Lecture 14

Part 3 Estimation and Hypothesis Test

Hypothesis Test

2. Test Statistic

- The test statistic is a sample statistic that we use as a criterion to determine whether or not to reject H_0.
- It is usually based on an estimator of the population parameter that we are testing.
- Bowling example:
  - $\bar{X} = 80$, i.e., sample mean bowling score from 10 games.

3. Decision Rule

- In order to make our decision about the hypotheses,we need to know the sampling distribution of the test statistic under H_0, called the **null distribution**.
- Remember that we always start by assuming H_0 is true.
- If the observed value of the test statistic is extreme (i.e., very unlikely to occur under the null distribution), then that is evidence against H_0.

Rejection Region

- Whether or not an observed test statistic is extreme is determined by the *rejection region* or *p-value* (more on *p*-values later).
- The **rejection region** is a range of values such that, if the test statistic falls within this range, we reject H_0.
- Bowling example:
  - The rejection region might be any sample mean less than 120, i.e., $\bar{X} < 120$.

Critical Values

- Related to the rejection region are the **critical values**, which are the values which represent the boundaries of the rejection region.
- That is, values of the test statistic *more extreme* than the critical values define the rejection region.
- Bowling example:
  - The critical value was $c = 120$ and any sample mean less than c lies in the rejection region.

4. Conclusion

- Final step of the hypothesis test.
- If the observed test statistic falls in the rejection region, we reject H_0 and conclude that H_1 is true.
- If the observed test statistic does not fall in the rejection region, we fail to reject H_0 and conclude that H_0 is true.
- Note that we do not "accept H_0".

Type I and Type II Errors

		Truth	
		H_0 is true	H_0 is false
Decision	Reject H_0	Type I error $P(\text{Type I error}) = \alpha$	Correct decision
	Fail to reject H_0	Correct decision	Type II error $P(\text{Type II error}) = \beta$

- **Type I error**: Rejecting H_0 when it is actually true.
- **Type II error**: Failing to reject H_0 when it is actually not true.

Type I and Type II Errors

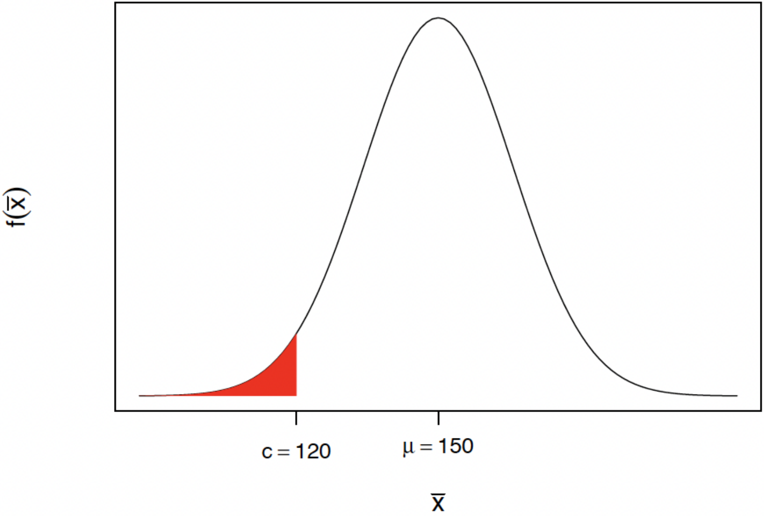
- For any hypothesis test, we would like both errors to be small.
- However, trying to make one small often causes the other to be large.
- Type I error considered more serious than type II error.
- $\alpha = P$ (Type I error) is also called the **significance level** of the test.

Type I Error

- The value of α is usually fixed beforehand and we try to keep it small.
- The value of α (together with H_1) determines the rejection region.
- The smaller the value of α, the more sure we can be of our decision if we end up rejecting H_0.

Type I Error

- For the bowling example, the shaded area is equal to α.



Summary

1. Hypotheses:
   - Establish H_0 and H_1.
   - Assume H_0 is true.
2. Test statistic:
   - Obtain a sample and calculate a test statistic.
3. Decision rule:
   - Determine the rejection region of the null distribution (based on H_1 and α).
4. Conclusion:
   - If the observed test statistic is extreme, i.e., falls in the rejection region, reject H_0.
   - If not, fail to reject H_0.

Hypothesis Test for μ when σ^2 is Known

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu (\neq, <, >) \mu_0$$

- Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Hypothesis Test for μ when σ^2 is Known

- Decision rule:
  - Under H_0, Z has a $N(0,1)$ distribution.
  - At significance level α, rejection regions are:
    - $Z > z_{\frac{\alpha}{2}}$ or $Z < -z_{\frac{\alpha}{2}}$ $(H_1 : \mu \neq \mu_0)$.
    - $Z < -z_{\alpha}$ $(H_1 : \mu < \mu_0)$.
    - $Z > z_{\alpha}$ $(H_1 : \mu > \mu_0)$.
  - NB: z_α is the value which cuts off an area of α in the upper tail of the $N(0,1)$ distribution.
- Conclusion:
  - If Z falls in the rejection region, reject H_0, otherwise, fail to reject H_0.

Solve the Bowling Example

- Your lecturer claims to have a bowling average of 150 or higher.
- You play 10 games with him, and he scores an average of 80.
- Suppose you know that the standard deviation for bowling scores is 50.
- Given a 5% significance level ($\alpha = 0.05$), do you reject your lecturer's claim?

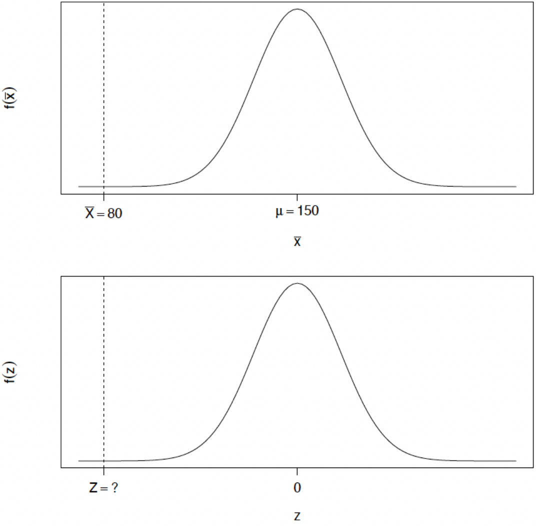
1. Hypotheses

- Set up our hypotheses:

$$H_0 : \mu = 150$$
$$H_1 : \mu < 150$$

- Assume that H_0 is true.
- Let's draw the null distribution.

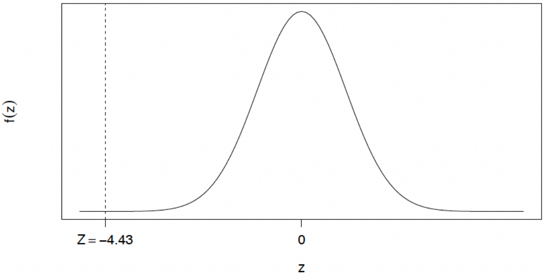
Null Distribution



2. Test Statistic

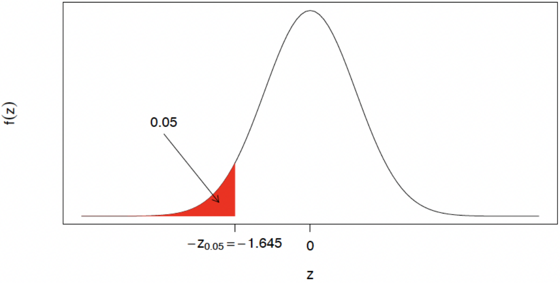
- Standardize $\bar{X} = 80$:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{80 - 150}{\frac{50}{\sqrt{10}}} = -4.43$$



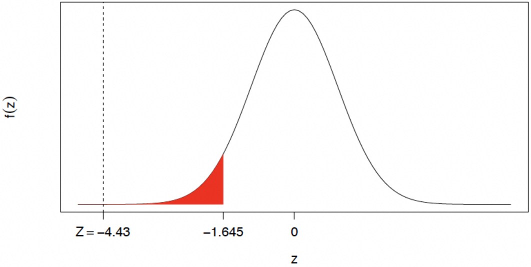
3. Decision Rule

- Rejection region:
  - Since $\alpha = 0.05$ and $H_1 : \mu < 150$, find critical value that cuts off 5% in the left tail of the $N(0,1)$ distribution.
  - From the z-tables, we know $P(Z < -1.645) = 0.05$.
  - Rejection region is $Z < -z_{0.05} = -1.645$.



4. Conclusion

- Since $-4.43 < -1.645$, we reject H_0 at the 5% significance level, and we conclude that my true bowling average is less than 150.



One-Tailed Test

- In a one-tailed test, we only care about one extreme tail of the distribution of the test statistic and the alternative hypothesis usually consists of a "$>$" or "$<$" sign.
- For example:
  - $H_1 : \mu < 150$.
  - We only care whether μ is smaller than 150.
  - We reject H_0 if the test statistic is too small.

Two-Tailed Test

- In a two-tailed test, we care about both extreme tails of the distribution of the test statistic and the alternative hypothesis usually consists of a "$\neq$" sign.
- For example:
  - $H_1 : \mu \neq 150.$
  - Now we care whether μ is smaller or larger than 150.
  - We reject H_0 if the test statistic is either too small or too large.

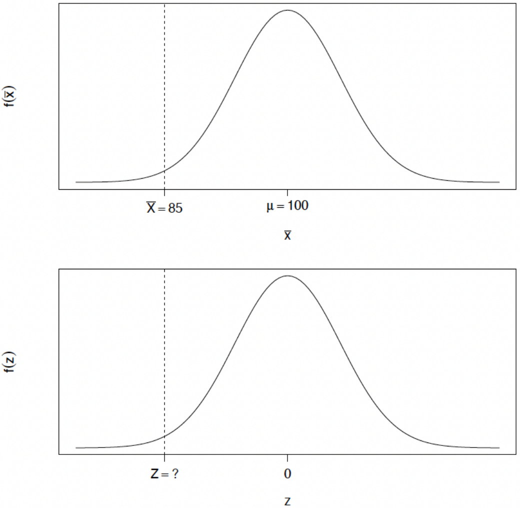
Example of a Two-Tailed Test

- Test the following hypothesis:

$$H_0 : \mu = 100$$
$$H_1 : \mu \neq 100$$

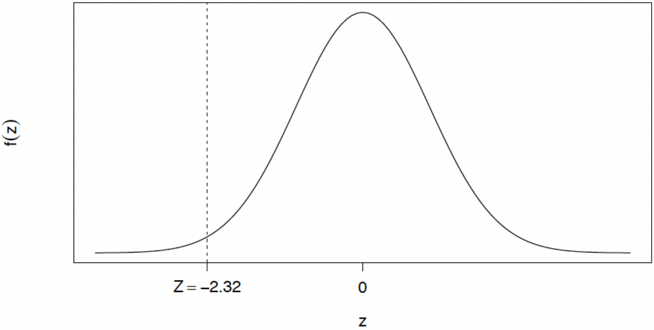
when $\bar{X} = 85$, $\sigma = 25$, $n = 15$ and $\alpha = 0.05$.

Null Distribution



2. Test Statistic

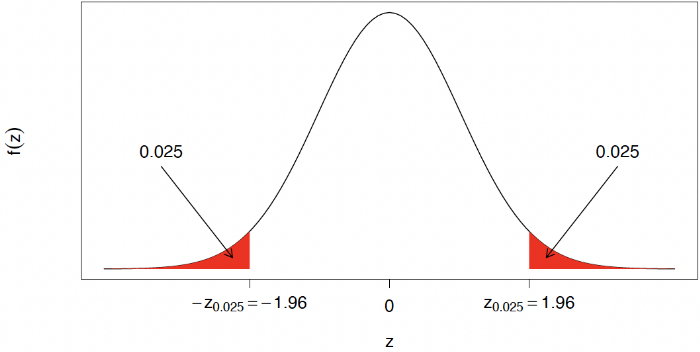
$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{85 - 100}{\frac{25}{\sqrt{15}}} = -2.32$$



3. Decision Rule

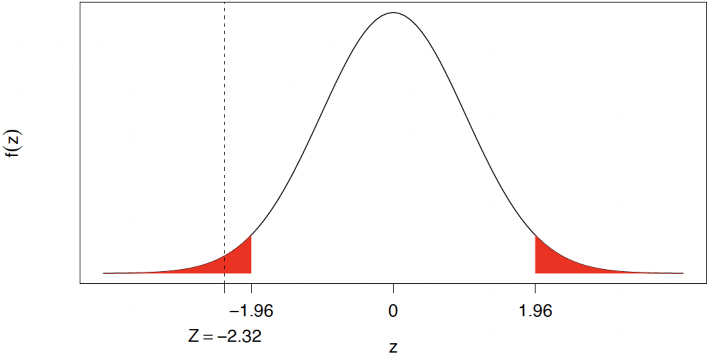
- There are two rejection regions for two-tailed tests!
- The critical values for the rejection regions are the values that cut off $100\left(\frac{\alpha}{2}\right)\%$ in each tail of the null distribution. Why?
- From the z-tables, we know that $P(Z > 1.96) = 0.025$.
- Therefore, by symmetry, the rejection regions are $Z < -1.96$ and $Z > 1.96$.

3. Decision Rule



4. Conclusion

- Since $-2.32 < -1.96$, we reject H_0.



Tips on Setting Up Hypotheses

- How we set up our hypotheses is very important.
- We need to choose H_0 and H_1 in a way that lets us decide between two distinct situations.
- The decision then has to help us answer the original question being asked.

Tips on Setting Up Hypotheses

- Remember that H_0 usually only includes an "$=$" sign.
- If the question has a claim that includes a "$<$", "$>$" or "$\neq$" sign, set that to be H_1.
- If the question has a claim that includes a "$\leq$" or "$\geq$" sign, let that claim be represented by H_0 and set the opposite claim to be H_1.

Example 1

- A school teacher believes they are an excellent teacher and that the average exam score of students in their class is greater than 85. Test the school teacher's claim.
- Let μ be the population mean exam score.

Example 1

- The question includes the claim $\mu > 85$. So we should set:

$$H_0 : \mu = 85$$
$$H_1 : \mu > 85$$

- If we reject H_0, we conclude the average exam score is greater than 85 and if we fail to reject H_0, we conclude the average exam score is not greater than 85.

Example 2

- A policeman believes that the average speed of motorists on a street that has a speed limit of 40 mph is at least 51 mph. Test the policeman's claim.
- Let μ denote the population mean speed.

Example 2

- The question includes the claim $\mu \geq 51$, and the opposite claim is $\mu < 51$. So we should set:

$$H_0 : \mu = 51$$
$$H_1 : \mu < 51$$

- If we reject H_0, we conclude that the mean speed is less than 51 and if we fail to reject H_0, we conclude that the mean speed is at least 51.

Example 3

- An office manager believes that the average number of sick days his staff take in a year is at most 13 days. Test the office manager's claim.
- Let μ denote the population mean number of sick days taken.

Example 3

- The question includes the claim $\mu \leq 13$, and the opposite claim is $\mu > 13$. So we should set:

$$H_0 : \mu = 13$$
$$H_1 : \mu > 13$$

- If we reject H_0, we conclude that μ is greater than 13 and if we fail to reject H_0, we conclude that μ is at most 13.

p-value

- There are two ways of conducting hypothesis tests:
  1. Using rejection regions.
  2. Using p-values.
- A ***p*-value** is the probability of observing a test statistic even more extreme than the one calculated from your sample, assuming that H_0 is true.

Example

- Suppose we are testing the following hypotheses:

$$H_0 : \mu = 140$$
$$H_1 : \mu > 140$$

- Further, suppose our standardized Z-statistic turns out to be $Z = 1.10$.
- What is the p-value?

Example

- The p-value is the probability of observing something more extreme than $Z = 1.10$.
- Because of our one-tailed alternative hypothesis, more extreme means greater than 1.10.
- The p-value:

$$\begin{aligned} P(Z > 1.10) &= 1 - P(Z < 1.10) \\ &= 1 - 0.8643 \\ &= 0.1357 \end{aligned}$$