9 Inferences on Two Samples

•• We may wish to compare two **treatment** groups in **experimental design**.

Example: In the agricultural setting, which type of seed produces a better yield per acre?

□

Example: Which of two drugs is better?

□

•• We may wish to compare two **populations** in **sample surveys**.

Example: Compare the heights of females in Mexico vs. the United States, or the likelihood of developing cancer between smokers and nonsmokers.

□

When comparing two **treatment groups** in experimental design or two **populations** in sample surveys, we may use

(a) **Independent samples** (sections 9.1 and 9.3), OR

(b) **Dependent samples (matched pairs – section 9.2 – IDEAL)**.

Example: *Matched pairs*

□

When **matched pairs** are not possible, use **independent samples**.

Example: *Independent samples.*

□

9.3 Inference About the Difference Between Two Proportions

Comparing two population proportions based on independent samples.

Z-test and Z-confidence interval on the difference between two population proportions, $(p_1 - p_2)$

Example: Let p_1 = *unknown, fixed* population proportion of **female** adults (at least 21-years-old) who have a high school diploma.

Let p_2 = *unknown, fixed* population proportion of **male** adults (at least 21-years-old) who have a high school diploma.

Are these two population proportions the same?

What is the difference between these two population proportions?

□

Example: Consider an **experiment** involving prostate cancer and surgery, as reported by the *New England Journal of Medicine*, 2002.

Does surgery reduce the death rate (due to prostate cancer, within 6.2 additional years) for prostate cancer patients?

From 1989 through 1999, 695 Scandinavian men with newly diagnosed prostate cancer were randomly assigned to surgery (radical prostatectomy) or control.

Treatment #	Group	died	survived	sample size	death rate
1	control	31	317	$n_1 = 348$	
2	surgery	16	331	$n_2 = 347$	

Let $\boldsymbol{p_1}$ be the **population** proportion of the **control** group who would die (within 6.2 years) from prostate cancer.

Let $\boldsymbol{p_2}$ be the **population** proportion of the **surgery** group who would die (within 6.2 years) from prostate cancer.

To be continued below.

□

What is a reasonable point estimate of $(p_1 - p_2)$?

$\mu_{\hat{p}_1 - \hat{p}_2} = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2$;

i.e., population mean difference between two sample proportions is the same as the difference between the two population proportions.

For independent or nearly independent observations,

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2,$$

and hence

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}.$$

Suppose all observations are independent or nearly independent (and the sample sizes are reasonably large). Then, by the Central Limit Theorem,

(1) $[\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)]/\sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2} \overset{approx.}{\sim} N(0, 1)$, if there are **at least 10 successes and at least 10 failures** in each of the two samples, and

(2) A confidence interval on *unknown, fixed* $(p_1 - p_2)$ is

$$\hat{p}_1 - \hat{p}_2 \pm z\,\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2},$$

if there are **at least 10 successes and at least 10 failures** in each of the two samples.

Confidence Interval on $(p_1 - p_2)$

Example: *Prostate cancer and surgery.*

(a) Determine the **point estimate** of $(p_1 - p_2)$.

(b) Interpret your above **point estimate** in regular English.

We estimate that for 4.3% of patients, surgery makes a positive difference in terms of surviving vs. not surviving at least an additional 6.2 years, but NOT for the remaining 95.7% of patients.

(c) Check the assumptions for constructing a confidence interval.

(d) Construct a **95%** confidence interval on $(p_1 - p_2)$.

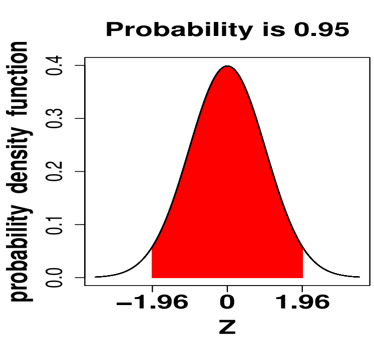
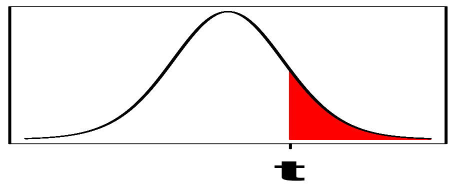


Table A.3 Critical Values for the Student's t Distribution, p. A-8



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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
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
	20%	50%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
	Confidence Level									

(e) State the Layman's interpretation and the mathematically rigorous interpretation of your above confidence interval.

Layman's interpretation: We are 95% confident that the difference in population death rates of control and surgery is between 0.58% and 8.02%.

Mathematically rigorous interpretation: If we repeat the sampling procedure many times to construct many 95% confidence intervals on $(p_1 - p_2)$, the difference in population death rates of control and surgery, then approximately 95% of these 95% confidence intervals will contain the true value of $(p_1 - p_2)$.

□

Hypothesis Testing on $(p_1 - p_2)$

Again, assume the observations are independent or nearly independent.

What is a reasonable point estimate of $(p_1 - p_2)$?

What is the **pooled** (i.e., overall) sample proportion of successes?

Under H_0, the standard deviation of $(\hat{p}_1 - \hat{p}_2)$ is
$\sqrt{p_0(1-p_0)/n_1 + p_0(1-p_0)/n_2}$,
which is estimated by $\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$.

Recall: If all observations are independent or nearly independent and the sample sizes are reasonably large, then by the Central Limit Theorem,
$[\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)]/\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} \overset{approx.}{\sim} N(0,1)$.

Determine the standardized test statistic.

Rule of thumb for hypothesis tests (on the difference between two proportions): If there are **at least 10 successes and at least 10 failures** in each of the two samples, then the **standardized test statistic** is approximately standard normal.

Example: Consider an **experiment** involving aspirin and heart attacks, as reported by *New England Journal of Medicine*, 1988.

Male physicians aged 40 to 84 in the United States in 1982 participated in the double-blinded randomized controlled experiment. Treatment was one 325 milligram aspirin tablet every other day. Results were determined about 5 years later. Test at level 0.05 whether or not aspirin reduces the likelihood of a heart attack in this population, in comparison to a placebo.

		heart attack			
Treatment #	Group	yes	no	sample size	sample proportion of heart attacks
1	placebo	189	10,845	$n_1 = 11,034$	
2	aspirin	104	10,933	$n_2 = 11,037$	
	total	293	21,778	22,071	

(a) State the notation.

Let $\boldsymbol{p_1}$ be the **population** proportion of **placebo** users who would suffer a heart attack.

Let $\boldsymbol{p_2}$ be the **population** proportion of **aspirin** users who would suffer a heart attack.

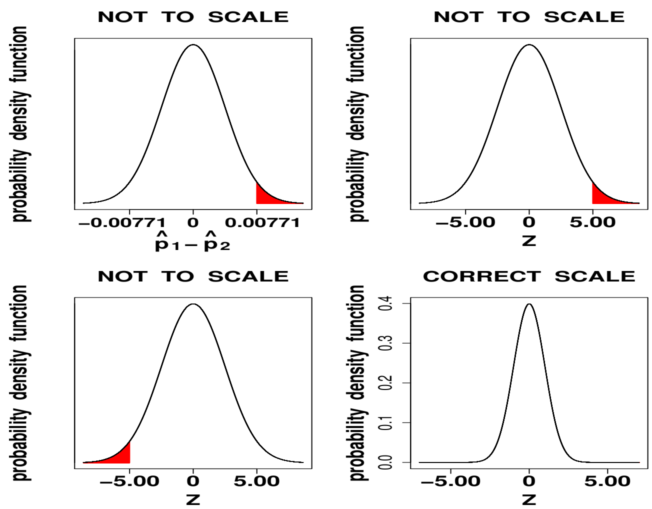
(b) State the hypotheses.

(c) Check the assumptions for performing a hypothesis test.

(d) Determine the **point estimate** of $(\boldsymbol{p_1} - \boldsymbol{p_2})$.

(e) Determine the value of the **standardized test statistic**.

(f) Determine the P-value.



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

We conclude that use of aspirin results in a lower likelihood of a heart attack in this population of male physicians aged 40 to 84 in the United States, in comparison to a placebo.

□

Note: For a **two**-sided test, the P-value is twice the tail probability of the appropriate **one**-sided test.

9.1 Inference About the Difference Between Two Means: Independent Samples

Comparing two population means based on

independent samples.

In this section, we focus on **independent observations**, not **matched pairs**.

Construct independent t-test and independent t-confidence interval.

Population #1: Take independent or nearly independent observations from a population with mean μ_1 and finite positive standard deviation σ_1.

Let $\bar{X}_1$ be the sample mean and s_1 be the sample standard deviation, based on a sample of size n_1.

Population #2: Take independent or nearly independent observations from a population with mean μ_2 and finite positive standard deviation σ_2.

Let $\bar{X}_2$ be the sample mean and s_2 be the sample standard deviation, based on a sample of size n_2.

Assume that the two samples are independent of each other.

Question: Is $\mu_1 = \mu_2$, OR is $\mu_1 - \mu_2 = 0$?

Estimate: $(\mu_1 - \mu_2)$

What is the **point estimate** of $(\mu_1 - \mu_2)$?

What is the mean of $(\bar{X}_1 - \bar{X}_2)$?

It can be shown that since the samples are independent or nearly independent, then

$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

For the rest of this section, assume that all observations in the samples are **independent** or **nearly independent**, and both σ_1 and σ_2 are positive and finite.

If n_1 and n_2 are both large (usually $n_1 > 30$ and $n_2 > 30$, if none of the tails of the two distributions are too heavy), or if the two populations are approximately normal, then

$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ **NOT PRACTICAL for inference**

is approximately standard normal, and

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad \textbf{PRACTICAL}$$

is approximately t distributed, so a **confidence interval** on $(\mu_1 - \mu_2)$ is

$$\bar{X}_1 - \bar{X}_2 \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Degrees of freedom: The **degrees of freedom** can be approximated conservatively by the **smaller** of $\boldsymbol{(n_1 - 1)}$ and $\boldsymbol{(n_2 - 1)}$.

Listed in your textbook is a very ugly but more accurate formula for degrees of freedom, so we simply will use the above approximation.

How can we verify the normality assumption?

Again, the t procedures are **robust**.

Example: A study of zinc-deficient mothers was conducted to determine whether zinc supplementation during pregnancy results in babies with increased mean weights at birth. {*Data are available at* Goldenberg et al., *JAMA* 1995 (August 9); 274 (6): 463-468.}

Treatment #1 Zinc supplement group	Treatment #2 Placebo group
$n_1 = 294$	$n_2 = 286$
$\bar{X}_1 = 3214$ g	$\bar{X}_2 = 3088$ g
$s_1 = 669$ g	$s_2 = 728$ g

Is there sufficient evidence to support the claim that zinc supplementation results in increased mean birth weight, in comparison to a placebo? Test at level $\alpha = 0.05$.

(a) Do we need to assume that the two populations for birth weight are approximately normally distributed?

(b) Define your notation.
Let $\boldsymbol{\mu_1}$= *unknown* **population** mean birth weight in the **zinc**-supplemented group.
Let $\boldsymbol{\mu_2}$= *unknown* **population** mean birth weight in the **placebo** group.

(c) State the hypotheses.

(d) Determine the value of the **standardized test statistic.**
Let $\boldsymbol{\bar{X}_1}$= **sample** mean birth weight in the **zinc**-supplemented group.
Let $\boldsymbol{\bar{X}_2}$= **sample** mean birth weight in the **placebo** group.

(e) Determine the estimated number of degrees of freedom.

(f) Determine the P-value.

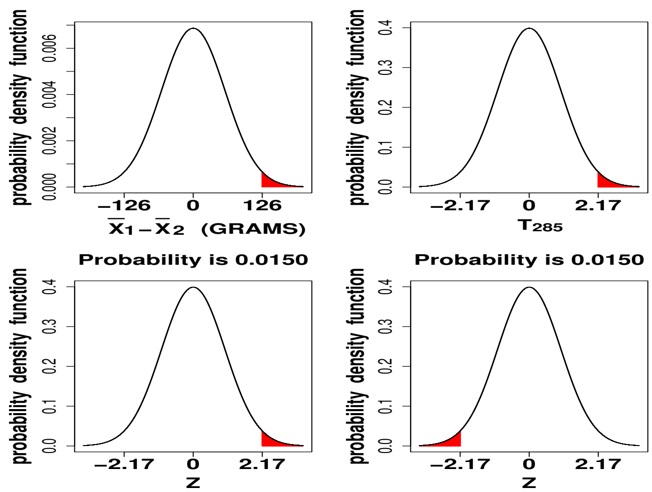
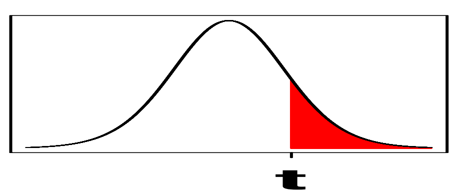
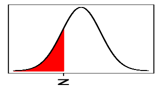


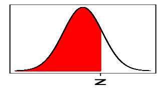
Table A.3 Critical Values for the Student's t Distribution, p. A-8



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**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
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
	20%	50%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
	Confidence Level									

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**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

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

We conclude that zinc supplementation during pregnancy among zinc-deficient mothers results in babies with increased mean weight at birth, in comparison to a placebo.

(h) Construct a **99%** confidence interval on $(\boldsymbol{\mu_1 - \mu_2})$.

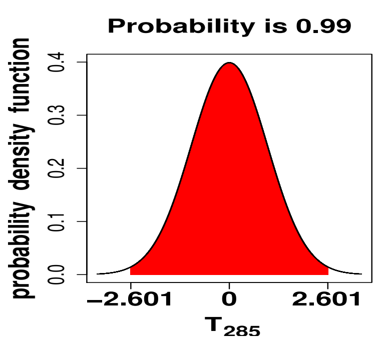
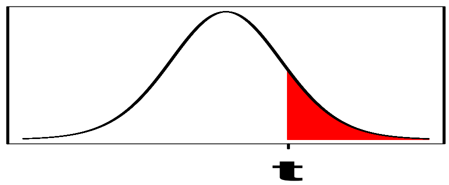


Table A.3 Critical Values for the Student's t Distribution, p. A-8



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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
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
	20%	50%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
	Confidence Level									

Layman's interpretation: We are 99% confident that the difference in population mean birth weights between zinc-users and placebo-users among zinc-deficient mothers lies between -25.1 grams and 277.1 grams.

Mathematically rigorous interpretation: If we repeat the sampling procedure many times to produce many 99% confidence intervals on $(\mu_1 - \mu_2)$, the difference in population mean birth weights between zinc-users and placebo-users among zinc-deficient mothers, then approximately 99% of these 99% confidence intervals will contain the true value of $(\mu_1 - \mu_2)$.

(i) Construct a **99%** confidence interval on $(\boldsymbol{\mu_2 - \mu_1})$.

□

9.2 Inference About the Difference

Between Two Means: Paired Samples

Comparing two population means based on paired (dependent) samples.

Here, we pair the observations.

Construct paired-t test and paired-t confidence interval.

What are some examples of **paired observations**?

We assume the pairs of observations are independent or nearly independent, but we do **NOT** necessarily have independence **within** a pair.

Let $\boldsymbol{X}$ be the (observation in sample #1) $-$ (observation in sample #2); hence, $X = Y_1 - Y_2$.

We make inferences on the **difference between two means**, $(\mu_1 - \mu_2)$, or the **mean difference**, μ_x.

What is a reasonable **point estimate** of μ_x?

Assumptions:

(1) The observations are reasonably paired.

(2) The **differences** are independent or nearly independent (and σ_x is positive and finite).

(3) $\boldsymbol{n}$ **is large** (usually $n > 30$, if neither tail of the distribution *of the* **differences** is too heavy), or the **differences** are **approximately normal.**

Then, the standardized test statistic is $(\bar{X} - \mu_x)/(s/\sqrt{n}) \overset{approx.}{\sim} t_{n-1}$.

Confidence interval on μ_x is $\bar{X} \pm t_{n-1}\ s/\sqrt{n}$.

Example: *Hypothetical data.* Test at level $\alpha = 0.05$ whether the **population** mean (systolic reading of) blood pressure is reduced by more than 10 when using a placebo. The data consist of the following *before* and *after* blood pressure readings of five patients: $\{(190, 180), (220, 205), (242, 214), (175, 156), (201, 177)\}$.

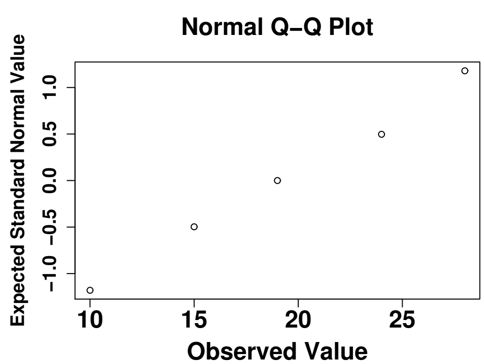
(a) Define your notation.

Let $\boldsymbol{X}$ be the difference in blood pressure, *before* minus *after*.

Let $\boldsymbol{\mu}$ (or $\boldsymbol{\mu_x}$) be the *unknown* **population** mean difference in blood pressure.

(b) State the hypotheses.

(c) Check the assumptions.



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

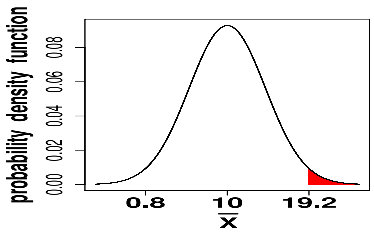
Let $\bar{\boldsymbol{X}}$ be the **sample** mean difference in blood pressure.

Let $\boldsymbol{s}$ be the **sample** standard deviation of the difference in blood pressure.

Goal: Construct a one-sample t test on μ.

(e) How many **degrees of freedom** are associated with this test?

(f) Determine the P-value.



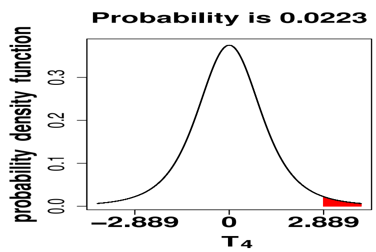
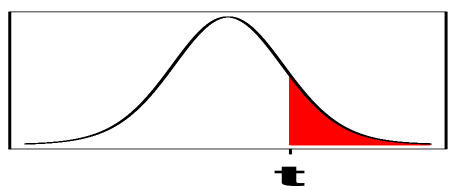


Table A.3 Critical Values for the Student's t Distribution, p. A-8



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
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

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

We conclude that the **population** mean (systolic reading of) blood pressure is reduced by more than 10 when using a placebo.

(h) Construct a **98%** confidence interval on μ_x.

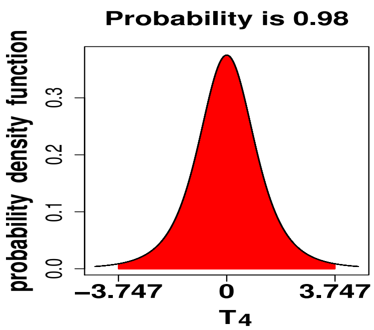
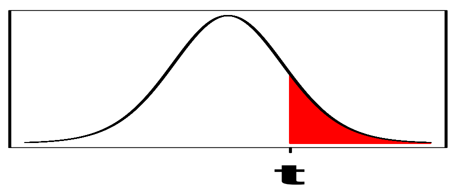


Table A.3 Critical Values for the Student's t Distribution, p. A-8



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
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
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 98% confident that μ_x, the population mean reduction in (systolic reading of) blood pressure due to the placebo effect, is between 7.27 and 31.13, when using a placebo.

Mathematically rigorous interpretation: If we repeat the sampling procedure many times to produce many 98% confidence intervals on μ_x, the population mean reduction in (systolic reading of) blood pressure due to the placebo effect, then approximately 98% of these 98% confidence intervals will contain the true value of μ_x.

□