Dugga 2: TMS063 Matematisk Statistik. Fredag 20 maj, 2016. Examinator: Marina Axelson-Fisk.
Tillåtna hjälpmed: typgodkänd miniräknare och bifogade formelblad och fördelningstabeller.

Namn:____ Personnr:________________________

****

Varje uppgift är värd 4 poäng. Duggan kan ge bonuspoäng på matstat-delen på den slutliga tentan. För 1 bonuspoäng krävs 6 poäng, för 2 bonuspoäng krävs 9 poäng (av totalt 12 poäng).

1. Vikten och längden mättes på 700 bergsgorillor i Kongo gav följande tabell, där varje cell anger hur många gorillor som tillhör varje kategori:

längd\vikt	lätt	mellan	tung
kort	170	70	30
lång	85	190	155

Låt $X =$ vikt där $X \in \{0,1,2\}$ för lätt, mellan och tung, och låt $Y =$ längd där $Y \in \{0,1\}$ för kort och lång, respektive.

a) Beräkna frekvensfunktionen för den gemensamma fördelningen och de båda marginalfördelningarna.
b) Beräkna kovariansen av X och Y.

Lösning:

a) Den gemensamma frekvensfunktionen

$$f_{XY}(x,y) = P(X = x, Y = y) = \frac{\text{antal i cell } (x,y)}{700}$$

Frekvensfunktionerna för marginalfördelningarna ges av
$f_X(x) = \sum_y f_{XY}(x,y)$ och $f_Y(y) = \sum_y f_{XY}(x,y)$ så:

$f_{XY}(x,y)$	$X = 0$	$X = 1$	$X = 2$	$f_Y(y)$
$Y = 0$	0.243	0.100	0.043	0.386
$Y = 1$	0.121	0.271	0.221	0.613
$f_X(x)$	0.364	0.371	0.265	0.999

b) Vi har att $Cov(X,Y) = E[XY] - E[X]E[Y]$
och vi får

$$E[X] = \sum_{x=0}^{2} x \cdot f_X(x) = 0 \cdot 0.364 + 1 \cdot 0.371 + 2 \cdot 0.265 = 0.901$$

$$E[Y] = \sum_{y=0}^{1} y \cdot f_Y(y) = 0 \cdot 0.386 + 1 \cdot 0.613 = 0.613$$

$$E[XY] = \sum_{x=0}^{2} \sum_{y=0}^{1} xy \cdot f_{XY}(x,y) = 0 \cdot 0 \cdot 0.243 + 0 \cdot 1 \cdot 0.100 + \cdots + 1 \cdot 2 \cdot 0.265 = 0.713$$

$$Cov(X,Y) = 0.713 - 0.901 \cdot 0.613 = 0.161$$

2. Antag att taxibilar anländer till ett central hotell enligt en Poissonprocess med i snitt 5 minuter mellan varje.

   a) Beräkna sannoliheten att det inte kommer någon taxi de närmaste 10 minuterna.
   b) Beräkna sannolikheten att en gäst får vänta i mer än 10 minuter.

Lösning:

a) Låt $N(t) =$ antal taxibilar i $[0, t]$. Vi använder tidsenhet "minuter" så med 5 minuter i snitt mellan två taxibilar får vi intensitet $\lambda = 1/5 \ = 0.20$. Vi får

$$P(N(10) = 0) = e^{-0.2 \cdot 10} \frac{(0.2 \cdot 10)^0}{0!} = e^{-2} = 0.135$$

b) Låt $X =$ tid mellan två taxibilar. Då är $X \sim Exp(0.2)$, och

$$P(X > 10) = \int_{10}^{\infty} 0.2 \cdot e^{-0.2x} dx = [e^{-0.2x}]_{10}^{\infty} = e^{-2} = 0.135$$

3. Antag att antal poäng ett basketlag får under en match är normalfördelad $N(\mu, \sigma^2)$. Antag att i de 10 sista matcherna fick laget följande poäng:

$$59, 62, 59, 74, 70, 61, 62, 66, 62, 75$$

a) Beräkna ett 95%-konfidensintervall för μ.
b) Hur många observationer skulle behövas för att få ett 99%-konfidensintervall av samma bredd?

Lösning:

a) Eftersom σ är okänd får vi använda t-fördelning och får ett $(1-\alpha)$-konfidensintervall genome

$$\mu = \bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

och

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 65\,, \qquad \sum_{i=1}^{n} x_i^2 = 42572$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{9}(42572 - 10 \cdot 65^2) = 35.78$$

$s = 5.98$

och t_9-tabell ger att $t_{\alpha/2,n-1} = t_{0.025,9} = 2.262$ så ett 95%-konfidensintervall blir

$$\mu = 65 \pm 2.262\frac{5.98}{\sqrt{10}} = 65 \pm 4.28$$

b) ***Obs! Fel i uppgift!***
Jag borde ha angivit en given varians σ^2 annars får man två okända: n och $t_{0.005,n-1}$.
Man kan visserligen pröva sig fram för olika n, men det var inte avsikten.

Om variansen hade angivits, t ex till $\sigma^2 = 35.78$ (som skattningen i a), skulle man för att få ett 99%-konfidensintervall ha slagit upp värdet $z_{0.995} = 2.58$ och fått att

$$2.58\frac{5.98}{\sqrt{n}} = 4.28 \Leftrightarrow n = \left(\frac{2.58 \cdot 5.98}{4.28}\right)^2 = 12.99 \approx 13$$

så för att få ett 99%-konfidensintervall av samma bredd som i (a) behöver vi minst $n = 13$ observationer.

Om man istället försökte lösa uppgiften som den faktiskt var angiven skulle man alltså behöva hitta det n så att

$$\frac{t_{0.005,n-1}}{\sqrt{n}} = \frac{4.28}{5.98} = 0.716$$

och man skulle behöva pröva sig fram

$$n = 15{:}\frac{2.947}{\sqrt{15}} \Rightarrow 0.761, \qquad n = 16{:}\frac{2.921}{\sqrt{16}} = 0.730, \qquad n = 17{:}\frac{2.898}{\sqrt{17}} = 0.702$$

så man behöver minst $n = 17$ observationer för att få ett intervall av (högst) samma bredd.