Kamonchanok Suban Na Ayudtaya
300471606
Tuesday 12pm- Oliver R

Question 1a

Row Labels	Count of Gender
Female	100
European	79
Maori	13
Other	3
Pacific	5
Male	100
European	76
Maori	13
Other	11
Grand Total	200

Question 1b
P(Female)=100/200=0.5
P(Maori ∩ male)=13/200=0.065
P(european | female)=79/100=0.79
P(maori | female)=13/100=0.13
P((Pacific ∩ other') U (Pacific' ∩ other) | female)=8/100=0.08
P(Pacific | female)=5/100=0.05
P(Other | female)=3/100=0.03
P(male | european)=76/155=0.49(2dp)

Question 1c
It is likely that gender and ethnicity are independent as a person being male or female does not influence their ethnicity (vice versa)
If events are independent, then: P(A) × P(B) = P(A ∩ B)
P(female) x P(european) = P(female and european)
(100/200)x(155/200)=15500/40000=0.3875=0.39(2dp)
P(female and european)=P(female and european)=79/200=0.395(3d.p)

Question 2a
P(Income>199.99)=187/200=0.935

Question 2b
Binomial random variable
x ~Bin (6 ,0.935)

Question 2c
n=6 p=0.935

success	n	p
6	6	0.658143373
5		0.278590819
4		0.04843557
3		0.004489572
2		0.000234082
1		6.50923E-06
0		7.54189E-08

Most likely number (the mode)= 6

Probability none in the sample have an income over $199.99 per week = $P(X=0)= 7.54189E-08$

What is the probability that more than half do?$=P(3 \geq X)=0.004730238$
$=P(x>3)=1-0.004730238=0.995269762$

What is the expected number of individuals?
$E[X]=(0 \times 0.000000075) + (1 \times 0.000006509) + (2 \times 0.000234082) + (3 \times 0.004489572) + (4 \times 0.04843557) + (5 \times 0.278690819) + (6 \times 0.668143373)=5.610000002$

what is the standard deviation of the number?
$\sigma= \sqrt{E[X^2] - (E[X])^2}$
$E[X^2]=(0^2 \times 0.000000075) + (1^2 \times 0.000006509) + (2^2 \times 0.000234082) + (3^2 \times 0.004489572) + (4^2 \times 0.04843557) + (5^2 \times 0.278690819) + (6^2 \times 0.668143373)=31.83675001$

$(E[X])^2=5.610000002^2=31.47210002244$

$\sigma= \sqrt{31.83675001 - 31.47210002244}$
$\sigma= 0.6039$ (4.dp)

<u>Question 2d</u>
If an individual could not be selected more than once into the sample, what implications does this have for the binomial assumptions?
The probability of drawing a person with an income over $199.99 per week decreases with each draw if replacement is not possible
e.g
$(187/200)\times(186/199)\times(185/198)\times(184/197)\times(183/196)\times(182/195)$

mean=

<u>Question 3a</u>
mean=34.65
std=15.77631008

<u>Question 3b</u>
What is the probability that this person works more than 40 hours per week?
$P(x>4)=1-P(x<40)$
$P(x>40)=1-0.63273885$
$P(x>40)=0.36726115$
=NORM.DIST(40,34.65,15.77631008,TRUE)

What is the probability they work less than 20 hours per week?
$P(x<20)=0.17654626$
=NORM.DIST(20,34.65,15.77631008,TRUE)

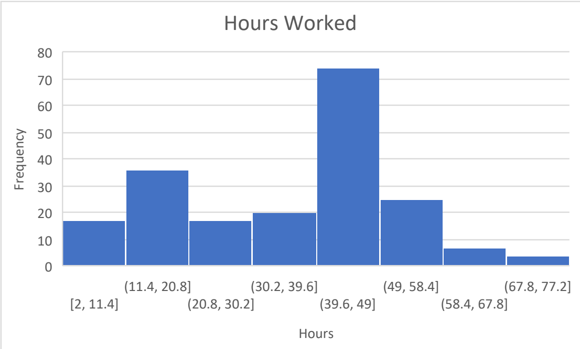
What is the probability they work between 35 and 45 hours per week?
$P(x<45)-P(x<35)=0.74410304-0.50884987$
$P(x<45)-P(x<35)=0.23525317$

=NORM.DIST(45,34.65,15.77631008,TRUE)
=NORM.DIST(35,34.65,15.77631008,TRUE)

<u>Question 3c</u>

40th percentile: 37.4
=PERCENTILE.EXC(G:G,0.4)

Question 3d



From the graph we can see that using normal distribution may not be valid in this case, as the sample is not central and appears to be skewed to the left.
Additionally there appears to be 2 modes in the sample.
As a result this may mean that the sample is unlikely to be well represented by normal distribution as the data is inconsistency scattered.