Lecture 22
Significance Level, Critical Value, and Making Decisions

Review From Monday

- **Statistical significance** – a statistical result is one that is decidedly not due to “ordinary variation” in the data (i.e., not due to chance or not a coincidence).

A hypothesis test is an evidence-based decision between two hypotheses:

  - **Null hypothesis** - H_0 - the hypothesis of “no effect” – it usually states that the population parameter is equal to some value
  - **Alternative hypothesis** - H_A - the hypothesis of “effect” – it usually states that the population parameter falls in some alternative range of values

- **The test statistic** - measures the distance between the point estimate of the parameter and the hypothesized value of the parameter.

- **The P-value** – the probability of observing a value of the test statistic as or more extreme than the observed value of the test statistic when H_0 is true.

Review: The five steps of a hypothesis test

- 1. Assumptions
- 2. State the null and alternative hypotheses
- 3. Compute the test statistic
- 4. Compute the *p*-value
- 5. Make a decision – reject or fail to reject the null
  - Is the result statistically significant?

How do we decide how significant a result needs to be to reject the null?

Stating Hypotheses

Example 1:

Does taking garlic supplements repel ticks? A study published in the Journal of the American Medical Association used a cross-over experimental design to determine if daily consumption of garlic would reduce tick bites. A total of 66 Swedish military personnel took 1200mg of garlic daily during one period, and a placebo during the other period. 37 subjects reported fewer tick bites during the period they took garlic supplements. Is there evidence that more than 50% of Swedish soldiers experienced fewer tick bites while taking the garlic supplement?

Example 2:

In the city of Las Vegas, by law cars are required to stop for pedestrians trying to cross the street. In practice many drivers do not stop. A study conducted by the University of Nevada, Las Vegas conducted trials by having pedestrians wait at the curb of a midblock crosswalk and observe whether the next car approaching would stop. Out of a total of 126 of such trials, in 76 trials the first car approaching stopped for the pedestrian. Is there evidence that fewer than 75% of drivers actually stop?

Example 3:

A study investigated if dogs could be trained to detect whether a person has lung cancer or breast cancer by smelling the subject's breath. The researchers trained five ordinary household dogs to distinguish, by scent alone, exhaled breath samples of 55 lung and 31 breast cancer patients from those of 83 healthy controls. Once trained, the dogs' ability to distinguish cancer patients from controls was tested using breath samples from 119 subjects not previously encountered (dogs and handlers were blinded from the treatments to avoid bias). The dogs correctly distinguished between cancer patients and controls for 101 out of 119 subjects. Is there any evidence that the probability of the dogs correctly selecting patients differs from random guessing?

Deciding between One-Sided and Two-Sided Tests

- A one-sided test implies we have some information about the direction of the effect (i.e the effect is lesser or greater)
  - A two-sided test allows the effect to be in either direction.

Things to consider:

1. The context of the question (“more than”, “less than”, “different”)
2. Most research articles use two-sided tests – this represents an objective approach to conducting research
3. Two-sided tests are essentially confidence intervals which are also two-sided

Technical Points of Significance Tests

- The null hypothesis usually has an equal sign (i.e $H_0\colon p_0 = 0.5$). The alternative hypothesis does not.

- You should never pick the alternative hypothesis based on looking at the data, this is a form of bias

- The hypotheses **always** refer to population parameters, not sample statistics:
  - Never state the null hypothesis as $H_0\colon \hat{p} = 0.5$

Significance Level α and Critical Value

The **significance level**, also known as alpha or α, is a measure of the strength of the evidence that must be present in your sample before you will reject the null hypothesis and conclude that the effect is statistically significant. The researcher determines the significance level before conducting the experiment.

The **critical value** of a significance test is a point on the distribution of the test statistic under the null hypothesis that defines a set of values that call for rejecting the null hypothesis.

The critical value determines the boundary of the **rejection region** – the area in the in tail (or tails) of the distribution of the test statistic for which a result is deemed significant

The Decision Rule

- **Significance level** α – of a significance test defines the boundary for the region in the distribution of test statistic for which we would consider the P-value “small enough” to reject the null hypothesis

  - by convention, we usually set alpha to be 5% ($\alpha = 0.05$)

  The decision rule:

  if $p\text{value} < \alpha$, reject H_0 (result is statistically significant)

  else, do not reject H_0 (result is NOT statistically significant)

- **Rejection region** - is a set of values (interval) for the test statistic for which the null hypothesis is rejected. It depends on the alternative hypothesis and the significance level

Ex.)

$$H_A: p > p_0 \quad \textbf{rejection region} \;\; Z_{1-\alpha} < Z < \infty$$

$$H_A: p < p_0 \quad \textbf{rejection region} \;\; -\infty < Z < Z_{\alpha}$$

$$H_A: p \neq p_0 \quad \textbf{rejection region} \;\; -\infty < Z < Z_{\alpha/2} \text{ or } Z_{1-\alpha/2} < Z < \infty$$

Parameter	Null Hypothesis	Alternative Hypothesis	Name	Test Statistic	Critical Value and Rejection Region	Pvalue
p	$H_0: p_0 = c$	$H_A: p > p_0$	"upper tail"	$Z_{obs} = \frac{\hat{p}-p}{SE(p)}$ $\sim N(01,)$	Z_α $-\infty < Z < Z_\alpha$	$P(Z \geq Z_{obs} \mid H_0\ True)$
p	$H_0: p_0 = c$	$H_A: p < p_0$	"lower tail"		$Z_{1-\alpha}$ $Z_{1-\alpha} < Z < \infty$	$P(Z \leq Z_{obs} \mid H_0\ True)$
p	$H_0: p_0 = c$	$H_A: p \neq p_0$	"two-tailed"		$\lvert Z_{1-\alpha/2} \rvert$ $\lvert Z_{1-\alpha} \rvert < \lvert Z \rvert < \infty$	$2 \times P(\lvert Z \rvert \geq \lvert Z_{obs} \rvert \mid H_0\ True)$
μ	$H_0: \mu_0 = c$	$H_A: \mu > \mu_0$	"upper tail"	$t_{obs} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ $\sim t(n-1)$	$t_{n-1,\,\alpha}$ $-\infty < t < t_\alpha$	$P(t \geq t_{obs} \mid H_0\ True)$
μ	$H_0: \mu_0 = c$	$H_A: \mu < \mu_0$	"lower tail"		$t_{n-1,1-\alpha}$ $t_{n-1,1-\alpha} < t < \infty$	$P(t \leq t_{obs} \mid H_0\ True)$
μ	$H_0: \mu_0 = c$	$H_A: \mu \neq \mu_0$	"two-tailed"		$\lvert t_{n-1,\,1-\alpha/2} \rvert$ $\lvert t_{n-1,1-\alpha/2} \rvert < \lvert t \rvert < \infty$	$2 \times P(\lvert t \rvert \geq \lvert t_{obs} \rvert \mid H_0\ True)$

Example 1:

Does taking garlic supplements repel ticks? A study published in the Journal of the American Medical Association used a cross-over experimental design to determine if daily consumption of garlic would reduce tick bites. A total of 66 Swedish military personnel took 1200mg of garlic daily during one period, and a placebo during the other period. 37 subjects reported fewer tick bites during the period they took garlic supplements. Is there evidence that more than 50% of Swedish soldiers experienced fewer tick bites while taking the garlic supplement?

Conduct a significance test at the 5% significance level and report your conclusion.

Example 2:

In the city of Las Vegas, by law cars are required to stop for pedestrians trying to cross the street. In practice many drivers do not stop. A study conducted by the University of Nevada, Las Vegas conducted trials by having pedestrians wait at the curb of a midblock crosswalk and observe whether the next car approaching would stop. Out of a total of 126 of such trials, in 76 trials the first car approaching stopped for the pedestrian. Is there evidence that fewer than 75% of drivers actually stop?

Conduct a significance test at the 5% significance level and report your conclusion.

Example 3:

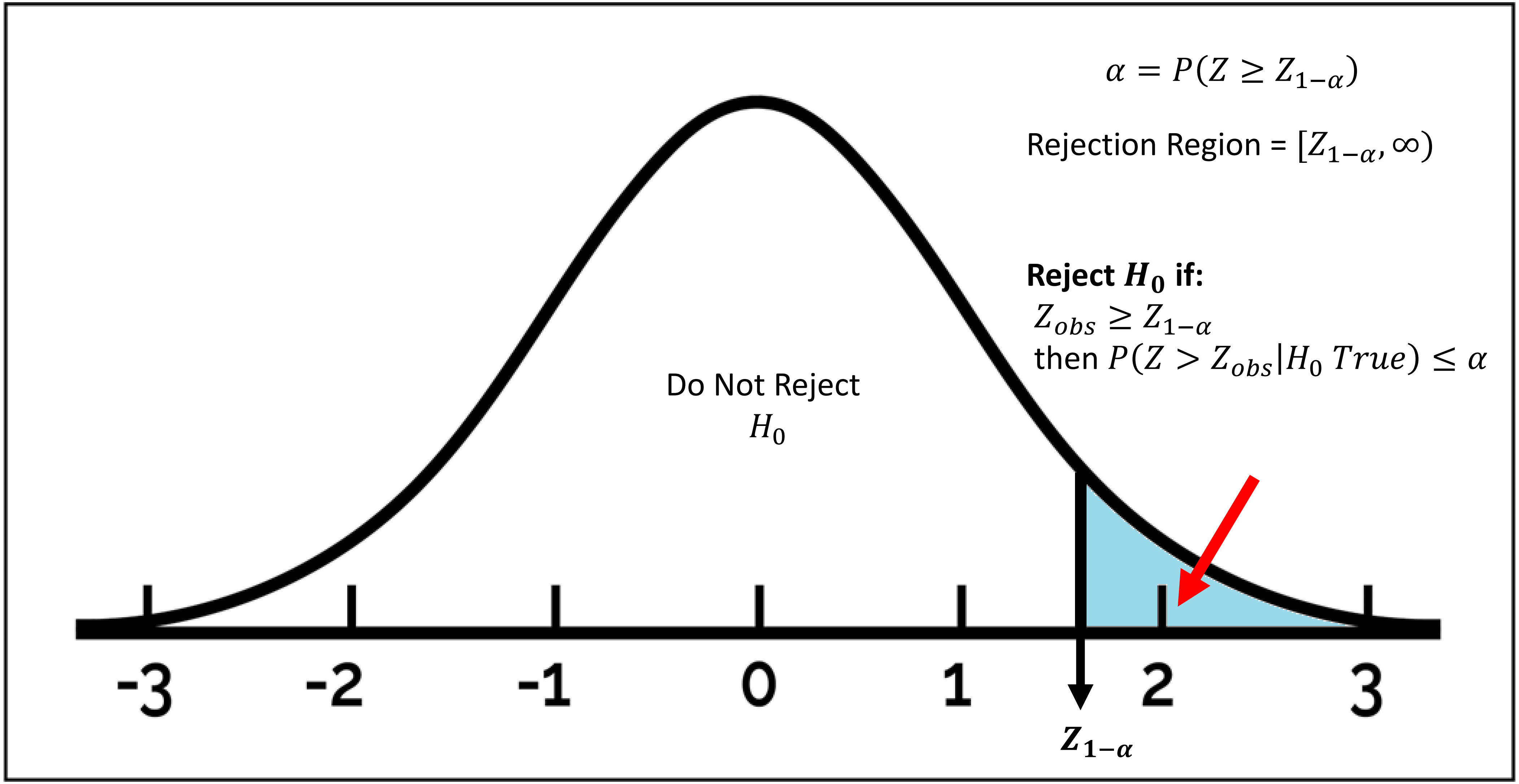
A study investigated if dogs could be trained to detect whether a person has lung cancer or breast cancer by smelling the subject's breath. The researchers trained five ordinary household dogs to distinguish, by scent alone, exhaled breath samples of 55 lung and 31 breast cancer patients from those of 83 healthy controls. Once trained, the dogs' ability to distinguish cancer patients from controls was tested using breath samples from 119 subjects not previously encountered (dogs and handlers were blinded from the treatments to avoid bias). The dogs correctly distinguished between cancer patients and controls for 101 out of 119 subjects. Is there any evidence that the probability of the dogs correctly selecting patients differs from random guessing?

Conduct a significance test at the 1% significance level and report your conclusion.

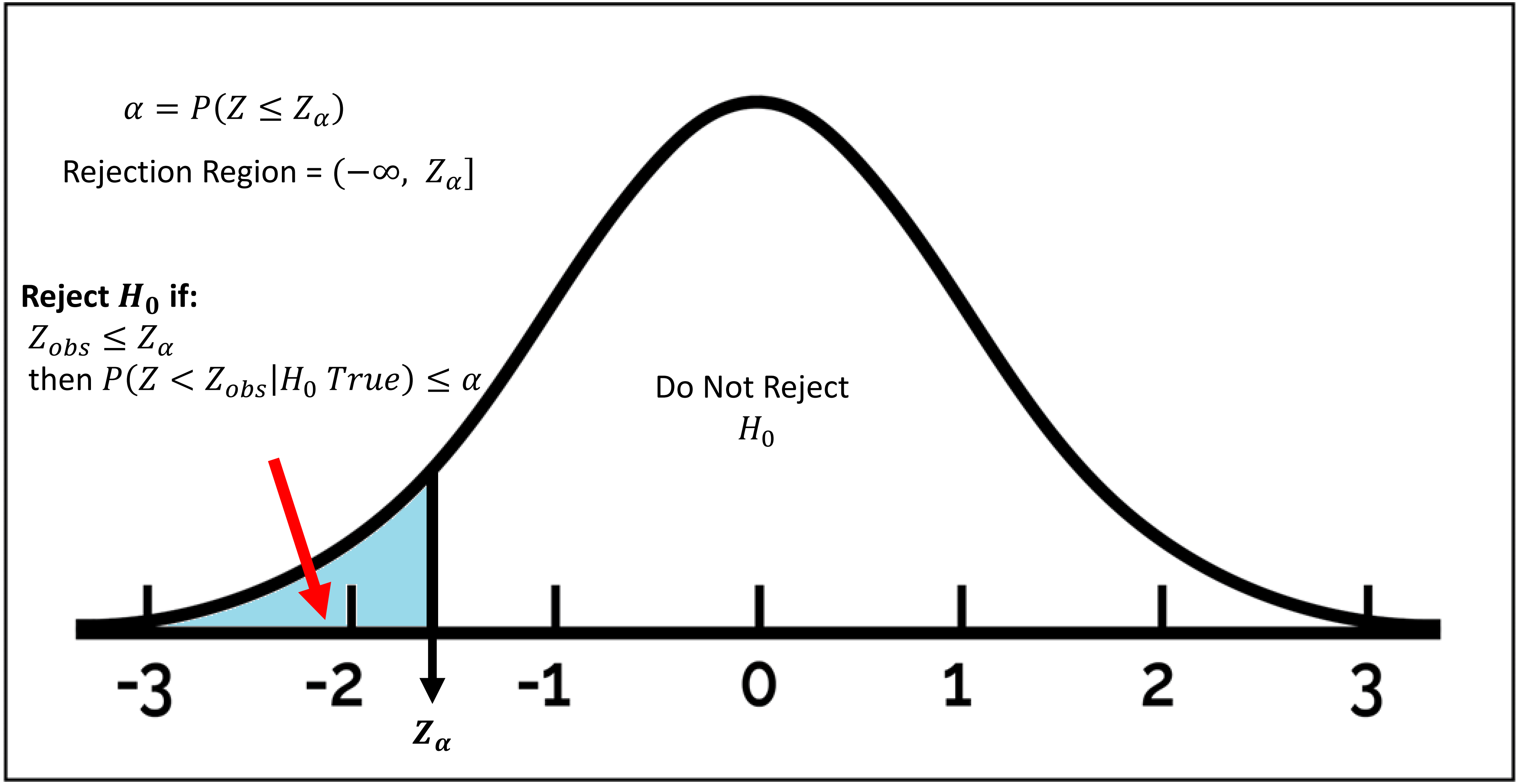
Example 4

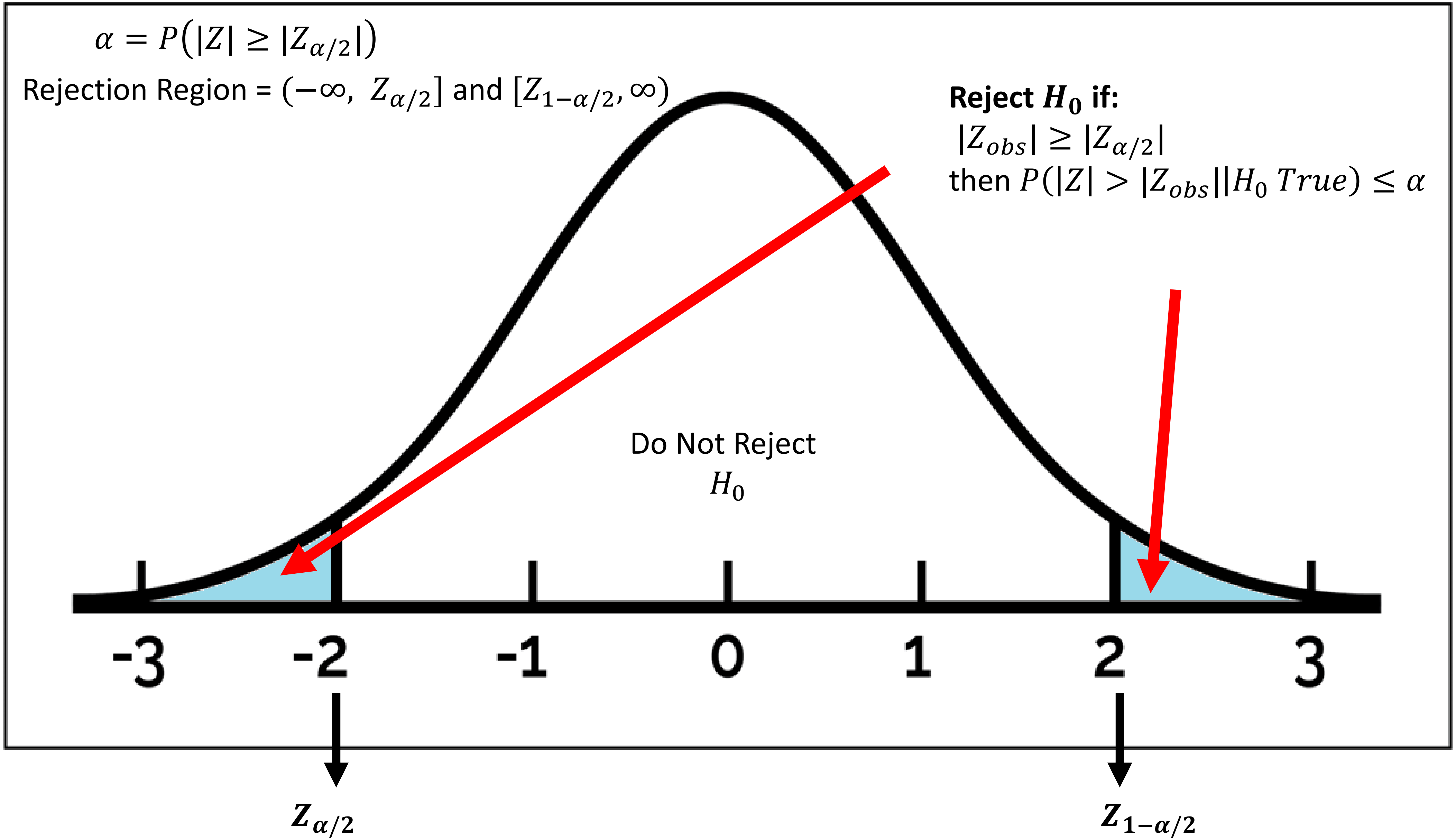
- Using fracking, the U.S has become the largest oil produce in the world. Despite its economic benefits, fracking has become controversial due to its environmental impacts. Survey was conducted to quantity public opinion about fracking. The survey interviewed 1,353 Americans and found that 637 reported being against fracking. The researchers are interested in whether or not there is evidence that most Americans are opposed to fracking.

Conduct a significance test at the $\alpha = 0.05$ significance level to determine if there is evidence for a majority opinion against fracking. Use the five steps for a hypothesis test



$\alpha = P(Z \geq Z_{1-\alpha})$
Rejection Region = $[Z_{1-\alpha}, \infty)$
Reject H_0 if:
$Z_{obs} \geq Z_{1-\alpha}$
then $P(Z > Z_{obs} | H_0\ True) \leq \alpha$
Do Not Reject
H_0
-3
-2
-1
0
1
2
3
$Z_{1-\alpha}$


$\alpha = P(Z \leq Z_\alpha)$
Rejection Region = $(-\infty,\ Z_\alpha]$
Reject H_0 if:
$Z_{obs} \leq Z_\alpha$
then $P(Z < Z_{obs}|H_0\ True) \leq \alpha$
Do Not Reject H_0
-3
-2
-1
0
1
2
3
Z_α



$\alpha = P(|Z| \geq |Z_{\alpha/2}|)$
Rejection Region = $(-\infty, Z_{\alpha/2}]$ and $[Z_{1-\alpha/2}, \infty)$
Reject H_0 if:
$|Z_{obs}| \geq |Z_{\alpha/2}|$
then $P(|Z| > |Z_{obs}|| H_0\ True) \leq \alpha$
Do Not Reject H_0
-3
-2
-1
0
1
2
3
$Z_{\alpha/2}$
$Z_{1-\alpha/2}$

Relationship between significance tests and confidence intervals

Consider a test of the hypotheses $H_0: \mu_0 = 0, \quad H_A: \mu \neq \mu_0$ how can we use a confidence interval to conduct a significance test on these hypotheses?

Decision Rule: Reject H_0 if

$$\bar{x} - t_\alpha \frac{s}{\sqrt{n}} < \mu_0 < \bar{x} + t_\alpha \frac{s}{\sqrt{n}}$$

Is <u>not</u> true

Significance tests are less useful than confidence intervals

- Significance tests have been overemphasized in practice

- A significance test only tells you whether or not a given parameter value in the null hypothesis (such as $\mu_0 = 0$, or $p_0 = 0.5$) is plausible given the data.

- When a P-value is small, it indicates the value specified by the null is not plausible but tells us little else about the possible values of the parameter.

- A confidence interval is more informative because it tells us the entire set of plausible values