

8 Statistical Inference: Significance Tests About Hypotheses

8.1 What Are the Steps for Performing a Significance Test?

Example: *Legal setting.* Test the claim that Ralph committed armed robbery.

H_0 : null hypothesis, status quo, conventional wisdom, old idea, accepted idea.

H_a : alternative hypothesis, the challenge to the conventional wisdom, new idea, proposed idea.

If we wish to reject H_0 (i.e., reject the idea that Ralph is innocent of armed robbery) in favor of H_a (i.e., in favor of the idea that Ralph is guilty of armed robbery), we need overwhelming evidence to support our claim, such as witnesses, video recordings, confession, or DNA evidence.

Otherwise, we fail to prove him guilty (i.e., not guilty).

Innocent until proved guilty.

We **reject** the null hypothesis in favor of the alternative hypothesis, or we **fail to reject** the null hypothesis.

Do **NOT** say “accept H_0 ” (which is equivalent to saying “proved innocent”), as a substitution for “**fail to reject H_0 .**”

What is the goal in hypothesis testing?

Regarding the goal of hypothesis testing, the **researcher** is analogous to whom in the legal setting?

□

Example: State the hypotheses for testing the claim that peanut oil **causes** colon cancer.

□

Example: State the hypotheses for testing the claim that peanut oil **prevents** colon cancer.

□

Statistical setting

Example: Suppose a particular politician's approval rating last month was 55%. You believe that this approval rating has decreased, due to a scandal. State the appropriate hypotheses. Let p be the *unknown* current population approval rating of this politician.

□

Example: Suppose the mean personal income of your community last year was \$41,000. You believe that mean personal income has increased, due to improved infrastructure. State the appropriate hypotheses. Let μ be the *unknown* population mean personal income this year.

□

Example: In a college's handbook, the mean

SAT score is listed as 1100. You believe that the information is outdated. State the appropriate hypotheses. Let μ be the *unknown* current population mean SAT score.

8.2 Significance Tests About Proportions

Let p = unknown population proportion.

Let \hat{p} = sample proportion.

We make inferences on p using the point estimate \hat{p} .

We use large samples and apply the Central Limit Theorem; i.e., \hat{p} is approximately normal for large n .

One-sample Z-test on a population proportion, p

Example: Suppose that the National Safety Council believes that more than 20% of all automobile ac-

cidents involve pedestrians. Test this claim at **significance level** $\alpha = 0.05$. Suppose a simple random sample of $n = 200$ automobile accidents results in $X = 46$ involving pedestrians.

(a) Define your notation.

Let p be the unknown **population** proportion of automobile accidents which involve pedestrians.

Let \hat{p} be the **sample** proportion of automobile accidents which involve pedestrians.

(b) State the hypotheses.

(c) Check the rule of thumb **under the null hypothesis**.

(d) Determine our specific value of \hat{p} , the point estimate of p .

(e) What is the approximate distribution of \hat{p} under H_0 ?

(f) Find the value of the **standardized test**

statistic.

(g) Find the P -value.

Standard normal table, pp. A1–A2										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Standard normal table, pp. A1–A2										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The **P -value** is the probability of obtaining a value of the standardized test statistic at least as extreme as the observed value, based on the assumption that H_0 is true.

The smaller the P -value, the stronger the evidence against H_0 .

When the P -value is small, we say that the data are **statistically significant**.

In this example, the P -value is 0.1446.

(h) State the conclusion in statistical terms and in regular English.

Is the P -value small enough that we should reject H_0 ?

□

Rule: Reject H_0 in favor of H_a if P -value $\leq \alpha$; otherwise, fail to reject H_0 .

When P -value = 0.1446, would H_0 be rejected if $\alpha = 0.05$, $\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.15$, and $\alpha = 0.1446$?

A mathematically rigorous definition: The **P -value** is the smallest value of α for which the null hypothesis would be rejected.

The **P -value** is also called the **observed** signifi-

cance level.

Do researchers typically prefer **small** or **large** P -values?

The P -value is NOT $P(H_0 \text{ is true})$.

For example, if the DNA match is 0.01 for the defendant, does this imply that there is a 1% chance that the defendant is innocent?

P -value is $P(\text{Such strong evidence would exist against the defendant, given that the defendant is innocent})$.

P -value is NOT $P(\text{Defendant is innocent})$.

Example: You visit a foreign country on vacation and get thrown into jail for no apparent reason. The blood at some crime scene is type A , and you (who unluckily have type A blood) become the defendant. Suppose that 42% of all people have type A blood.

Is there a 42% chance that you are innocent?

Is there an 58% chance that you are guilty?

□

Example: Suppose that two summers ago 60% of *then-recent* high school graduates enrolled in college. We are interested in whether or not the college enrollment rate changed since two summers ago.

Test the claim at **significance level** $\alpha = 0.1$.

Suppose a simple random sample of 500 most recent high school graduates results in 275 enrolled in college.

(a) Define your notation.

Let p be the unknown **population** proportion of most recent high school graduates who are enrolled in college.

Let \hat{p} be the **sample** proportion of most recent high school graduates who are enrolled in college.

(b) State the null and alternative hypotheses.

(c) Check the rule of thumb **under the null hypothesis**.

- (d) Under H_0 , what is the approximate distribution of \hat{p} , the point estimate of p ?
- (e) Find the value of the **standardized test statistic**.
- (f) Find the P -value.

Standard normal table, pp. A1–A2										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- (g) State the conclusion in statistical terms and in regular English.

We conclude that the **population** proportion of most recent high school graduates who are enrolled in college DIFFERS from 60%.

□

Remark: Commonly used values of α are 0.01, **0.05**, and 0.1.

8.3 Significance Tests About Means

Let μ = unknown population mean.

Let \bar{X} = sample mean.

We make inferences on μ using the **point estimate** \bar{X} .

We use large samples and apply Central Limit Theorem; i.e., \bar{X} is approximately normal for **large** n . Alternatively, we start with an approximately normal population, in which case \bar{X} is approximately normal for **any** n .

One-sample t -test on a population mean, μ

Recall: For independent or nearly independent observations (and finite σ), if **the original population is approximately normal OR n is large**, then

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \underset{\text{approx.}}{\sim} t_{n-1}.$$

Example: The manufacturer of Chimney cigarettes lists the mean nicotine content as being 2.5 mg. Volcano Incorporated invented a new brand of cigarettes with the same great taste as Chimneys, but claims that this new brand has a lower mean nicotine content. To test the claim of Volcano Incorporated at significance level $\alpha = 0.1$, a researcher samples the nicotine content of 41 Volcano cigarettes and finds $\bar{X} = 2.47$ mg and $s = 0.09$ mg. Let $\mu =$ population mean nicotine content of Volcano cigarettes.

- (a) State the null and alternative hypotheses.
- (b) Find the value of the **standardized test statistic**.
- (c) Find the P -value.

<i>t</i> -table, p. A3						
Confidence Level						
	80%	90%	95%	98%	99%	99.8%
Right-Tail Probability						
<i>df</i>	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
⋮	⋮	⋮	⋮	⋮	⋮	⋮

(d) State the conclusion in statistical terms and in regular English.

We conclude that the **population** mean nicotine content of Volcano cigarettes is less than 2.5 mg.

□

Example: Suppose the mean lifetime of mice is 26 months. Test at level 0.01 if a new strain of mice has mean lifetime different from 26 months. Seven mice are independently sampled, and their lifetimes in months are {20, 23, 13, 7, 17, 15, 10}.

- (a) Define your notation.
- (b) Is the original population approximately normal, or is the sample size large?
- (c) State the null and alternative hypotheses.
- (d) Find the value of the **standardized test statistic**.
- (e) Find the P -value.

<i>t</i> -table, p. A3						
	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
	Right-Tail Probability					
<i>df</i>	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
⋮	⋮	⋮	⋮	⋮	⋮	⋮

- (f) State the conclusion in statistical terms and in regular English.

We conclude that this new strain of mice has **population** mean lifetime **different** from 26 months.

(g) Now, construct a 99% confidence interval on μ .

t-table, p. A3						
Confidence Level						
	80%	90%	95%	98%	99%	99.8%
Right-Tail Probability						
<i>df</i>	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Layman's interpretation: We are 99% confident that the population mean lifetime of this new strain of mice is between 7.20 months and 22.80 months.

Mathematically rigorous interpretation: If we repeat the sampling procedure many times to construct many 99% confidence intervals on μ , the population mean lifetime of this new strain of mice, then approximately 99% of these 99% confidence intervals will contain the true value of μ .

- (h) Is our 99% confidence interval consistent with the conclusion of our 2-sided hypothesis test of level $\alpha = 0.01$?
- (i) Suppose we had tested $H_0 : \mu = 20$ months against $H_a : \mu \neq 20$ months at level 0.01.

□

Remark: A strong connection exists between 2-sided hypothesis tests of level α and $(1 - \alpha)$ level confidence intervals.

Remark: The t -procedures are **robust**. Hence, even under some violations of the assumptions (i.e., n is large, or the original population is approximately normal), the t -test and the t -confidence interval often produce accurate results anyway.

Example: Suppose 20 observations are sampled from the following **Uniform** population.

□

8.4 Decisions and Types of Errors in Significance Tests

Recall: The **P-value** is the probability of obtaining a value of the standardized test statistic at least as extreme as the observed value, based on the assumption that H_0 is true.

Recall: We reject H_0 if and only if $P\text{-value} \leq \alpha$.

What is the likelihood that we (erroneously) reject H_0 , when in fact H_0 is true?

Definition: A **Type I error** occurs if we reject H_0 , when H_0 is true.

What is $P(\text{Type I error})$; i.e.,
what is $P(\text{We reject } H_0 | H_0 \text{ is true})$?

Definition: A **Type II error** occurs if we fail to reject H_0 , when H_a is true.

Define $\beta = P(\text{Type II error})$
 $= P(\text{We fail to reject } H_0 | H_a \text{ is true})$

Example: Consider the hypotheses:

H_0 : Ralph is **innocent** of armed robbery.

H_a : Ralph is **guilty** of armed robbery.

Describe the Type I error.

Describe the Type II error.

Describe α and β in terms of probabilities and proportions.

□

What are the possible values of α ?

What are the possible values of β ?

Do we want α to be *large* or *small*?

Do we want β to be *large* or *small*?

For the *defendants-in-court* example, how can α be made small (near zero)?

What happens to β , as α gets small (near zero)?

For the *defendants-in-court* example, how can β be made small (near zero)?

What happens to α , as β gets small (near zero)?

Example: State the null and alternative hypotheses, when a person is tested for a disease.

Describe the **Type I error** in regular English and in medical terminology.

Describe the **Type II error** in regular English and in medical terminology.

□

In the **statistical** setting, how can we minimize both α and β ?

8.5 Limitations of Significance Tests

“Statistical significance does not mean practical significance.”

Definition: Data are **statistically significant** when the P -value is **small**; i.e., P -value $\leq \alpha$, so the data suggest rejecting H_0 in favor

of H_a .

Definition: Data are **practically significant** when the conclusion is of **practical** value.

Example: *Revisit Volcano cigarettes.* $\bar{X} =$
(sample mean nicotine content) = 2.47 mg.

$$H_0 : \mu = 2.5 \text{ mg}$$

$$H_a : \mu < 2.5 \text{ mg}$$

$$P\text{-value} < 0.025 < 0.1 = \alpha$$

We concluded that the **population** mean nicotine content of Volcano cigarettes is less than 2.5 mg.

Are the data **statistically** significant?

Are the data **practically** significant?

□

Example: *Revisit lifetime of mice.* $\bar{X} =$ (sample mean lifetime) = 15 months.

$$H_0 : \mu = 26 \text{ months}$$

$H_a : \mu \neq 26$ months

P -value $< 0.002 < 0.01 = \alpha$

We concluded that this new strain of mice has

population mean lifetime **different** from 26 months.

Are the data **statistically** significant?

Are the data **practically** significant?

□