Chapter 20 Testing Hypotheses About Proportions
Jury Trial

Does the reasoning of hypothesis tests seem backward? That could be because we usually prefer to think about getting things right rather than getting them wrong. You have seen this reasoning before in a different context. This is the logic of jury trials.

Let's suppose a defendant has been accused of robbery. In British common law and those systems derived from it (including U.S. law), the null hypothesis is that the defendant is innocent. Instructions to juries are quite explicit about this.

The evidence takes the form of facts that seem to contradict the presumption of innocence. For us, this means collecting data. In the trial, the prosecutor presents evidence. (“If the defendant were innocent, wouldn’t it be remarkable that the police found him at the scene of the crime with a bag full of money in his hand, a mask on his face, and a getaway car parked outside?”)

The next step is to judge the evidence. Evaluating the evidence is the responsibility of the jury in a trial, but it falls on your shoulders in hypothesis testing. The jury considers the evidence in light of the presumption of innocence and judges whether the evidence against the defendant would be plausible if the defendant were in fact innocent.

Like the jury, you ask, “Could these data plausibly have happened by chance if the null hypothesis were true?” If they are very unlikely to have occurred, then the evidence raises a reasonable doubt about the null hypothesis.

Ultimately, you must make a decision. The standard of “beyond a reasonable doubt” is wonderfully ambiguous because it leaves the jury to decide the degree to which the evidence contradicts the hypothesis of innocence. Juries don’t explicitly use probability to help them decide whether to reject that hypothesis. But when you ask the same question of your null hypothesis, you have the advantage of being able to quantify exactly how surprising the evidence would be were the null hypothesis true.

How unlikely is unlikely? Some people set rigid standards, like 1 time out of 20 (0.05) or 1 time out of 100 (0.01). But if you have to make the decision, you must judge for yourself in each situation whether the probability of observing your data is small enough to constitute “reasonable doubt.”
If the evidence is not strong enough to reject the defendant’s presumption of innocence, what verdict does the jury return? They say “not guilty.” Notice that they do not say that the defendant is innocent. All they say is that they have not seen sufficient evidence to convict, to reject innocence. The defendant may, in fact, be innocent, but the jury has no way to be sure.

Said statistically, the jury’s null hypothesis is $H_0$: innocent defendant. If the evidence is too unlikely given this assumption, the jury rejects the null hypothesis and finds the defendant guilty. But—and this is an important distinction—if there is insufficient evidence to convict the defendant, the jury does not decide that $H_0$ is true and declare the defendant innocent. Juries can only fail to reject the null hypothesis and declare the defendant “not guilty.”

In the same way, if the data are not particularly unlikely under the assumption that the null hypothesis is true, then the most we can do is to “fail to reject” our null hypothesis. We never declare the null hypothesis to be true (or “accept” the null), because we simply do not know whether it’s true or not. (After all, more evidence may come along later.)

In the trial, the burden of proof is on the prosecution. In a hypothesis test, the burden of proof is on the unusual claim. The null hypothesis is the ordinary state of affairs, so it’s the alternative to the null hypothesis that we consider unusual and for which we must marshal evidence.
**JUST CHECKING**

1. A research team wants to know if aspirin helps to thin blood. The null hypothesis says that it doesn’t. They test 12 patients, observe the proportion with thinner blood, and get a P-value of 0.32. They proclaim that aspirin doesn’t work. What would you say?

2. An allergy drug has been tested and found to give relief to 75% of the patients in a large clinical trial. Now the scientists want to see if the new, improved version works even better. What would the null hypothesis be?

3. The new drug is tested and the P-value is 0.0001. What would you conclude about the new drug?

1. $H_0$: aspirin doesn’t thin blood  
   $H_A$: aspirin helps to thin blood  
   P-value: 0.32

2. $H_0$: $p = 0.75$  
   $H_A$: $p > 0.75$
Hₐ: p \neq p₀. We use a two-sided alternative because we are interested in deviations on either side of the null hypothesis value. For two-sided alternatives, the P-value is the probability of deviating in either direction from the null hypothesis value. Add the probabilities in both tails.
$H_A: p < p_0$ we use a one-sided alternative that focuses on deviations from the null hypothesis value in only one direction. The P-value considers the probability of the values beyond the test statistic value in the specified direction.
13. Absentees. The National Center for Education Statistics monitors many aspects of elementary and secondary education nationwide. Their 1996 numbers are often used as a baseline to assess changes. In 1996 34% of students had not been absent from school even once during the previous month. In the 2000 survey, responses from 8302 students showed that this figure had slipped to 33%. Officials would, of course, be concerned if student attendance were declining. Do these figures give evidence of a change in student attendance?

\[ H_0: \hat{p} = .34 \quad \alpha = .05 \]
\[ H_A: \hat{p} \neq .34 \]

\[ \text{Steps: } Z = \frac{\hat{p} - \hat{p}_0}{\text{SD} (\hat{p})} = \frac{\hat{p} - \hat{p}_0}{\sqrt{\frac{\hat{p}_0 (1-\hat{p}_0)}{n}}} = -1.923 \]
A.P. Statistics
Hypothesis Testing Worksheet

Step 1:
Identify test used.

Step 2:
State hypotheses - Define parameter.

Step 3:
State alpha value (α) if given.

Step 4:
Check conditions necessary for normal distribution to be appropriate.

Problem:

Name ____________________________
Step 5: 
Calculate test statistic
show/shade curve
show formula

Step 6: 
Find p-value
associated with test statistic

Step 7: 
Based on the p-value of .054, fail to reject the null hypothesis.

Step 8: 
Write a non-math conclusion in context of problem about $H_a$

There is insufficient evidence to suggest that student attendance has changed.
16. Take the offer, part II. In Exercise 19.16 you learned that First USA, a major credit card company, is planning a new offer for their current cardholders. First USA will give double airline miles on purchases for the next 6 months if the cardholder goes online and registers for this offer. To test the effectiveness of this campaign, the company recently sent out offers to a random sample of 50,000 cardholders. Of those, 184 registered. A staff member suspects that the success rate for the full campaign will be comparable to the standard 2% rate that they are used to seeing in similar campaigns. What do you predict?

a. What are the hypotheses?

\[ H_0: \hat{p} = 0.02 \quad g = 0.98 \]

\[ H_A: \hat{p} \neq 0.02 \]

b. Are the assumptions and conditions for inference met?

\[ \hat{p} = \frac{1,184}{50,000} = 0.02368 \]
c. Do you think the rate would change if they use this fundraising campaign? Explain.
11. Dowsing.

a) \( H_0: \) The percentage of successful wells drilled by the dowser is 30%. \( (p = 0.30) \)
\( H_A: \) The percentage of successful wells drilled by the dowser is greater than 30%. \( (p > 0.30) \)

b) Independence assumption: There is no reason to think that finding water in one well will affect the probability that water is found in another, unless the wells are close enough to be fed by the same underground water source.

Randomization condition: This sample is not random, so hopefully the customers you check with are representative of all of the dowser's customers.

10% condition: The 80 customers sampled may be considered less than 10% of all possible customers.

Success/Failure condition: \( np = (80)(0.30) = 24 \) and \( nq = (80)(0.70) = 56 \) are both greater than 10, so the sample is large enough.

c) The sample of customers may not be representative of all customers, so we will proceed cautiously. A Normal model can be used to model the sampling distribution of the proportion, with \( \mu_p = p = 0.30 \) and \( \sigma(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.30)(0.70)}{80}} = 0.0512 \).

We can perform a one-proportion z-test.

The observed proportion of successful wells is \( \hat{p} = \frac{27}{80} = 0.3375 \).

\[
\begin{align*}
z &= \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \\
&= \frac{0.3375 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{80}}} \\
&= \frac{0.3375 - 0.30}{0.0512} \\
&= 0.73
\end{align*}
\]

If his dowsing has the same success rate as standard drilling methods, there is more than a 23% chance of seeing results as good as those of the dowser, or better, by natural sampling variation.

d) If his dowsing has the same success rate as standard drilling methods, there is more than a 23% chance of seeing results as good as those of the dowser, or better, by natural sampling variation.

e) With a \( P \)-value of 0.232, we fail to reject the null hypothesis. There is no evidence to suggest that the dowser has a success rate any higher than 30%.

a) \( H_0 \): The percentage of children with genetic abnormalities is 5\%. \((p = 0.05)\)
\( H_A \): The percentage of children with genetic abnormalities is greater than 5\%. \((p > 0.05)\)

b) Independence assumption: There is no reason to think that one child having genetic abnormalities would affect the probability that other children have them.
Randomization condition: This sample may not be random, but genetic abnormalities are plausibly independent. The sample is probably representative of all children, with regards to genetic abnormalities.
10\% condition: The sample of 384 children is less than 10\% of all children.
Success/Failure condition: \( np = (384)(0.05) = 19.2 \) and \( nq = (384)(0.95) = 364.8 \) are both greater than 10, so the sample is large enough.

c) The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with \( \mu_p = p = 0.05 \) and \( \sigma_p(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.05)(0.95)}{384}} \approx 0.0111 \).

We can perform a one-proportion \( z \)-test. The observed proportion of children with genetic abnormalities is \( \hat{p} = \frac{46}{384} \approx 0.1198 \).

\[
\begin{align*}
z &= \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \\
&= \frac{0.1198 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{384}}} \approx 6.28
\end{align*}
\]
The value of \( z \) is approximately 6.28, meaning that the observed proportion of children with genetic abnormalities is over 6 standard deviations above the hypothesized proportion. The \( P \)-value associated with this \( z \) score is \( 2 \times 10^{-10} \), essentially 0.

d) If 5\% of children have genetic abnormalities, the chance of observing 46 children with genetic abnormalities in a random sample of 384 children is essentially 0.

e) With a \( P \)-value of this low, we reject the null hypothesis. There is strong evidence that more than 5\% of children have genetic abnormalities.

f) We don’t know that environmental chemicals cause genetic abnormalities. We merely have evidence that suggests that a greater percentage of children are diagnosed with genetic abnormalities now, compared to the 1980s.
15. Contributions, please, part II.

a) $H_0$: The contribution rate is 5% ($p = 0.05$)  
$H_A$: The contribution rate is less than 5% ($p < 0.05$)

b) **Independence assumption**: There is no reason to believe that one randomly selected potential donor’s decision will affect another’s decision.

**Randomization condition**: The sample was 100,000 randomly selected potential donors.

**10% condition**: We will assume that the entire mailing list has over 1,000,000 names.

**Success/Failure condition**: $np = 5000$ and $nq = 95,000$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_\hat{p} = p = 0.05$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.05)(0.95)}{100,000}} = 0.0007$.

We can perform a one-proportion z-test. The observed contribution rate is $\hat{p} = \frac{4,781}{100,000} = 0.04781$.

c) Since the $P$-value = 0.0006 is low, we reject the null hypothesis. There is strong evidence that contribution rate for all potential donors is lower than 5%.

```
z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}} = \frac{0.048 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{100,000}}} = -3.25

P = 0.0006
```

```
z = -3.25
```

```
z = -3.25
```

```
0.05
```

```
0.048
```

```
0.05
```

```
0.048
```