STAT 7203 : Assignment 2

name : Peng Yu

id : 46635884

1.(a) if Poisson distribution with mean 1,

then $\lambda = 1$

$$\mathbb{E}[e^{t\bar{x}}] = \mathbb{E}[e^{t\sum_{i=1}^{n} x_i \cdot \frac{1}{n}}]$$
$$= \mathbb{E}[e^{t\frac{x_1}{n}} \cdot e^{t\frac{x_2}{n}} \cdots e^{t\frac{x_n}{n}}]$$
$$= \mathbb{E}[e^{t\frac{x_1}{n}}] \cdot \mathbb{E}[e^{t\frac{x_2}{n}}] \cdots \mathbb{E}[e^{t\frac{x_n}{n}}]$$

$$M_X(t) = \mathbb{E}[e^{tx}] = \exp((e^t - 1)\lambda) = \exp(e^t - 1)$$

$$\mathbb{E}[e^{t\bar{x}}] = (\exp(e^{\frac{t}{n}} - 1))^n = e^{ne^{\frac{t}{n}} - n}$$

$$H(t;a) := e^{-at}\,\mathbb{E}[e^{t\bar{x}}] = e^{-at} \cdot e^{ne^{\frac{t}{n}} - n}$$
$$= e^{ne^{\frac{t}{n}} - n - at}$$

e^x is monotonically increasing, so we just need to compute the minimum of $(ne^{\frac{t}{n}} - n - at)$, which is the minimum of the original function

$$f(t) = ne^{\frac{t}{n}} - n - at$$
$$f'(t) = ne^{\frac{t}{n}} \cdot \frac{1}{n} - a = e^{\frac{t}{n}} - a$$

if $f(t)$ is minimum

then $f'(t) = 0$

$$e^{\frac{ta}{n}} - a = 0$$

$$t_a = n \ln a$$

So. ~~when~~ $t_a = n\ln a$. which mimimises $H(t:a)$

(b). $X \sim \text{Poisson}(\lambda)$ $\lambda = 1$.

$\mathbb{E}[X] = \lambda = 1$ $\quad \mathbb{E}[\bar{X}] = \mathbb{E}[\sum_{i=1}^{n} X_i \cdot \frac{1}{n}] = \frac{\sum_{i=1}^{n} \mathbb{E}[X_i]}{n} = 1$

$Var(X) = \lambda = 1$ $\quad Var(\bar{X}) = Var(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{1}{n}$

$n = 100,\ a = 1.1$ $\quad = \frac{1}{100}$

$$\mathbb{P}(\bar{X} \geq 1.1) = \mathbb{P}\left(\frac{\bar{X}-1}{\sqrt{\frac{1}{100}}} \geq \frac{1.1-1}{\sqrt{\frac{1}{100}}}\right) = \mathbb{P}\left(\frac{\bar{X}-1}{\sqrt{\frac{1}{100}}} \geq 1\right) = 0.159$$

$$\cancel{\frac{\mathbb{E}[e^{t\bar{X}}]}{e^{at}}} \quad \mathbb{P}(\bar{X} \geq a) \leq \frac{\mathbb{E}[e^{t\bar{X}}]}{e^{at}} = e^{-at}\,\mathbb{E}[e^{t\bar{X}}]$$

if $t = n\ln a$

then $e^{-at}\mathbb{E}[e^{t\bar{X}}]$ is mimimun.

$$\mathbb{P}(\bar{X} \geq 1.1) = e^{ne^{\frac{t}{n}} - n - at}$$

$$= e^{100 \times e^{\frac{100 \times \ln 1.1}{100}} - 100 - 1.1 \times 100 \times \ln 1.1}$$

$$= 0.616$$

Thus, the bound in (a) $= 0.616$ is longer than (b) $= 0.159$.

2. H_0: age and the perceived likelihood of their message being edited is independent

H_1: Some association between age and the perceived likelihood of their message being edited.

221	46	113	380
201	30	147	378
422	76	260	758

$\frac{422 \times 380}{758} = 211.6$	$\frac{76 \times 380}{758} = 38.1$	$\frac{260 \times 380}{758} = 130.3$
$\frac{422 \times 378}{758} = 210.4$	$\frac{76 \times 378}{758} = 37.9$	$\frac{260 \times 378}{758} = 129.7$

$$X^2 = \frac{(221-211.6)^2}{211.6} + \frac{(46-38.1)^2}{38.1} + \frac{(113-130.3)^2}{130.3}$$

$$+ \frac{(201-210.4)^2}{210.4} + \frac{(30-37.9)^2}{37.9} + \frac{(147-129.7)^2}{129.7}$$

$$= 8.72679.$$

$$\mathbb{P}(\chi^2_{(2-1)\times(3-1)} \geqslant X^2)$$

$$= \mathbb{P}(\chi^2_{(2)} \geqslant 8.72679.$$

thus: p-value is between 0.01 and 0.025

This is moderate evidence against ~~the null hypotheoio~~ ~~to~~ the null hypotheoio, suggesting some association between age and the perceived likeihood of their message being edited.

The p-value is between 0.01 and 0.025.

3. (a) $\mathbb{P}(X=\text{"Yes"}) = \mathbb{P}(X_{First} = \text{"Yes"}) + \mathbb{P}(X_{second} = \text{"Yes"})$

$$= \frac{1}{2} \times \frac{4}{6} + \frac{1}{2} \times P$$

thus $\mathbb{P}(X=\text{"Yes"}) = \frac{1}{3} + \frac{1}{2}P$

(b) $\cancel{\mathbb{E}[\sum_{i=1}^{220} X_i] =}$

$$\mathbb{E}[\sum_{i=1}^{220} X_i = \text{"Yes"}] = \sum_{i=1}^{220} \mathbb{E}[X_i = \text{"Yes"}]$$

$$= 220 \times \mathbb{P}(X = \text{"Yes"})$$

$$= 110 \times (\frac{2}{3} + P)$$

if $\sum_{i=1}^{220} X_i = \text{"Yes"} = 90$

then $\mathbb{E}[90] = 110 \times (\frac{2}{3} + P)$

$$\mathbb{E}[\frac{90}{110} - \frac{2}{3}] = P$$

$$\mathbb{E}[0.1515] = P$$

thus, 0.1515 is an unbiased estimator for p

(c). $\bar{\mathbb{P}}(X=\text{"Yes"}) = \frac{1}{3} + \frac{1}{2}\bar{P} = \frac{90}{220} = 0.4091$

$\alpha = 1-0.95 = 0.05 \quad \bar{P} = \frac{5}{33} = 0.1515$

$$\mu_{(X=\text{"Yes"})} = \bar{\mathbb{P}}_{(X=\text{"Yes"})} \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\bar{\mathbb{P}}(X=\text{"Yes"}) \cdot (1-\bar{\mathbb{P}}(X=\text{"Yes"}))}{220}}$$

$$= 0.4091 \pm 1.96 \times \sqrt{\frac{0.4091 \times (1-0.4091)}{220}}$$

$$\mu_{(X=\text{"Yes"})} = 0.4091 \pm 0.0650$$

$$\mu_{(X=\text{"Yes"})} \in [0.3441, 0.4741]$$

$$\mu_p = \bar{P} \pm Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\bar{P}\cdot(1-\bar{P})}{220}}$$

$$= 0.1515 \pm 1.96 \times \sqrt{\frac{0.1515 \times (1-0.1515)}{220}}$$

$$= 0.1515 \pm 0.0474$$

$$\mu_p \in [0.1041, 0.1989]$$

4. $H_0: \mu_{use} = \mu_{not}$ $\quad H_1: \mu_{use} < \mu_{not}$

$n_{use} = 320 \quad n_{not} = 873.$

$\bar{P}_{use} = \frac{194}{320} = 0.60625 \quad \bar{P}_{not} = \frac{614}{873} = 0.70332$

$$\bar{P} = \frac{194+614}{320+873} = 0.67728$$

$$P\left(Z_\alpha \geq \frac{(\bar{P}_{not} - \bar{P}_{use}) - (\mu_{not} - \mu_{use})}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_{use}} + \frac{\bar{P}(1-\bar{P})}{n_{not}}}}\right)$$

$$P\left(Z_\alpha \geq \frac{0.70332 - 0.60625 - 0}{\sqrt{\frac{0.67728 \times (1-0.67728)}{320} + \frac{0.67728 \times (1-0.67728)}{873}}}\right)$$

$$P(Z_\alpha \geq 3.177)$$

$$0.0005 < \alpha < 0.001$$

thus p-value is between 0.0005 and 0.001

This is strong evidence against the null hypotheoio, suggest the data provide evidence that dog owners who use electric collars are less likely to take their dog to formal obedience training.

The p-value is between 0.0005 and 0.001

5. $n_A = 18$ $\bar{X}_A = 460$ $S_A = 32$

$n_B = 22$ $\bar{X}_B = 478$ $S_B = 46$

(a) $\alpha = 1 - 90\% = 0.1$

$$\mu_B = \bar{X}_B \pm t_{(1-\frac{\alpha}{2})}(n_B - 1)\sqrt{\frac{S_B^2}{n_B}}$$

$$= 478 \pm t_{(1-0.05)}(22-1) \times \sqrt{\frac{46^2}{22}}$$

$$= 478 \pm 1.721 \times 9.807$$

$$= 478 \pm 16.878$$

$$\mu_B \in [461.122, 494.878]$$

(b) $H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$.

$$S_p^2 = \frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A + n_B - 2}$$

$$S_p^2 = \frac{(18-1)\times 32^2 + (22-1)\times 46^2}{18+22-2}$$

$$= 1627.47$$

$$\mathbb{P}\left(t_{\frac{\alpha}{2}}(n_A+n_B-2) \geqslant \frac{(\bar{X}_B-\bar{X}_A)-(\mu_B-\mu_A)}{\sqrt{\frac{S_p^2}{n_A}+\frac{S_p^2}{n_B}}}\right)$$

$$\mathbb{P}\left(t_{\frac{\alpha}{2}}(18+22-2) \geqslant \frac{22-18-0}{\sqrt{\frac{1627.47}{18}+\frac{1627.47}{22}}}\right)$$

$$\mathbb{P}\left(t_{\frac{\alpha}{2}}(38) \geqslant 1.404\right).$$

$$0.05 < \frac{\alpha}{2} < 0.1$$

$$0.1 < \alpha < 0.2$$

thus p-value is between 0.1 and 0.2

This is inconclusive evidence against ~~the null~~ H_0 ~~hypothesis~~, suggesting the data provide evidence of a same in mean PEF between the two drugs.

The p-value is between 0.1 and 0.2