## Homework 2: Can dogs sniff out cancer?

Inference and Confidence Intervals for a Single Categorical Variable

Can dogs sniff out cancer? In 2011, an article published by the medical journal Gut - An International Journal of Gastroenterology and Hepatology (Sonoda et al.) reported the results of a study conducted in Japan in which a dog was tested to see whether she could detect colorectal cancer. The dog used was an eight-year-old black Labrador named Marine. (As her name might suggest, she was originally trained for water rescues.) The study was designed so that the dog first smelled a bag that had been breathed into by a patient with colorectal cancer. This was the standard that the dog would use to judge the other bags. Marine then smelled the breath in five different bags from five different patients, only one of which contained breath from a colorectal cancer patient (not the same as the original patient); the others contained breath from noncancer patients. The dog was then trained to sit next to the bag which she thought contained breath from a cancer patient (i.e., had the cancer scent). If she sat down next to the correct bag, she was rewarded with a tennis ball.

Marine completed 33 attempts of this experimental procedure, with a different set of five patients each time: four noncancer patients and one cancer patient. And, each time, the researchers recorded whether or not she correctly identified the bag with the breath of the cancer patient.

In the 33 attempts, Marine chose the correct bag 30 times.

**Research Question** Is there statistical evidence that Marine is xxxx?

## Setup

- 1. Identify the following in context of the scenario:
- Population of interest:
- Sample of interest:

• Variable of interest (and data type):
• Parameter of interest (assign a symbol):
• Statistic of interest (assign a symbol and value):
Visualize and Explore
2.
Carry out the formal hypothesis test to address the research question.
3. Write (first in words, and then using symbols) the null hypothesis and the alternative hypothesis.
• Null hypothesis
• Alternative hypothesis
4. Carry out a simulation study with 1000 replicated simulations to investigate the research question (sketch or paste your simulation results here).
Use the simulation results to estimate the p-value:

6.	Based on the p-value, state a conclusion about this study.
7.	Is it <i>possible</i> that in general Marine just randomly picks a bag from the five bags, and the observed number of proportion of times she picked the correct bag was by chance (i.e. luck) alone?
	Choose one: Yes No
8.	Suppose that instead of 30, Marine had picked the correct bag 10 out of 33 times. What would the conclusion be then?
	rval Estimate of Marine's Proportion of Success  What is the <i>point estimate</i> (or "best guess") for the proportion of times Marine accurately identifies the bag with the breath of the cancer patient? (assign a symbol)
10.	Calculate the $standard\ error$ (using Normal Approximation Method) associated with this point estimate.
11.	Calculate the $margin\ of\ error$ for a 95% confidence interval.

5. Interpret the p-value.

12.	Find the $endpoints$ of the 95% confidence interval.
Sketo	ch the following values for our example on the normal curve shown below:
•	the point estimate the lower and upper endpoints the margin of error.
13.	Where does the proportion of times Marine accurately identifies the bag with the breath of the cancer patient $under$ the assumption that Marine is just "randomly picking" a bag (i.e., null value) fall within the 95% confidence interval?
	Sketch this value on the plot in Question 12.
	Why does this make sense?
14.	Suppose instead of Marine completes 99 attempts and chooses the correct bag 90 times.
•	What parts of your 95% confidence interval calculation changed?
	Select all that apply:
	Point Estimate Standard Error Multiplier for the Margin of Error
•	What happens to your $95\%$ confidence interval compared to the original confidence interval based on $33$ trials? The confidence interval:
	Choose one: Narrows Stays the same Widens
	Why does this make sense?

15.	Suppose instead of calculating a $95\%$ confidence interval, scientists want to report a $99\%$ confidence interval for the original $33$ trials.
•	What parts of your confidence interval calculation changed?
	Select all that apply:
	Point Estimate Standard Error Multiplier for the Margin of Error
•	What happens to your confidence interval compared to the original $95\%$ confidence interval based on $33$ trials? The confidence interval:
	Choose one: Narrows Stays the same Widens
	Why does this make sense?
16.	Check the conditions necessary to use Normal Approximation of binomial probabilities
	to calculate a confidence interval (i.e., Wald interval) for the proportion of times Marine accurately identifies the bag with the breath of the cancer patient.
	Does the Wald interval seem to be appropriate here?
4	Canvas Quiz
Ma	ake sure to complete the Homework Quiz on Canvas.