Question 1

Allotted time: 25 minutes (plus 5 minutes to submit)

An online company is hoping to sell a new product by placing a pop-up advertisement on their website. There are two different designs for the advertisement, and the company would like to determine which one is more effective, as measured by clicks and purchases. For the first 200 visitors to the updated website, half were randomly assigned to receive advertisement 1 and half to receive advertisement 2. The two-way table summarizes the results:

Customer Behavior

	Did not click	Clicked, no purchase	Clicked, made purchase	Total
Advertisement 1	70	8	22	100
Advertisement 2	64	20	16	100
Total	134	28	38	200

(a) Is this study an observational study or an experiment? Explain your answer.

Experiment. The customers were randomly assigned to a treatment of advertisement 1 or advertisement 2.

(b) i. Provide one piece of evidence for why advertisement 1 is more effective.

Advertisement 1 had a higher proportion of customers who made a purchase ($22/100 = 0.22$) than advertisement 2 ($\frac{16}{100} = 0.16$).

ii. Provide one piece of evidence for why advertisement 2 is more effective.

Advertisement 2 had a higher proportion of customers who clicked ($\frac{20+16}{100} = 0.36$) than advertisement 1 ($\frac{8+22}{100} = 0.30$).

(c) One customer from the study will be selected at random. Are the events "advertisement 1" and "made purchase" independent? Justify your answer based on probabilities calculated from the table above.

$P(\text{purchase}) = \frac{38}{200} = 0.19$

$P(\text{purchase} \mid \text{Ad } 1) = \frac{22}{100} = 0.22$

$P(\text{purchase} \mid \text{Ad } 2) = \frac{16}{100} = 0.16$

No, the two events are not independent because $0.19 \neq 0.22 \neq 0.16$. Knowing which treatment the customer received changes the probability they purchased.

$\hat{p}_1 = \frac{22}{100} = 0.22$

$\hat{p}_2 = \frac{16}{100} = 0.16$

The online company conducted a test of the hypotheses

$$H_0 : p_1 - p_2 = 0$$
$$H_a : p_1 - p_2 \neq 0,$$

$\hat{p}_1 - \hat{p}_2 = 0.22 - 0.16 = 0.06$

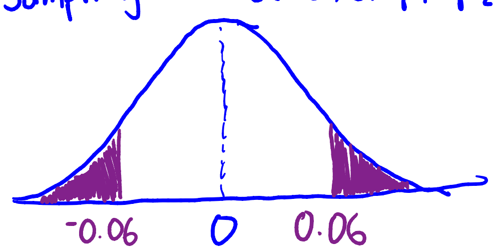
where p_1 is the proportion of customers similar to those in the study given advertisement 1 that would click and make a purchase and p_2 is the proportion of customers similar to those in the study given advertisement 2 that would click and make a purchase. The conditions for inference have been met.

(d) One of the conditions for inference that was met is that $n\hat{p}_c \geq 10$ and $n(1-\hat{p}_c) \geq 10$ for each group. Explain why it is necessary to satisfy this condition.

So we can assume the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal and we can use a z-test statistic to calculate a p-value.

(e) The test resulted in a p-value of 0.2795. Interpret the p-value.

Sampling Distribution of $\hat{p}_1 - \hat{p}_2$



Assuming the proportion that make purchases for each treatment group is the same ($p_1 - p_2 = 0$), there is a 0.2795 probability of getting a difference of sample proportions of 0.06 or greater OR -0.06 or lower, purely by chance.

(f) Based on the p-value, what conclusion should the online company make?

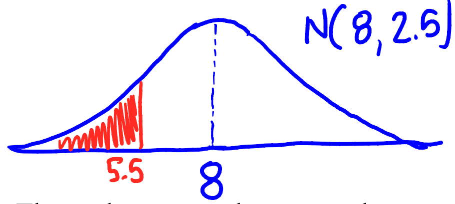
Because the p-value of 0.2795 is greater than $\alpha = 0.05$, the company should fail to reject H_0. They do not have convincing evidence that there is a difference in the proportion of customers like these who would make purchases between the two advertisements.

Question 2

Allotted time: 15 minutes (plus 5 minutes to submit)

A park ranger at a large national park wants to estimate the mean diameter of all the aspen trees in the park. The park ranger believes that due to environmental changes, the aspen trees are not growing as large as they were in 1975.

(a) Data collected in 1975 indicate that the distribution of diameter for aspen trees in this park was approximately normal with a mean of 8 inches and a standard deviation of 2.5 inches. Find the approximate probability that a randomly selected aspen tree in this park in 1975 would have a diameter less than 5.5 inches.

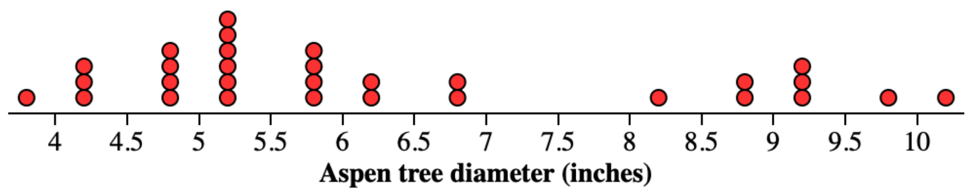


$N(8, 2.5)$ $z = \frac{5.5-8}{2.5} = -1 \xrightarrow{\text{Table A}} \boxed{0.1587}$

OR

Using the empirical rule $100 - 68 = 32$ $\frac{32}{2} = \boxed{16\%}$

The park ranger selects a random sample of 30 aspen trees from the park in 2020 and measures their diameters. A dotplot of the diameters is shown below.



(b) Describe the distribution of aspen tree diameters from the sample.

The shape of the distribution of aspen tree diameters is skewed right with a gap between 7 & 8 inches. The center of the distribution is between 5.5 & 6 inches with no outliers. The range of the distribution is around $10.2 - 3.8 = 6.4$ inches.

(c) Assume the conditions for inference have been met. The park ranger uses the sample data to construct a 95% confidence interval for the mean diameter of all aspens in the park in 2020 as 5.6 to 7.0 inches. Calculate the point estimate and the margin of error.

POINT ESTIMATE $= \frac{5.6+7.0}{2} = \frac{12.6}{2} = \boxed{6.3 \text{ inches}}$ MARGIN OF ERROR $= \frac{7.0-5.6}{2} = \frac{1.4}{2} = \boxed{0.7 \text{ inches}}$

(d) Based on the confidence interval, does the park ranger have convincing evidence that the mean diameter for all aspen trees in the park is different than 8 inches, as it was in 1975?

Yes. All of the plausible values for the mean diameter in 2020 are less than 8 inches (8 is not contained in the interval) so the park ranger has convincing evidence (at $\alpha = 0.05$) that the 2020 mean diameter does not equal 8.

(e) Aspen trees tend to be smaller in the highlands of the park because they are subject to strong winds. While the number of aspen trees in the highlands of the park is about the same as the number of aspen trees in the lowlands of the park, the park ranger is concerned that 24 out of the 30 trees in the random sample came from the highlands. Propose a sampling method to address this concern and describe the benefit of using that sampling method.

Use a stratified random sample by taking an SRS of 15 trees from the highlands and a separate SRS of 15 trees from the lowlands. This will reduce the variability of the estimates of the mean (increasing precison) and will make it easier to determine if the mean diameter has changed since 1975.