

TDAB01 SANNOLIKHETSLÄRA OCH STATISTIK

TENTAMEN 2020-10-26

LÄRARE

Jose M. Peña. Nås via Teams (eller på `jose.m.pena@liu.se` om Teams inte fungerar).

BETYG

För full poäng i varje delfråga krävs tydliga och väl motiverade svar.

Maximalt antal poäng: 20 poäng

Betyg 5 = 18-20 poäng

Betyg 4 = 14-17 poäng

Betyg 3 = 10-13 poäng

Betyg U = 0-9 poäng

TILLÅTNA HJÄLPMEDEL

Miniräknare med tomt minne. Tabell- och formelsamling (ingår i tentamen). Slides på kurswebbsidan.

INSTRUKTIONER

- Ladda ner tentan från LISAM från kl 14.00. Lämna in dina svar i LISAM senaste kl 18.15. Om du har fått förlängd tid, v.v. och mejla mig dina svar.
- Om du har frågor, v.v. och använd Teams för att kontakta mig (eller `jose.m.pena@liu.se` om Teams inte fungerar).
- Tentan är individuell, dvs hjälp från eller kommunikation med andra är inte tillåten.
- Tentan är anonym, dvs skriv ej ditt namn någonstans.
- Du rekommenderas lösa uppgifterna med papper och penna, fota lösningarna och ladda upp dem. Det är också OK att använda Word eller LaTeX.

UPPGIFTER

- (1) (5 p = 1 + 2 + 2) 10 personer deltar i en fest. 5 av de 10 deltagarna har covid. (a) Om en slumpmässigt utvald deltagare testas, vad är sannolikheten att hen är smittad ? (b) Om två slumpmässiga utvalda deltagare testas, vad är sannolikheten att båda är smittade ? (c) Hur många slumpmässiga utvalda deltagare måste testas för att hitta en covid smittad deltagare med sannolikhet högre än 0.9 ?

Lösning:

$$(a) P(1 \text{ av } 1 \text{ smittad}) = \frac{\binom{5}{0}\binom{5}{1}}{\binom{10}{1}} \text{ där } \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

$$(b) P(2 \text{ av } 2 \text{ smittade}) = \frac{\binom{5}{0}\binom{5}{2}}{\binom{10}{2}}$$

$$(c) P(0 \text{ av } k \text{ smittade}) = \frac{\binom{5}{k}\binom{5}{0}}{\binom{10}{k}} = \frac{5!(10-k)!}{10!(5-k)!} < 0.1. \text{ Då, } k = 3.$$

- (2) (2 p) X och Y är två binära variabler så att $0 < P(X = 0) < 1$ och $0 < P(Y = 0) < 1$. Bevisa att om $P(X = x, Y = y) = 0$ för några x och y , då är X och Y beroende.

Lösning:

Om de är oberoende, då $P(X = x, Y = y) = P(X = x)P(Y = y)$ för alla x och y . Då, $P(X = x, Y = y) > 0$ för alla x och y . Då, de måste vara beroende.

- (3) (3 p) Lillemor deltar i ett covid vaccin test. Hon vet inte om hon fick vaccinet eller placebon. Kroppstemperaturen för de i kontrolgruppen (dvs, de som fick placebo) är $N(\mu = 37, \sigma = 1)$ fördelad, medan för dem i behandlingsgruppen (dvs, de som fick vaccinet) är $N(\mu = 39, \sigma = 2)$ fördelad. Behandlingsgruppen är två gånger större än kontrolgruppen. Lillemor har en temperatur som är lägre än 38 grader. Beräkna sannolikheten att hon fick vaccinet.

Lösning:

$$P(\text{vaccin}) = 2/3 \text{ och } P(\text{placebo}) = 1/3 \text{ och}$$

$$P(\text{temp} < 38 | \text{vaccin}) = P\left(\frac{\text{temp} - 37}{1} < \frac{38 - 37}{1} | \text{vaccin}\right) = \Phi\left(\frac{38 - 37}{1}\right) = \Phi(1) = 0.8413$$

och

$$P(\text{temp} < 38 | \text{placebo}) = P\left(\frac{\text{temp} - 39}{2} < \frac{38 - 39}{2} | \text{vaccin}\right) = \Phi\left(\frac{38 - 39}{2}\right) = \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915$$

och då

$$P(\text{vaccin} | \text{temp} < 38) = \frac{P(\text{temp} < 38 | \text{vaccin})P(\text{vaccin})}{P(\text{temp} < 38 | \text{vaccin})P(\text{vaccin}) + P(\text{temp} < 38 | \text{placebo})P(\text{placebo})}$$

- (4) (3 p = 1 + 2) (a) Använd Chebyshevs olikhet för att bevisa att sannolikheten att en variabel X tillhör intervallet $[\mu - 2\sigma, \mu + 2\sigma]$ är lika eller större än 0.75. (b) Anta att X är standard normal fördelad. Visa att 0.75 är en mycket konservativ sannolikhet, dvs den sanna sannolikheten att X tillhör intervallet $[\mu - 2\sigma, \mu + 2\sigma]$ är betydligt större än 0.75.

Lösning:

$$(a) P(|X - \mu| > \epsilon) \leq (\sigma/\epsilon)^2 \text{ och då } P(|X - \mu| > 2\sigma) \leq (\sigma/2\sigma)^2 = 1/4 \text{ och då } P(|X - \mu| \leq 2\sigma) \geq 3/4.$$

$$(b) P(|X - \mu| \leq 2\sigma) = P(|X| \leq 2) = P(-2 \leq X \leq 2) = \Phi(2) - \Phi(-2) = \Phi(2) - 1 + \Phi(2) = 0.9772 - 1 + 0.9772$$

- (5) (6 p = 1 + 1 + 2 + 2) Några svenska företag har haft de följande budgetarna under åren:

år (20xx)	03	04	05	06	07
budget (MSEK)	17	23	31	29	33

- (a) Bygg en linjär regression modell från datan. (b) Anta att $\sigma = 5$. Bygg en 95 % konfidensintervall för budgeten för 2017. (c) Anta att $\sigma = 5$. Beräkna p -värdet för

hypotestest "budgeten ökar med exakt 3.5 MSEK per år i genomsnitt". (d) Några finska företag har haft de följande budgetarna under åren:

år (20xx)	03	04	05	06	07
budget (MSEK)	19	20	21	24	27

Beräkna p -värdet för hypotestest "den genomsnittliga årliga ökningen av budgeten är lika för svenska och finska företagen". Anta att $\sigma = 5$ för båda svenska och finska företagen.

Lösning:

(a) $b_1 = S_{xy}/S_{xx} = 3.8$ och $b_0 = \bar{y} - \bar{x}b_1 = 7.6$.

(b) $\hat{y}_* = b_0 + b_1x_* = 7.6 + 3.8 \cdot 17$ och då 95 % konfidensintervallet är $\hat{y}_* \pm z_{0.025}\sigma\sqrt{\frac{1}{5} + \frac{(17-\bar{x})^2}{S_{xx}}}$.

(c) Kör ett Z -test med $H_0 : \beta_1 = 3.5$ och $H_A : \beta_1 \neq 3.5$. Då, $Z_{obs} = (b_1 - \beta_1)/(\sigma/\sqrt{S_{xx}}) = (3.8 - 3.5)/(5/\sqrt{10})$ och p -värdet $= 2[1 - \Phi(Z_{obs})]$.

(d) För de finska företagen, $b_1 = 2$ och $b_0 = 12.2$. Kör ett test med $H_0 : \beta_1^s = \beta_1^f$ och $H_A : \beta_1^s \neq \beta_1^f$, dvs $H_0 : \beta_1^s - \beta_1^f = 0$ och $H_A : \beta_1^s - \beta_1^f \neq 0$. Under H_0 , $b_1^s - b_1^f$ följer $N(\beta_1^s - \beta_1^f, \sigma/\sqrt{S_{xx}^s} + \sigma/\sqrt{S_{xx}^f})$. Då, kör ett Z -test med $Z_{obs} = (b_1^s - b_1^f - \beta_1^s + \beta_1^f)/(\sigma/\sqrt{S_{xx}^s} + \sigma/\sqrt{S_{xx}^f}) = (3.8 - 2)/(5/\sqrt{10} + 5/\sqrt{10})$. Då, p -värdet $= 2[1 - \Phi(Z_{obs})]$.

(6) (1 p) Skriv R (pseudo)kod för att dra 100 X -värden från den följande modellen.

$$\begin{aligned} p_1 &\sim \text{Beta}(1, 1) \\ p_2 &\sim \text{Beta}(1, 10) \\ X &\sim \text{Binomial}(10, \max(p_1, p_2)) \end{aligned}$$

Lösning:

```
for(i in 1:100){
  p1 <- rbeta(1,1)
  p2 <- rbeta(1,10)
  rbinom(1,10,max(p1,p2))
}
eller
p1 <- rbeta(100,1,1)
p2 <- rbeta(100,1,10)
rbinom(100,10,pmax(p1,p2))
```