Mittuniversitetet
MID SWEDEN UNIVERSITY

Försättsblad Prov Original

Kurskod	Provkod	Tentamensdatum
MA087G	T100	2018-08-25

Kursnamn	Matematik GR (B), Matematisk statistik
Provnamn	Tentamen
Ort	Sundsvall
Termin	H18
Ämne	Matematik

8982538474

0027893

Tentamen i matematik **Mittuniversitetet**
Matematisk statistik (MA087G/MA097G) MOD
25 augusti 2018 Stefan Borell

Skrivtid: 5 timmar
Hjälpmedel: Godkänd, ej symbolhanterande miniräknare, Matematisk formelsamling (Upplaga 4) samt bifogade formler och tabeller.

Till alla uppgifter skall fullständiga lösningar lämnas. Resonemang, ekvationslösningar och uträkningar får inte vara så knapphändiga att de blir svåra att följa. En uppgift per blad, skriv endast på en sida.

Betyg sätts efter hur väl lärandemålen är uppfyllda. Riktvärde för att bli godkänd är 18 p (av maximalt 42 p).

---

1. a) En enhet produceras vid två olika fabriker. För produktionsfel görs en klassificering med avseende på typ av fel. En felproducerad enhet har antingen fel A, fel B eller fel C. På sistone har antalet felproducerade enheter registrerats enligt tabellen nedan.

	Fel A	Fel B	Fel C
Fabrik 1	15	45	40
Fabrik 2	75	30	45

Av de felproducerade enheterna väljs en enhet ut på måfå. Är händelserna *"enheten har fel A"* och *"enheten är tillverkad i fabrik 2"* oberoende?

b) Stickprovet nedan kommer från en Poissonfördelning $\mathrm{Po}(\mu)$.

3 2 3 1 1 5 0 3 1 2

Bestäm ML-skattningen μ^*_{obs} av μ. (5 p)

2. Ur en fångst av 120 stycken torskar har 10 stycken valts på måfå och vägts (enhet: kg). Resultatet visas nedan.

2,7 2,8 1,3 6,0 3,1 2,9 2,8 3,1 2,3 4,9

a) Beräkna stickprovets medelvärde och standardavvikelse.

b) Antag att vikten hos en slumpmässigt vald torsk är normalfördelad och bestäm ett 95 % konfidensintervall för väntevärdet med hjälp av stickprovet.

c) Använd normalapproximation för att uppskatta sannolikheten att hela fångstens vikt överstiger 400 kg.

OBS: I den här deluppgiften görs inget antagande om att torskarnas vikt är normalfördelad. Motivera hur Centrala gränsvärdessatsen används! (6 p)

3. På ett cigarettpaket kan man läsa att "9 av 10 som drabbas av strupcancer är rökare". Låt det vara givet att 51 % av befolkningen är kvinnor, av kvinnorna röker 11,7 %, av männen röker 12,1 % och sannolikheten att drabbas av strupcancer är 1 %.

   a) Vad är sannolikheten att en slumpmässigt vald person är rökare?

   b) Vad är sannolikheten att få strupcancer om man är rökare? (5 p)

4. I en tävling i skidskytte finns fyra stationer där de tävlande stannar och skjuter. Vid varje sådan station finns fem måltavlor och de tävlande har åtta försök att träffa de fem måltavlorna (fem ordinarie skott samt tre extraskott). För varje måltavla som ej har träffats vid en station får man åka en straffrunda.

   En viss tävlande träffar en måltavla med 80 % sannolikhet och alla skott på måltavlor sker oberoende av varandra. Vad är sannolikheten att denna tävlande

   a) får åtminstone en straffrunda vid en viss station?

   b) slipper straffrundor under loppet?

   c) får straffrundor vid fler än en av de fyra stationerna? (6 p)

5. Massan hos två preparat ska jämföras. Vardera preparats massa mäts ett antal gånger med ett och samma instrument. Instrumentets mätfel antas vara normalfördelat och utan systematiskt fel, dvs. mätfelet är $N(0, \sigma)$.

	Skattade värden (gram)		
	$\bar{x}$	s	Antal mätningar
Preparat 1	8,68	0,048	10
Preparat 2	8,72	0,044	15

   Gör ett hypotesprövning på signifikansnivån 1 % för att testa om det är någon skillnad i massa mellan de två preparaten. (7 p)

6. I en badrumslampa sitter två stycken LED-lampor som lyser samtidigt när badrumslampan tänds och släcks och är oberoende av varandra. LED-lamporna är av en sort som har en livslängd på 50 000 h (timmar) och en brinntid som är exponentialfördelad. Brinntid är det antal timmar en lampa kan lysa innan den går sönder och livslängd är brinntidens väntevärde.

   Antag att båda LED-lamporna är nya. Hur många timmar kan man förvänta sig att badrumslampan kan vara tänd till dess att

   a) någon av LED-lamporna går sönder?

   b) båda LED-lamporna har gått sönder? (7 p)

7. Vid kontroll av ett parti stenkol uppmättes askhalten i 12 olika prover till i genomsnitt $\bar{x} = 8{,}6$ %. Mätvärdena (enhet: procentenheter) vid detta mätförfarande på denna kolsort kan betraktas som observationer av en $N(\mu, 1)$-fördelning. Leverantören har garanterat att den genomsnittliga askhalten skall vara högst 8 % i partiet. Kan man påstå att leverantören har fel? Besvara frågan genom att formulera lämpliga hypoteser och bestämma ett lämpligt test på signifikansnivån 1 % respektive 5 %. (6 p)

Lycka till!

1 Sannolikhetsteori

1.1 Stokastiska variabler

Varians $V(X) = E(X^2) - (E(X))^2$

Standardavvikelse $D(X) = \sqrt{V(X)}$

Kovarians $C(X,Y) = E\big((X - E(X))(Y - E(Y))\big) = E(XY) - E(X)E(Y)$

Korrelationskofficient $\rho(X,Y) = \dfrac{C(X,Y)}{D(X)D(Y)}$

1.2 Diskreta fördelningar

Binomialfördelningen X är $\mathrm{Bin}(N,p)$ där $0 < p < 1$ och $N \in \mathbb{N}$ om

$$p_X(k) = \binom{N}{k} p^k (1-p)^{N-k}, \qquad k = 0, 1, \ldots, N.$$

$E(X) = Np$, $V(X) = Np(1-p)$

"För-första-gången"-fördelningen X är $\mathrm{ffg}(p)$ där $0 < p < 1$ om

$$p_X(k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \ldots.$$

$E(X) = \dfrac{1}{p}$, $V(X) = \dfrac{1-p}{p^2}$

Hypergeometriska fördelningen Låt $0 < p < 1$ och $N, n \in \mathbb{N}$ sådana att $2 \le N$, $n < N$ och $Np \in \mathbb{N}$. X är $\mathrm{Hyp}(N,n,p)$ om

$$p_X(k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}, \qquad 0 \le k \le Np, \quad 0 \le n-k \le N(1-p).$$

$E(X) = np$, $V(X) = np(1-p)\frac{N-n}{N-1}$

Poissonfördelningen X är $\mathrm{Po}(\mu)$ där $\mu > 0$ om

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \qquad k = 0, 1, 2, \ldots.$$

$E(X) = \mu$, $V(X) = \mu$

1.3 Kontinuerliga fördelningar

Likforming fördelning X är $\mathrm{U}(a, b)$ där $a < b$ om

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{för } a < x < b \\ 0 & \text{annars} \end{cases}$$

$E(X) = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12}$

Exponentialfördelningen X är $\mathrm{Exp}(\lambda)$ där $\lambda > 0$ om

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{för } x > 0 \\ 0 & \text{annars} \end{cases}$$

$E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$

Normalfördelningen X är $\mathrm{N}(\mu, \sigma)$ där $\mu \in \mathbb{R}$, $\sigma > 0$ om

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$E(X) = \mu$, $V(X) = \sigma^2$

1.4 Centrala gränsvärdesatsen

Om $X_1, X_2, \ldots, X_n$ är oberoende likafördelade stokastiska variabler med väntevärde μ och standardvvikelse σ så är

$$\overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$$

approximativt $\mathrm{N}(\mu, \sigma/\sqrt{n})$ om n är stort.

1.5 Approximation

$\mathrm{Hyp}(N, n, p)$ är approximativt $\mathrm{Bin}(n, p)$ om $\frac{n}{N} \leq 0.1$

$\mathrm{Bin}(N, p)$ är approximativt $\mathrm{Po}(Np)$ om $p \leq 0.1$

$\mathrm{Bin}(N, p)$ är approximativt $\mathrm{N}(Np, \sqrt{Np(1-p)})$ om $Np(1-p) \geq 10$

$\mathrm{Po}(\mu)$ är approximativt $\mathrm{N}(\mu, \sqrt{\mu})$ om $\mu \geq 15$

2 Statistikteori

2.1 Beskrivande statistik

$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{j=1}^{n} x_j \\
s^2 &= \frac{1}{n-1}\sum_{j=1}^{n}(x_j-\overline{x})^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n}x_j^2-\frac{1}{n}\Big(\sum_{j=1}^{n}x_j\Big)^2\right)
\end{aligned}$$

2.2 Punktskattningar

Låt X vara en stokastisk variabel där fördelningen för X beror på en okänd parameter θ. Låt $x_1,\ldots,x_n$ vara en observation på stokastiska variabler $X_1,\ldots,X_n$ som är oberoende och som alla är lika fördelade som X.

Maximum-likelihood-metoden. Det värde θ^*_{obs} som maximerar L-funktionen

$$L(\theta) = \begin{cases} p_X(x_1;\theta)\ldots p_X(x_n;\theta), & \text{(diskreta fallet)} \\ f_X(x_1;\theta)\ldots f_X(x_n;\theta), & \text{(kontinuerliga fallet)} \end{cases}$$

kallas *maximum-likelihood-skattningen (ML-skattningen)* av θ.

Minsta-kvadrat-metoden. Antag att $E(X) = m(\theta)$. Det värde θ^*_{obs} som minimerar kvadratsumman

$$Q(\theta) = \sum_{j=1}^{n}(x_j - m(\theta))^2$$

kallas *minsta-kvadrat-skattningen (MK-skattningen)* av θ.

2.3 Vanliga stickprovsvariabler

Ett stickprov. Låt $X_1,\ldots,X_n$ vara oberoende stokastiska variabler som alla är lika fördelade.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j-\overline{X})^2.$$

Om $X_1,\ldots,X_n$ är $\mathrm{N}(\mu,\sigma)$ gäller

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \text{ är } \mathrm{N}(0,1), \quad \frac{\overline{X}-\mu}{S/\sqrt{n}} \text{ är } t(n-1), \quad \frac{n-1}{\sigma^2}S^2 \text{ är } \chi^2(n-1).$$

Poolad variansskattning från två stickprov. Låt $X_1, \dots, X_{n_1}$ vara likafördelade och $Y_1, \dots, Y_{n_2}$ vara likafördelade. Samtliga stokastiska variabler antas oberoende och med samma varians σ^2.

$$S^2 = \frac{1}{n_1 + n_2 - 2}\Big(\sum_{j=1}^{n_1}(X_j - \overline{X})^2 + \sum_{j=1}^{n_2}(Y_j - \overline{Y})^2\Big).$$

Om $X_1, \dots, X_n$ är $\mathrm{N}(\mu_1, \sigma)$ och $Y_1, \dots, Y_{n_2}$ är $\mathrm{N}(\mu_2, \sigma)$ gäller

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S\sqrt{1/n_1 + 1/n_2}} \quad \text{är} \quad t(n_1 + n_2 - 2).$$

2.4 Konfidensintervall

Låt $x_1, \dots, x_n$ vara en observation på stokastiska variabler $X_1, \dots, X_n$ som är oberoende och som alla är $\mathrm{N}(\mu, \sigma)$.

Konfidensintervall för μ. Ett två-sidigt konfidensintervall med konfidensgraden $1 - \alpha$ är:

a) σ känd: $\quad \overline{x} - \frac{\sigma}{\sqrt{n}}\lambda_{\alpha/2} \le \mu \le \overline{x} + \frac{\sigma}{\sqrt{n}}\lambda_{\alpha/2}$

b) σ okänd: $\quad \overline{x} - \frac{s}{\sqrt{n}}t_{\alpha/2}(n-1) \le \mu \le \overline{x} + \frac{s}{\sqrt{n}}t_{\alpha/2}(n-1)$

Konfidensintervall för σ^2. Ett två-sidigt konfidensintervall med konfidensgraden $1 - \alpha$ är:

$$\frac{n-1}{\chi^2_{\alpha/2}(n-1)}s^2 \le \sigma^2 \le \frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}s^2$$

2.5 Hypotesprövning

Konfidensmetoden Förkasta $H_0 : \theta = \theta_0$ på nivån α om θ_0 ej faller inom ett lämpligt valt konfidensintervall med konfidensgraden $1 - \alpha$.

χ^2-test

- Test av fördelning: Ett försök ger något av resultaten $A_1, \dots, A_r$ med respektive sannolikheter $P(A_1), \dots, P(A_r)$. Man har n observationer där frekvensen för händelse A_j är x_j.

$H_0 : P(A_1) = p_1, \ldots, P(A_r) = p_r$.

Om H_0 är sann blir

$$Q = \sum_{j=1}^{r} \frac{(x_j - np_j)^2}{np_j}$$

ett utfall av en approximativt $\chi^2(r-1)$-fördelad stokastisk variabel.

Tumregel för god approximation: $np_j \geq 5$

- Homogenitetstest: Ett försök ger något av resultaten $A_1, \ldots, A_r$. Man har s försöksserier. Inom den i:te serien har man n_i observationer och frekvensen x_{ij} för händelsen A_j.

serie	A_1	A_2	$\ldots$	A_r	antal försök
1	x_{11}	x_{12}	$\ldots$	x_{1r}	n_1
2	x_{21}	x_{22}	$\ldots$	x_{2r}	n_2
$\vdots$	$\vdots$			$\vdots$	$\vdots$
s	x_{s1}	x_{s2}	$\ldots$	x_{sr}	n_s
summa	m_1	m_2	$\ldots$	m_r	n

H_0: Sannolikheterna för $A_1, \ldots, A_r$ är desamma i alla försöksserier.

Om H_0 är sann blir

$$Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n_i m_j / n)^2}{n_i m_j / n}$$

ett utfall av en approximativt $\chi^2\big((r-1)(s-1)\big)$-fördelad stokastisk variabel.

Tumregel för god approximation: $n_i p^*_{j,\text{obs}} \geq 5$

2.6 Linjär regression

Låt Y_j vara $\mathrm{N}(\alpha + \beta x_j, \sigma)$, $j = 1, \ldots, n$, och oberoende.

Skattade regressionslinjen $\beta^*_{\text{obs}} = \dfrac{S_{xy}}{S_{xx}}, \quad \alpha^*_{\text{obs}} = \overline{y} - \beta^*_{\text{obs}}\overline{x}$,

$$\begin{aligned}
S_{xy} &= \Big(\sum_{j=1}^{n} x_j y_j\Big) - n\overline{x}\overline{y} = \sum_{j=1}^{n}(x_j y_j - x_j\overline{y} - \overline{x}y_j + \overline{x}\overline{y}) = \sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y}), \\
S_{xx} &= \Big(\sum_{j=1}^{n} x_j^2\Big) - n\overline{x}^2 = \sum_{j=1}^{n}(x_j^2 - 2x_j\overline{x} + \overline{x}^2) = \sum_{j=1}^{n}(x_j - \overline{x})^2
\end{aligned}$$

Fördelningar

a) $\beta^* = \dfrac{\sum_{j=1}^{n}(x_j - \overline{x})(Y_j - \overline{Y})}{\sum_{j=1}^{n}(x_j - \overline{x})^2}$ är $\mathrm{N}\Big(\beta, \dfrac{\sigma}{\sqrt{S_{xx}}}\Big)$

b) $\alpha^* = \overline{Y} - \beta^*\overline{x}$ är $\mathrm{N}\Big(\alpha, \sigma\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{S_{xx}}}\Big)$