limaida_300536130_assignment1

1a.
No, micrograms per millilitre can take decimal number form, and will not be able to be categorised.

b.
It would be categorial, if you considered all people between 18 years old and 19 years old as an "18-year-old", then this variable would be categorical as 18-year-old would be the category.
However, if you considered people 18.14159265 years old, it would not be categorical.

c.
Yes, there are only two categories (favour and oppose)

d.
Yes, there are only two categories (yes or no)

e.
Yes, there are only five categories (1="None",2="Mild",3="Moderate",4="Severe",5="Very severe")

f.
No, millimetres can take decimal number form, and will not be able to be categorised.
However, if all values were rounded to the nearest millimetre, then it would be categorical.

2a.
We are assuming that there is an equal chance that these people do not have a job, and that each person has no impact on another's (All individuals are independent on others chances).

b.
Expected value = 3.00 (gotten from n*p = 12*0.25)
Variance = 2.25 (gotten from n*p*(1-p) = 4*0.75)

c.
P(Y=2) = $(12/2)*0.25^2*(1-0.25)^{12-2}$= $(12!/(10!*2!))*0.25^2*0.75^{10}$
Approx. = (12!/(10!*2!)) *0.0625*0.0563135147
Approx. = 0.23229324817

di.
```
> 1-(pbinom(3,12,0.25))
[1] 0.3512214
```

P(Y ≥ 4) = P(Y=4) + P(Y=5) + P(Y=6) + P(Y=7) + P(Y=8) + P(Y=9) + P(Y=10) + P(Y=11) + P(Y=12) =
.1935777+.10324144+.04014945+.01147127+23898e-3+3.5405e-4+3.5405e-5+2.1458e-6+5.9605e-8
= 0.351221320405

ii.

μ = (0.25*12) = 3

σ^2 = (12*0.25*(1-0.25)) = 2.25

σ = (sqrt(2.25)) = 1.5

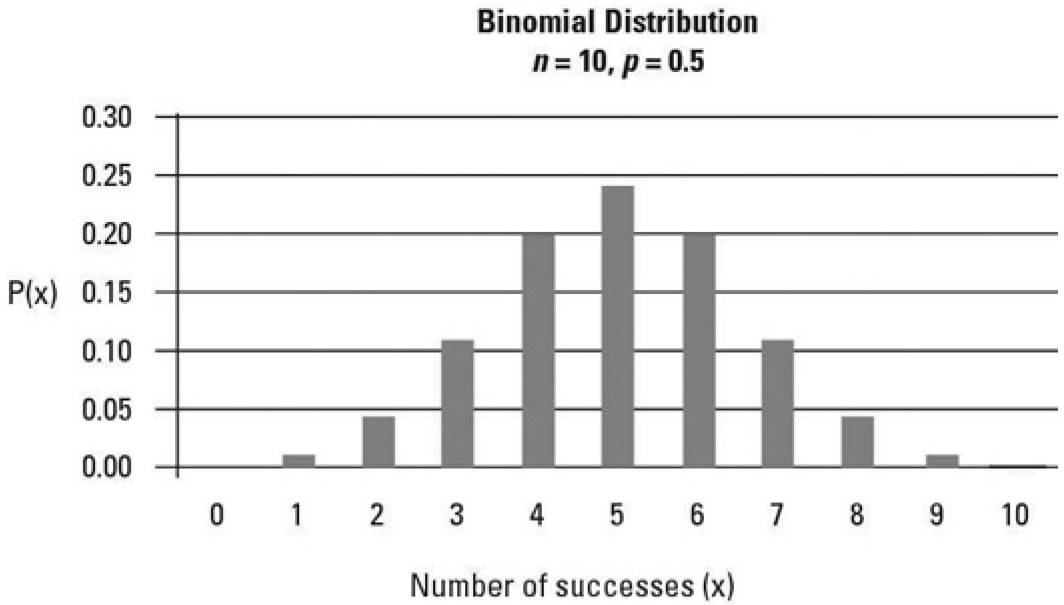
approx. = P(Z≥(3.5-3)/1.5) = 0.33

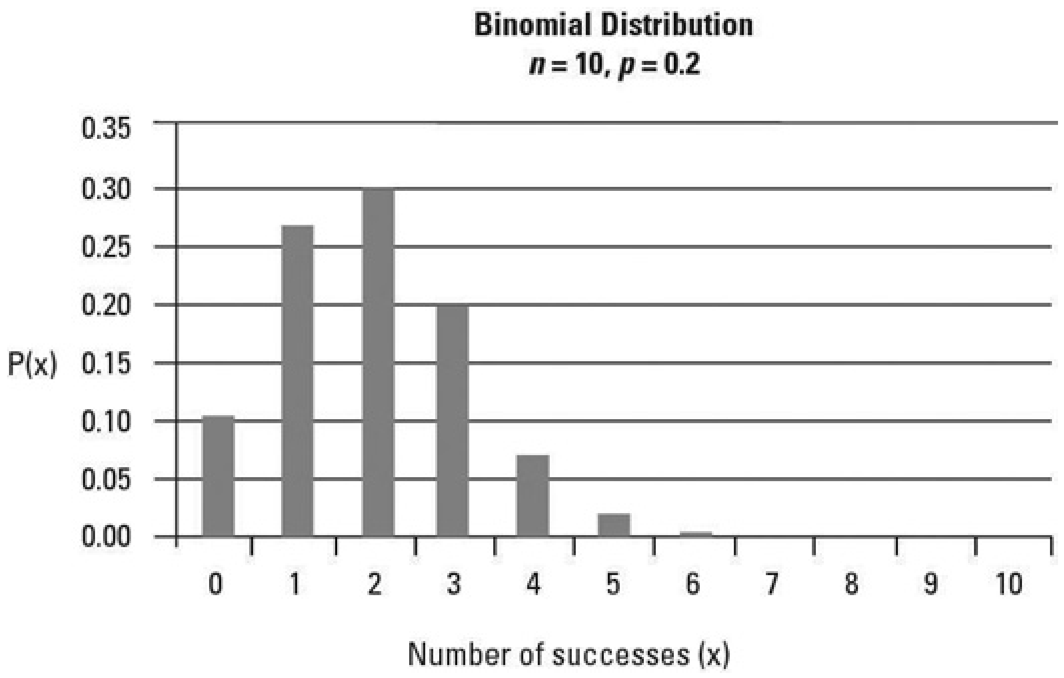
Z = 0.33 on the Standard Normal Probability Table Results in 0.5-0.1293 = 0.3707

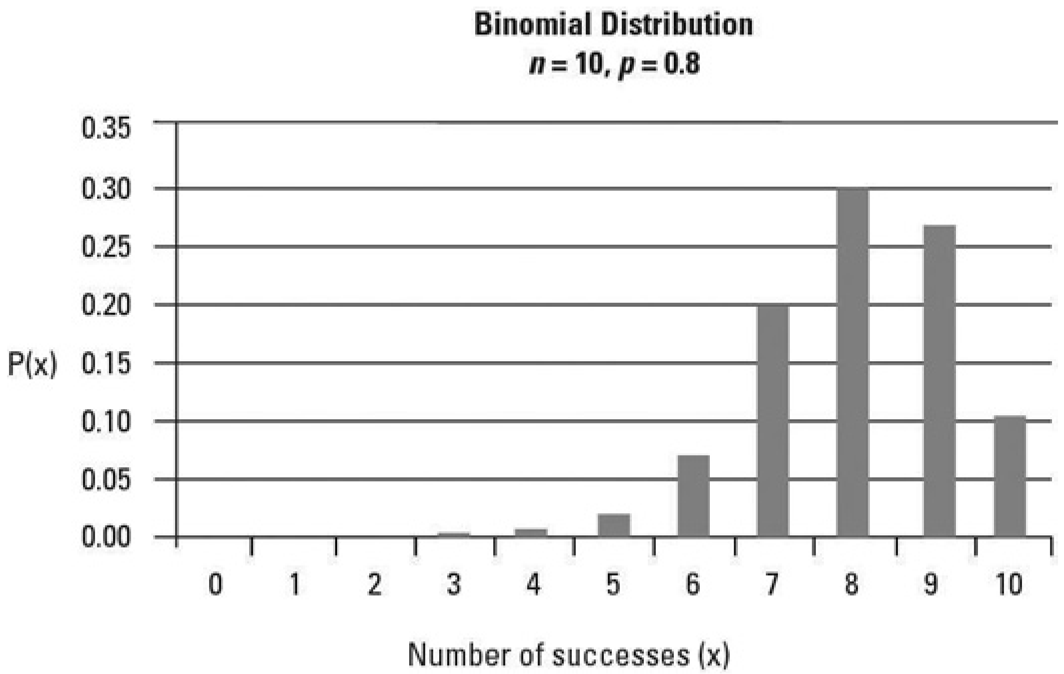
e.

Because for the normal approximation to be reasonable, as a rough guideline, we need both np≥5 (12*0.25=3, which is not greater than or equal to ten) and for n(1-p) ≥10 (12*(1-0.25)=9, which is also not greater than or equal to ten). Because neither of these two required conditions are met, the approximation will not be giving a better approximation.

The mean needs to be near half the number of trials in order for it to approximate a normal distribution, because if it is near the edges of a graph, because you can not get above the upper limit, or below the lower limit; therefore the bars near the edge will get abnormally high, therefore it will no longer be symmetrical, and no longer able to depict a normal distribution.







Source: https://www.dummies.com/education/math/business-statistics/how-to-graph-the-binomial-distribution/#:~:text=One%20way%20to%20illustrate%20the,probability%20of%20each%20value%20occurring

3.

Error = 0.01

p (or $\hat{p}$) = 0.5

Confidence = 0.99

Z=2.576

$n \geq (\frac{2.576}{0.01})^2 * 0.5 * (1-0.5)$

n≥16590 (16589.44 rounded up to 16590)

16590 is the minimum sample size required.

4a.

Final answer needs to be 3.d.p

$\hat{p}$ = 62/110 = 0.56363
Z (With 95% confidence interval) = 1.960

Standard

$0.56363 \pm 1.960 * \sqrt{\frac{0.56363*(1-0.5636)}{110}}$

$\sqrt{\frac{0.56363*(1-0.5636)}{110}} = 0.04729$

0.56363 ± (0.04729 * 1.960)

0.56363 ± 0.09269

The Standard 95% confidence interval will be ranging between 0.471 to 0.656 (0.471,0.656).

$\hat{p}*$ = 0.56140
Z (With 95% confidence interval) = 1.960

Agresti-Coull

$0.56140 \pm 1.960 * \sqrt{\frac{0.56140*(1-0.5140)}{114}}$

$\sqrt{\frac{0.56140*(1-0.56140)}{114}} = 0.04647$

0.56140 ± (0.04647 * 1.960)

The Agresti-Coull 95% confidence interval will be ranging between 0.470 to 0.652 (0.470,0.652).

b.

$\hat{p}_1$ = 0.56363
$\hat{p}_2$ = 0.34286
$\hat{p}$ = 0.47778
$S^2\hat{p}1\text{-}\hat{p}2$ = 0.47778 * (1-0.47778) * (1/110 + 1/70) = 0.00583

$z* = \frac{0.22077}{\sqrt{0.24950*(0.02338)}} = 2.89056$

p-value = 2*P(Z > |z*|)
2*P(Z > 2.89056)
2*(0.50 – P(0 ≤ Z ≤ 2.9))
2*(0.5-0.4981) = 0.0038

p-value = 0.0038

For the significance level of a = 0.05, we get p-value < a, we reject our hypothesis which is p1 = p2. Therefore, they must be different.

5.

a.

1.457 ((0*20+1*34+2*21+3*12+4*4+5*0+6*1)/92)

b.

The Poisson distribution requires that all events that happen are independent of another's. People will group together. Cons

There is no reason to assume that the probability of a no-show changes. Pros

Mean (1.45652) ~= Variance (1.42202) - Pros

Hypothesis H_0: The population is consistent with a Poisson distribution.

H_1: The population is not consistent with a Poisson distribution.

We assume that the 92 days are completely independent of other days chances.

We assume that every person is individually independent of others.

We assume that 92 days is a big enough sample size.

Number of no shows	Freq	Poisson prob		Expected Frequency.
0	20	0.23304586	x 92	21.44022188
1	34	0.33943596	x 92	31.22810832
2	21	0.24719763	x 92	22.74218196
3	12	0.1200161	x 92	11.0414812
4	4	0.04370146	x 92	4.02053432
5	0	0.01273041	x 92	1.17119772
6	1	0.0030903	x 92	0.2843076

Grouping those expected frequencies into groups of 5 or more…

0	20	0.23304586	x 92	21.44022188
1	34	0.33943596	x 92	31.22810832
2	21	0.24719763	x 92	22.74218196
3	12	0.1200161	x 92	11.0414812
4 and up	5	0.05952217	x 92	5.47603964

$$\frac{(20-21.44022188)*(20-21.44022188)}{21.44022188}+\frac{(34-31.22810832)*(34-31.22810832)}{31.22810832}+\frac{(21-22.74218196)*(21-22.74218196)}{22.74218196}$$

$+\frac{(12-11.0414812)*(12-11.0414812)}{11.0414812}$ approx. = 0.559456710561 (q=0.559456710561)

$X^2_{5-1-1} \rightarrow X^2_3$ (df = 3)

```
> pchisq(q = 0.559456710561, df = 3, lower.tail = FALSE)
[1] 0.9056478
```

0.9056478 is much bigger than the significance level (a= 0.05), there we do not reject H_0, and conclude that the population distribution is consistent with the Poisson distribution.