Math 122
Introduction to Statistics
Hypothesis Testing - Binomial Distribution

HYPOTHESIS TESTING IDEA

Informal Idea: To test a claim using statistics, we perform an experiment or make an observation to collect data.

We then make a *working assumption* (which we will call the *null hypothesis* below) based on the claim and use this assumption to find the probability of getting results at least as extreme as our observations. This probability – called our P-value – becomes a measure of the compatibility of the working assumption and our observations.

If the P-value is small, then our observations are not compatible with the working assumption. In this case, the working assumption is probably false, so we reject it (the working assumption) in favor of an *alternative* (called the *alternative hypothesis* below).

If the P-value is not small, then the observations are compatible with our working assumption. In this case we do not reject the working assumption.

The cut-off for when P is considered "small" is called our *significance level* and is denoted α. The significance level is often something like 5%. If the certainty of results are essential (as in medical testing) then a smaller significance level may be used. If the certainty is less essential, a larger significance level may be used.

Parameters: When we test a claim given in words, we must translate that claim into symbols. The symbols involved are usually parameters associated with the population being studied. For our tests in this section, the primary parameter that we will be testing is a proportion p. Since we will be using a binomial distribution to calculate probabilities, we will also have a sample size n (or number of trials) and a number of successes x. The variable p is the only real parameter here since it relates to a population.

The Claim: Once symbols have been selected to perform our calculations with, we can translate the claim being tested into a symbolic form. For proportions, the symbolic form of the claim will be one of these:

$$\begin{array}{ll} p = a & p \neq a \\ p \leq a & p > a \\ p \geq a & p < a \end{array}$$

These possible claims come in pairs of opposites:

The opposite of this	is this
$p = a$	$p \neq a$
$p \leq a$	$p > a$
$p \geq a$	$p < a$

In each hypothesis test, we will begin by writing down in symbols the claim we are testing and its opposite. One of these two will include an equal sign. The other will not. The one with the equal sign we will call the Null Hypothesis. The one without the equal sign is the

Alternative Hypothesis.

The Working Hypothesis or Null Hypothesis: To calculate a probability (in this case using a binomial distribution) we have to have a value for the parameter p. Therefore, we must assume that p is equal to something. This assumption is called the null hypothesis and is denoted H_0. The null hypothesis in a test will always be either the claim or its opposite – whichever includes an equal sign.

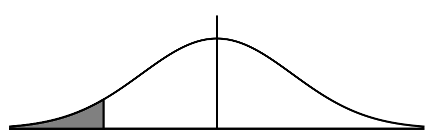
The Alternative Hypothesis: At the end of the hypothesis test, we will (with some varying degree of commitment) side either with the null hypothesis or something else. The "something else" we will call the alternative hypothesis. The alternative hypothesis is denoted H_1. The alternative hypothesis in a hypothesis test will always be either the claim or its opposite – whichever does not include an equal sign.

Distribution: Once the parameters have been defined, we can determine what distribution and what random variable we will use to calculate probabilities. The random variable we use in these tests is x, the number of successes in our observations. The value of x in our observations will be called our test statistic. For now, we will use O to represent our observered number of successes. We will assume that x follows a binomial distribution.

Significance Level: If no significance level is given, use $\alpha = 0.05$.

Tailedness and P-values: Based on the null hypothesis (working assumption) we are to find the probability of seeing results *at least as extreme* as our observations. What "at least as extreme" means depends on the alternative hypothesis H_1.

If H_0 is $p \geq a$ and H_1 is $p < a$ and O is our observed number of successes, then we will reject H_0 in favor of H_1 if the observed number of successes O is too small. In this case, results "at least as extreme" as our observations are numbers x of successes which are less than or equal to our observed number O of successes. Here is a histogram for the number x of successes:

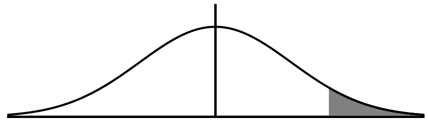


The shaded region corresponds to values of x which are less than or equal to O. In this case,

$$P = P(x \leq O) = binomcdf(n, p, O).$$

Because of the picture, this is called a *left-tailed test*.

If H_0 is $p \leq a$ and H_1 is $p > a$ and O is our observed number of successes, then we will reject H_0 in favor of H_1 if the observed number of successes O is too big. In this case, results "at least as extreme" as our observations are numbers x of successes which are greater than or equal to our observed number O of successes. Here is a histogram for the number x of successes:

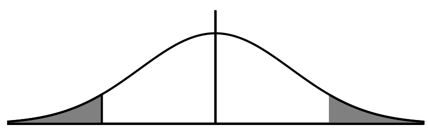


The shaded region corresponds to values of x which are greater than or equal to O. In this case,

$$P = P(x \geq O) = 1 - \textit{binomcdf}(n, p, O-1).$$

Because of the picture, this is called a *right-tailed test*.

If H_0 is $p = a$ and H_1 is $p \neq a$ and O is our observed number of successes, then we will reject H_0 in favor of H_1 if the observed number of successes O is either too small or two big. In this case, results "at least as extreme" as our observations are numbers x of successes which are at least as far from the expected value of x as O is. Recall that the expected value for x is $\mu = np$. Here is a histogram for the number x of successes:



The shaded region corresponds to values of x which are at least as far from np as O. Let $d = |np - O|$. Then $P = P(x \leq np - d) + P(x \geq np + d)$. Because of the picture, this is called a *two-tailed test*.

Formal Conclusion: Remember that P is a measure of consistency between our observations and H_0. If P is small (that is, less than or equal to α), then our observations are not consistent with H_0. In this case, we reject H_0 (and support H_1). If P is big (that is, greater than α) then our observations are consistent with H_0 so we fail to reject H_0 (and do not support H_1).

Final Conclusion: Our final conclusions should be phrased in non-technical terms which refer to the original claim. In some tests, our claim is H_0. In these cases, if we reject H_0, then we reject the claim. If we fail to reject H_0, then we do not reject the claim (but we do not support it either). In other tests, our claim is H_1. In these cases, if we reject H_0, then we support H_1 and our claim. If we fail to reject H_0, then we are not supporting H_1. Here, we do not support the claim (but we do not reject it either).

Summary:

	Claim	H_0	H_1	$P \leq \alpha$ (Reject H_0)	$P > \alpha$ (Do Not Reject H_0)
	$p \leq a$	$p \leq a$	$p > a$	Reject Claim	Do not reject claim
Claim$\approx$ H_0	$p \geq a$	$p \geq a$	$p < a$	Reject Claim	Do not reject claim
	$p = a$	$p = a$	$p \neq a$	Reject Claim	Do not reject claim
	$p > a$	$p \leq a$	$p > a$	Support Claim	Do not support claim
Claim$\approx$ H_1	$p < a$	$p \geq a$	$p < a$	Support Claim	Do not support claim
	$p \neq a$	$p = a$	$p \neq a$	Support Claim	Do not support claim

Notice that claims in the top half of this table – those that contain equality – can only be rejected or not rejected by a hypothesis test. These types of claims cannot be supported by a hypothesis test. Claims in the bottom of the table – those that do not contain equality – can only be supported or not supported by a hypothesis test. These claims cannot be rejected.

Outline: These are the steps we will follow in a hypothesis test:

1. **Parameters:** Define the relevant parameters.

2. **Symbolic Claim:** State in symbols the claim being tested and its opposite.

3. **Null Hypothesis:** Determine the null hypothesis H_0. This will include equality.

4. **Alternative Hypothesis:** Determine the alternative hypothesis H_1. This will not include equality.

5. **Distribution and Test Statistic:** Determine the appropriate random variable, its distribution, and the relevant test statistic.

6. **Significance Level:** Establish a significance level. This is most usually $\alpha = 0.05$.

7. P**-value:** Determine P-value. We will usually use technology for this.

8. **Formal Conclusion:** Decide whether to reject H_0 or not. If $P \leq \alpha$ reject H_0. Otherwise, fail to reject H_0.

9. **Final Conclusion:** Phrase the formal conclusion as a statement about the original claim.

Example: Bob believes that a coin he has lands on T more often than it lands on H. He flips the coin 103 times and sees an H 49 times. Use this information to test the claim that the probability that this coin lands on H is less than one half.

1. **Parameters:** Let p be the probability that this coin lands on H. Let $n = 103$, and let x be the number of H's that appear when this coin is flipped 103 times.

2. **Symbolic Claim:** The claim is $p < 0.5$. The opposite of the claim is $p \geq 0.5$.

3. **Null Hypothesis:** H_0 is $p \geq 0.5$.

4. **Alternative Hypothesis:** H_1 is the same as the claim $P < 0.5$.

5. **Distribution and Test Statistic:** Each flip of the coin is an independent trial that ends in either H or T. We are assuming that the probability of an H on each trial is $p = 0.5$. The variable x is the number of H's that appear in 103 flips. Thus, x has a binomial distribution. Our observed value of x is $O = 49$.

6. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.

7. P**-value:** Since H_1 involves $<$, this is a left-tailed test. Then

$$P = P(x \leq 49) = \mathit{binomcdf}(103, 0.5, 49) = 0.3468.$$

8. **Formal Conclusion:** Since $P > \alpha$, we fail to reject H_0.

9. **Final Conclusion:** Since our claim is the same as H_1, we first rephrase our formal conclusion to refer to H_1. Since we are not rejecting H_0, we are not supporting H_1. This means that we are not supporting our claim. Our conclusion is:

   > There is not enough sample evidence to support the claim that the probability that this coin lands on H is less than one half.

10. **Interpretation:** The observation of 49 out of 103 flips ending in H is not extreme enough to conclude that the coin is not fair. The coin is probably fair.

Example: In a sample of 100 of a certain type of parrot, 59 were found to have red wings while 41 had green wings. Use this information to test the claim that most of these birds have red wings.

1. **Parameters:** Let p be the proportion of this type of bird with red wings. Let $n = 100$, and let x be the number of red-winged birds out of 100 of this kind of bird.

2. **Symbolic Claim:** The claim is $p > 0.5$. The opposite of the claim is $p \leq 0.5$.

3. **Null Hypothesis:** H_0 is $p \leq 0.5$.

4. **Alternative Hypothesis:** H_1 is the same as the claim $P > 0.5$.

5. **Distribution and Test Statistic:** The variable x has a binomial distribution with $n = 100$ and $p = 0.5$. Our observed value of x is $O = 59$.

6. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.

7. P**-value:** Since H_1 involves $>$, this is a right-tailed test. Then

$$P = P(x \geq 59) = 1 - P(x < 58) = 1 - \mathit{binomcdf}(100, 0.5, 58) = 0.0443.$$

8. **Formal Conclusion:** Since $P < \alpha$, we reject H_0 (just barely).

9. **Final Conclusion:** Since our claim is the same as H_1, we first rephrase our formal conclusion to refer to H_1. Since we are rejecting H_0, we are supporting H_1. This means that we are supporting our claim. Our conclusion is:

   > The sample evidence supports the claim that most of these birds have red wings.

10. **Interpretation:** It appears as if most of these birds have red wings. However, note that the P value is quite close to our significance level. A lower significance level would have resulted in not supporting the claim.

Example: In a laboratory experiment, of 580 pea pods with green/yellow genes, 428 were green. Use this information to test Gregor Mendel's claim that the proportion of pods of this type which are green is 0.75.

1. **Parameters:** Let p be the proportion of pea pods with green/yellow genes which are green. Let $n = 580$, and let x be the number of green pods out of 580.

2. **Symbolic Claim:** The claim is $p = 0.75$. The opposite of the claim is $p \neq 0.75$.

3. **Null Hypothesis:** H_0 is $p = 0.75$. Note that this is the same as the claim being tested.

4. **Alternative Hypothesis:** H_1 is $p \neq 0.75$.

5. **Distribution and Test Statistic:** We treat the 580 pods as 580 independent trials that end in either "green" or "not green." We are assuming that the probability of green in each trial is $p = 0.75$. If x is the number of green pods out of $n = 580$, then x has a binomial distribution. In the lab experiment, we encountered a value of $O = 428$.

6. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.

7. **P-value:** Since H_1 is $p \neq 0.75$, this is a two-tailed test. We will find the probability in one tail and then double it. To see if we need to look at the left or right tail, we need to compare our test statistic $O = 428$ to the expected value of x. The expected value is: $\mu = np = 580 \times 0.75 = 435$. Let $d = 435 - 428 = 7$. Then

$$P = P(x \leq 435 - 7) + P(x \geq 435 + 7) = 0.5331.$$

8. **Formal Conclusion:** Since $P > \alpha$, we fail to reject H_0.

9. **Final Conclusion:** Since H_0 is the same as our claim, and since we are not rejecting H_0, we are not rejecting the claim. Our conclusion is:

   > There is not enough sample evidence to warrant rejection of the claim that the proportion of pea pods with green/yellow genes which are green is 0.75.

10. **Interpretation:** The laboratory results are consistent with Mendel's claim. Mendel may be correct.