

Svar till tentamen i Matematisk statistik D, IT (TMA290, TMS155), 23 augusti 2005

OBS: Detta är bara kortfattade, ibland ofullständiga, svar, dvs långt ifrån hur lösningarna på en tenta ska se ut!

1. a)

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$$

där $\mu = E[X]$.

b) Det linjära sambandet.

Om $\text{korr}(X, Y) = 1$ så $Y = a + bX$ där b är positiv, och om $\text{korr}(X, Y) = -1$ så $Y = a - bX$ där b är positiv.

Om X och Y är oberoende så är korrelationen 0.

2. a) $f(-9) = f(9) = 1/8$, $f(-3) = f(3) = 3/8$

b)

$$F(x) = \begin{cases} 0, & -\infty < x < -9 \\ 1/8, & -9 \leq x < -3 \\ 1/2, & -3 \leq x < 3 \\ 7/8, & 3 \leq x < 9 \\ 1, & 9 \leq x < \infty \end{cases}$$

3. a) $\frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}}$ som har fördelning $N(0, 1)$

b) $\frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}$ som har fördelning t_{n-1}

c) $\frac{(n-1)S^2}{\sigma_0^2}$ som har fördelning χ_{n-1}^2

d) $\frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}}$ som har approx fördelning $N(0, 1)$ pga CGS.

d) $\frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}$ som har approx fördelning t_{n-1} pga CGS.

4. Låt E= klave alla 4 ggr; F= fuskmynt valt.

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} = \frac{1/4}{(1 + 3 \cdot 2^{-4})/4} = 16/19 \approx 0.842.$$

5. Antag beroende observationer. Tillräckligt stort stickprov för att medelvärdet ska vara approx normalfördelat, vilket gör att ett 95 % konfint är

$$[\bar{X} - t_{\alpha/2, n-1} S/\sqrt{n}, \bar{X} + t_{\alpha/2, n-1} S/\sqrt{n}] \Rightarrow [0.53 - 2.026 \cdot 0.08/\sqrt{38}, 0.53 + 2.026 \cdot 0.08/\sqrt{38}] = [0.504, 0.556]$$

6. a) $E[X] = 3 \cdot 0.4 + 2 \cdot 0.3 + 1 \cdot 0.2 = 2$; $\text{Var}(X) = (3 - 2)^2 0.4 + \dots + (0 - 2)^2 \cdot 1 = 1$

b) Centrala gränsvärdessatsen ger att summan är approximativt normalfördelat med väntevärde $E[\sum_{i=1}^{100} X_i] = 100E[X_1] = 200$, och varians $\text{Var}(\sum_{i=1}^{100} X_i) = 100\text{Var}(X_1) = 100$.

7. a) Typ I fel: Att förkasta H_0 då H_0 är sann. Typ II fel: Att inte förkasta H_0 då H_0 är falsk.

b) Oberoende observationer, vilken är populationen och täcker stickprovet in hela pop, är data normalfördelat, tillräckligt stort stickprov om ej normalförd, vad beror bortfall på, mm.

b) Man garderar sig främst mot typ I fel. Det innebär att om man kan förkasta kan man sluta sig till att mothyp är sann med den säkerhet som anges av konfidensgraden. Kan man inte förkasta kan man däremot INTE dra slutsatsen att nollhyp är sann, bara att man inte har belägg för att förkasta den.

8. $X =$ antal patienter vars blodtryck sänks $\sim Bin(50, p)$, där $p =$ slh för sänkt blodtryck.

$$H_0 : p = 0.6$$

$$H_a : p > 0.6.$$

Testnivå 99 %.

X är approximativt normalfördelad $N(50 \cdot 0.6, \sqrt{50 \cdot 0.6 \cdot 0.4})$ under H_0 .

$$P(X \geq 35) \approx 1 - \Phi\left(\frac{35 - 50 \cdot 0.6}{\sqrt{50 \cdot 0.6 \cdot 0.4}}\right) \approx 1 - \Phi(5/\sqrt{12}) \approx 0.075 > 0.01.$$

Kan ej förkasta.

Med kontinuitetskorrektur fås $P(X \geq 35) \approx 0.097$.

9. a) Om $\hat{\theta}$ är en skattning av θ så $E[\hat{\theta}] = \theta$.
b) $E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n E[(X_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2$.
c) $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 (= S^2)$

10. a) $X =$ tot antal bilar under 2 minuter $= Y + Z$, där $Y \sim Po(2)$ och $Z \sim Po(4)$ och oberoende $\Rightarrow X \sim Po(6)$.

$$P(X = 5) = e^{-6} 6^5 / 5! \approx 0.161.$$

- b) $X_i =$ antal bilar minut $i \sim Po(3)$, oberoende. CGS ger att $(\sum_{i=1}^{90} X_i - 270) / \sqrt{270}$ approx $N(0,1)$.

$$P\left(\sum_{i=1}^{90} X_i \geq 300\right) \approx 1 - \Phi(1.826) \approx 0.03.$$