Hypothesis testing

Lecture 4

Quantitative Methods for Political Economy
7SSPN107

Recap: estimator and estimate

An estimator is a function that uses sample data to estimate a population parameter.

For example, we can use the sample mean $\bar{H} = \frac{1}{n}\sum_{i=1}^{n} H_i$ to estimate the population mean height, μ_H.

An estimate is the numerical value of the estimator for a particular sample.

Hypothetical sample: $\bar{H} = 164.0$ cm.

Recap: properties of estimators

Unbiasedness: the expected value of the estimator equals the population parameter:

$$E(\widehat{\mu}_H) = \mu_H.$$

Consistency: as the sample size ***n*** grows, the probability that the estimator is close to the true parameter approaches one.

Efficiency: among unbiased estimators, an efficient estimator is the one with the smallest variance.

Four measures of spread

σ_H	**Population standard deviation** Spread of individual heights in the population.
s_H	**Sample standard deviation** Spread of observed heights.
$\text{SD}(\bar{H}) = \frac{\sigma_H}{\sqrt{n}}$	**Standard deviation of the sample mean** Spread of sample means across repeated samples.
$\text{SE}(\bar{H}) = \frac{s_H}{\sqrt{n}}$	**Standard error of the sample mean** Estimates the standard deviation of the sample mean.

Recap: normal approximation and confidence intervals

For samples with large n, the sample mean is approximately normally distributed:

$$\frac{\bar{H} - \mu_H}{\mathrm{SE}(\bar{H})} \approx N(0, 1).$$

The corresponding 95% confidence interval is

$$\bar{H} \pm 1.96\,\mathrm{SE}(\bar{H}).$$

Across repeated samples, about 95% of intervals made this way contain μ_H.

Today: is a particular proposed population mean compatible with the sample evidence?

3.2

Hypothesis tests concerning the population mean

Hypothesis tests

Sometimes we are interested in a population mean, but observe only a sample mean.

Suppose we want to ask UK voters: “On a scale from 0 (strongly dislike) to 10 (strongly like), how much do you like or dislike the Labour Party leader?”

Hypothesis tests

Sometimes we are interested in a population mean, but observe only a sample mean.

Suppose we want to ask UK voters: “On a scale from 0 (strongly dislike) to 10 (strongly like), how much do you like or dislike the Labour Party leader?”

We might want to test whether the population mean rating is 5:

$$H_0 : \mu_Y = 5 \quad \text{null hypothesis},$$
$$H_1 : \mu_Y \neq 5 \quad \text{two-sided alternative}.$$

The p-value

Suppose that in a sample of $n = 100$ voters, the mean rating is $\bar{Y}^{act} = 5.3$.

Clearly, 5.3 is not 5, but does this suggest that the population mean is not truly 5 or does the discrepancy merely reflect sampling variation?

The p-value

Suppose that in a sample of $n = 100$ voters, the mean rating is $\bar{Y}^{act} = 5.3$.

Clearly, 5.3 is not 5, but does this suggest that the population mean is not truly 5 or does the discrepancy merely reflect sampling variation?

The p-value is the probability of drawing a sample statistic at least as adverse to the null hypothesis as the one actually computed in the sample, assuming the null hypothesis is correct.

$$
\begin{aligned}
p &= \Pr_{H_0}\left(|\bar{Y} - \mu_{Y,0}| \geq |\bar{Y}^{act} - \mu_{Y,0}|\right) \\
&= \Pr_{\mu_Y = 5}\left(|\bar{Y} - 5| \geq |5.3 - 5|\right).
\end{aligned}
$$

When the population standard deviation is known

To compute the p-value, we need to know how the sample mean is distributed under H_0.

Suppose the population standard deviation is known: $\sigma_Y = 2$. The standard deviation of the sample mean is then equal to $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$.

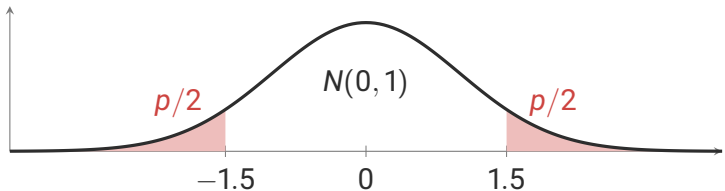
We can then compute the p-value using the fact that the sample mean is approximately normally distributed:

$$p \approx 2\Phi\left(-\left|\frac{\bar{Y}^{act} - \mu_{Y,0}}{\sigma_{\bar{Y}}}\right|\right) = 2\Phi\left(-\left|\frac{5.3 - 5}{0.2}\right|\right) = 2\Phi(-1.5) \approx 0.1336.$$

If H_0 were true, we would expect to find a sample mean at least this far from 5 in about 13.36% of samples.

A two-sided p-value counts both tails

Graphically, the p-value corresponds to the shaded areas in the graph below: the probability that the sample mean lies at least 1.5 sampling standard deviations from the null mean.



Exercise 1: a p-value for mean ratings

A new sample of 144 voters gives a mean rating of the Labour Party leader of 6.4 out of 10. The population standard deviation σ_Y is 3.

$$H_0 : \mu_Y = 7 \qquad \text{versus} \qquad H_1 : \mu_Y \neq 7.$$

Calculate the two-sided p-value using the normal approximation.

Exercise 1: a p-value for mean ratings

A new sample of 144 voters gives a mean rating of the Labour Party leader of 6.4 out of 10. The population standard deviation σ_Y is 3.

$$H_0 : \mu_Y = 7 \qquad \text{versus} \qquad H_1 : \mu_Y \neq 7.$$

Calculate the two-sided p-value using the normal approximation.

$$\sigma_{\bar{Y}} = \frac{3}{\sqrt{144}} = 0.25, \qquad z^{act} = \frac{6.4 - 7}{0.25} = -2.40.$$

Exercise 1: a p-value for mean ratings

A new sample of 144 voters gives a mean rating of the Labour Party leader of 6.4 out of 10. The population standard deviation σ_Y is 3.

$$H_0 : \mu_Y = 7 \qquad \text{versus} \qquad H_1 : \mu_Y \neq 7.$$

Calculate the two-sided p-value using the normal approximation.

$$\sigma_{\bar{Y}} = \frac{3}{\sqrt{144}} = 0.25, \qquad z^{act} = \frac{6.4 - 7}{0.25} = -2.40.$$

$$p \approx 2\Phi(-|z^{act}|) = 2\Phi(-2.40) \approx 0.0164 \quad (1.64\%).$$

Estimating the unknown spread

In practice, the population standard deviation σ_Y is typically unknown and we need to estimate it.

The sample variance, $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, is an unbiased and consistent estimator of σ_Y^2.

Similarly, the standard error $\text{SE}(\bar{Y}) = \hat{\sigma}_{\bar{Y}} = s_Y/\sqrt{n}$ is an estimator of the standard deviation of the sample mean $\sigma_{\bar{Y}}$.

Suppose that in our sample, $s_Y = 2.1$, so $\text{SE}(\bar{Y}) = 2.1/\sqrt{100} = 0.21$.

The t-statistic

The t-statistic measures the distance of an estimate to the null value measured in standard errors.

$$t = \frac{\bar{Y} - \mu_{Y,0}}{\text{SE}(\bar{Y})}, \qquad t^{act} = \frac{5.3 - 5}{2.1/\sqrt{100}} = \frac{0.3}{0.21} \approx 1.43.$$

Our sample mean is about 1.43 standard errors above the null value of 5.

The t-statistic

The t-statistic measures the distance of an estimate to the null value measured in standard errors.

$$t = \frac{\bar{Y} - \mu_{Y,0}}{\mathrm{SE}(\bar{Y})}, \qquad t^{act} = \frac{5.3 - 5}{2.1/\sqrt{100}} = \frac{0.3}{0.21} \approx 1.43.$$

Our sample mean is about 1.43 standard errors above the null value of 5.

When n is large, s_Y^2 is close to σ_Y^2 with high probability and so the distribution of the t-statistic is approximately normal under the null hypothesis,

$$t \approx N(0,1), \qquad p \approx 2\Phi(-|t^{act}|) = 2\Phi(-0.3/0.21) \approx 0.1531.$$

Exercise 2: a two-sided test of mean earnings

Suppose we are interested in the mean hourly earnings of recent UK university graduates.

We have a sample of $n = 225$ graduates with $\bar{Y}^{act} = £17.50$ per hour and $s_Y = £9$ per hour.

$$H_0 : E(Y) = 16 \qquad \text{versus} \qquad H_1 : E(Y) \neq 16.$$

Calculate the standard error, t-statistic and two-sided p-value.

Exercise 2: a two-sided test of mean earnings

Suppose we are interested in the mean hourly earnings of recent UK university graduates.

We have a sample of $n = 225$ graduates with $\bar{Y}^{act} = £17.50$ per hour and $s_Y = £9$ per hour.

$$H_0 : E(Y) = 16 \qquad \text{versus} \qquad H_1 : E(Y) \neq 16.$$

Calculate the standard error, t-statistic and two-sided p-value.

$$\text{SE}(\bar{Y}) = \frac{9}{\sqrt{225}} = 0.60, \qquad t^{act} = \frac{17.50 - 16}{9/\sqrt{225}} = 2.50.$$

Exercise 2: a two-sided test of mean earnings

Suppose we are interested in the mean hourly earnings of recent UK university graduates.

We have a sample of $n = 225$ graduates with $\bar{Y}^{act} = £17.50$ per hour and $s_Y = £9$ per hour.

$$H_0 : E(Y) = 16 \qquad \text{versus} \qquad H_1 : E(Y) \neq 16.$$

Calculate the standard error, t-statistic and two-sided p-value.

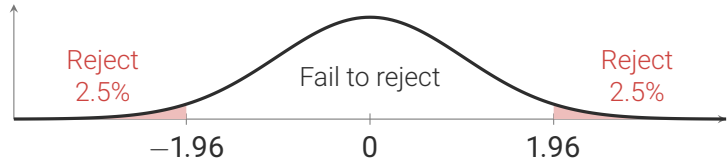
$$\text{SE}(\bar{Y}) = \frac{9}{\sqrt{225}} = 0.60, \qquad t^{act} = \frac{17.50 - 16}{9/\sqrt{225}} = 2.50.$$

$$p \approx 2\Phi(-|t^{act}|) \approx 0.0124 \quad (1.24\%).$$

A prespecified significance level

When testing hypotheses, we can follow a rule that says that we should reject a hypothesis if the p-value is less than a prespecified significance level, such as 5%.

At $\alpha = 0.05$, reject if $p < 0.05$, equivalently if $|t^{act}| > 1.96$.



The points that mark the rejection regions are known as critical values.

In our example, $p \approx 0.1531 > 0.05$ and $|t^{act}| \approx 1.43 < 1.96$: we fail to reject $H_0 : \mu_Y = 5$ at 5%.

Errors, size and power

	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct rejection
Fail to reject H_0	Correct decision	Type II error

Size: the probability of a Type I error: approximately equal to α by construction.

Power: the probability that a test correctly rejects the null hypothesis when a particular alternative is true. A high power means fewer Type II errors.

Exercise 3: interpreting the p-value

In our rating example the p-value was $p \approx 0.1531$.

Suppose that a student says: “There is a 15.31% probability that the population mean rating is 5.”

How would you correct this statement?

Exercise 3: interpreting the p-value

In our rating example the p-value was $p \approx 0.1531$.

Suppose that a student says: “There is a 15.31% probability that the population mean rating is 5.”

How would you correct this statement?

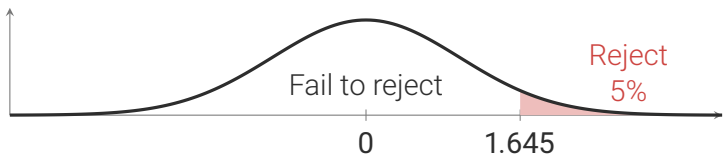
Explanation: The p-value is the probability of obtaining a test statistic at least as extreme as the one observed assuming that H_0 is true. The randomness arises because we draw a specific sample (rather than another sample). In our framework, population parameters are fixed and it makes no sense to talk about them in probability terms.

A one-sided alternative

We might be interested in whether a politician's rating *exceeds* the proposed value:

$$H_0 : E(Y) = \mu_{Y,0}, \qquad H_1 : E(Y) > \mu_{Y,0}.$$

In this case, we only want to reject the null hypothesis if the t-statistic is large and positive.



$p \approx 1 - \Phi(t^{act})$, reject at 5% if $t^{act} > 1.645$.

One-sided and two-sided tests

For the same null $H_0 : E(Y) = \mu_{Y,0}$ and the same t-statistic:

Alternative	Approximate p-value	Reject at 5% if
$E(Y) \neq \mu_{Y,0}$	$2\Phi(-\lvert t^{act} \rvert)$	$\lvert t^{act} \rvert > 1.96$
$E(Y) > \mu_{Y,0}$	$1 - \Phi(t^{act})$	$t^{act} > 1.645$
$E(Y) < \mu_{Y,0}$	$\Phi(t^{act})$	$t^{act} < -1.645$

Exercise 4: one-sided or two-sided?

Suppose we run a test that gives $t^{act} = 1.80$.

1. Find the p-value for $H_1 : E(Y) \neq \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?
2. Find the p-value for $H_1 : E(Y) > \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?

Exercise 4: one-sided or two-sided?

Suppose we run a test that gives $t^{act} = 1.80$.

1. Find the p-value for $H_1 : E(Y) \neq \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?
2. Find the p-value for $H_1 : E(Y) > \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?

$$p_{\text{two-sided}} \approx 2(1 - 0.96407) = 0.07186.$$

Fail to reject at 5%.

Exercise 4: one-sided or two-sided?

Suppose we run a test that gives $t^{act} = 1.80$.

1. Find the p-value for $H_1 : E(Y) \neq \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?
2. Find the p-value for $H_1 : E(Y) > \mu_{Y,0}$. Do you reject at $\alpha = 0.05$?

$$p_{\text{two-sided}} \approx 2(1 - 0.96407) = 0.07186.$$

Fail to reject at 5%.

$$p_{\text{one-sided}} \approx 1 - 0.96407 = 0.03593.$$

Reject at 5%.

3.4

Comparing means from different populations

Estimating a difference between population means

We may be interested in the question of whether the means from two different populations are the same.

Do men and women like the Labour Party leader equally?

$$H_0 : \mu_m - \mu_w = 0 \qquad \text{versus} \qquad H_1 : \mu_m - \mu_w \neq 0.$$

To test this hypothesis, we need to think about the distribution of $(\bar{Y}_m - \bar{Y}_w)$.

The distribution of the difference of means

If $\bar{Y}_m$ and $\bar{Y}_w$ are constructed from different independent random samples, the variance of the difference is simply given by:

$$\text{Var}(\bar{Y}_m - \bar{Y}_w) = \text{Var}(\bar{Y}_m) + \text{Var}(\bar{Y}_w) = \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w}.$$

If both $\bar{Y}_m$ and $\bar{Y}_w$ are approximately normal, then since a linear combination of two independent normal random variables is itself normally distributed, the difference $(\bar{Y}_m - \bar{Y}_w)$ is also approximately normally distributed :

$$\bar{Y}_m - \bar{Y}_w \approx N\left(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w}\right).$$

The standard error of the difference

In practice, σ_m and σ_w are usually unknown and need to be estimated.

The standard error of the difference $(\bar{Y}_m - \bar{Y}_w)$ is then given by

$$\text{SE}(\bar{Y}_m - \bar{Y}_w) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}.$$

Note that the two populations are allowed to have different variances and different sample sizes.

Testing a hypothesised difference

The t-statistic is constructed in the same way as before. Let
$H_0 : \mu_m - \mu_w = d_0$

$$t = \frac{(\bar{Y}_m - \bar{Y}_w) - d_0}{\mathrm{SE}(\bar{Y}_m - \bar{Y}_w)}.$$

For two large independent samples, under H_0 this is approximately $N(0, 1)$.

A confidence interval for the population difference

We construct the 95% confidence interval in the same way as before:

$$\text{estimate} \ \pm \ 1.96 \times \text{standard error}.$$

Here, the estimate is the difference between sample means. For $d = \mu_m - \mu_w$:

$$(\bar{Y}_m - \bar{Y}_w) \pm 1.96 \, \text{SE}(\bar{Y}_m - \bar{Y}_w).$$

Exercise 5: comparing ratings among men and women

The British Election Study asked respondents in 2023 how much they liked Labour Party leader Keir Starmer (0-10):

Group	Mean	Sample SD	n
Men (m)	3.99	3.02	12,937
Women (w)	4.10	2.88	15,106

Test $H_0 : \mu_m - \mu_w = 0$ against a two-sided alternative at 5%.

Exercise 5: comparing ratings among men and women

The British Election Study asked respondents in 2023 how much they liked Labour Party leader Keir Starmer (0-10):

Group	Mean	Sample SD	n
Men (m)	3.99	3.02	12,937
Women (w)	4.10	2.88	15,106

Test $H_0 : \mu_m - \mu_w = 0$ against a two-sided alternative at 5%.

$$\bar{Y}_m - \bar{Y}_w \approx 3.99 - 4.10 = -0.11, \quad \text{SE} \approx \sqrt{\frac{3.02^2}{12937} + \frac{2.88^2}{15106}} \approx 0.0354.$$

Exercise 5: comparing ratings among men and women

The British Election Study asked respondents in 2023 how much they liked Labour Party leader Keir Starmer (0-10):

Group	Mean	Sample SD	n
Men (m)	3.99	3.02	12,937
Women (w)	4.10	2.88	15,106

Test $H_0 : \mu_m - \mu_w = 0$ against a two-sided alternative at 5%.

$$\bar{Y}_m - \bar{Y}_w \approx 3.99 - 4.10 = -0.11, \quad \text{SE} \approx \sqrt{\frac{3.02^2}{12937} + \frac{2.88^2}{15106}} \approx 0.0354.$$

$$t^{act} \approx \frac{-0.11}{0.0354} \approx -3.11, \quad |t^{act}| > 1.96 : \quad \text{reject } H_0.$$

Exercise 6: a confidence interval for the difference

Using the rounded BES summaries, $\bar{Y}_m - \bar{Y}_w \approx -0.11$ points and SE ≈ 0.0354 points, under our sampling assumption.

1. Construct the 95% confidence interval for $d = \mu_m - \mu_w$.

2. Would a two-sided 5% test reject $d_0 = 0$? What about $d_0 = -0.1$ points?

Exercise 6: a confidence interval for the difference

Using the rounded BES summaries, $\bar{Y}_m - \bar{Y}_w \approx -0.11$ points and SE ≈ 0.0354 points, under our sampling assumption.

1. Construct the 95% confidence interval for $d = \mu_m - \mu_w$.

$$-0.11 \pm 1.96(0.0354) \approx [-0.18, -0.04] \text{ rating points.}$$

2. Would a two-sided 5% test reject $d_0 = 0$? What about $d_0 = -0.1$ points?

Exercise 6: a confidence interval for the difference

Using the rounded BES summaries, $\bar{Y}_m - \bar{Y}_w \approx -0.11$ points and SE ≈ 0.0354 points, under our sampling assumption.

1. Construct the 95% confidence interval for $d = \mu_m - \mu_w$.

$$-0.11 \pm 1.96(0.0354) \approx [-0.18, -0.04] \text{ rating points.}$$

2. Would a two-sided 5% test reject $d_0 = 0$? What about $d_0 = -0.1$ points?
   - Zero is outside the interval: **reject** $d_0 = 0$.
   - -0.1 is inside: **fail to reject** $d_0 = -0.1$.