Chapter 8: Inference for means

Yonghyun Kwon

Department of Mathematics, Korea Military Academy

Confidence interval for a population mean

Notations

Recall: Sample mean, $\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

where $x_1, x_2, \cdots, x_n$ represent the n observed values.

Recall: Sample variance, s^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

where $x_1, x_2, \cdots, x_n$ represent the n observed values.

$s = \sqrt{s^2}$ is the sample standard deviation.

Sample Mean from a Normal Population

Recall: Sample mean from $N(\mu, \sigma^2)$

If $X_1, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, then

$$E[\bar{X}] = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n},$$

and since a linear combination of independent normals is normal,

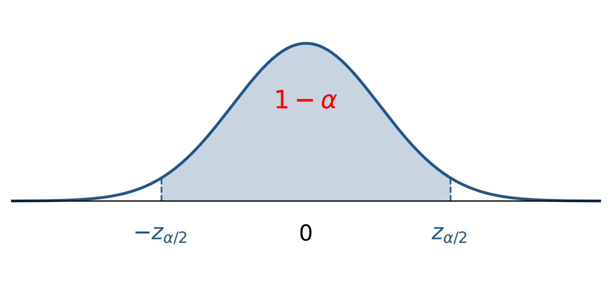
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Standardizing $\bar{X}$ gives

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1).$$

From Sampling Distribution to Confidence Interval

Since $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, the middle $100(1-\alpha)\%$ of the distribution is captured between $-z_{\alpha/2}$ and $z_{\alpha/2}$:



$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha.$$

Derivation of the z-Confidence Interval

Substituting $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ and rearranging to isolate μ:

$$
\begin{aligned}
1 - \alpha &= P\left(-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \bar{X} - \mu \le z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(-\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right).
\end{aligned}
$$

A z-Confidence Interval for the Mean (known σ)

Recall: Confidence interval using any confidence level (Ch. 5)

If a point estimate closely follows a normal model, a $100(1-\alpha)\%$ CI for the population parameter is

$$\text{point estimate} \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \times \underbrace{SE}_{\text{std. error}},$$

where $z_{\alpha/2} \times SE$ is called the *margin of error*.

The z-CI for the mean is a **special case** with point estimate $= \bar{x}$ and $SE = \sigma/\sqrt{n}$:

z-confidence interval for the *mean* (known σ)

$$\bar{x} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}.$$

Exercise: z-Confidence Interval for the Mean

A quality control engineer measures the fill volume (in ml) of 49 randomly selected bottles from a production line. The sample mean is $\bar{x} = 500.4$ ml, and the population standard deviation is known to be $\sigma = 2.1$ ml. Construct a 95% confidence interval for the true mean fill volume μ.

- Standard error:

$$SE =$$

- Critical value at $\alpha = 0.05$: $z_{0.025} =$
- 95% confidence interval:

$$\bar{x} \pm z_{0.025} \times SE \ \rightarrow$$

Introducing the t-Distribution

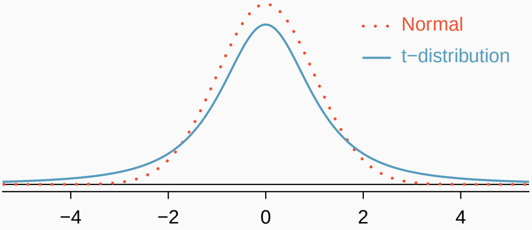
- When σ is unknown, we estimate it with s, giving

$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}.$$

- This estimate is reliable for large n, but introduces extra uncertainty for small n.
- To account for this, we replace the z-distribution with the *t-distribution*.

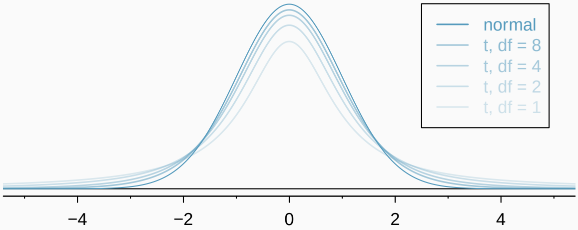
Comparison of the t-Distribution and Normal Distribution

- The t-distribution has the same bell shape as $N(0, 1)$ but with **thicker tails**.



Degrees of Freedom (*df*)

- The t-distribution is always centered at zero and has a single parameter: **degrees of freedom (*df*)**.
- Write $T \sim t(df)$.
- As df increases, $t(df) \to N(0, 1)$.



Remarks on the t-Distribution and T-Statistic

Note: Definition of t-distribution: Suppose $Z \sim N(0,1)$ and $V \sim \chi^2(k)$ are independent. Then

$$\frac{Z}{\sqrt{V/k}} \sim t(k).$$

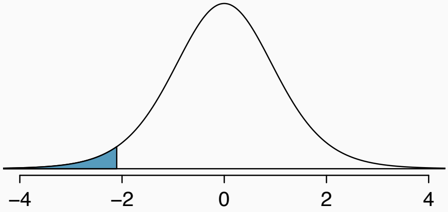
Note: Derivation of T-statistic: If $X_1, X_2, \cdots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$,

$$Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ and } V := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

and $\bar{X}$, S^2 are independent. From the definition of t-distribution,

$$T := \frac{\bar{X} - \mu}{S/\sqrt{n}} = \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{Z} \times \left(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{V} \frac{1}{n-1} \right)^{-1/2} = \frac{Z}{\sqrt{V/(n-1)}} \sim t(n-1).$$

What proportion of the *t*-distribution with 18 degrees of freedom falls below –2.10?




```
> pt(q = -2.10, df = 18, lower.tail = TRUE)
[1] 0.0250452
```

A t-Confidence Interval for the Mean

t-confidence interval for the *mean* (unknown σ)

Based on a sample of n independent and nearly normal observations, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} \pm t_{\alpha/2}(df) \times \frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}(df)$ is the $\alpha/2$-th upper quantile of $t(df)$ with $df = n - 1$.

	z-interval	t-interval
Condition	σ known	σ unknown
SE	$\sigma/\sqrt{n}$	$s/\sqrt{n}$
Critical value	$z_{\alpha/2}$	$t_{\alpha/2}(n-1)$

Example: Mercury Content in Risso's Dolphins

Photo by Mike Baird (*www.bairdphotos.com*)

- Dolphins accumulate mercury from the ocean, which can be harmful to them and to humans who consume them.
- Goal: construct a confidence interval for the average mercury content in dolphin muscle.

Sample Data and Preliminary Calculations

n	$\bar{x}$	s	min	max
19	4.4	2.3	1.7	9.2

Table 1: Mercury content in 19 Risso's dolphins (μg/wet g).

- Standard error ($n = 19$, $s = 2.3$):

$$SE =$$

- Degrees of freedom: $df =$

Computing the 95% Confidence Interval

- Critical value at $\alpha = 0.05$: $t_{0.025}(18) =$

```
> qt(p = 0.025, df = 18, lower.tail = FALSE)
[1] 2.100922
```

- 95% confidence interval:

$$\bar{x} \pm t_{0.025}(18) \times SE \;\rightarrow$$

- We are 95% confident that the true mean mercury content is between 3.29 and 5.51 μg/wet g.

Hypothesis test for a population mean

Hypothesis Testing Framework

- *Null Hypothesis (H_0)*: The default skeptical assumption.
- *Alternative Hypothesis (H_A)*: The new claim being tested.

It is believed that the mean human body temperature is 36.5°C. A researcher suspects the true mean for newborns may be **higher** than this.

Example hypothesis test

$$H_0 : \mu = 36.5, \qquad H_A : \mu > 36.5$$

One-Sided vs. Two-Sided Hypotheses

- *Two-sided* – interested in whether μ differs from μ_0 in either direction:

$$H_0 : \mu = \mu_0, \qquad H_A : \mu \neq \mu_0$$

- *One-sided* – interested in a specific direction only:
  - Lower-tailed(Left-sided): $H_0 : \mu = \mu_0, \quad H_A : \mu < \mu_0$
  - Upper-tailed(Right-sided): $H_0 : \mu = \mu_0, \quad H_A : \mu > \mu_0$

Decision Errors

We make a decision about whether to reject H_0, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 error*
	H_A true	*Type 2 error*	✓

- *Type 1 error*: Reject H_0 when H_0 is true.
  $P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$
- *Type 2 error*: Fail to reject H_0 when H_A is true.
  $P(\text{Type 2 error} \mid H_A \text{ true}) = \beta$

Significance Level

- The *significance level* α is the threshold for the Type 1 error rate we are willing to tolerate:

$$P(\text{reject } H_0 \mid H_0 \text{ is true}) = \alpha.$$

- Common choice: $\alpha = 0.05$.
- Lowering α reduces Type 1 errors but increases Type 2 errors, and vice versa.

Power of a Test

Power

The *power* of a test is the probability of correctly rejecting H_0 when H_A is true:

$$\text{Power} = P(\text{reject } H_0 \mid H_A \text{ true}) = 1 - \beta.$$

- Higher power $\Rightarrow$ better ability to detect a true effect.
- Power increases with: larger n, larger effect size, larger α.
- **Trade-off:** $\downarrow \alpha$ (fewer Type 1 errors) $\Rightarrow \uparrow \beta$ (lower power).

Example: Body Temperature of Newborns

It is believed that the human body temperature is normally distributed with mean $\mu_0 = 36.5°$C. A researcher suspects the true mean for newborns may be **higher** than this. Assume the population standard deviation is known to be $\sigma = 0.9°$C.

- A random sample of $n = 36$ newborns is collected.
- Hypotheses:

$$H_0 : \mu = 36.5, \qquad H_A :$$

Computing the Test Statistic

Since $X_1, \cdots, X_n$ is a random sample from a normal distribution,

$$\bar{X} \overset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \Longrightarrow \quad Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0, 1).$$

A *test statistic* is a statistic used to conduct a hypothesis test.

From the sample, we have: $\bar{x} = 36.8$.

Also, $\mu_0 = 36.5$, $\sigma = 0.9$, $n = 36$.

- Observed test statistic:

$$z =$$

Finding the p-Value

Since $H_A : \mu > \mu_0$, we ask: how likely is it to observe $z \geq 2.00$ if H_0 is true? This probability, computed under H_0, is the *p-value*:

$$p\text{-value} = P(Z \geq 2.00) = 0.023.$$

- Since p-value (greater / less) than $\alpha = 0.05$, we (reject / fail to reject) H_0.
- The data (provide / do not provide) strong evidence that the mean body temperature of newborns exceeds 36.5°C.

The *p*-Value

p-value

The *p-value* is the probability of observing data at least as extreme as the current data, *assuming H_0 is true.*

- Smaller p-value $\Rightarrow$ stronger evidence against H_0.
- **Decision rule:**
  - p-value $< \alpha \Rightarrow$ **reject** H_0.
  - p-value $\geq \alpha \Rightarrow$ **fail to reject** H_0.

Rejection Region Approach

Since $H_A : \mu > \mu_0$, large values of z provide evidence against H_0.
We reject H_0 when $z > c$ for some constant c.

At significance level $\alpha = 0.05$, we choose c such that

$$P(Z > c) = 0.05 \quad \Longrightarrow \quad c = z_{0.05} = 1.645.$$

This c is called the *critical value*.

- Rejection region: $z >$
- Observed test statistic: $z = 2.00 >$ so we
- This is consistent with the p-value conclusion.

Rejection Region

Rejection Region

The *rejection region* is the set of test-statistic values for which H_0 is rejected at level α.

Note: The rejection-region approach and the p-value approach always yield the same conclusion.

One-Sample Z-Test for Population Mean (known σ)

Z-test for the mean with *known* σ

For $H_0 : \mu = \mu_0$ with known σ, the test statistic and its null distribution are

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1).$$

Alternative	Rejection region	p-value
$H_A : \mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$	$2\,P(Z > \lvert z \rvert)$
$H_A : \mu > \mu_0$	$z > z_{\alpha}$	$P(Z > z)$
$H_A : \mu < \mu_0$	$z < -z_{\alpha}$	$P(Z < z)$

Example: Cherry Blossom Race (unknown σ)

The Cherry Blossom Run is a 10-mile road race held annually in Washington, D.C. In 2006, the mean finishing time among all runners was $\mu_0 = 93.29$ minutes. A researcher wonders whether the average run time has **changed** in 2017. Unlike the previous example, the population standard deviation σ is **unknown**.

- A random sample of $n = 100$ runners from 2017 is collected.
- Hypotheses:

$$H_0 : \mu = 93.29, \qquad H_A :$$

Computing the Test Statistic

Since $X_1, \cdots, X_n$ is a random sample from a normal distribution,

$$\bar{X} \overset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \Longrightarrow \quad T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(n-1).$$

From the sample, we have: $\bar{x} = 97.32$, $s = 16.98$.
Also, $\mu_0 = 93.29$, $n = 100$.

- Observed test statistic:

$$t =$$

Finding the p-Value

Since $H_A : \mu \neq \mu_0$, both large and small values of t provide evidence against H_0. We ask: how likely is it to observe $|t| \geq 2.37$ if H_0 is true?

$$p\text{-value} = 2 \times P(T \geq 2.37) = 2 \times 0.0099 = 0.020, \quad T \sim t(99).$$

- Since p-value (greater / less) than $\alpha = 0.05$, we (reject / fail to reject) H_0.
- The data (provide / do not provide) strong evidence that the mean run time in 2017 differed from 2006.

Rejection Region Approach

Since $H_A : \mu \neq \mu_0$, we reject H_0 when $|t| > c$ for some constant c. At $\alpha = 0.05$, we choose c such that

$$P(|T| > c) = 0.05 \quad \Longrightarrow \quad c = t_{0.025}(99) = 1.984.$$

- Rejection region: $|t| >$
- Observed test statistic: $|t| = 2.37 >$ so we
- This is consistent with the p-value conclusion.

Using `t.test()` in R

```
> t.test(time, alternative = "two.sided", mu = 93.29)


 One Sample t-test


data:  time
t = 2.3701, df = 99, p-value = 0.019721
alternative hypothesis: true mean is not equal to 93.29
95 percent confidence interval:
  93.93546 100.70783
sample estimates:
mean of x
 97.32167
```

One-Sample t-Test for Population Mean (unknown σ)

T-test for the mean with *unknown* σ

For $H_0 : \mu = \mu_0$ with unknown σ, the test statistic and its null distribution are

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(n-1).$$

Alternative	Rejection region	p-value
$H_A : \mu \neq \mu_0$	$\lvert t \rvert > t_{\alpha/2}(n-1)$	$2\,P(T > \lvert t \rvert)$
$H_A : \mu > \mu_0$	$t > t_{\alpha}(n-1)$	$P(T > t)$
$H_A : \mu < \mu_0$	$t < -t_{\alpha}(n-1)$	$P(T < t)$

Summary: Z-test vs. t-test for the Mean

	Z-test	t-test
σ	Known	Unknown
Test statistic	$Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$T = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$
Null distribution	$N(0, 1)$	$t(n-1)$
Critical value (two-sided)	$z_{\alpha/2}$	$t_{\alpha/2}(n-1)$

- The t-critical value $t_{\alpha/2}(n-1)$ is always larger than $z_{\alpha/2}$ for any finite n.
- As $n \to \infty$, $t(n-1) \to N(0, 1)$, so the two procedures converge.

Power calculations for a population mean

Example: Delivery Time of a Courier Service

A courier service has historically delivered packages with mean time $\mu_0 = 30$ minutes, with known population standard deviation $\sigma = 5$ minutes. After implementing a new training program, the manager suspects that the true mean delivery time is now **lower** than 30 minutes.

- A random sample of $n = 25$ deliveries is collected.
- Hypotheses:

$$H_0 : \mu = 30, \qquad H_A :$$

Computing the Test Statistic

Under H_0, since $X_1, \cdots, X_n$ is a random sample from a normal distribution with mean $\mu_0 = 30$ and variance $\sigma^2 = 5^2$, the test statistic is the sample mean

$$\bar{X} \overset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right), \quad \text{that is,} \quad \bar{X} \overset{H_0}{\sim}$$

The recorded delivery times (in minutes) are:

i	1	2	3	$\cdots$	24	25
x_i	25	27	30	$\cdots$	28	29

- Observed test statistic:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 28$$

Rejection Region in Terms of $\bar{X}$

Since $H_A : \mu < \mu_0$, small values of $\bar{x}$ provide evidence against H_0. We reject H_0 when $\bar{x} < c$ for some constant c.

At significance level $\alpha = 0.05$, we choose c such that $P(\bar{X} < c) = \alpha$ under H_0. Standardizing, $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$, and

$$P\left(Z < \frac{c - \mu_0}{\sigma/\sqrt{n}}\right) = 0.05 \implies \frac{c - \mu_0}{\sigma/\sqrt{n}} = -z_{0.05} = -1.645.$$

Hence

$$c =$$

- Rejection region: $\bar{x} <$
- Observed: $\bar{x} = 28$ so we

Computing the Power

Suppose H_A is true and the true mean is $\mu = 27$. Under $\mu = 27$,

$$\bar{X} \overset{H_A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right), \text{ or } \bar{X} \overset{H_A}{\sim} N(27, 1^2).$$

The *power* of the test is the probability of rejecting H_0 under H_A ($\mu = 27$):

$$\begin{aligned} \text{Power} = P(\bar{X} < 28.355 \mid \mu = 27) &= P\left(\frac{\bar{X} - 27}{1} < \frac{28.355 - 27}{1}\right) \\ &= P(Z < 1.355) \approx 0.912. \end{aligned}$$

So the test has about a 91.2% chance of detecting a true mean as small as 27 minutes.

Paired data

Amazon vs UCLA Bookstore

- We investigate whether Amazon prices were different from UCLA Bookstore prices.

	Subject	Bookstore Price	Amazon Price	Price Difference
1	American Indian Studies	47.97	47.45	0.52
2	Anthropology	14.26	13.55	0.71
3	Arts and Architecture	13.50	12.53	0.97
⋮	⋮	⋮	⋮	⋮
68	Jewish Studies	35.96	32.40	3.56

Paired Observations

Paired data

Two sets of observations are **paired** if each observation in one set corresponds with exactly one observation in the other set.

- Each textbook has two prices: UCLA Bookstore and Amazon.
- Differences computed as:

$$\text{UCLA Bookstore price} - \text{Amazon price}$$

- Example differences:
  - $47.97 - 47.45 = 0.52$
  - $14.26 - 13.55 = 0.71$
  - $13.50 - 12.53 = 0.97$

Histogram of Price Differences

Does UCLA Bookstore prices seem different from Amazon?

Paired t-test

Consider a hypothesis test on the difference of two means μ_{diff}:

$$H_0 : \mu_{diff} = \delta_0, \quad H_A : \mu_{diff} \neq \delta_0$$

Test statistic for paired data

The test statistic and its null distribution are

$$T = \frac{\bar{X}_{diff} - \delta_0}{SE(\bar{X}_{diff})} = \frac{\bar{X}_{diff} - \delta_0}{S_{diff} / \sqrt{n_{diff}}} \overset{H_0}{\sim} t(n_{diff} - 1),$$

where $\bar{X}_{diff}$, S_{diff}, and n_{diff} are the sample mean, the standard deviation, and the number of differences respectively.

Hypothesis test for Price Differences

- Hypotheses:

$$H_0 : \mu_{diff} = 0 \text{ (No difference in average textbook prices)}$$
$$H_A : \mu_{diff} \neq 0 \text{ (Difference in average prices)}$$

- To analyze a paired data set, we analyze the differences.
- Summary statistics for price differences

n_{diff}	$\bar{x}_{\text{diff}}$	s_{diff}
68	3.58	13.42

Computing the Test Statistic

- Sample statistics:

$$n_{diff} = 68, \quad \bar{x}_{diff} = 3.58, \quad s_{diff} = 13.42$$

- Compute the observed T-score:

$$t =$$

Finding the P-value

P-value of the two sided test $= 2 \times P(T > 2.20)$, where $T \sim$

```
> pt(2.20, df = 67, lower.tail = FALSE)
[1] 0.01562998
```

Therefore, the P-value is

- Since P-value is (greater / less) than $\alpha = 0.05$, we (reject / don't reject) H_0.
- The data (provide / do not provide) strong evidence that Amazon prices are different from UCLA bookstore prices.

```
> t.test(UCLA, Amazon, alternative = "two.sided",
         mu = 0, paired = TRUE)

 Paired t-test

data:  UCLA and Amazon
t = 2.2012, df = 67, p-value = 0.03117
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.3340641 6.8324064
sample estimates:
mean difference
       3.583235
```

Practice

(Exercise 7.19) We analyzed temperature data from 197 NOAA stations where records were available for both 1948 and 2018, comparing the number of days exceeding 90℉. The average difference (2018 - 1948) was 2.9 days with a standard deviation of 17.2 days. We seek evidence that there were more days in 2018 that exceeded 90℉ from NOAA's weather stations using **a paired t-test**.

1. State the null and alternative hypotheses.(Use a one-sided test.)

2. Find the null distribution of the test statistic.

Practice

(Exercise 7.19) We analyzed temperature data from 197 NOAA stations where records were available for both 1948 and 2018, comparing the number of days exceeding 90℉. The average difference (2018 - 1948) was 2.9 days with a standard deviation of 17.2 days. We seek evidence that there were more days in 2018 that exceeded 90℉ from NOAA's weather stations using **a paired t-test**.

3. Compute the observed test statistic.

4. Compute the p-value and complete the hypothesis test.

Difference of two means: unequal variances

Stem cells and heart function

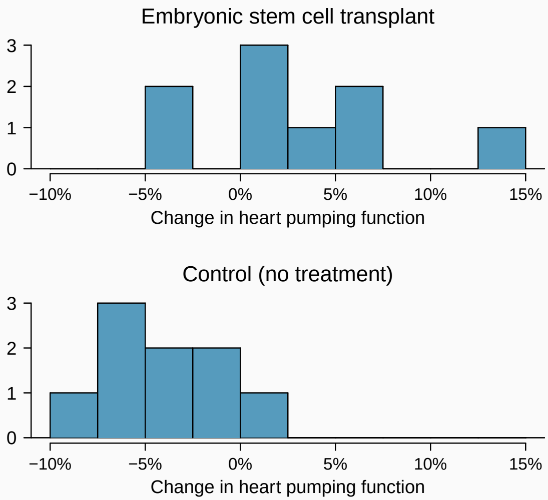
Study on embryonic stem cells (ESCs) and heart function in sheep.

- Sheep were *randomly assigned* to the ESC or control group.
- Change in heart pumping capacity was measured.
- Positive value corresponds to increased pumping capacity.

Group	n	$\bar{x}$	s
ESCs	9	3.50	5.17
Control	9	-4.33	2.76

Table 2: Summary statistics of the ESC study.

Are the ESC group and control group paired?



Can we use *t*-test to compare the two groups?

Suppose $\bar{X}_1$ and $\bar{X}_2$ are independent sample means of size n_1 and n_2 from the two distributions with mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. From Exercise 3.47,

$$Var(\bar{X}_1) = \frac{\sigma_1^2}{n_1}, \quad Var(\bar{X}_2) = \frac{\sigma_2^2}{n_2}.$$

Since $\bar{X}_1$ and $\bar{X}_2$ are independent,

$$Var(\bar{X}_1 - \bar{X}_2) =$$

which leads to

$$SE(\bar{X}_1 - \bar{X}_2) =$$

Two-sample t-test

Consider a hypothesis test on a difference in two means $\mu_1 - \mu_2$.

$$H_0 : \mu_1 - \mu_2 = \delta_0, \quad H_A : \mu_1 - \mu_2 \neq \delta_0$$

Test statistic for a difference in means

The test statistic and its null distribution are

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{SE(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \overset{H_0}{\sim} t(\psi),$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means of groups 1 and 2, S_1^2 and S_2^2 are the sample variances, and n_1 and n_2 are the sample sizes.

Note: The Satterthwaite's df: $\psi = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$

Back to the ESC example

We want to determine whether ESC help improve heart function.

- Using one-sided test, hypotheses are:

$$H_0 : \mu_{esc} - \mu_{control} = 0$$
$$H_A :$$

Computing the Test Statistic

- Sample statistics:

$$n_{esc} = 9, \quad \bar{x}_{esc} = 3.50, \quad s_{esc} = 5.17$$
$$n_{control} = 9, \quad \bar{x}_{control} = -4.33, \quad s_{control} = 2.76$$

- Compute the observed T-score:

$$t =$$

- Compute the Satterthwaite's degrees of freedom:

$$\psi = \frac{(5.17^2/9 + 2.76^2/9)^2}{(5.17^2/9)^2/(9-1) + (2.76^2/9)^2/(9-1)} = 12.22$$

Finding the P-value

P-value of the one-sided test is

```
> pt(4.02, df = 12.22, lower.tail = FALSE)
[1] 0.0008202582
```

- P-value is less than $\alpha = 0.05$, so we (reject / don't reject) H_0.
- The data (provide / do not provide) convincing evidence that ESCs help improve heart function following a heart attack.

t-test using Satterthwaite's degrees of freedom

```
> t.test(esc, ctrl, alternative = "greater", mu = 0)

 Welch Two Sample t-test

data:  esc and ctrl
t = 4.0073, df = 12.225, p-value = 0.0008388
alternative hypothesis: true difference
                        in means is greater than 0
95 percent confidence interval:
 4.35469     Inf
sample estimates:
mean of x mean of y
 3.500000 -4.333333
```

Practice

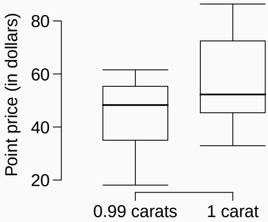
(Exercise 7.24) Diamond prices are known to increase with carat weight—1 carat diamonds are often priced much higher than 0.99 carat diamonds. We compare two random samples of 23 diamonds each (0.99 and 1 carat) using standardized prices, calculated by dividing the price by 100 times the carat weight. Conduct a hypothesis test to determine if 1 carat diamonds are more expensive than 0.99 carat diamonds.

1. State the null and alternative hypotheses.(Use one-sided test.)

2. Find the null distribution of the test statistic.

Practice

	0.99 carats	1 carat
Mean	\$44.51	\$56.81
SD	\$13.32	\$16.13
n	23	23



3. Compute the observed test statistic.

4. Compute the p-value and complete the hypothesis test. (Use $\psi \approx 42.48$.)

Difference of two means: equal variances

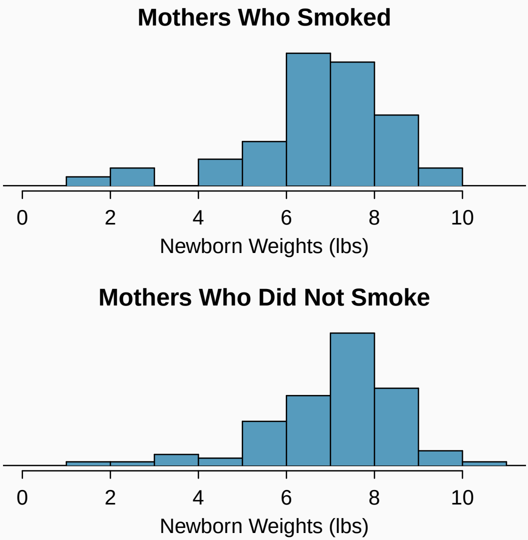
Smoking and birth weights of newborns

- We analyze the difference in birth weights of newborns based on whether the mother smoked during pregnancy.
- Data set: **ncbirths** - a random sample of 150 cases from North Carolina.
- Variables of interest:
    - **weight**: newborn's weight (in pounds)
    - **smoke**: whether the mother smoked (yes/no)

Sample Data

	fage	mage	weeks	weight	sex	smoke
1	NA	13	37	5.00	female	nonsmoker
2	NA	14	36	5.88	female	nonsmoker
3	19	15	41	8.13	male	smoker
⋮	⋮	⋮	⋮	⋮	⋮	
150	45	50	36	9.25	female	nonsmoker

Table 3: Four cases from the **ncbirths** data set.



Careful statistical analysis suggested that the population variances of the two groups(smokers and non-smokers) are the same.

Suppose $\bar{X}_1$ and $\bar{X}_2$ are independent sample means of size n_1 and n_2 from the two distributions with mean μ_1 and μ_2 respectively and with *equal* variance σ^2. From Exercise 3.47,

$$Var(\bar{X}_1) = \frac{\sigma^2}{n_1}, \quad Var(\bar{X}_2) = \frac{\sigma^2}{n_2}.$$

Since $\bar{X}_1$ and $\bar{X}_2$ are independent,

$$Var(\bar{X}_1 - \bar{X}_2) =$$

To estimate σ^2, we use pooled sample variance defined as

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

where S_1^2 and S_2^2 are sample variances of two groups. This leads to

$$SE(\bar{X}_1 - \bar{X}_2) =$$

Two-sample t-test with equal variances

Suppose $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and consider a hypothesis test on $\mu_1 - \mu_2$:

$$H_0 : \mu_1 - \mu_2 = \delta_0, \quad H_A : \mu_1 - \mu_2 \neq \delta_0$$

Test statistic for a difference in means with equal variances

The test statistic and its null distribution are

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{SE(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_p \sqrt{1/n_1 + 1/n_2}} \overset{H_0}{\sim} t(df),$$

where the pooled sample variance is defined as

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

$\bar{X}_1$ and $\bar{X}_2$ are the sample means of groups 1 and 2, S_1^2 and S_2^2 are the sample variances, n_1 and n_2 are the sample sizes, and $df = n_1 + n_2 - 2$.

Remarks on the pooled sample variance

Note: Derivation of pooled sampled variance: If $X_{11}, X_{12}, \cdots, X_{1n_1}$ are independent observations from a normal distribution with mean μ_1 and variance σ^2 and $X_{21}, X_{22}, \cdots, X_{2n_2}$ are from a normal distribution with mean μ_2 and the same variance σ^2, we have

$$Z := \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1),$$

$$V_1 := \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1), \text{ and } V_2 := \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1),$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means , S_1^2 and S_2^2 are the sample variances of groups 1 and 2. V_1 and V_2 are independent, so the additivity of χ^2 distribution gives

$$V := V_1 + V_2 = \frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2).$$

It follows that

$$T := \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} \times \left(\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \frac{1}{n_1 + n_2 - 2}\right)^{-1/2}$$

$$= \frac{Z}{\sqrt{V/(n_1 + n_2 - 2)}} \sim t(n_1 + n_2 - 2)$$

Back to the example

We want to determine whether newborns whose mothers did not smoke weigh **more** than newborns whose mothers smoked under the assumption of equal variance ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).
Using one-sided test, hypotheses are:

$H_0 : \mu_N - \mu_S = 0$ (No difference in average birth weight)

H_A :

Computing the Test Statistic

- Sample statistics:

$$n_N = 100, \quad \bar{x}_N = 7.18, \quad s_N = 1.60$$
$$n_S = 50, \quad \bar{x}_S = 6.78, \quad s_S = 1.43$$

- Compute pooled variance:

$$s_p^2 = \frac{(n_N - 1)s_N^2 + (n_S - 1)s_S^2}{n_N + n_S - 2} =$$

- Compute the observed T-score:

$$t =$$

Rejection Region Approach

Since $H_A : \mu_N - \mu_S > 0$, large values of t provide evidence against H_0. We reject H_0 when $t > c$ for some constant c. At significance level $\alpha = 0.05$, we choose c such that

$$P(T > c) = 0.05, \quad T \sim \qquad\qquad \implies \quad c =$$

- Rejection region:
- Observed test statistic: $t = 1.48$, so we (reject / don't reject) H_0.
- There is (sufficient / insufficient) evidence to conclude that babies born to mothers who smoke have lower birth weight on average.

```
> t.test(nonsmoker, smoker, alternative = "greater",
+       mu = 0, var.equal = T)

 Two Sample t-test

data:  nonsmoker and smoker
t = 1.4824, df = 148, p-value = 0.07050
alternative hypothesis: true difference
                        in means is greater than 0
95 percent confidence interval:
 -0.046407       Inf
sample estimates:
mean of x mean of y
   7.1795    6.7790
```

Practice

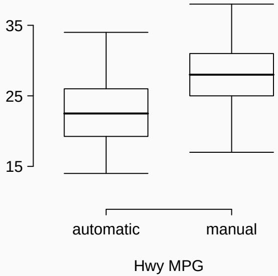
(Exercise 7.30*) The US Environmental Protection Agency (EPA) collects fuel economy data annually. Below are summary statistics on highway fuel efficiency (MPG) from random samples of cars with automatic and manual transmissions. Conduct a hypothesis test to determine if there is a difference in average highway mileage between the two transmission types. **Assume that the variances of the two types MPG are equal.**

1. State the null and alternative hypotheses.

2. Find the null distribution of the test statistic.

Practice

	Hwy MPG	
	Automatic	Manual
Mean	22.92	27.88
SD	5.29	5.01
n	26	26



3. Compute the observed test statistic.

4. Compute the p-value and complete the hypothesis test.

Power calculations for a difference of means

Example: Blood Pressure (BP)

A pharmaceutical company is testing a new drug for lowering blood pressure. Patients taking a standard medication are randomly assigned to:

- Control group: continues with the current medication.
- Treatment group: receives the new drug.

The researchers want to test whether the new drug performs **differently** from the standard medication.

Example: BP — Setup and Hypotheses

Previously published studies suggest the population standard deviation of blood pressure is *known* to be $\sigma = 12$ mmHg in both groups, and the test is conducted at significance level $\alpha = 0.05$.

- Random samples of $n_{trmt} = n_{ctrl} = 100$ patients per group.
- Hypotheses:

$$H_0 : \mu_{trmt} - \mu_{ctrl} = 0, \quad H_A :$$

Computing the Test Statistic

Under H_0, since $\bar{X}_{trmt}$ and $\bar{X}_{ctrl}$ are independent sample means from normal populations with $\mu_{trmt} = \mu_{ctrl}$ and *known* variances $\sigma^2_{trmt} = \sigma^2_{ctrl} = 12^2$, the test statistic is the difference of sample means

$$\bar{X}_{trmt} - \bar{X}_{ctrl} \overset{H_0}{\sim} N\left(0, \frac{\sigma^2_{trmt}}{n_{trmt}} + \frac{\sigma^2_{ctrl}}{n_{ctrl}}\right), \quad \text{that is,}$$

$$\bar{X}_{trmt} - \bar{X}_{ctrl} \overset{H_0}{\sim}$$

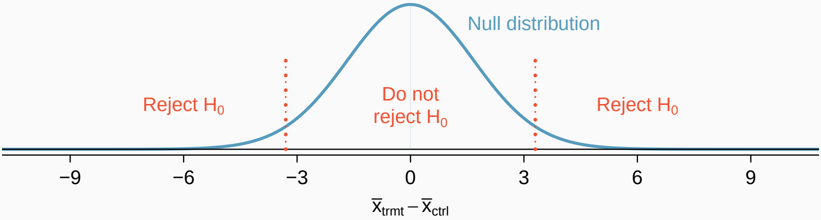
- Observed test statistic: With $\bar{x}_{trmt} = 130.5$ and $\bar{x}_{ctrl} = 134$,

$$\bar{x}_{trmt} - \bar{x}_{ctrl} =$$

Rejection Region in Terms of $\bar{X}_{trmt} - \bar{X}_{ctrl}$

Since $H_A : \mu_{trmt} - \mu_{ctrl} \neq 0$, large $|\bar{x}_{trmt} - \bar{x}_{ctrl}|$ provides evidence against H_0. We reject H_0 when $|\bar{x}_{trmt} - \bar{x}_{ctrl}| > c$ for some $c > 0$. At $\alpha = 0.05$, choosing c such that $P(|\bar{X}_{trmt} - \bar{X}_{ctrl}| > c) = \alpha$ under H_0 gives

$$c = z_{0.025} \times SE = 1.96 \times 1.70 \approx 3.33 \text{ mmHg}.$$



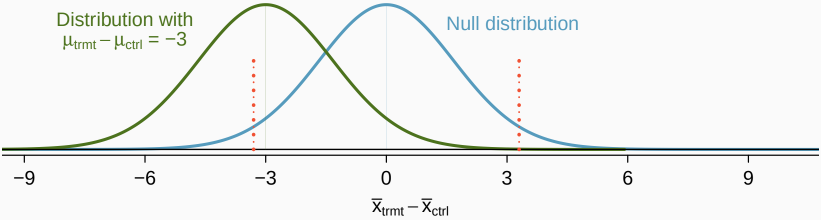
- Rejection region: $|\bar{x}_{trmt} - \bar{x}_{ctrl}| >$
- Observed: $|\bar{x}_{trmt} - \bar{x}_{ctrl}| = 3.5$ so we

Sampling Distribution under H_A

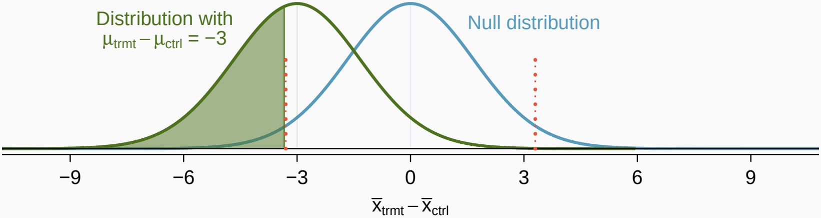
Suppose H_A is true and the true mean difference is $\mu_{trmt} - \mu_{ctrl} = -3$. Then

$$\bar{X}_{trmt} - \bar{X}_{ctrl} \overset{H_A}{\sim} N(-3, 1.70^2).$$

Compared with the null, the sampling distribution **shifts left by** 3, while the standard deviation is **unchanged**.



Computing the Power

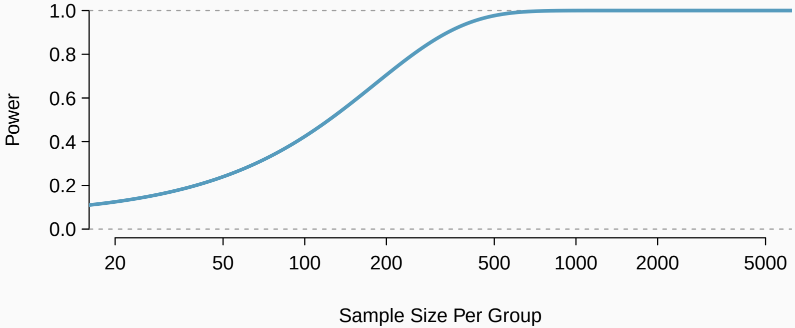


The *power* is the probability of rejecting H_0 under H_A ($\mu_{trmt} - \mu_{ctrl} = -3$):

$$\begin{aligned} \text{Power} &= P\left(\left|\bar{X}_{trmt} - \bar{X}_{ctrl}\right| > 3.33 \mid \mu_{trmt} - \mu_{ctrl} = -3\right) \\ &= P\left(Z < \frac{-3.33-(-3)}{1.70}\right) + P\left(Z > \frac{3.33-(-3)}{1.70}\right) \\ &= P(Z < -0.20) + P(Z > 3.72) \;\approx\; 0.42. \end{aligned}$$

So the test has **about a** 42% **chance** of detecting a true mean difference of −3 mmHg.

As power increases with a larger sample size, we may calculate the required sample size for a desired level of power.



Practice

D Electronics' mobile phones use batteries from Company A. Company B claims their batteries have a longer lifespan. To test this at $\alpha = 0.05$, D Electronics samples $n_A = n_B = 16$ batteries from each company. Let $X_{A1}, \cdots, X_{A16} \sim N(\mu_A, 64)$ and $X_{B1}, \cdots, X_{B16} \sim N(\mu_B, 80)$ denote the lifespans of batteries from Companies A and B, respectively.

1. State the null and alternative hypotheses(Use **one-sided** test).

2. Find the null distribution of the test statistic.

Practice

D Electronics' mobile phones use batteries from Company A. Company B claims their batteries have a longer lifespan. To test this at $\alpha = 0.05$, D Electronics samples $n_A = n_B = 16$ batteries from each company. Let $X_{A1}, \cdots, X_{A16} \sim N(\mu_A, 64)$ and $X_{B1}, \cdots, X_{B16} \sim N(\mu_B, 80)$ denote the lifespans of batteries from Companies A and B, respectively.

3. Find the rejection region.

4. Given that $\bar{x}_A = 60$, $\bar{x}_B = 66$, complete the hypothesis test.

Practice

D Electronics' mobile phones use batteries from Company A. Company B claims their batteries have a longer lifespan. To test this at $\alpha = 0.05$, D Electronics samples $n_A = n_B = 16$ batteries from each company. Let $X_{A1}, \cdots, X_{A16} \sim N(\mu_A, 64)$ and $X_{B1}, \cdots, X_{B16} \sim N(\mu_B, 80)$ denote the lifespans of batteries from Companies A and B, respectively.

5. Find the power of this test assuming $\mu_A - \mu_B = -6$.

Summary of Chapter 7

	Null hypothesis	Test statistic	df
One-sample t-test	$H_0 : \mu = \mu_0$	$T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(df)$	$n - 1$
One-sample Z-test (with known variance)	$H_0 : \mu = \mu_0$	$\bar{X} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$	
Paired t-test	$H_0 : \mu_{diff} = \delta_0$	$T = \dfrac{\bar{X}_{diff} - \delta_0}{S_{diff}/\sqrt{n_{diff}}} \sim t(df)$	$n_{diff} - 1$
Two-sample t-test	$H_0 : \mu_1 - \mu_2 = \delta_0$	$T = \dfrac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t(df)$	ψ
Two-sample t-test (with equal variances)	$H_0 : \mu_1 - \mu_2 = \delta_0$	$T = \dfrac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(df)$	$n_1 + n_2 - 2$
Two-sample Z-test (with known variances)	$H_0 : \mu_1 - \mu_2 = \delta_0$	$\bar{X}_1 - \bar{X}_2 \sim N\left(\delta_0, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$	