STAT2203 - Assignment 3 solutions

Q1 (a). The joint pdf of (X,Y) is

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) \cdot f_X(x)$$

$$= \begin{cases} e^{-(y+x)} \cdot e^{-x}, & y \geq -x,\ x \geq 0 \\ 0, & \text{else} \end{cases}$$

$$= \begin{cases} e^{-y-2x}, & y \geq -x,\ x \geq 0 \\ 0, & \text{else} \end{cases}$$

We integrate out x to get the marginal pdf of Y

$$f_Y(y) = \int f_{X,Y}(x,y)\,dx$$

$(0 \wedge -y) = \min(0,-y)$

$$= \int_{(0\wedge -y)}^{\infty} e^{-y-2x}\,dx$$

$$= \begin{cases} e^{-y}\int_0^{\infty} e^{-2x}\,dx, & y > 0 \\ e^{-y}\int_{-y}^{\infty} e^{-2x}\,dx, & y < 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2}e^{-y}, & y \geq 0 \\ \frac{1}{2}e^{y}, & y < 0 \end{cases}$$

$$= \frac{1}{2}e^{-|y|}, \quad y \in \mathbb{R}$$

(ok if left in piece-wise form)

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

As $f_Y(y)$ is symmetric about 0, $\mathbb{E}[Y] = 0$.

$$\mathbb{E}[XY] = \mathbb{E}[X\,\mathbb{E}[Y|X]]$$

$$\mathbb{E}[Y|X=x] = \int y\, f_{Y|X}(y|x)\, dy$$

$$= \int_{-x}^{\infty} y\, e^{-(y+x)}\, dy \qquad \text{let } u = y+x,\ du = dy$$

$$= \int_{0}^{\infty} (u-x)\, e^{-u}\, du$$

$$= \int_{0}^{\infty} u e^{-u}\, du - x\int_{0}^{\infty} e^{-u}\, du$$

$$= 1 - x$$

(Do $\int_0^\infty u e^{-u}\,du$ by parts or recognise that this is just the expectation of a $\mathrm{Exp}(1)$ r.v)

$$\mathbb{E}[XY] = \mathbb{E}[X(1-X)]$$

$$= \mathbb{E}[X] - \mathbb{E}[X^2]$$

$$= \mathbb{E}[X] - \mathrm{Var}(X) - \mathbb{E}[X]^2$$

$$= 1 - 1 - 1 = -1$$

So $\mathrm{cov}(X,Y) = -1$

2(a) Let μ_M be mean time males spend playing video games / week.
" μ_F " " " females " " " "

We want a 99% CI for $\mu_M - \mu_F$. This has the form

estimate $\pm\ t_{\text{critical}} \times$ s.e(estimate)

We need the pooled sample variance

$$S_p^2 = \frac{(103-1)\times 4.55^2 + (97-1)\times 6.42^2}{103+97-2} = 30.64863$$

So the 99% CI is

$$(6.96 - 2.16) \pm t_{198;0.995} \sqrt{30.64863}\sqrt{1/103 + 1/97}$$

$$4.8 \pm 2.576 \times 5.5361 \times 0.14149$$

$$4.8 \pm 2.018 \text{ (hours)}$$

2.576 is from tables taking df $= \infty$
exact value for $t_{198;0.995} = 2.6008\ldots$

(b) Let μ_M be the mean response from males and
" μ_F " " " " " females

Test $H_0: \mu_M = \mu_F$ against $H_1: \mu_M \neq \mu_F$

test statistic $t = \dfrac{\text{estimate} - \text{hypothesised} ~~\text{value}~~}{\text{s.e.(estimate)}}$

$$= \frac{(2.92 - 2.74) - 0}{S_p\sqrt{1/103 + 1/97}}$$

$$S_p^2 = \frac{(103-1) \times 0.94^2 + (97-1) \times 1.05^2}{103 + 97 - 2} = 0.9897$$] should have done this first.

$$t = \frac{0.18}{\sqrt{0.9897}\sqrt{1/103 + 1/97}} = 1.2788$$

$$\text{p-value} = 2 \times \min\{\mathbb{P}(T_{198} > 1.2788), \mathbb{P}(T_{198} < 1.2788)\}$$

$$= 2 \times \mathbb{P}(T_{198} > 1.2788) \simeq 2 \times 0.1 \quad \text{(using table)}$$

$$= 0.2$$

There is no evidence / inconclusive evidence against the null hypothesis suggesting the mean response is the same for males + females.

(c) Let p_M be the proportion of males that did not play video games
and " p_F " " " " females " " " " " "

We want a 95% CI for $p_M - p_F$.

$\hat{p}_M = 0.113$ $\hat{p}_F = 0.214$ $n_M = 97$ $n_F = 103$

$$\hat{p}_M - \hat{p}_F \pm Z_{0.975} \times \sqrt{\frac{\hat{p}_M(1-\hat{p}_M)}{n_m} + \frac{\hat{p}_F(1-\hat{p}_F)}{n_F}}$$

$$(0.113 - 0.214) \pm 1.96 \times \sqrt{\frac{0.113 \times 0.887}{97} + \frac{0.214 \times 0.786}{103}}$$

$$-0.101 \pm 0.1012$$

(d) Test H_0: Gender and preferred medium are independent.
against H_1: There is an association between gender and preferred medium.

Expected counts

	Console	Portable	Comp.	Other	
Female	$\frac{81 \times 39}{167} = 18.9$	$\frac{81 \times 13}{167} = 6.3$	$\frac{79 \times 81}{167} = 38.3$	$\frac{36 \times 81}{167} = 17.5$	81
Male	$\frac{86 \times 39}{167} = 20.1$	$\frac{86 \times 13}{167} = 6.7$	$\frac{79 \times 86}{167} = 40.7$	$\frac{36 \times 86}{167} = 18.5$	86
	39	13	79	36	167

test statistic
$$X^2 = \frac{(18.9-13)^2}{18.9} + \frac{(20.1-26)^2}{20.1} + \frac{(6.3-6)^2}{6.3} + \frac{(6.7-7)^2}{6.7}$$
$$+ \frac{(38.3-36)^2}{38.3} + \frac{(40.7-43)^2}{40.7} + \frac{(17.5-26)^2}{17.5} + \frac{(18.5-10)^2}{18.5}$$
$$= 12.00$$

Compare the test statistic to the χ^2_3 distribution to get the p-value

$3 = (\text{rows}-1) \times (\text{columns}-1) = (2-1) \times (4-1)$

$\mathbb{P}(\chi^2_3 \geq 12.84) = 0.005$ and $\mathbb{P}(\chi^2_3 \geq 11.34) = 0.01$

The p-value is between 0·005 and 0·01. This is strong evidence against the null hypothesis, suggesting an association between gender and preferred medium for playing video games.