STAT 7203 Applied Probability and Statistics

Final Exam

name: Peng Yu

id: 46635884

1. (a) if $X \sim$ ~~Pass~~ Poisson (2) if $Y \sim$ Binomial $(5, 0.2)$

then $\mathbb{E}[X] = \lambda = 2$ then $\mathbb{E}[Y] = np = 5 \times 0.2 = 1$

$\mathbb{E}[W] = \mathbb{E}[X+2Y] = \mathbb{E}[X] + 2\mathbb{E}[Y] = 2 + 2 \times 1 = 4$

(b) X and Y are two independent random variables

$Var(X) = 2$ $Var(Y) = np(1-p) = 5 \times 0.2 \times 0.8 = 0.8$

$Var(W) = Var(X+2Y) = Var(X) + 4\,Var(Y)$
$= 2 + 4 \times 0.8 = 5.2$

(c). $Cov(X, W) = \mathbb{E}[XW] - \mathbb{E}[X]\,\mathbb{E}[W]$
$= \mathbb{E}[X(X+2Y)] - 2 \times 4$
$= \mathbb{E}[X^2] + \mathbb{E}[2XY] - 8$
$= Var(X) + (\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] - 8$
$= 2 + 2^2 + 2 \times 2 \times 1 - 8$
$= 2$

(d) $\mathbb{E}[W|X] = \mathbb{E}[(X+2Y)|X] = \mathbb{E}[X|X] + \mathbb{E}[2Y|X]$

$= X + 2\mathbb{E}[Y] = X + 2\times 1 = X+2$

(e) $\cancel{M_X(t) = e^{2(e^t-1)}}$

$M_X(t) = \mathbb{E}[e^{Xt}] = e^{2(e^t-1)}$

$M_Y(t) = \mathbb{E}[e^{Yt}] = (0.8+0.2e^t)^5$

$M_W(t) = \mathbb{E}[e^{Wt}] = \mathbb{E}[e^{(X+2Y)t}] = \mathbb{E}[e^{Xt}]\,\mathbb{E}[e^{2Yt}]$

$= e^{2(e^t-1)} \cdot (0.8+0.2e^t)^5$

2. if $X \sim Exp(2)$ then $f_X(x) = \lambda e^{-\lambda x} = 2e^{-2x}$

$\mathbb{P}(X_1X_2X_3 \geq t) = 1 - \mathbb{P}(X_1X_2X_3 < t) = 1 - \mathbb{P}(X_1<t) \cap \mathbb{P}(X_2<t) \cap \mathbb{P}(X_3<t)$

$= 1 - \int_0^t 2e^{-2x_1}dx_1 \cdot \int_0^t 2e^{-2x_2}dx_2 \cdot \int_0^t 2e^{-2x_3}dx_3$

$= 1 - (1 - \frac{1}{e^{2t}})^3$

$= \frac{(e^{2t}-1)^2}{e^{6t}} + \frac{2}{e^{2t}} - \frac{1}{e^{4t}}$

3 (a) $f_{X|Y=y}(x|Y=y) = \begin{cases} \frac{1}{y-0} = \frac{1}{y} & \begin{cases} x \in (0,y) \\ y \in (0,1) \end{cases} \\ = 0 & \text{else.} \end{cases}$

$f_{XY}(xy) = f_{X|Y=y}(x|Y=y) \cdot f_Y(y) \begin{cases} = \frac{1}{y} \cdot 6y(1-y) = 6(1-y) & \begin{cases} x \in (0,y) \\ y \in (0,1) \end{cases} \\ = 0 & \text{else.} \end{cases}$

(b) $f_X(x) = \int_x^1 6(1-y)\,dy = \begin{cases} 3(x-1)^2 & x \in (0,y) \\ 0 & \text{else} \end{cases}$

(c) $\mathbb{E}[X|Y] = \int_0^y \frac{x}{y}\,dx = \frac{y}{2}$

$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[Y\,\mathbb{E}[X|Y]] - \mathbb{E}[\mathbb{E}[X|Y]]\,\mathbb{E}[Y]$

$= \mathbb{E}[\frac{Y^2}{2}] - \mathbb{E}[\frac{Y}{2}]\,\mathbb{E}[Y]$

$\mathbb{E}[\frac{Y^2}{2}] = \int_0^1 \frac{y^2}{2} \cdot 6y(1-y)\,dy = \frac{3}{20}$

$\mathbb{E}[\frac{y}{2}] = \int_0^1 \frac{y}{2} \cdot 6y(1-y)\,dy = \frac{1}{4}$

$\mathbb{E}[y] = \int_0^1 y \cdot 6y(1-y)\,dy = \frac{1}{2}$

$Cov(X,Y) = \frac{3}{20} - \frac{1}{4} \times \frac{1}{2} = \frac{1}{40} = 0.025$

(d) $f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{d}{dz}\mathbb{P}(Z \le z)$

$= \frac{d}{dz}\mathbb{P}(-\log Y \le z) = \frac{d}{dz}\mathbb{P}(Y \ge e^{-z})$

$= \frac{d}{dz}(1-\mathbb{P}(Y < e^{-z})) = \frac{d}{dz}(1-F_Y(e^{-z}))$

$= e^{-z} \cdot 6e^{-z} \cdot (1-e^{-z}) = 6e^{-2z} - 6e^{-3z}$

~~$f_Z(z) =$~~

$y \in (0,1)$

$e^{-z} \in (0,1)$

$z \in (0,+\infty)$

$$f_Z(z) = \begin{cases} 6e^{-2z} - 6e^{-3z} & (z>0) \\ 0 & \text{else} \end{cases}$$

(e) $\mathbb{E}[Z] = M_Z'(s) = \frac{d}{ds}\left(\frac{6}{6-5s+s^2}\right)$ $\cancel{(s \ne 0)}$ $(s<2)$

$= \frac{6(5-2s)}{((s-3)(s-2))^2}$

if $s=0$

then $\mathbb{E}[Z] = \frac{6\times 5}{(-3\times -2)^2} = \frac{5}{6}$

4(a) ~~$\mu_G = \mu$~~ $n_a = 12$ $\bar{X}_a = 3.2$ $S_a = 4.5$

$n_m = 12$ $\bar{X}_m = 6.9$ $S_m = 5.7$

$H_0: \mu_m = \mu_a \qquad H_1: \mu_m > \mu_a$

$$\mathbb{P}\left(t_{\alpha}(n_m+n_a-2) \geq \frac{\bar{X}_m - \bar{X}_a - (\mu_m - \mu_a)}{\sqrt{\frac{s^2}{n_m} + \frac{s^2}{n_a}}}\right)$$

$$s^2 = \frac{s_m^2(n_m-1) + s_a^2(n_a-1)}{n_m+n_a-2} = 26.37.$$

$$\mathbb{P}\left(t_{\alpha}(22) \geq \frac{6.9-3.2}{\sqrt{\frac{26.37}{12} + \frac{26.37}{12}}}\right)$$

$$\mathbb{P}(t_{\alpha}(22) \geq 1.7649)$$

~~P~~ p-value $\in (0.025, 0.05)$

This is moderate evidence against the null hypotheoio, suggesting. M-rated video games raise the heart rate more than G-rated video games, p-value is between 0.025 and 0.05.

(b) $\alpha = 1 - 0.95 = 0.05$

$$CI = \bar{X}_a \pm t_{1-\frac{\alpha}{2}}(n_a-1)\sqrt{\frac{s_a^2}{n_a}}$$

$= 3.2 \pm 2.201 \times 1.299$

$= 3.2 \pm 2.859$

CI ~~[illegible]~~ $\in (0.341, 6.059)$

5. (a) $n_M = 211$ $\qquad n_F = 188$

$$P_m = \frac{28}{211} = 0.1327 \qquad P_F = \frac{17}{188} = 0.0904$$

$\alpha = 1 - 0.99 = 0.01$

$$CI = P_m - P_F \pm t_{(1-\frac{0.01}{2}),(n_m+n_F-2)} \cdot \sqrt{\frac{P_m(1-P_m)}{n_m} + \frac{P_F(1-P_F)}{n_F}}$$

$$= \cancel{0.0363}\ 0.0423 \pm 2.576 \times \cancel{0.0306}\ 0.03135$$

$$= \cancel{0.0363}\ 0.0423 \pm \cancel{0.0789}\ 0.08076$$

$CI \in (-0.03846, 0.12306)$

~~CI ∈ (−0.0426, 0.1152)~~

~~CI ∈ (−0.04446, 0.11706)~~

As the interval covers 0. there is no evidence of the difference in the proportion of males and females affected

(b) ~~12 136~~

H_0: age and experiencing loss or damage due to a computer virus are independent

H_1: Some association between age and experiencing loss or damage due to a computer virus

12	136	148
12	120	132
19	101	120
43	357	400

$\frac{43 \times 148}{400} = 15.91$	132.09
14.19	117.81
12.90	107.1

$$X^2 = \frac{(12-15.91)^2}{15.91} + \frac{(136-132.09)^2}{132.09} + \frac{(12-14.19)^2}{14.19}$$

$$+ \frac{(120-117.81)^2}{117.81} + \frac{(19-12.90)^2}{12.90} + \frac{(101-107.1)^2}{107.1}$$

$= 4.68728$

$\mathbb{P}(\chi^2_{(2-1)(3-1)} \geq 4.68728)$

p-value $\in (0.05, 0.10)$

This is weak evidence against the null hypotheoio, suggesting some association between age and experiencing loss or damage due to a computer virus, p-value is between 0.05 and 0.1

6 (a) One additional Index of Difficulty is associated with an increase of 0.19801 in Target Acquisition Time

(b) $\cancel{CI = 0.42225 \pm t_{(1-\frac{1-0.95}{2}, (29-2))} \cdot 0.09725}$

$\cancel{= 0.42225 \pm 2.052 \times 0.09725}$

$\cancel{= 0.42225 \pm 0.199557}$

$\cancel{CI \in (0.222693, 0.621807)}$

$CI = \cancel{0.19801}\ 0.42225 \pm t_{(1-\frac{1-0.95}{2}, (29-2))} \times \cancel{0.02526}\ 0.09725$

$= \cancel{0.19801}\ 0.42225 \pm 2.052 \times \cancel{0.02526}\ 0.09725$

$= \cancel{0.19801}\ 0.42225 \pm \cancel{0.05183}\ 0.199557$

$\cancel{CI \in (0.14618, 0.24984)}$

$CI \in (0.22269, 0.621807)$

(c)

~~tStat~~

Let β_1 be the slope in the line relationship between IoD and TAT.

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$$tStat = \frac{0.19801 - \beta_1}{0.02526} = 7.83887$$

$$tStat_{\frac{\alpha}{2}}(29-2) = 7.83887$$

$$\frac{\alpha}{2} = 9.932 \times 10^{-9}$$

$$\alpha = 1.9864 \times 10^{-8}$$

$$p\text{-value} = 1.9864 \times 10^{-8}.$$

This is strong evidence against the null hypotheoio, suggesting some association between IoD and TAT

(d) The normal probability plot is relatively straight indicating normal distribution for residuals. The. residual vs fitted value shows some ~~fanning~~ increasing variability as fitted values suggesting the assumption of constant variance is violated. There is no clear trend in the residual plot so linearity assumption maybe is ok