SIT191 Problem Solving Task 2 solutions

2.1

a)

i) P(not A positive) = 1 – P(A positive) = 1 - 0.34=0.66 or 66%

ii) P(B positive or AB) = 0.09+0.04 =0.13 or 13%

iii) P(type O or Rh-positive) = P(type O)+P(Rh-positive) – P(O positive) =0.45+0.84-0.38=0.91 or 91%

b)

i) P(all type A) = 0.4^3 = 0.064 or 6.4%

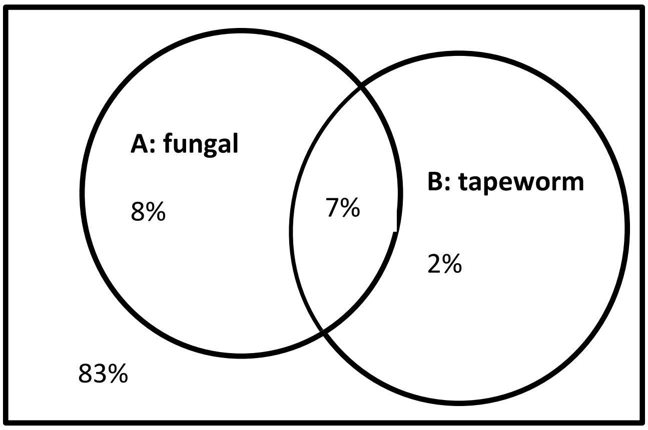
ii) P(none Rh-negative) = P(all not Rh-negative) = 0.84^3 = 0.5927 or 59.27%

iii) P (at least one O positive)= 1 – P(none O positive) = 1- 0.62^3 =0.7617 or 76.17%

iv) P (third only is type B)= 0.89×0.89×0.11 = 0.0871 or 8.71%

[3+5=8 marks]

2.2



a) P(neither) = 83% or 0.83

b) P(A or B) = 17% or 0.17 (8%+7%+2%, or 100%-83% or addition rule: 15%+9%-7%)

c) P (B|A) = $\frac{\text{P(A and B)}}{\text{P(A)}} = \frac{0.07}{0.15} = 0.467$ or 46.7%

(i.e. the probability of a fish having a tapeworm infection given that it has a fungal infection)

d) For independent events, the multiplication rule holds: P(A and B) = P(A) × P(B)

P(A and B) = 0.07 and P(A)×P(B) = 0.15×0.09=0.0135.

Since these answers are not equal, fungal and tapeworm infections are not independent events.

(Or P(B|A) = P(B) if independent, and 0.467 is not equal to 0.09, so not independent)

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

2.3

A false positive is the conditional probability that the result is positive given that the patient doesn't have the condition. i.e. P(positive result|condition absent) = $\frac{P(positive\ result\ and\ condition\ absent)}{P(condition\ absent)} = \frac{20/200}{90/200}$ or $\frac{20}{90} = 0.2222$

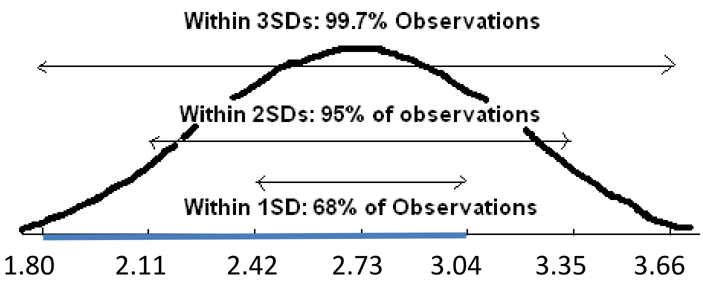
A false negative is the conditional probability that the result is negative given that the patient has the condition. i.e. P(negative result|condition present) = $\frac{P(negative\ result\ and\ condition\ present)}{P(condition\ present)} = \frac{10/200}{110/200}$ or $\frac{10}{110} = 0.0909$

Add for the final answer: 0.2222+0.0909=0.313 or 31.3%

[3 marks]

2.4 $\mu = 2.73 \quad \sigma = 0.31$

a)



99.7% - ½(99.7%-68%) = 83.85% or 68%+½(95%-68%)+ ½(99.7%-95%)

b)

i) $P(x < 3.15) = P(z < \frac{3.15-2.73}{0.31})$

$= P(z < 1.35)$

=0.9115 or 91.2% (or with technology: 0.9123)

ii) $P(2 < x < 3) = P\left(\frac{2-2.73}{0.31} < z < \frac{3-2.73}{0.31}\right)$

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

= 0.8078-0.0094=0.7984 or 79.84% (or with technology: 0.7988)

c) Solve $z = \frac{x-2.73}{0.31}$ for *x*, with z=0.84 for the largest 20%

$0.84 = \frac{x-2.73}{0.31}$

$x = 2.99$ cm

d) Solve $z = \frac{Q_1-2.73}{0.31}$ for Q_1, with z= -0.67 for the lowest 25% and

Solve $z = \frac{Q_3-2.73}{0.31}$ for Q_3, with z= 0.67 for the highest 25%

Q_1 = 2.73-0.67×0.31=2.52

Q_3 = 2.73+0.67×0.31=2.94

IQR = $Q_3 - Q_1$ = 2.94 – 2.52= 0.42 cm

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

2.5

p = 0.7, q = 1- p = 0.3

a) i) $P(X = 0) = \binom{7}{0}0.7^0 0.3^7 = 0.0002187$ or $0.3^7 = 0.0002187$

ii) $P(X = 3) = \binom{7}{3}0.7^3 0.3^4$

$= 0.0972$

iii) $P(no\ more\ than\ 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

or $1 - (P(X = 6) + P(X = 7))$ using the complement rule

$= 1 - (\binom{7}{6}0.7^6 0.3^1 + \binom{7}{7}0.7^7 0.3^0)$

$= 1 - (0.24706 + 0.08235)$

$= 0.6706$

b) i) $E(x) = \mu = np = 300 \times 0.7 = 210$

ii) $\sigma = \sqrt{npq} = \sqrt{300 \times 0.7 \times 0.3} = 7.93725$

$P(X < 190) = P(z < \frac{190-210}{7.93725})$ (using a normal approximation)

$= P(z < -2.52)$

= 0.0059 (Note the exact answer using the binomial distribution is 0.0055)

[7+4 = 11 marks]

2.6

a) $H_0: p = 0.85$
$H_a: p > 0.85$

b) **Independence assumption**: since the patients were randomly selected, we can assume the observations are independent regarding individual pain reduction

Success/failure condition:

$np = 216.75 > 10$, $nq = 38.25 > 10$, so the sample proportion is normally distributed

c) $\hat{p} = \frac{225}{255} = 0.8824$ $\hat{q} = 1 - \hat{p} = 0.1176$

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

$$= \frac{0.8824-0.85}{\sqrt{\frac{0.85\times0.15}{255}}} = \frac{0.0324}{0.02236} = 1.45$$

∴ P-value = P(z > 1.45)= 1 – 0.9265 = 0.0735

d) Since the P-value is $> \alpha = 0.05$, we fail to reject H_0.
There is no evidence that the PRP treatment relieves pain in significantly more than 85% of patients with shoulder pain caused by rotator cuff tears.

e) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$= 0.8824 \pm 1.645 \sqrt{\frac{0.8824 \times 0.1176}{255}}$$

$= 0.8824 \pm 1.645 \times 0.02017$

$= 0.8824 \pm 0.03318 = (0.8492, 0.9156)$

or 84.9% to 91.6%

We are 90% confident that between 84.9% and 91.6% of all patients with shoulder pain caused by rotator cuff tears will experience at least a 50% reduction in pain due to the PRP treatment.

f) The confidence interval supports the decision in d) to fail to reject H_0 (that p = 85%) since 0.85 is included in the interval.

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

2.7

a) $H_0 : p_1 = p_2$ (or $p_1 - p_2 = 0$) for 1: placebo, 2: new drug
$H_a : p_1 \neq p_2$

OR:
H_o : There is no difference in the proportions of subjects who experience side effects between either drug
H_a : The proportions of subjects who experience side effects are different for the new drug and placebo

b) $\hat{p}_1 = \frac{65}{210} = 0.3095$ $\hat{q}_1 = 0.6905$
$\hat{p}_2 = \frac{85}{374} = 0.2273$ $\hat{q}_2 = 0.7727$

$\hat{p}_{\text{pooled}} = \frac{65+85}{210+374} = 0.2568$ $\hat{q}_{\text{pooled}} = 0.7432$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}}$$

$$= \frac{0.3095 - 0.2273}{\sqrt{\frac{0.2568 \times 0.7432}{210} + \frac{0.2568 \times 0.7432}{374}}}$$

$= 2.18$ (or -2.18 if 1:new drug and 2:placebo) (or using the online calculator, z =2.1832)

P-value = $2 \times P(Z > 2.18) = 2 \times 0.0146 = 0.0292$ (or using the online calculator, P-value = 0.02926)

Since the P-value < $\alpha = 0.05$, reject H_0.
There is evidence that the proportion of subjects who experience side effects on the new drug is significantly different to the proportion who experience side effects on the placebo.

c) $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

$= (0.3095 - 0.2273) \pm 1.96 \sqrt{\frac{0.3095 \times 0.6905}{210} + \frac{0.2273 \times 0.7727}{374}}$
$= 0.0822 \pm 0.0756$
= (0.0066, 0.1578) (or, -0.1578 to -0.0066 for new drug - placebo)

We are 95% confident that the proportion of subjects who experience side effects on the placebo is between 0.7% and 15.8% higher than the proportion who experience side effects on the new drug.

d) The confidence interval was not expected to contain 0 (and it doesn't, since it is 0.7% to 15.8%).
This is consistent with the decision to reject H_0 in b) that there is no difference ($p_1 - p_2 = 0$) in the proportions. The difference is significantly different to 0.

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

2.8

a) Expected counts:

Transport method	Expected Count
Car	68% of 150 = 102
Public transport	17% of 150 = 25.5
Bike	11% of 150 = 16.5
Walk	4% of 150 = 6

b) H_0 : the distribution of transportation methods for the sampled workers matches the council's survey results
H_a : the distribution of transportation methods for the sampled workers doesn't match the council's survey

c) $\chi^2 = \sum \frac{(obs-exp)^2}{exp}$

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

= 2.88

d) *df* = 4 - 1 = 3
P-value > 0.3 (or, exact: 0.41055)

e) Fail to reject H_0 (P-value > α = 0.05).
There is no significant evidence that the distribution of transportation methods for the sampled workers is different to the council's survey results.

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

2.9

H_0 : alcohol consumption and gender are independent
H_a : alcohol consumption and gender are not independent

We must have counted data, independent observations and each expected count is at least 5. These are all met.

Expected values:

	At least weekly	Less than weekly	None	TOTAL
Female	45.4	22.0	18.6	86
Male	49.6	24.0	20.4	94
TOTAL	95	46	39	180

$$\chi^2 = \sum \frac{(obs-exp)^2}{exp} = \frac{(38-45.4)^2}{45.4} + \frac{(24-22)^2}{22} + \frac{(24-18.6)^2}{18.6} + \frac{(57-49.6)^2}{49.6} + \frac{(22-24)^2}{24} + \frac{(15-20.4)^2}{20.4}$$

= 5.66 (exact: 5.6194)

df = (R-1) (C-1) = (2-1) (3-1) = 2

0.05<P-value < 0.1 (exact: 0.060222)
Reject H_0 since the P-value <α=0.1
There is significant evidence that alcohol consumption and gender are not independent.
Using row percentages, it appears that a greater proportion of males than females drink at least weekly and a greater proportion of females don't drink.

[9 marks]