan de båda tidskrifterna, en skillnad som inte förklaras av antalet helsidesannonser? Besvara frågan med hjälp av ett lämpligt 95% konfidensintervall och redovisa din slutsats beträffande den eventuella skillnaden. (1.5p)

c) Låt oss nu anta att modellen ovan kan tillämpas för 2014 och att man för tidskrift B till det första numret har fått beställning på 60 helsidesannonser. Konstruera ett lämpligt 95% intervall som beskriver antalet sidor i detta nummer av tidskriften. Är du nöjd med intervallet? (2p)

5. Låt $(X_1, ..., X_n)$ är ett stickprov på den stokastiska variabeln X som är $N(\mu, \sigma)$, där standardavvikelsen σ är känt. Man önskar testa hypotesen

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu \neq \mu_0,$$

på signifikansnivån 1%.

Formulera ett lämpligt test och teckna styrkefunktionen uttryckt i den standardiserade normalfördelnings fördelningsfunktion, Φ. (3p)

6. Antag att $(X_1, ..., X_n)$ är ett stickprov på den stokastiska variabeln X som är likformigt fördelad på intervallet $[0.5-\theta, 0.5+\theta]$. där θ är okänt. Härled en punktskattning av θ enligt maximum-likelihood-metoden och ge den ett numeriskt värde, då det observerade stickprovet är följande: (0.15, 0.38, 0.71, 0.45, 0.11, 0.23, 0.61). (3p)