Hypothesis Testing for Population Mean and Proportion

STA 032: Gateway to data science Lecture 22

Jingwei Xiong

May 22, 2023

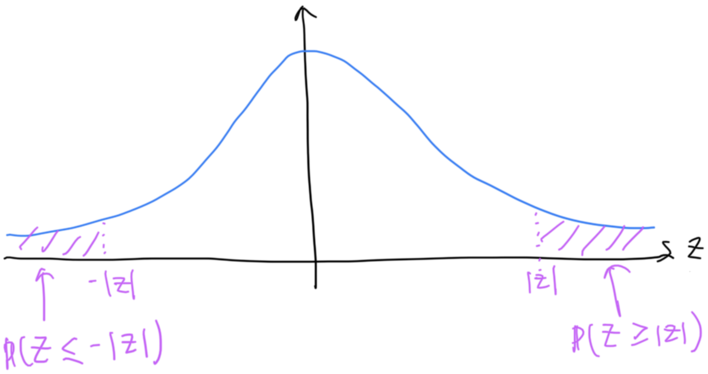
Reminders/announcements

- Midterm 2 on May 24 at 9pm (in person, Wellman 6!)
  - Practice problems from homework
  - Read through the lecture notes
- Book problems: T/F and multiple choice questions on statistic
- R problems: T/F, multiple choice and code sort problems.

Today

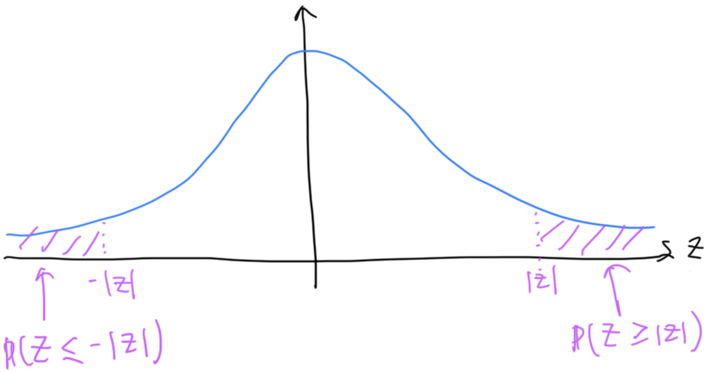
- Hypothesis tests for population mean and proportion
  - p-value approach
  - Critical value approach

More about p-values



- Note that all computations are done assuming that H_0 is true, i.e., to be precise, the decision rule is reject H_0 if the p-value $P(|Z| \geq |z| \mid H_0) = P(Z \geq |z| \text{ or } Z \leq -|z| \mid H_0) < \alpha$
- The blue distribution is the distribution under the null hypothesis
- $P(|Z| \geq |z|) = P(Z \geq |z|) + P(Z \leq -|z|)$ (shaded area)

More about p-values

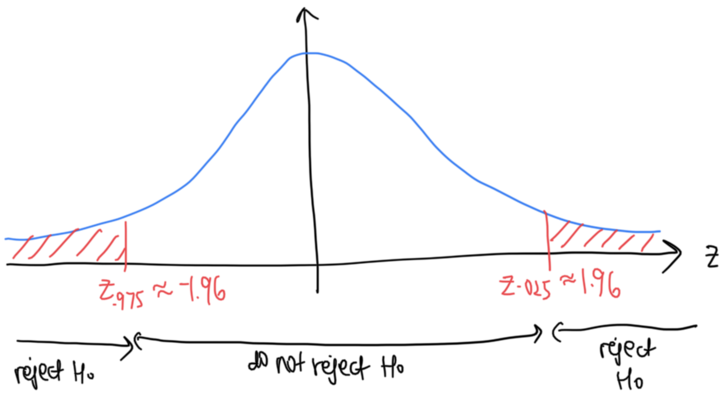


- The value of the **test statistic** is z (value on horizontal axis)
- **p-value** = $P(|Z| \geq |z|)$ under H_0 (for a two-sided test; more details coming). It is the probability of getting a result as extreme as what we got, if H_0 were true.
- Recall the decision rule: reject H_0 if $P(|Z| \geq |z|) = P(Z \geq |z| \text{ or } Z \leq -|z|) < \alpha$. Alternatively, the p-value can be interpreted as the smallest significance level that we would reject H_0.

More about p-values

- The p-value is the probability of getting data like ours or more extreme data if H_0 were true
- Common misinterpretation: "p-value is the probability that H_0 is true". The p-value is calculated *assuming* that H_0 is true. It cannot be used to tell us how likely it is that assumption is correct.
- **Decision rule**: reject H_0 if p-value $< \alpha$
  - We will demonstrate that this produces the required property that P(reject H_0 | H_0 true) $= \alpha$

Critical value approach



- The **rejection region** is $|z| > z_{\alpha/2}$ or $|z| > 1.96$ when $\alpha = .05$. This is a **portion of the x-axis**.
- The boundaries of the rejection region are called **critical values**.
- **Significance level** is the probability over the rejection region, the red area: $P(|Z| > z_{\frac{\alpha}{2}}) = \alpha$

Hypothesis Testing for the Population Mean (σ known)

Say X_i has mean μ and standard deviation σ. The test statistic we will use is $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. By CLT, $Z \approx N(0,1)$ when n large.

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

Under H_0, $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \approx N(0,1)$

Value of test statistic: $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$

Decision rule:

- p-value approach: reject H_0 if $P(|Z| \geq z) = P(Z \geq |z| \text{ or } Z \leq -|z|) < \alpha$
- Critical value approach: reject if $|z| > z_{\alpha/2}$ or $|z| > 1.96$ when $\alpha = .05$

Hypothesis Testing for the Population Mean (σ unknown)

Say X_i has mean μ and standard deviation σ. The test statistic we will use is $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$. $Z \approx N(0,1)$ when n large. (Here notice that σ has been replaced by S)

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

Under H_0, $Z = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \approx N(0,1)$ (Here notice that σ has been replaced by S)

Value of test statistic: $z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ (replace σ by s)

Decision rule:

- p-value approach: reject H_0 if $P(|Z| \geq z) = P(Z \geq |z| \text{ or } Z \leq -|z|) < \alpha$
- Critical value approach: reject if $|z| > z_{\alpha/2}$ or $|z| > 1.96$ when $\alpha = .05$

Hypothesis Testing for the Population Mean (σ unknown, X_i i.i.d normal)

Say X_i are i.i.d normally distributed mean μ and standard deviation σ. The test statistic we will use is $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$. $T \sim t_{n-1}$. (Here notice that σ has been replaced by S)

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

Under H_0, $T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ (Here notice that σ has been replaced by S)

Value of test statistic: $t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$ (replace σ by s)

Decision rule:

- p-value approach: reject H_0 if $P(|T| \geq t) = P(T \geq |t| \text{ or } T \leq -|t|) < \alpha$
- Critical value approach: reject if $|T| > t_{n-1,\alpha/2}$.

```
-qt(0.05/2, 20-1)
```

```
[1] 2.093024
```

Hypothesis Testing for the Population Proportion

Say $X_i \sim$ Bernoulli(p). The test statistic we will use is $Z = \frac{\hat{P}-p}{\sqrt{p(1-p)}/\sqrt{n}}$. By CLT, $Z \approx N(0,1)$ when n large.

H_0: $p = p_0$
H_A: $p \neq p_0$

Under H_0, $Z = \frac{\hat{P}-p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} \approx N(0,1)$ (Here notice that p is replaced by p_0)

Value of test statistic: $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$

Decision rule:

- p-value approach: reject H_0 if $P(|Z| \geq z) = P(Z \geq |z|$ or $Z \leq -|z|) < \alpha$
- Critical value approach: reject if $|z| > z_{\alpha/2}$ or $|z| > 1.96$ when $\alpha = .05$

Example 1: test for population mean, σ known

Assume that the heights of redwood trees in California follow a distribution with standard deviation 25 feet. Let the random variable X_i denote the height of the ith redwood tree.

We guess that the unknown population mean is 230, and would like to test this hypothesis against the alternative that $\mu \neq 230$. We collect data on the heights of 300 randomly sampled redwood trees. Assume the samples are indepedent. We get a sample mean of 220. Construct a hypothesis test at a 5% significance level.

Example 1: test for population mean, σ known (critical value approach)

H_0: $\mu = 230$
H_A: $\mu \neq 230$

Test statistic: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. By CLT, $Z \approx N(0,1)$ when n large.

Under H_0, $Z = \frac{\overline{X}-230}{25/\sqrt{300}} \approx N(0,1)$

Value of test statistic: $z = \frac{220-230}{25/\sqrt{300}} = -6.928203$

The rejection region is $|z| > 1.96$ when $\alpha = .05$

$|z| = 6.928203 > 1.96$. The test statistic is in the rejection region, so we reject H_0 that $\mu = 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is 230 feet.

Example 1: test for population mean, σ known (p-value approach)

(Same set up as last slide)

Value of test statistic: $z = \frac{220-230}{25/\sqrt{300}} = -6.928203$

The p-value is $P(|Z| \geq |z|)$, in this case
$P(|Z| \geq 6.928203) = P(Z \geq 6.928203 \text{ or } Z \leq -6.928203)$

```
2*pnorm(-6.928203)
```

```
[1] 4.262199e-12
```

The p-value is less than .05, so we reject H_0 that $\mu = 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is 230 feet.

Example 2: test for population mean, σ unknown

Assume that the heights of redwood trees in California follow a distribution with unknown mean and standard deviation. Let the random variable X_i denote the height of the ith redwood tree.

We guess that the unknown population mean is 230, and would like to test this hypothesis against the alternative that $\mu \neq 230$. We collect data on the heights of 300 randomly sampled redwood trees. Assume the samples are indepedent. We get a sample mean of 220 and sample standard deviation of 24. Construct a hypothesis test at a 5% significance level.

Example 2: test for population mean, σ unknown

H_0: $\mu = 230$
H_A: $\mu \neq 230$

Test statistic: $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$. $Z \approx N(0,1)$ when n large.

Under H_0, $Z = \frac{\overline{X}-230}{S/\sqrt{300}} \approx N(0,1)$

Value of test statistic: $z = \frac{220-230}{24/\sqrt{300}} = -7.216878$

The rejection region is $|z| > 1.96$ when $\alpha = .05$

$|z| = 7.216878 > 1.96$. The test statistic is in the rejection region, so we reject H_0 that $\mu = 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is 230 feet.

Example 3: test for population proportion

We are interested in the population proportion of likely voters that approve of President Biden. We guess that this is .4 and would like to test this hypothesis against the alternative that it is different from .4. We conduct a random sample of 1500 likely voters, and the proportion among them that approve of President Biden is .3. Construct a hypothesis test at a 5% significance level to determine if our hypothesis is plausible.

Example 3: test for population proportion

Let X_i be a binary random variable denoting whether or not the ith sampled voter approves of President Biden. Now, $X_i \sim$ Bernoulli(p), and by CLT, $Z = \frac{\hat{P}-p}{\sqrt{p(1-p)}/\sqrt{n}} \approx N(0,1)$ when n large.

H_0: $p = .4$
H_A: $p \neq .4$

Under H_0, $Z = \frac{\hat{P}-.4}{\sqrt{.4(1-.4)}/\sqrt{1500}} \approx N(0,1)$

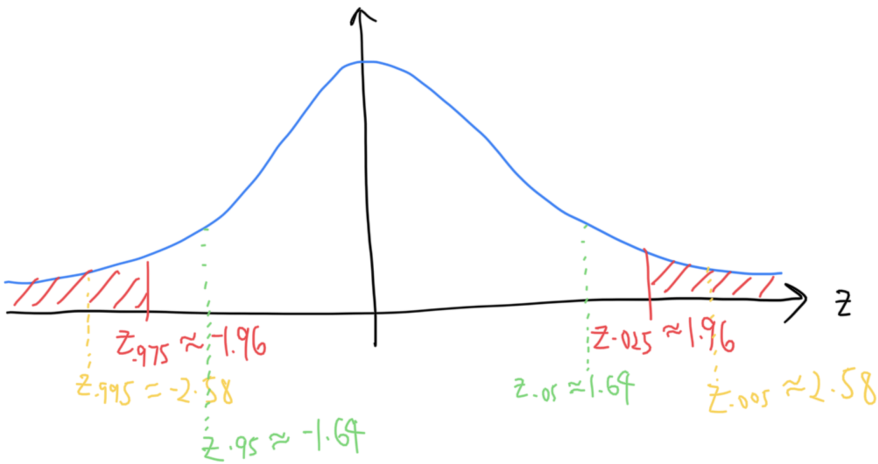
Value of test statistic: $z = \frac{.3-.4}{\sqrt{.4(1-.4)}/\sqrt{1500}} = -7.91$

The rejection region is $|z| > 1.96$ when $\alpha = .05$

$|z| = 7.91 > 1.96$. The test statistic is in the rejection region, so we reject H_0 that $p = .4$. There is sufficient evidence at a 5% level to reject the null hypothesis that the population proportion of likely voters that approve of President Biden is .4.

Different critical values

Recall: The boundaries of the rejection region are called critical values




```r
qnorm(.95) # alpha = .1 (5% in each tail)
```

```
[1] 1.644854
```

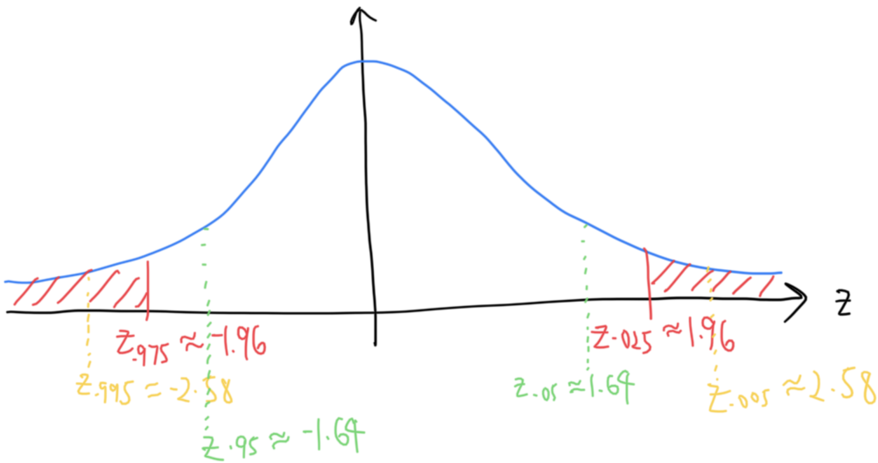
```r
qnorm(.975) # alpha = .05 (2.5% in each tail)
```

```
[1] 1.959964
```

```r
qnorm(.995) # alpha = .01 (.5% in each tail)
```

Rejection regions

Decision rule: Reject H_0 if $|z| > z_{\alpha/2}$



- $|z| > 2.58$ for $\alpha = .01$
- $|z| > 1.96$ for $\alpha = .05$
- $|z| > 1.64$ for $\alpha = .1$

Summary of rejection rules

For two-sided z-tests:

α	**Critical value approach**	**p-value approach**
.01	$\|z\| > z_{\alpha/2} \approx 2.58$	$P(\|Z\| \geq \|z\| \mid H_0) < .01$
.05	$\|z\| > 1.96$	$P(\|Z\| \geq \|z\| \mid H_0) < .05$
.1	$\|z\| > 1.64$	$P(\|Z\| \geq \|z\| \mid H_0) < .1$

Recall: Significance level

- We defined the significance level, α, when discussing confidence intervals:
  - Confidence level = $100(1-\alpha)$%, i.e., a 95% confidence interval will need $\alpha = .05$
  - P(CI contains true parameter) = $1-\alpha$.
- The significance level is also an important ingredient in a hypothesis test
- It defines the tolerable **Type I error**: the probability of rejecting H_0 **when H_0 is actually true**. $\alpha = 0.05$ is often used.
- There is a direct correspondence between hypothesis tests and confidence intervals

Summary

- Hypothesis tests for population mean and proportion
  - p-value approach: reject H_0 if $P(|Z| \geq |z| \mid H_0) = P(Z \geq |z|$ or $Z \leq -|z| \mid H_0) < \alpha$
  - Critical value approach: reject if $|z| > z_{\alpha/2}$
- Test statistics (all approximately standard normal):
  - Population mean when σ known: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$
  - Population mean when σ unknown, n is large: $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$
  - Population mean when σ unknown, normally distributed sample: $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  - Population proportion: $Z = \frac{\hat{P}-p}{\sqrt{p(1-p)}/\sqrt{n}}$

Correspondence between Hypothesis Tests and Confidence Intervals

Correspondence between Hypothesis Tests and Confidence Intervals

- Consider the confidence interval $\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
    - We proved that P(CI contains true parameter) = $1-\alpha$, i.e., $P(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$
- Consider a hypothesis test with $H_0 : \mu = \mu_0$. We want a test with level α. This means we need P(reject H_0 when H_0 true) $= \alpha$.
- Proposal: reject H_0 if $\mu_0 \notin \left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
- For this to be a valid hypothesis test (controls type I error), we need to show that P(reject H_0 when H_0 true) = $.05$ using our proposed rejection rule

Proof

We have $P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

- Proposal: reject H_0 if $\mu_0 \notin \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$
- Want to show: P(reject H_0 when H_0 true) $= \alpha$
- When H_0 is true, $\mu = \mu_0$, so our first line becomes $P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$, i.e., $P(\mu_0 \in \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)) = 1 - \alpha$

Hence,

$$\begin{aligned} P(\text{reject } H_0 \text{ when } H_0 \text{ true}) &= P(\mu_0 \notin \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)) \\ &= 1 - (1 - \alpha) \\ &= \alpha \end{aligned}$$

Correspondence between Hypothesis Tests and Confidence Intervals

Hence our decision rule for an α-level test for $H_0 : \mu = \mu_0$ is reject H_0 if $\mu_0 \notin \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$

In other words, given a sample, we construct a $100(1-\alpha)$% confidence interval for μ.

Given a hypothesis about μ that we wish to test, $\mu = \mu_0$, we simply check if μ_0 falls within the constructed confidence interval. If it does, do not reject our hypothesis. If it doesn't, reject. Some intuition:

- For a hypothesis test, we assume $\mu = \mu_0$. $100(1-\alpha)\%$ of CIs will contain $\mu = \mu_0$, so if we reject if μ_0 is not in the CI, P(reject) = α.
- Confidence intervals can be thought of as representing the range of population means that are "compatible with the data". Looking at a confidence interval can hence help us decide if a proposed value for the population mean is reasonable.

Previous Example

Let $X_1, X_2, \ldots, X_{200}$ be independent $N(\mu, 4)$ random variables. We collect the sample of size 200, and the resulting sample mean, $\overline{x}$, is $\overline{x} = 24$. What is a 95% confidence interval for μ?

We are 95% confident that μ falls within the interval (23.72, 24.28).

Consider the hypothesis that $\mu = 24$. Do we have sufficient evidence to reject this hypothesis at a 5% level?

How about $\mu = 25$?

"Sidedness" of tests

The one sided part will not be tested in midterm 2.

- Typically, hypothesis tests are **two-sided**, meaning that we are looking for any difference from a hypothesized value, whether smaller or larger.
- All the examples that we have seen so far are two-sided tests
- Sometimes, it is appropriate to conduct a **one-sided** test. For example, in the low-dose contraception example, we might only be interested in whether the dose is less than the required level, which would cause breakthrough pregnancies.

Null and Alternative Hypotheses

Consider all the possibilities for tests for μ:

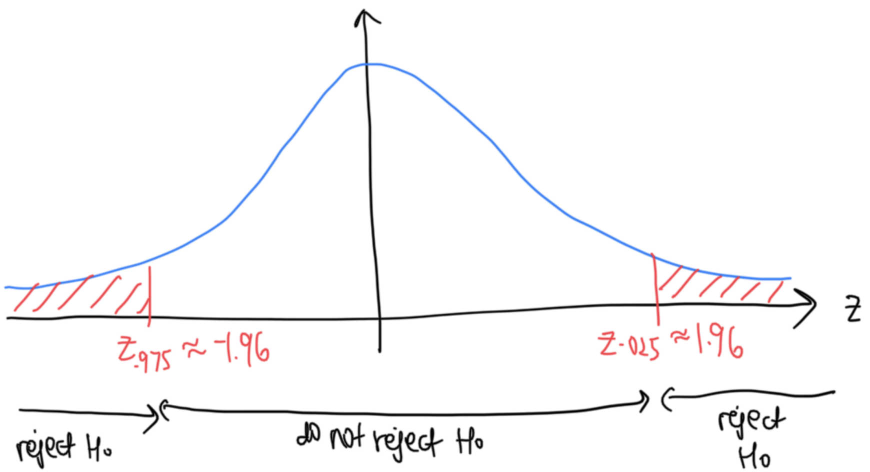
	H_0	H_A
Two-sided	$\mu = \mu_0$	$\mu \neq \mu_0$
One-sided	$\mu \leq \mu_0$	$\mu > \mu_0$
One-sided	$\mu \geq \mu_0$	$\mu < \mu_0$

Similarly for p:

	H_0	H_A
Two-sided	$p = p_0$	$p \neq p_0$
One-sided	$p \leq p_0$	$p > p_0$
One-sided	$p \geq p_0$	$p < p_0$

Recall: rejection regions for two-sided tests

Decision rule: Reject H_0 if $|z| > z_{\alpha/2}$



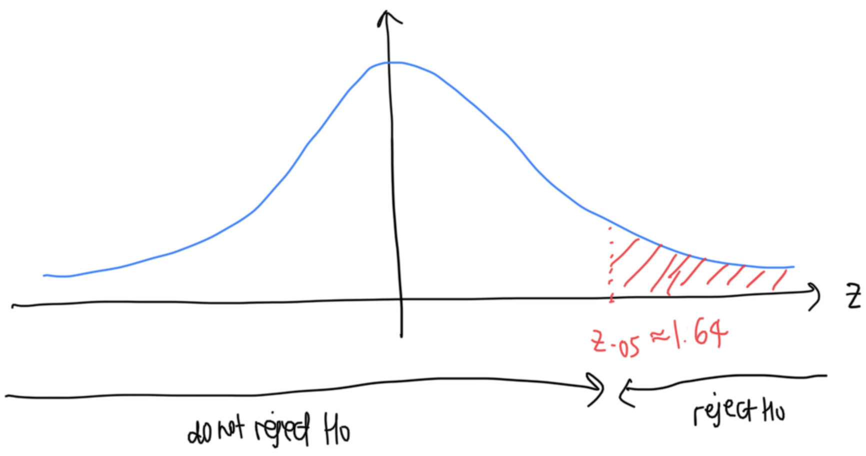
- $|z| > 2.58$ for $\alpha = .01$
- $|z| > 1.96$ for $\alpha = .05$
- $|z| > 1.64$ for $\alpha = .1$

For one-sided tests, we only reject H_0 if the test statistic is either larger than expected (e.g., for a test where $H_A : \mu > \mu_0$), or smaller than expected

One-sided tests: critical value approach

Consider $H_A : \mu > \mu_0$

Decision rule: Reject H_0 if $z > z_\alpha$ or $z > 1.64$ for $\alpha = .05$. The probability over the rejection region is still the significance level α, but all of the probability is now in the right tail.



How do we get the critical value in R?

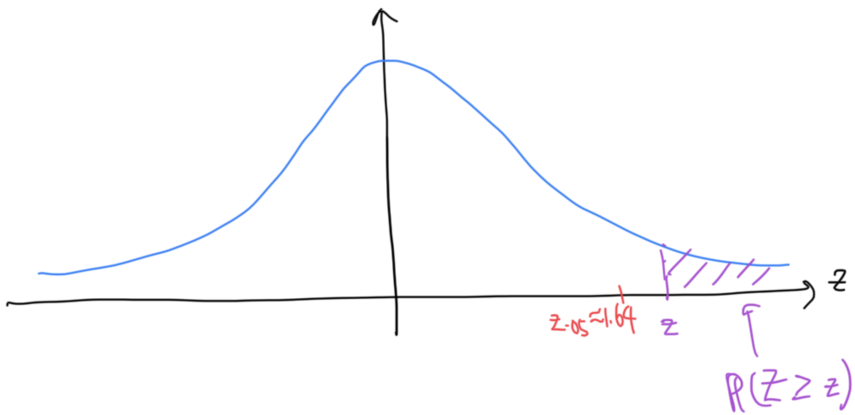
```
qnorm(.95) # alpha = .05
```

```
[1] 1.644854
```

One-sided tests: p-value approach

Consider $H_A : \mu > \mu_0$

Decision rule: Reject H_0 if $P(Z \geq z) < \alpha$



How do we get the p-value in R? `1 - pnorm(z)`, where `z` is the value of the test statistic

One-sided tests: critical value approach

Consider $H_A : \mu < \mu_0$

Decision rule: Reject H_0 if $z < z_{1-\alpha}$ or $z < -1.64$ for $\alpha = .05$

How do we get the critical value in R?

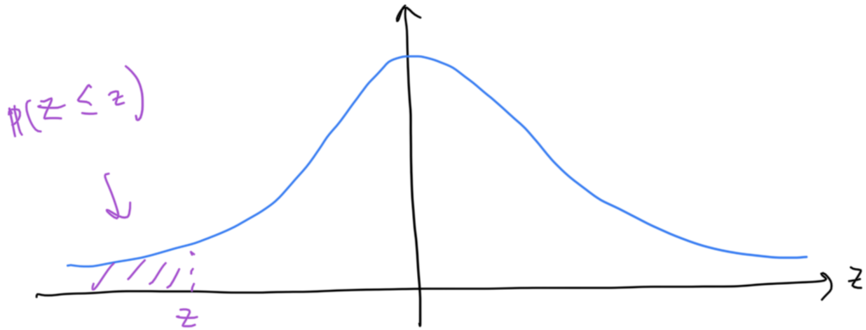
```
qnorm(.05) # alpha = .05
```

```
[1] -1.644854
```

One-sided tests: p-value approach

Consider $H_A : \mu < \mu_0$

Decision rule: Reject H_0 if $P(Z \leq z) < \alpha$



How do we get the p-value in R? `pnorm(z)`, where `z` is the value of the test statistic

Rejection regions and p-values

	H_0	H_A	**Critical value approach**	**p-value approach**
Two-sided	$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$	$P(\lvert Z \rvert \geq \lvert z \rvert \mid H_0) < \alpha$
One-sided	$\mu \leq \mu_0$	$\mu > \mu_0$	$z > z_{\alpha}$	$P(Z \geq z \mid H_0) < \alpha$
One-sided	$\mu \geq \mu_0$	$\mu < \mu_0$	$z < z_{1-\alpha}$	$P(Z \leq z \mid H_0) < \alpha$

Example 1: test for population mean, σ known

Assume that the heights of redwood trees in California follow a distribution with standard deviation 25 feet. Let the random variable X_i denote the height of the ith redwood tree.

Now, we are only interested in tall trees, so we would like to do a test with the alternative that $\mu > 230$. We collect data on the heights of 300 randomly sampled redwood trees. Assume the samples are indepedent. We get a sample mean of 240. Conduct a hypothesis test at a 5% significance level.

Example 1: one-sided z-test for population mean, σ known (critical value approach)

H_0: $\mu \leq 230$
H_A: $\mu > 230$

Test statistic: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. By CLT, $Z \approx N(0,1)$ when n large.

Under H_0, $Z = \frac{\overline{X}-230}{25/\sqrt{300}} \approx N(0,1)$

Value of test statistic: $z = \frac{240-230}{25/\sqrt{300}} = 6.928203$

The rejection region is $z > 1.64$ when $\alpha = .05$

$z = 6.928203 > 1.64$. The test statistic is in the rejection region, so we reject H_0 that $\mu \leq 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is less than or equal to 230 feet.

Example 1: one-sided z-test for population mean, σ known (p-value approach)

(Same set up as last slide)

Value of test statistic: $z = \frac{240-230}{25/\sqrt{300}} = 6.928203$

The p-value is $P(Z \geq z)$, in this case $P(Z \geq 6.928203)$

```
1 - pnorm(6.928203)
```

```
[1] 2.131073e-12
```

The p-value is less than .05, so we reject H_0 that $\mu \leq 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is less than or equal to 230 feet.

Example 2: one-sided z-test for population mean, σ unknown

Assume that the heights of redwood trees in California follow a distribution with unknown mean and standard deviation. Let the random variable X_i denote the height of the ith redwood tree.

Now, we are only interested in tall trees, so we would like to do a test with the alternative that $\mu > 230$. We collect data on the heights of 300 randomly sampled redwood trees. Assume the samples are indepedent. We get a sample mean of 240 and sample standard deviation of 24. Conduct a hypothesis test at a 5% significance level.

Example 2: one-sided z-test for population mean, σ unknown

H_0: $\mu \leq 230$
H_A: $\mu > 230$

Test statistic: $Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$. $Z \approx N(0, 1)$ when n large.

Under H_0, $Z = \frac{\overline{X} - 230}{S/\sqrt{300}} \approx N(0, 1)$

Value of test statistic: $z = \frac{240 - 230}{24/\sqrt{300}} = 7.216878$

The rejection region is $z > 1.64$ when $\alpha = .05$

$z = 7.216878 > 1.64$. The test statistic is in the rejection region, so we reject H_0 that $\mu \leq 230$. There is sufficient evidence at a 5% level to reject the null hypothesis that the mean height of a Californian redwood tree is less than or equal to 230 feet.

Example 3: test for population proportion

We are interested in the population proportion of likely voters that approve of President Biden. We only care about low approval ratings, so our alternative hypothesis is $p < .4$. We conduct a random sample of 1500 likely voters, and the proportion among them that approve of President Biden is .3. Conduct a hypothesis test at a 10% significance level.

Example 3: test for population proportion

Let X_i be a binary random variable denoting whether or not the ith sampled voter approves of President Biden. Now, $X_i \sim$ Bernoulli(p), and by CLT, $Z = \frac{\hat{P}-p}{\sqrt{p(1-p)}/\sqrt{n}} \approx N(0,1)$ when n large.

H_0: $p \geq .4$
H_A: $p < .4$

Under H_0, $Z = \frac{\hat{P}-.4}{\sqrt{.4(1-.4)}/\sqrt{1500}} \approx N(0,1)$

Value of test statistic: $z = \frac{.3-.4}{\sqrt{.4(1-.4)}/\sqrt{1500}} = -7.91$

The rejection region is $z < -1.281552$ when $\alpha = .1$ (we get this using `qnorm(.1)`)

$z = -7.91 < -1.281552$. The test statistic is in the rejection region, so we reject H_0 that $p \geq .4$. There is sufficient evidence at a 10% level to reject the null hypothesis that the population proportion of likely voters that approve of President Biden is greater than or equal to .4.

Summary

- Hypothesis tests for population mean and proportion
  - p-value approach: reject H_0 if $P(|Z| \geq |z| \mid H_0) = P(Z \geq |z|$ or $Z \leq -|z| \mid H_0) < \alpha$
  - Critical value approach: reject if $|z| > z_{\alpha/2}$
  - Using a confidence interval: reject H_0 if $\mu_0 \notin \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$
- One-sided vs. two-sided tests
  - Decision rules using critical values and p-values