8 Hypothesis Testing

8.1 Basic Principles of Hypothesis Testing

Example: *Legal setting.* Test the claim that Athena shoplifted at Walmart.

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

H_1: 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 Athena is innocent of shoplifting) in favor of H_1 (i.e., in favor of the idea that Athena is guilty of shoplifting), we need overwhelming evidence to support our claim, such as witnesses, video recordings, confession, or DNA evidence.

Otherwise, we fail to prove her 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 of 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 coffee **causes** colon cancer.

□

Example: State the hypotheses for testing the claim that coffee **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 number of credit hours which full-time students take at the beginning of the Fall semester is listed as 15.2. You believe that the information is outdated. State the appropriate hypotheses. Let μ be the *unknown* current population mean number of credit hours which full-time students take at the beginning of the Fall semester.

□

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

Definition: A **Type II error** occurs if we fail to reject H_0, when H_1 is true.

Example: Consider the hypotheses:

H_0 : Athena is **innocent** of shoplifting at Walmart.

H_1 : Athena is **guilty** of shoplifting at Walmart.

Describe the Type I error.

Describe the Type II error.

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.

□

8.4 Hypothesis Tests for Proportions, p

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

Example: Suppose that the National Safety Council believes that more than 20% of all automobile accidents 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 $\boldsymbol{p}$ be the unknown **population** proportion of automobile accidents which involve pedestrians.

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

(b) State the hypotheses.

(c) Check the rule of thumb **under the null hypothesis,** and discuss any other necessary assumptions.

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

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

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

(g) Find the P-value.

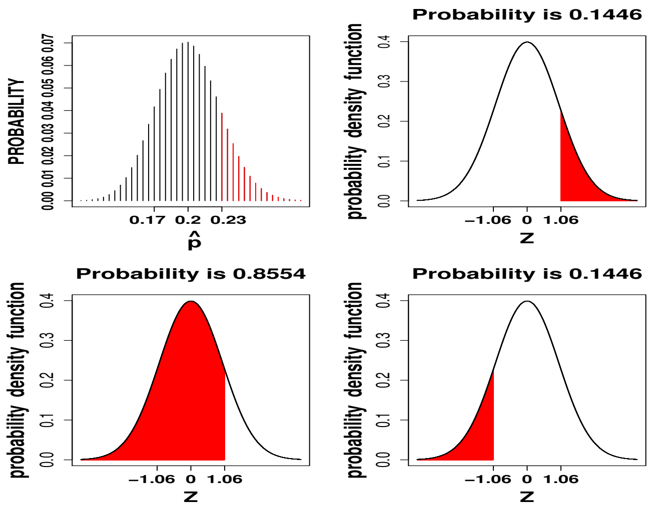
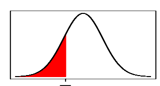
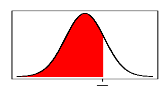


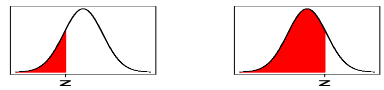
Table A.2 Cumulative Normal Distribution, pp. A-6 and A-7





z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table A.2 Cumulative Normal Distribution, pp. A-6 and A-7



z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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_1 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.

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

□

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 $\boldsymbol{p}$ be the unknown **population** proportion of most recent high school graduates who are enrolled in college.

Let $\hat{\boldsymbol{p}}$ 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,** and discuss any other necessary assumptions.

(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.

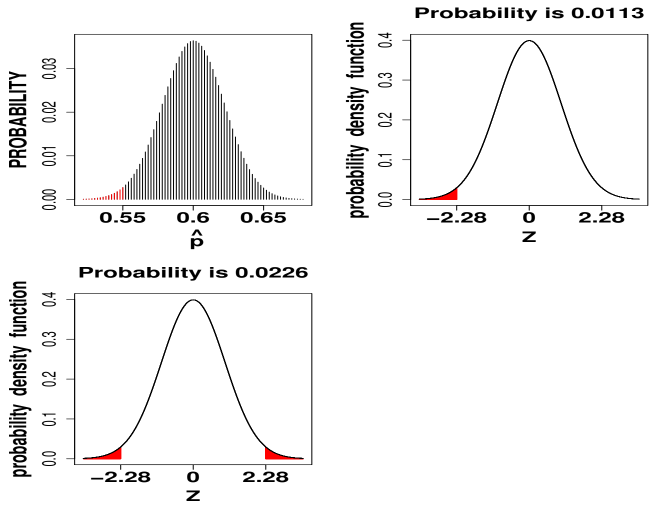
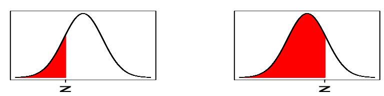


Table A.2 Cumulative Normal Distribution, pp. A-6 and A-7



z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−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.

8.2 Hypothesis Tests for a Population Mean μ, Standard Deviation σ Known

Scenario: μ is unknown, but σ is known.

What is needed in order for σ to be **known**?

8.3 Hypothesis Tests for a Population Mean μ, Standard Deviation σ Unknown

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 positive finite σ), if **the original population is approximately normal OR n is large**, then

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

Example: The manufacturer of Cardio-delight Inc. lists the mean saturated fat content as being 2.5 grams per chocolate chip cookie. Sugarland Inc. produces a new brand of chocolate chip cookies with the same great taste (and weight/mass) as Cardio-delight, but claims that this new brand has a lower mean saturated fat content. To test the claim of Sugarland at significance level $\alpha = 0.1$, a researcher samples the saturated fat content of 41 Sugarland cookies and finds $\bar{X} = 2.47$ g and $s = 0.09$ g. Let μ = population mean saturated fat content of Sugarland cookies.

(a) State the null and alternative hypotheses.

(b) Find the value of the **standardized test statistic**.

(c) Find the P-value.

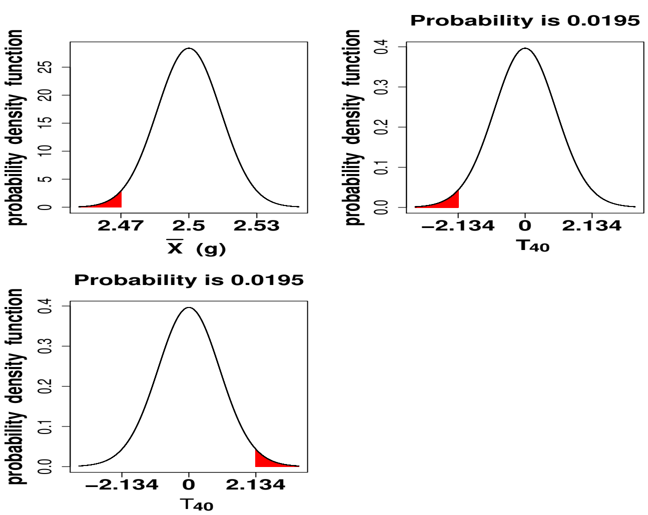
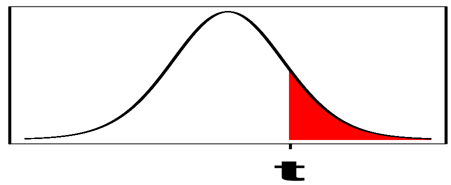


Table A.3 Critical Values for the Student's t Distribution, p. A-8



Degrees of Freedom	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
					Area in Right Tail					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
39	0.255	0.681	1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.255	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

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

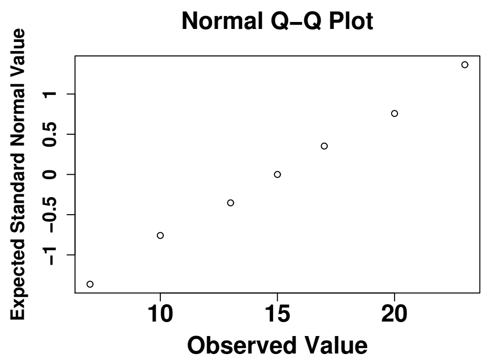
We conclude that the **population** mean saturated fat content of Sugarland cookies is less than 2.5 grams.

□

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? Also, discuss any other necessary assumptions.



(c) State the null and alternative hypotheses.

(d) Find the value of the **standardized test statistic**.

(e) Find the P-value.

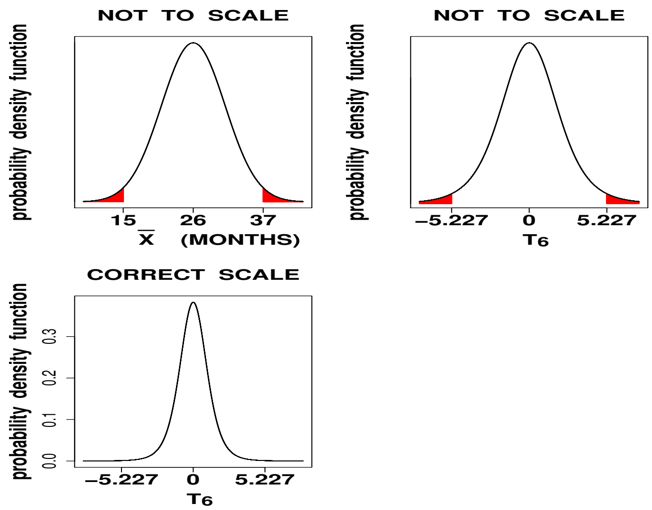
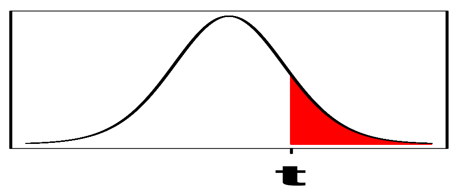


Table A.3 Critical Values for the Student's t Distribution, p. A-8



Degrees of Freedom	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
					Area in Right Tail					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(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.

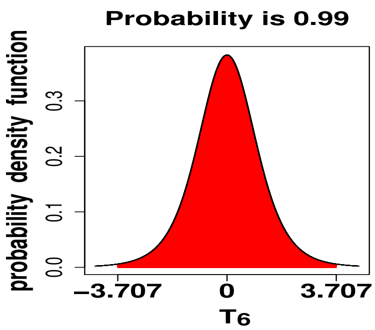
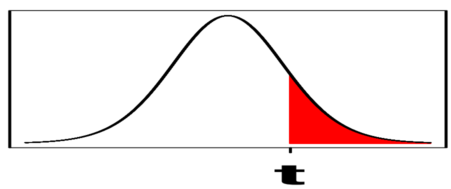


Table A.3 Critical Values for the Student's t Distribution, p. A-8



Degrees of Freedom	Area in Right Tail 0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
200	0.254	0.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340
z	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
Confidence Level	20%	50%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

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 $\alpha = 0.01$?

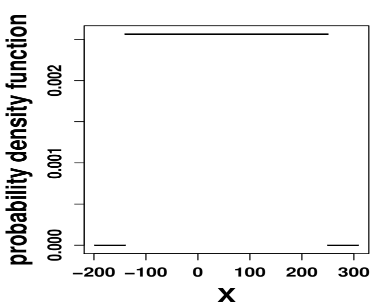
(i) Suppose we had tested $H_0 : \mu = 20$ months against $H_1 : \mu \neq 20$ months at level 0.01.

□

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.



□

Statistical significance does not imply practical significance.

(from section 8.2)

A large sample size often can produce statistical significance without 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_1.

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

Example: *Revisit Sugarland cookies.* $\bar{X} =$ (sample mean saturated fat content) $= 2.47$ g.

$H_0 : \mu = 2.5$ g

$H_1 : \mu < 2.5$ g

P-value $< 0.025 < 0.1 = \alpha$

We concluded that the **population** mean saturated fat content of Sugarland cookies is less than 2.5 grams.

Are the data **statistically** significant?

Are the data **practically** significant?

□

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

$H_0 : \mu = 26$ months

$H_1 : \mu \neq 26$ 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?

□