9 Inferences Based 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 females and males.

□

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

(a) **Independent samples** (sections 9.1, 9.2, 9.4, 9.5), OR

(b) **Paired samples – IDEAL** (section 9.3).

Example: *Matched pairs.*

□

When **paired samples** are not possible, use **independent samples**.

Example: *Independent samples.*

□

9.1 Z Tests and Confidence Intervals for a Difference Between Two Population Means

In this section, we focus on **independent samples**, not **paired samples**.

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

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

Let $\bar{X}$ be the sample mean, based on a sample of size m.

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

Let $\bar{Y}$ be the sample mean, based on a sample of size n.

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 $(\bar{X} - \bar{Y})$?

It can be shown that since the samples are independent or nearly independent, then $\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma_1^2/m + \sigma_2^2/n}$.

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 m and n are both large (usually $m \geq 30$ and $n \geq 30$, if none of the tails of the two distribution are too heavy), or if the two populations are approximately normal, then

$Z = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}$ **NOT ALWAYS PRACTICAL for inference**

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

$$\bar{X} - \bar{Y} \pm z\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}.$$

How can we verify the normality assumption?

Example: Suppose that in a large country, based on a simple random sample, the sample mean height of **400** women in year **1950** was **63.2** inches with a standard deviation of **2.4** inches. In the same country, based on another simple random sample, the sample mean height of **500** women **this year** is **64.5** inches with a standard deviation of **2.5** inches. Test at level 0.05 if the average height of women increased by more than **one** inch between **1950** and **this year**.

(a) Define your notation.

(b) Are the original populations approximately normal, or are the sample sizes large?

(c) State the null and alternative hypotheses.

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

(e) Find the P-value.

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

□

Example: *Unrealistic.* Consider butterfat production (in pounds) for a cow during the 305-day milk production period following birth of a calf. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be butterfat production for cows on a farm in **Wisconsin** and a farm in **Michigan**, respectively, where the population standard deviations of X and Y are known to be $\sigma_1 = 175$ and $\sigma_2 = 140$, but the population means, μ_1 and μ_2, are unknown. All 28 observations were sampled independently, and two Q-Q plots confirm the normality assumption. The sample means and sample sizes are $\bar{X} = 712$, $m = 11$, $\bar{Y} = 630$ and $n = 17$.

(a) Why is this example "unrealistic"?

(b) Construct the **95%** confidence interval on $(\boldsymbol{\mu_1 - \mu_2})$.

(c) Test at level 0.05 for the inequality of the population mean butterfat production for **Wisconsin** and the population mean butterfat production for **Michigan**.

- State the null and alternative hypotheses.

- Determine the value of the **standardized test statistic**.

- Determine the **P-value**.

- State the conclusion in statistical terms and in regular English.

- Determine $\boldsymbol{\beta(\mu_1 - \mu_2 = 150}$ pounds$)$.

- Determine power($\boldsymbol{\mu_1 - \mu_2 = 150}$ pounds).

□

9.2 The Two-Sample t Test and Confidence Interval

In this section, we again focus on **independent samples**, not **paired samples**.

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 positive finite standard deviation σ_1.

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

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

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

Assume that the two samples are independent of each other.

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 m and n are both large (usually $m \geq 30$ and $n \geq 30$, if none of the tails of the two distribution are too heavy), or if the two populations are approximately normal, then

$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$ **NOT ALWAYS 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 are estimated by

$$\nu = \frac{(s_1^2/m + s_2^2/n)^2}{(s_1^2/m)^2/(m-1) + (s_2^2/n)^2/(n-1)}$$

You need NOT memorize this formula.

What happens to ν and the distribution of T as m and n get large?

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
$m = 294$	$n = 286$
$\bar{X}_1 = 3214$ 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 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.

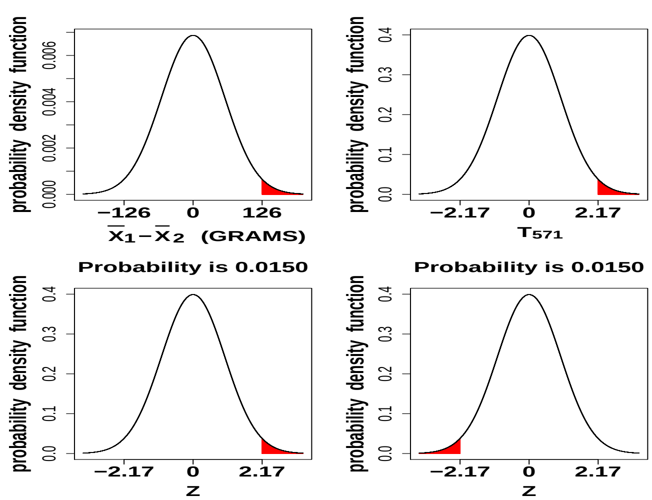
(d) Determine the value of the **standardized test statistic.**

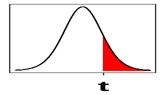
Let $\bar{\boldsymbol{X}}_{\boldsymbol{1}}$= **sample** mean birth weight in the **zinc**-supplemented group.

Let $\bar{\boldsymbol{X}}_{\boldsymbol{2}}$= **sample** mean birth weight in the **placebo** group.

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

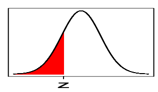
(f) Determine the P-value.

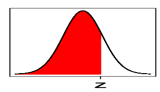




t-table, pp. 728–729

$t \backslash \nu$	29	30	35	40	60	120	$\infty (= z)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.1	.022	.022	.022	.021	.020	.019	.018
2.2	.018	.018	.017	.017	.016	.015	.014
2.3	.014	.014	.014	.013	.012	.012	.011
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮





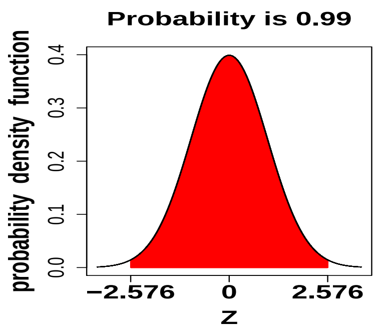
Standard normal table, pp. 722–723

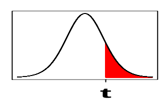
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(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 $(\mu_1 - \mu_2)$.





t-table, p. 725

	α						
ν	**.10**	**.05**	**.025**	**.01**	**.005**	**.001**	**.0005**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

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 -23.7 grams and 275.7 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 $(\boldsymbol{\mu_2 - \mu_1})$.

□

Pooled t Procedures

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

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

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

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

Assume that the two samples are independent of each other.

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

Hence, for **pooled t procedures**, the **new** assumption is a **common standard deviation, $\boldsymbol{\sigma}$**, for the two populations.

Estimate $\boldsymbol{\sigma^2}$ by the **pooled** sample variance,

$$s_p^2 = \frac{(m-1)\ s_1^2 + (n-1)\ s_2^2}{m+n-2}.$$

Degrees of freedom are $\boldsymbol{m+n-2}$.

$$\sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} = \frac{\sigma^2}{m} + \frac{\sigma^2}{n} = \sigma^2\ \left(\frac{1}{m} + \frac{1}{n}\right)$$

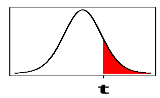
$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_p\ \sqrt{1/m + 1/n}} \sim t_{m+n-2}$$

Example: Suppose that $X_1, \ldots, X_m \sim N(\mu_1, \sigma)$ and $Y_1, \ldots, Y_n \sim N(\mu_2, \sigma)$, such that all observations are mutually independent. Consider the data $m = 9$, $\bar{X} = 195$, $s_1 = 12$, $n = 6$, $\bar{Y} = 160$, and $s_2 = 11$.

(a) Test at level 0.01,

$H_0 : \mu_1 - \mu_2 \geq 50$

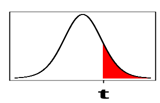
$H_a : \mu_1 - \mu_2 < 50$



t-table, pp. 728–729

$t \setminus \nu$	11	12	13	14	15	16	17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.3	.021	.020	.019	.019	.018	.018	.017
2.4	.018	.017	.016	.015	.015	.014	.014
2.5	.015	.014	.013	.013	.012	.012	.011
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

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



t-table, p. 725

	α						
ν	**.10**	**.05**	**.025**	**.01**	**.005**	**.001**	**.0005**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

□

Comparing Pooled t and Non-pooled t procedures

(1) If $\sigma_1 = \sigma_2$, then **pooled** t procedures are preferable.

(2) The **pooled** t procedures are NOT **robust** to departures from $\sigma_1 = \sigma_2$. If $\sigma_1 \neq \sigma_2$, then **non-pooled** t procedures typically are preferred.

(3) t-procedures are **robust** to departures from non-normality, in general.

(4) Equality of σ_1 and σ_2 often is **difficult** to establish, especially for small sample sizes.

(5) Only use **pooled** t if the exercise explicitly states that $\sigma_1 = \sigma_2$, OR if you can verify that $\sigma_1 = \sigma_2$.

9.3 Analysis of Paired Data

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 $\boldsymbol{D}$ be the (observation in sample #1) $-$ (observation in sample #2).

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

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

Assumptions:

(1) The observations are reasonably paired.

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

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

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

Confidence interval on μ_D is $\bar{D} \pm t_{n-1}\ s_D/\sqrt{n}$.

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: $\{(190, 180), (220, 205), (242, 214), (175, 156), (201, 177)\}$.

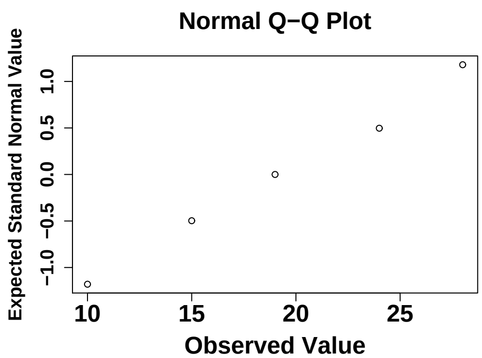
(a) Define your notation.

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

Let $\boldsymbol{\mu_D}$ be the **population** mean difference in blood pressure.

(b) State the hypotheses.

(c) Check the assumptions.



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

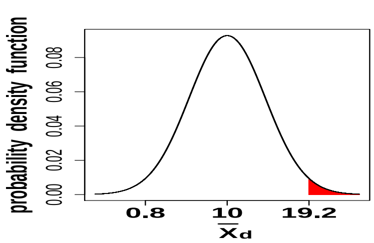
Let $\bar{D}$ (or $\bar{X} - \bar{Y}$) be the **sample** mean difference in blood pressure.

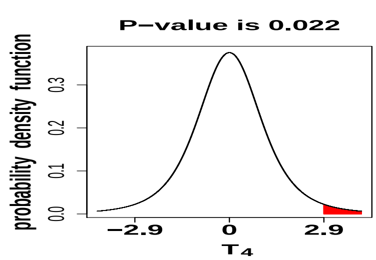
Let s_D be the **sample** standard deviation of the difference in blood pressure.

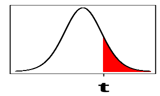
Goal: Construct a one-sample t test on μ_D.

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

(f) Determine the P-value.







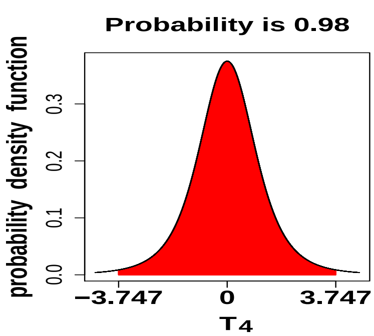
t-table, pp. 728–729

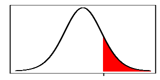
$t \setminus \nu$	**1**	**2**	**3**	**4**	**5**	**6**	**7**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.8	.109	.054	.034	.024	.019	.016	.013
2.9	.106	.051	.031	.022	.017	.014	.011
3.0	.102	.048	.029	.020	.015	.012	.010
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(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_D.





t-table, p. 725

	$\boldsymbol{\alpha}$						
ν	**.10**	**.05**	**.025**	**.01**	**.005**	**.001**	**.0005**
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Layman's interpretation: We are 98% confident that μ_D, 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_D, 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_D.

□

9.4 Inferences Concerning a Difference Between Population Proportions

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 $\boldsymbol{p_2}$ = *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	$m = 348$	
2	surgery	16	331	$n = 347$	

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,

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

and hence

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

If all observations are independent or nearly independent and the sample sizes are reasonably large, then by the Central Limit Theorem,

(1) $[\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)]/\sqrt{p_1(1-p_1)/m + p_2(1-p_2)/n} \overset{approx.}{\sim} N(0,1)$, and

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

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

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

The number of **successes** *in sample* **#1** *is* $m\hat{p}_1$.
The number of **successes** *in sample* **#2** *is* $n\hat{p}_2$.
The number of **failures** *in sample* **#1** *is* $m(1-\hat{p}_1)$.
The number of **failures** *in sample* **#2** *is* $n(1-\hat{p}_2)$.

Confidence Interval on $(p_1 - p_2)$

Example: *Prostate cancer and surgery.*

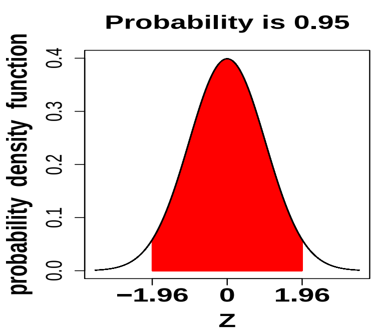
(a) Determine the **point estimate** of $(p_1 - p_2)$.

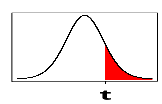
(b) Interpret your above **point estimate** in regular English.

We estimate that for 4.3% of patients, surgery makes a positive difference in terms of surviving vs. not surviving 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 $(p_1 - p_2)$.





t-table, p. 725

	α						
ν	**.10**	**.05**	**.025**	**.01**	**.005**	**.001**	**.0005**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

(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 **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.

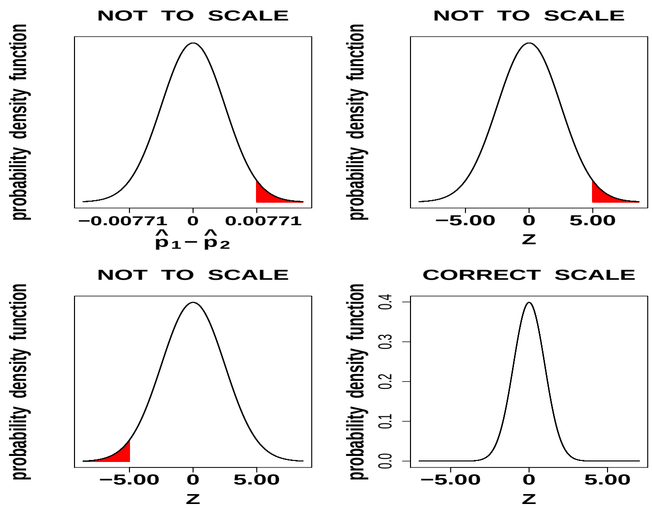
(b) State the hypotheses.

(c) Check the assumptions for performing a significance test (i.e., hypothesis test).

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

(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.

□

9.5 Inferences Concerning Two Population Variances

Population #1: Take independent or nearly independent observations from a **normal** population with positive finite standard deviation σ_1.

Let s_1 be the sample standard deviation, based on a sample of size m.

Population #2: Take independent or nearly independent observations from a **normal** population with positive finite standard deviation σ_2.

Let s_2 be the sample standard deviation, based on a sample of size n.

Assume that the two samples are independent of each other.

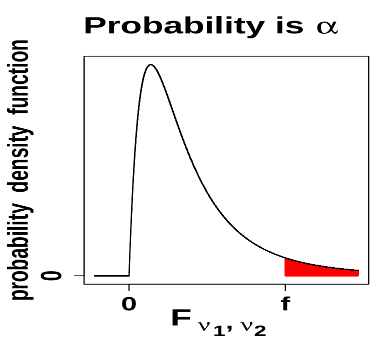
For the rest of this section, assume that all observations in the samples are **independent** or **nearly independent**, and both σ_1 and σ_2 are positive and finite.

Test at level α,

$H_0: \ \sigma_1^2 = \sigma_2^2$

$H_a: \ \sigma_1^2 \neq \sigma_2^2$

Then, $s_1^2/s_2^2 \sim F_{m-1,n-1}$ under H_0, and $s_2^2/s_1^2 \sim F_{n-1,m-1}$ under H_0.



Example: Prove that $f_{\alpha,\nu_1,\nu_2} = 1/f_{1-\alpha,\nu_2,\nu_1}$.

□

For the **two-tailed** test at level $\boldsymbol{\alpha}$:

(a) Reject H_0 if and only if $\max\{s_1^2/s_2^2,\ s_2^2/s_1^2\}$ exceeds the critical value, $f_{\alpha/2}$.

(b) For degrees of freedom, use **numerator** degrees of freedom and **denominator** degrees of freedom corresponding to **max** in the above expression.

Example: Assume that $X_1, \ldots, X_m \sim N(\mu_1, \sigma_1)$, and $Y_1, \ldots, Y_n \sim N(\mu_2, \sigma_2)$, such that all observations are mutually independent.

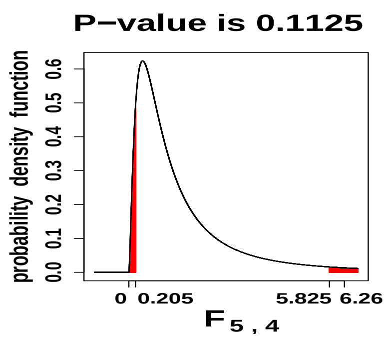
The data set from population **#1** consists of {15, 21, 11, 16, 26}.

The data set from population **#2** consists of {35, 50, 44, 62, 75, 55}.

Test at level $\alpha = 0.1$,

$H_0: \ \sigma_1^2 = \sigma_2^2$

$H_a: \ \sigma_1^2 \neq \sigma_2^2$



□

Remark: The F test is **sensitive** to moderate departures from normality, whereas the t test is NOT.

Hence, the F test is NOT reliable when populations deviate much from normality assumptions, unless sample sizes are huge.

Remark: The F test sometimes is used to decide whether to use a **pooled** t test or an **independent** t test.

However, the **pooled** t test should NOT be used if the two variances differ a lot, since the F test is NOT too powerful (i.e., difficult to reject $H_0 : \sigma_1^2 = \sigma_2^2$), especially for **small** sample sizes.

Thus, even when we fail to reject $H_0 : \sigma_1^2 = \sigma_2^2$, the **pooled** t test still might be **inappropriate** since the likelihood of committing a type **II** error (where $\sigma_1^2 \neq \sigma_2^2$) is often high.