Chapter 8 Tests of Hypotheses Based on One Sample

In this chapter, we will look at how to decide between two claims about a parameter.

8.1 Hypotheses and Test Procedures:

Often we are interested in testing a statistical hypothesis which is a claim about a parameter, several parameters, or the entire probability distribution.

ex:
- $\mu = 10$
- Population distribution is $N(0,1)$
- $\mu_1 > \mu_2$

The structure of a hypothesis test:

- We have two contradictory hypotheses under consideration

  ex: $\mu = 10$ vs. $\mu \neq 10$

- Based on a random sample, we want to decide which is correct.
- One claim, called the null hypothesis H_0, is the claim that is initially assumed to be true. (A prior belief claim.)
- The other claim is called the alternative hypothesis H_a. It contradicts the null hypothesis.
- If the sample provides strong evidence the null hypothesis is false, then we reject H_0 and now favor the alternative hypothesis H_a.
- If the sample does not strongly contradict H_0, then we continue to believe the null hypothesis is plausible.
- So the only two possible conclusions are
  - Reject H_0, or
  - Fail to reject H_0.

Example 1: Identify the null hypothesis H_0 and alternative hypothesis H_a for each situation.

a) A company says their tonor cartridge lasts for 2000 pages of text on average. A competitor tries to prove it's actually less than 2000.

H_0: $\mu = 2000$

H_a: $\mu < 2000$

b) The current scientific literature states that a physical constant w is 4.62. A researcher explores the possibility that this estimate is incorrect.

H_0: $w = 4.62$

H_a: $w \neq 4.62$

The Test Procedure and P-Value:

A company wants to test if consumers like brand C more than brand D. They give a blind taste test to a random sample of 100 consumers and record that 72 people prefer brand C to D.

The hypothesis test comes in 4 parts.

① Setup Hypotheses

Let p be the proportion of consumers who like brand C more than D.

$H_0: p = 0.5$
$H_a: p > 0.5$

② Test Statistic

In this case, our test statistic is X, the number of consumers in the sample who prefer brand C to D.

In this case, if H_0 is true then

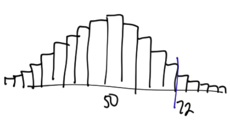
$X \sim Bin(100, 0.5)$

③ P-Value

Data: $X = 72$

If H_0 is true, what is the probability that X is at least as extreme as 72?

$P(X \geq 72)$



in R: pbinom(71.5, 100, 0.5, lower.tail = FALSE)
$\approx 6.29 \times 10^{-6}$

P-value = 0.00000629

④ Conclusion

This sample provides strong evidence against H_0, and in favor of H_a.
We reject H_0 and adopt H_a.

Defns: The Test Statistic is a function of the sample data used as a basis to decide if H_0 should be rejected.

The P-Value is the probability, calculated assuming H_0 is true, that of obtaining a test statistic at least as contradictory to H_0 as the value from this sample.

The significance level α is the number used as a cutoff. If the P-value $\leq \alpha$, then we reject H_0. Instead if P-value $> \alpha$, then we fail to reject H_0.

Example 2: The drying time for a type of paint under specific conditions is known to be normally distributed with mean 75 minutes and sd 9 min. A new additive is tested to see if it decreases drying time. (We'll assume $\sigma = 9$ does not change.) The significance level is $\alpha = 0.05$.

- Form the hypotheses:

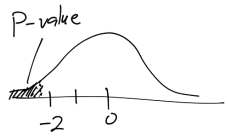
$H_0: \mu = 75$

$H_a: \mu < 75$

- We use the test statistic $z = \dfrac{\bar{X} - 75}{9/\sqrt{n}}$ so that if H_0 is true, $z \sim N(0,1)$

- A sample of $n=25$ gives $\bar{x} = 71$. Find the P-value.

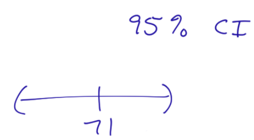
$$z = \frac{71-75}{9/\sqrt{25}} = \frac{-4}{9/5} = -4 \cdot \frac{5}{9} = -\frac{20}{9}$$



P-value $=$ pnorm$(-20/9) \approx 0.013$.

- Therefore, our conclusion is

Since P-value is less than $\alpha = 0.05$, then we reject H_0.
We adopt H_a that the mean drying time is now less than 75.



Errors in Hypothesis Testing:

A **Type I error** is made if we reject H_0 when it is true.

A **Type II error** is made if we fail to reject H_0 when it is false.

	we reject H_0	we fail to reject H_0
H_0 true	Type I error	correct
H_0 false	correct	Type II error

Example 3: A protein bar is advertised to have 10g of protein per serving. A random sample of bars is tested to see if the true average protein per serving is less than 10g. What is a Type I error in this context? What about Type II?

Type I error: There actually is 10g protein per serving, but we rejected H_0.

Type II error: There are actually less than 10g per serving, but we failed to reject H_0.

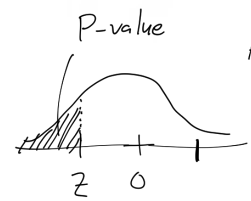
Example 4: As in Example 2, say we are testing the drying time of the paint.

$H_0: \mu = 75$ We use significance level $\alpha = 0.05$, and as before $\sigma = 9$, $n = 25$.
$H_a: \mu < 75$

a) What is a Type I error in this situation?

$\mu = 75$ (additive makes no difference) but we rejected H_0.

$H_0: \mu = 75$
$H_a: \mu < 75$



b) What is P(Type I error)?

Type I error is rejecting H_0 when it's actually true.

If H_0 is true,

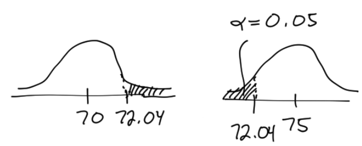
$$P(\text{Type I error}) = P(\text{P-value} \leq 0.05) = 0.05$$

$$z = \frac{71-75}{9/\sqrt{25}}$$

c) What is a Type II error in this situation?

The drying time is less than 75 minutes but we failed to reject H_0.

d) What is the probability we make a Type II error if the true mean drying time with the chemical additive is $\mu = 70$?



If the true mean drying time is $\mu = 70$,

qnorm$(0.05, \text{mean}=75, \text{sd}=\frac{9}{\sqrt{25}})$

$P(\text{type II error}) = P(\text{fail to reject } H_0)$

$= P(\bar{X} > 72.04) = 1 - \text{pnorm}(72.04, \text{mean}=70, \text{sd}=\frac{9}{5})$

$\approx \boxed{0.129}$

Prop: • The test procedure that rejects H_0 if P-value $\leq \alpha$ and otherwise does not reject H_0 has $P(\text{type I error}) = \alpha$.

• After we decide our sampling procedure, sample size, and test statistic, as α increases, the probability of making a type II error decreases.

Comments on P-values:

- The P-value provides more information than just whether or not we reject H_0. So it's a good practice to always report the P-value.
- The P-value is a probability calculated assuming H_0 is true.
- Smaller P-values provide stronger evidence against H_0 and in favor of H_a.
- The P-value is not the probability H_0 is true or false or that our conclusion is an error.

8.2 z Tests for Hypotheses about a Population Mean.

The general hypothesis test structure is

1) Form the hypotheses H_0 and H_a.
2) Compute the appropriate test statistic
3) Determine the P-value, the probability, assuming H_0 is true, that we observe a test statistic at least as extreme as what resulted from the data.
4) Reject H_0 if P-value $\leq \alpha$
   Fail to reject H_0 if P-value $> \alpha$.

Steps (2) and (3) are the steps that vary based on the test statistic used.
In this section, we cover tests for a population mean using a z test statistic.

Normal Population with Known σ:

It is unrealistic to know σ in practice, but this is a good starting point.

1) Our test will involve $H_0: \mu = \mu_0$ and one of these alternatives:

$H_a: \mu > \mu_0$, $H_a: \mu < \mu_0$, or $H_a: \mu \neq \mu_0$.

2) If $X_1, \ldots, X_n$ is sampled from the normal population, then if H_0 is true,

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right).$$

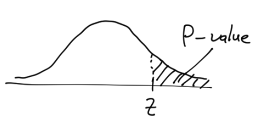
Our test statistic is $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$.

If H_0 is true, $Z \sim N(0,1)$.

3) Finding the P-value depends on the form of H_a:

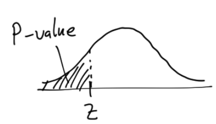
If $H_a: \mu > \mu_0$, then the P-value is the area under the standard normal curve to the right of the test statistic z.

In R, P-value = 1 − pnorm(z)



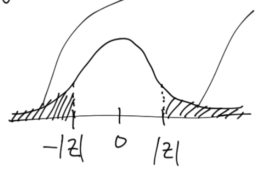
If $H_a: \mu < \mu_0$, then the P-value is the area under the standard normal curve to the left of the test statistic z.

In R, P-value = pnorm(z).



Lastly if $H_a: \mu \neq \mu_0$, then the P-value is the area under the standard normal curve which is left of $-|z|$ or right of $|z|$.

P-value is the sum



$$\text{P-value} = 2 \cdot \text{pnorm}(-|z|)$$

4) As usual, if P-value $\leq \alpha$ we reject H_0.
If P-value $> \alpha$ we fail to reject H_0.

<u>Example 1</u>: A particular type of tire is supposed to be filled to 30 psi.

Assume we know the psi is normally distributed with $\sigma = 2$ psi.

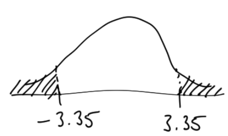
We obtain a random sample of 20 tires and measure their psi to test if the mean psi is not equal to 30. We get $\bar{x} = 28.5$ psi. Perform the hypothesis test at the $\alpha = 0.05$ significance level.

Let μ be the population mean tire psi.

<u>Hypotheses</u>: $H_0: \mu = 30$, $H_a: \mu \neq 30$

<u>Test Statistic</u>: $z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$. We get $z = \dfrac{28.5 - 30}{2/\sqrt{20}} \approx -3.35$

<u>P-value</u>: in R: $2 \cdot \text{pnorm}(-3.35)$



P-value ≈ 0.00081

<u>Conclusion</u>: Since P-value is less than $\alpha = 0.05$, we reject H_0.

So we conclude that the true mean psi of these tires is not 30.

Large Sample Tests

When the sample size is large ($n > 40$), we can slightly modify the test statistic to use a Z-test. (It is not required that the population is normal, and we do not need to know σ.)

As in chapter 7 with CIs, when n is large,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ is approximately standard normal.}$$

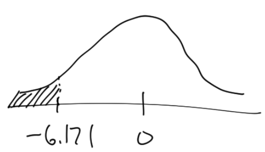
So we can use $Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ as our test statistic. When H_0 is true, this test statistic is approximately standard normal.

Example 2: The recommended daily dietary zinc for a certain cohort is 15 mg/day. In a random sample of $n = 115$ individuals' zinc intake, $\bar{x} = 11.3$ and $s = 6.43$. Does this data indicate the average daily zinc intake for this cohort is below 15 mg?

Hypotheses: $H_0: \mu = 15$ vs $H_a: \mu < 15$

Test Statistic: $Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$. $\quad z = \dfrac{11.3 - 15}{6.43/\sqrt{115}} \approx -6.171$

P-value:



$$\text{P-value} = \text{pnorm}(-6.171) \approx 3.39 \times 10^{-10}$$

Conclusion: We reject H_0, and conclude this cohort's daily zinc intake is on average less than 15 mg.