SIT191 Problem Solving Task 2

Due: by 8pm Monday September 13th

Total marks: 80, Weighting: 20%

2.1 People have one of four blood types – O, A, B or AB and each blood type is further classified as Rh-positive or Rh-negative. For a particular country, the percentages of the population having each blood type are shown below.

Rh	O	A	B	AB
positive	38%	34%	9%	3%
negative	7%	6%	2%	1%

a) For a randomly selected person from this country, what is the probability that their blood type is

i) not A positive? $1 - 0.34 = 0.66$

ii) B positive or AB? $0.09 + 0.04 = 0.13$

iii) type O or Rh-positive? $0.45 + 0.84 - 0.38 = 0.91$

b) Among 3 random blood donors, what is the probability (to 4 decimal places) that

i) all are type A? $0.4^3 = 0.064$

ii) none of them are Rh-negative? $0.84^3 = 0.5927$

iii) at least one person is type O positive? $1 - 0.38^3 = 0.9451$

iv) the third donor only has blood type B? $0.89 * 0.89 * 0.11 = 0.0871$

[3+5 = 8 marks]

2.2 Most wild freshwater fish carry parasites which usually don't affect the health of the fish. Due to environmental changes, parasites can increase and result in fungal and tapeworm infections causing visible symptoms on fish.
In fish observed at one location, 15% had fungal infections, 7% had fungal and tapeworm infections and 9% were infected with tapeworm. Let A be the event that a fish has a fungal infection, and let B be the event that a tapeworm infection is present. Using a Venn diagram or otherwise, answer the following:

a) What is the probability that a fish doesn't show any signs of a fungal or tapeworm infection?
$1 - 0.17 = 0.83$

b) Calculate P(A or B).
$0.15 + 0.09 - 0.07 = 0.17$

c) Find P(B|A).
$\frac{0.07}{0.15} = 0.4667$

d) Are fungal and tapeworm infections independent events? Explain.
$P(A\ and\ B) = 0.07$
$P(A) * P(B) = 0.0135$
$Since\ P(A\ and\ B)\ and\ P(A) * P(B)\ are\ not\ the\ same, they\ are\ not\ indpendent.$

[2+2+2+2 = 8 marks]

2.3 Diagnostic tests of medical conditions can be positive or negative whether or not a patient has the condition. A false negative occurs when the test result is negative given that the patient does in fact have the condition. A positive test can indicate that the patient has the condition or is a false positive when the patient does not actually have the condition. Consider a sample of 200 randomly selected patients, some of whom have a medical condition. Results for a new diagnostic test for the condition are shown.

	Condition present	Condition absent	**Totals**
Positive test result	100	20	120

Negative test result	10	70	80
Totals	110	90	200

What is the probability of a false positive or false negative result?
Include any calculations or diagrams used.

$P(false\ positive) + P(false\ negative)$

$\frac{20}{90} + \frac{10}{80} = 0.3472$ [3 marks]

2.4 For healthy adults, the main pulmonary artery (MPA) diameter is 2.73 cm, with standard deviation 0.31 cm. Assume MPA diameter is normally distributed.

a) Use the **68-95-99.7% rule** to find the percentage of healthy adults expected to have an MPA diameter between 1.80 cm and 3.04 cm.

$100 - (16 + 0.15) = 0.83$

b) What proportion of healthy adults have an MPA diameter

i) no greater than 3.15 cm?

$\frac{3.15-2.73}{0.31} = 1.355$

$P(z < 1.355) = 0.9123$

ii) between 2 cm and 3 cm?

$\frac{2 - 2.73}{0.31} = -2.35$

$\frac{3 - 2.73}{0.31} = 0.87$

$P(-2.35 < z < 0.87) = 0.7988$

c) What is the cut off length for the largest 20% of diameters?

$z > 0.8416 = \frac{x - 2.73}{0.31} => 2.99$

$2.99cms$

d) Give the interquartile range of the MPA diameters.

$z\ score\ for\ 25th\ perecentile = -0.674$

$z\ score\ for\ 75th\ perecentile = 0.674$

$z = \frac{Q3 - \mu}{\sigma} => 0.674 = \frac{Q3 - 2.73}{0.31} => Q3 = 2.94$

$z = \frac{Q1 - \mu}{\sigma} => -0.674 = \frac{13 - 2.73}{0.31} => Q1 = 2.52$

$IQR = 2.94 - 2.52 = 0.42$

[1+3+2+3 = 9 marks]

2.5 Assume 70% of the adult population are vaccinated for a particular virus.

a) Amongst 7 randomly selected adults, what is the probability that:

i) no-one is vaccinated?

$\frac{7!}{0!\,(7-0)!} * (0.70)^0 * (0.30)^7 = 0.0002187$

ii) exactly 3 people are vaccinated?

$\frac{7!}{3!\,(7-3)!} * (0.70)^3 * (0.30)^4 = 0.0972405$

iii) no more than 5 people have been vaccinated?

$1 - [P(x = 6) + P(x = 7)]$

$1 - (0.2471 + 0.08235)$

$P(x \leq 5) = 0.67058$

b) For 300 randomly selected adults,

i) how many would you expect to be vaccinated?

$300 * 0.7 = 210$

ii) what is the probability that less than 190 people have been vaccinated?

$P(x < 190) = P\left(\frac{x - \mu}{\sigma} < \frac{190 - 210}{7.94}\right)$

$P(z < -2.52)$

$= 0.00587$

[7+4 = 11 marks]

2.6 Platelet rich plasma (PRP) therapy is used to accelerate the healing of injured tendons, ligaments and joints, using plasma derived from the patient's own blood. A study involving 255 randomly selected patients with shoulder pain caused by rotator cuff tears reported that after receiving a PRP injection, 225 patients experienced at least a 50% reduction in pain.
Is there evidence that PRP therapy relieves pain in more than 85% of patients with such shoulder pain?

a) Write appropriate hypotheses.
$H_0 = 0.85, H_A > 0.85$

b) Check the assumptions and conditions.
$\textit{The sample size is big enough that } np \geq 10 \textit{ and } n(1-P) \geq 10, \textit{meaning it is simple random sample.}$

c) Calculate the test statistic and find the P-value.
$$z = \frac{0.88 - 0.85}{\sqrt{\frac{0.85 * 0.15}{255}}} = 1.34$$

d) State your conclusion in plain English. Use a significance level of 5%.
Since the p value is greater than the significance level, we fail to reject the null hypothesis. Therefore, we can conclude that there isn't enough evidence that PRP therapy relieves pain in more than 85% of patients with such shoulder pain.

e) Calculate a 90% confidence interval for the true proportion of patients with shoulder pain caused by rotatorcuff tears whose pain is reduced by the PRP therapy and interpret your interval in context.

$$0.88 \pm 1.645 * \sqrt{\frac{0.85(1-0.85)}{255}}$$

$$0.88 \pm 0.04$$

$$(0.84, 0.92)$$

The 90% confidence shows that the proportion of the patients with shoulder pain caused by rotator cuff tears whose pain is reduced by the PRP therapy lies between 0.84 to 0.92

f) Explain how the confidence interval supports your conclusion from part d).
The confidence interval supports my conclusion because 0.85 lies between 90% confidence interval 0.84 to 0.92.

[1+2+3+1+3+2 = 12 marks]

2.7 A clinical trial for a potential new drug to treat attention deficit hyperactivity disorder (ADHD) was performed. While the drug appeared to reduce the symptoms of ADHD, there was concern over possible side effects. Researchers reported that 85 of the 374 patients who took the drug experienced side effects, compared to 65 of the 210 subjects who were given a placebo. Patients didn't know which treatment they were given.
Is there evidence of a significant difference between the treatment groups for the proportions of subjects who experience side effects?

$$\widehat{P_A} = \frac{85}{374} = 0.22727$$
$$\widehat{P_B} = \frac{65}{210} = 0.30952$$

a) Write appropriate hypotheses.
$H_0: P_A = P_B$

$H_A: P_A \neq P_B$

b) Calculate the test statistic, find the P-value and state your conclusion in plain English. Use α = 0.05.

$$\frac{\widehat{P_A} - \widehat{P_B} - (P_A - P_B)}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{P} = \frac{85 + 65}{374 + 210} = 0.25684$$

$$z = \frac{(0.22727 - 0.30952) - (0)}{\sqrt{0.25684(1 - 0.25684)\left(\frac{1}{374} + \frac{1}{210}\right)}}$$

$$z = -2.18324$$

$p - value: 0.029036$

$p - value < 0.05$

Since the p value is less than the significant level, we reject the null hypothesis. Therefore, there is a significant difference between the treatment groups for the proportions of subjects who experience side effects.

c) Create a 95% confidence interval for the difference in the proportions of side effects and interpret your interval in context.

$$\left(\widehat{P_A} - \widehat{P_B}\right) \pm z^* \times \sqrt{\frac{\widehat{p_A} * \widehat{q_A}}{n_1} + \frac{\widehat{p_B} * \widehat{q_B}}{n_2}}$$

$$(0.22727 - 0.30952) \pm 1.96 \times \sqrt{\frac{(0.22727)(1 - 0.22727)}{374} + \frac{(0.30952)(1 - 0.30952)}{210}}$$

$$(-0.1578, -0.0066)$$

95% confident that the mean difference between the two treatment is between $-0.1578\ and\ -0.0066$

d) Did you expect 0 to be in your confidence interval? Explain briefly.

Confidence interval doesn't contain 0 so we reject eh null hypothesis

[1+5+4+2 = 12 marks]

2.8 The council of a regional town conducted a survey to determine how their workers usually travel to work. The study found that 68% of workers travelled using their cars, 17% used public transport, 11% rode bicycles and 4% walked.
A local large employer ran a similar survey at her workplace, with 150 responses recorded. Below are the observed counts for each method of transportation used by the sampled workers:

Method of Transport	Observed Count
Car	98
Public Transport	29
Bicycle	20
Walk	3

a) If the sampled workers' transportation methods match the findings of the council's survey, what are the expected numbers for each method of transport?

Method of Transport	Observed Count	Expected count
Car	98	$150 * \frac{68}{100} = 102$
Public Transport	29	$150 * \frac{17}{100} = 25.5$
Bicycle	20	$150 * \frac{11}{100} = 16..5$
Walk	3	$150 * \frac{4}{100} = 6$

b) State your hypotheses for the appropriate test.
H_0: Workers usually travel to work are independent
H_A: Workers usually travel to work are dependent

c) Calculate the test statistic value.

$$\chi^2 = \frac{(98-102)^2}{102} + \frac{(29-25.5)^2}{25.5} + \frac{(20-16.5)^2}{16.5} + \frac{(3-6)^2}{6}$$
$$\chi^2 = 2.8796$$

d) State the degrees of freedom and P-value.

$$Number\ of\ degrees\ of\ freedom = 4 - 1 = 3$$
$$P(\chi^2 > 2.879) = 0.410563$$

e) Give your conclusion (use α = 0.05) in the context of the question.
Since p value is greater than alpha, we fail to reject the null hypothesis, hence there isn't enough evidence to conclude workers usually travelling to work are dependent.

[1+1+2+2+2 = 8 marks]

2.9 Alcohol consumption patterns were recorded for a random sample of 180 university students to explore a possible relationship between gender and alcohol consumption, with the results shown below.

	At least weekly	**Less than weekly**	**None**
Female	38	24	24
Male	57	22	15

Carry out the appropriate hypothesis test and include all the steps (including checking the assumptions and conditions). Use a significance level of 10%.

H_0: gender and alcohol consumption are independent
H_A: gender and alcohol consumption are dependent

Observed

	At least weekly	**Less than weekly**	**None**	**Total**
Female	38	24	24	86
Male	57	22	15	94
Total	95	46	39	180

Expected

	At least weekly	**Less than weekly**	**None**
Female	$\frac{86*95}{180} = 45.39$	$\frac{86*46}{180} = 21.98$	$\frac{86*39}{180} = 18.63$
Male	$\frac{94*95}{180} = 49.61$	$\frac{94*46}{180} = 24.02$	$\frac{94*39}{180} = 20.37$

$$\frac{(38-45.39)^2}{45.49} + \frac{(24-21.98)^2}{21.98} + \frac{(24-18.63)^2}{18.63} + \frac{(57-49.61)^2}{49.61} + \frac{(22-24.02)^2}{24.02} + \frac{(15-20.37)^2}{20.37}$$

$= 5.6194$

$d.f = 2, \chi^2 = 5.6194$
$P - value = 0.060222$

Since p value is less than alpha, therefore we reject the null hypothesis, hence there is significant evidence that alcohol consumption and gender are not independent.

[9 marks]