A DUCK'S STORY
INTRODUCING THE IDEA OF TESTING (STATISTICAL) HYPOTHESES

This is a short story that will introduce you to the ideas and vocabulary of hypothesis testing.

Please read the story and questions carefully and fill in the blanks.

A. INTERACTIVE LECTURE PART
I The research question:

ARE FEMALE MALLARDS ATTRACTED TO THE COLOR GREEN?

A student is taking a biology class that studies animal behavior and is assigned the following research:

In a certain species (mallards), male ducks have green heads and females are a plain color. Probably the purpose of the green coloring of the male heads is to attract the females. The question is: are female ducks also attracted to the green color in food, for example in bread?

II Writing statistical hypotheses

We basically want to know if female ducks are indifferent to green bread versus plain bread or if they prefer green bread. The research question can be translated into the confrontation of two opposite ideas:

Idea 1:   Female ducks are indifferent to plain versus green bread.
Idea 2:   Female ducks prefer green bread.

When a female duck of the above-mentioned species is confronted with two pieces of bread, one plain and one green, the probability of picking the green one will be called **p**. Write the two previous ideas in terms of p.

Idea 1: $p =$
Idea 2: $p >$

We call these confronting ideas **'statistical hypotheses'**. The first one states that the ducks equally like the green and the plain bread. This statement is called the **'null hypothesis'** because it represents an idea of no difference and is labeled by the symbol 'H_0'. The second idea says that the ducks prefer the green bread and states something different than the first one, so it is called the **'alternative hypothesis'**. The symbol used for the alternative hypothesis is 'H_a'.

We must decide which of the two hypotheses is more likely to be true. The decision between the two hypotheses is usually expressed in terms of H_0 (idea # 1). If we favor H_a (idea # 2), we usually say that **'we reject H_0'**.

III Gathering evidence to make the decision.

The student designs a study in order to be able to make a decision about the two statistical hypotheses. She will go to a lake near campus where mallards are quite abundant and will randomly select 10 female ducks. Each duck will be offered two pieces of bread: one plain and one dyed green. The student will write down which piece of bread each duck approaches first. Then she will summarize her information reporting how many ducks approach the green bread first.

		x	p(x)
Think about the variable x = # of ducks in the sample that prefer the green bread. Think of 'picking green first' as 'success'. Note that the sample size is n=10. If the ducks are truly indifferent to plain versus green bread, what is the distribution of the variable x?		0	0.000977
		1	0.009766
		2	0.043945
		3	0.117188
		4	0.205078
Name of the distribution:		5	0.246094
_____		6	0.205078
Parameters: n= p=		7	0.117188
		8	0.043945
The values of P(x) appear to the right.		9	0.009766
		10	0.000977

IV Arriving at a conclusion.

If female ducks were truly indifferent between green and plain bread, about how many ducks, of the ten that were observed, would you have expected to choose the green bread first? _____. Of course, even if the null hypothesis was true we are not always going to get that result in reality due to sampling variability or just chance. Suppose the biology student finds that 9 of the 10 female ducks sampled prefer the green bread. So, p=0.5 and $\hat{p} = 0.9$

If female ducks are really indifferent to plain versus green bread, what is the probability that 9 female ducks in a sample of 10 would pick the green bread first just by chance? _____.

Nine out of 10 seems to indicate that female ducks tend to prefer green bread to plain. If more than 9 had picked the green bread first, it would be a situation even farther from what was expected under the null hypothesis. A number higher than 9 would have given us even a clearer idea that female ducks tend to prefer the green color. That is why we are interested in knowing what is the probability that 9 (the value the student observed) or more female ducks pick the green bread first. We want to know not only what the chances are of getting the result that we got, but also what the chances are of getting a result that is farther from what the null hypothesis indicates, provided the null hypothesis is true. What is the probability that, assuming that in general female ducks are really indifferent between green and plain bread, <u>9 or more </u>female ducks in a sample of 10 would pick the green bread first just by chance? _____.

To summarize our results we would say that the probability of getting a result as the one we got (9 ducks picking the green bread first) or a more extreme one when the null hypothesis (p=0.5 , meaning ducks are indifferent between green and plain) is true is **0.0107430** *(This probability of getting the result we got or a more extreme one is called 'p-value'.)*

So, becoming aware that the probability of getting the result we got when the null hypothesis Ho is true is very small, would you feel like believing Ho is true?   YES   NO

So, which hypothesis, H_0 or H_a, do you favor? _____
So, which of these conclusions seem more reasonable? (Circle one)

REJECT Ho   DO NOT REJECT Ho
Now write your answer to the research question posed at the beginning of this worksheet:

The question is: are female ducks also attracted to the green color in food, for example in bread?

 YES   NO

Note: How do we decide if the p-value is small or large?
At the beginning of the study, before the data are collected, we fix the desired value of α ('significance level') the most common value is α =0.05. We will explain later what α means. The value
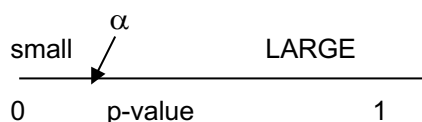
V Reviewing the thinking process.

Read sections I-V again and notice that the way we thought in order to arrive at a conclusion can be summarized in the following steps:
   a) Identifying the research question (Do female ducks prefer green to plain bread?)
   b) Identifying a quantity related to the research question whose value we don't know. In this case the quantity of interest is the probability of a hypothetical female duck picking the green bread (or the proportion of all female ducks that would pick the green bread). In general, that quantity is called a '**parameter**'.
   c) Writing the statistical hypotheses in terms of that parameter of interest. In the example the statistical hypotheses are H_0: p=0.5 and H_a: p>0.5.
   d) Collecting data and calculating a statistic (A study was conducted and it was observed that 9 out of 10 ducks preferred the green bread)
   e) Finding the p-value (probability that the result we got or a more extreme one happens just by chance given that the null hypothesis is true).
   f) Deciding if the p-value is small or large. In the example of ducks, we felt like rejecting the null hypothesis because the p-value was small.

This thinking procedure is called '**hypothesis testing**' and can be applied to many situations in which a research question is asked, and data are collected (through a survey or experiment) in order to answer the research question.   Here we have done a test of hypothesis for a population proportion using a small sample. Common examples of test of hypotheses in introductory statistics courses are test of hypotheses about proportions with large samples, test about the mean of a population, matched pairs tests, tests for the means of two populations. The main difference among those cases will be the probability distribution (or 'sampling distribution' because is the distribution of a sample statistic) that we use to find the p-values but the steps a) to f) are similar.

VI In how many different ways can we make a wrong decision?

In hypothesis testing we need to pick either H_0 or H_a. Obviously, we would like to make the correct decision, but we can sometimes make the wrong decision. How would you describe in words (in terms of what the ducks prefer and what we say they prefer) each one of these situations?

   1) We select H_a but it is the wrong decision because H_0 is true.

_____

   2) We select H_0 but it is the wrong decision because H_0 is not true.

_____

We call these situations: **type I error** and **type II error**. Of course, we would like to keep the chances of making a mistake very small. We already mentioned that we usually express our decision in terms of H_0 (reject or not reject H_0). In the same way we usually focus on the probability of making type I error (rejecting the null hypothesis when it is true). This is because the null hypothesis reflects a 'status quo' or neutrality situation, and if we reject it, we are making a statement saying that something is better or preferred, or worse, or different, depending on the situation. When two medicines are being compared in a pharmaceutical study a ' type I error' would mean to ascertain that one medicine is better when they have actually similar effectiveness. Type I error is usually considered a serious error and we like to have some control over it.

VII How 'small' is small (in the 'p-value' world)?

When we made the decision about the null hypotheses, we had a figure to help us decide if the 'p-value' was small or large, but we did not mention how the value of α had been decided at the beginning of the study.
The probability of making type I error is called α (or **'significance level'.**) We set the value we want for α at the beginning of a study. A very common value is 0.05 but in studies (such as medical research) where the consequences of type I error are very serious, we like to have a smaller α such as 0.01. If you doubt whether the p-value is small or not, you can compare it to α in order to decide if it is small or large. If the p-value is too close to α you may say the test is inconclusive and ask for more evidence (data).

α

small   LARGE

0   p-value   1
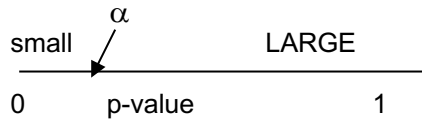
It is very important to fix the desired value of α at the beginning of the study, before the data are collected and the results are observed. To do it later could lead to accommodating the situation to get a result (reject or not reject H_0) that we want instead of trying to find out the truth.

Now think about this: if you exaggerate the caution to avoid making a type I error, which involves 'rejecting H_0', you will try never to reject H_0 and the probability of doing type II error (accepting Ho when it is false) would grow uncontrollably. The probability of 'type II error' is called β. To keep β at a reasonable level we should not exaggerate, making α extremely small, unless we have enough data (large n) as to make a very sound decision.

B. ACTIVITY PART- Using your knowledge to apply to other examples.

You are now familiar with the general ideas and vocabulary of hypotheses testing. You also know how to translate a research question into statistical hypotheses about a population proportion and how to test the hypotheses using small samples. Now you can apply your knowledge in other to answer other research questions.

Come up with an interesting research question about a probability or a population proportion. Some simple examples are:

- *Is the probability of getting a six in a slanted die still 1/6? Or is it higher?*
- *Does telepathy (like one person thinking in one of 10 digits and the other person guessing it) work?*
- *Is the dominant hand faster than the non-dominant hand (using a reaction time ruler)?*
- *When 2 people enter a restaurant (could be changed for other type of location like a bank), one female and one male, is the man the one that most frequently opens the door?*
- *Do more people wear snickers than shoes when going to the grocery store?*

I am sure you can produce more interesting and original research questions. Data can be obtained either by experimentation or through a survey.

Write the research question:

Write the null and alternative hypotheses:

Ho:

Ha:

Decide the value of α you will work with.   _____

What type of data will you collect? What is considered a 'success' in this case?

Collect the data.

Discuss if in the case you are working it makes sense to put all the data together or not (i.e., if you are all talking about the same population). Put the data collected by the whole group together (if applicable). Pay attention to the fact that probably the observations collected by each one of the students differ even if you all worked with the same

population. After the data are collected, you will go through the process of arriving at a decision and answering the research question.

Look for a probability table that will help you finding the p-value.

How many 'successes' you found in the n trials? _____
What is the value of $\hat{p}$?_____

Calculate the p-value.

What is your conclusion about the null hypothesis?

Answer the research question.

Now write a paragraph in plain English, telling a friend the small research your class conducted and their conclusions.