STAT 22000 Lecture Slides
Hypothesis Testing About Population Means

Yibi Huang
Department of Statistics
University of Chicago

Outline

- Hypothesis Testing About Population Means (Section 4.3)
- Relationships Between Confidence Intervals and Hypothesis Tests (Section 4.3.2)

Hypothesis Tests about Population Means

Example: Number of College Applications

To know how many colleges students applied to, the dean of a certain university took a random sample of size 106 from their newly admitted students. This sample yielded an average of 9.7 college applications with a standard deviation of 7. College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all freshmen in this university apply to is <u>higher</u> than recommended?

http://www.collegeboard.com/student/apply/the-application/151680.html

Example: Number of College Applications – Hypotheses

- *Population*: all freshmen in this university
- The *parameter of interest* μ is the average number of schools applied to by all freshmen in this university

Example: Number of College Applications – Hypotheses

- *Population*: all freshmen in this university
- The *parameter of interest* μ is the average number of schools applied to by all freshmen in this university
- Two possible explanations of why the sample mean is higher than the recommended 8 schools.
  - The true population mean is higher than 8.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability.

Example: Number of College Applications – Hypotheses

- *Population*: all freshmen in this university
- The *parameter of interest* μ is the average number of schools applied to by all freshmen in this university
- Two possible explanations of why the sample mean is higher than the recommended 8 schools.
  - The true population mean is higher than 8.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability.
- $H_0 : \mu = 8$ (Freshmen in this university have applied 8 schools on average, as recommended)

Example: Number of College Applications – Hypotheses

- *Population*: all freshmen in this university
- The *parameter of interest* μ is the average number of schools applied to by <u>all</u> freshmen in this university
- Two possible explanations of why the sample mean is higher than the recommended 8 schools.
  - The true population mean is higher than 8.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability.
- $H_0 : \mu = 8$ (Freshmen in this university have applied 8 schools on average, as recommended)
- $H_A : \mu > 8$ (Freshmen in this university have applied *over 8* schools on average)

Wrong Ways to State H_0 and H_A

H_0 and H_A are **ALWAYS** stated in terms of population parameters, not sample statistics

Neither

$$\mathrm{H}_0: \ \bar{x} = 8, \quad \mathrm{H}_A: \ \bar{x} > 8$$

nor

H_0 : Freshmen in the sample have applied 8 schools on average
H_A : freshmen in the sample have applied 9.7 schools on average

is correct. The correct statements should be

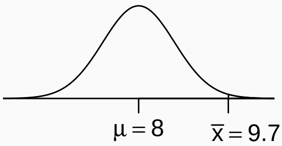
$$\mathrm{H}_0: \ \mu = 8, \quad \mathrm{H}_A: \ \mu > 8$$

Also please **clearly specify what is** μ.
e.g., μ is the average number of colleges freshmen in this university have applied to.

Number of College Applications — Test Statistic

By CLT, under H_0: $\mu = 8$, the sampling distribution of the sample mean is

$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{106}} = 0.68\right)$$



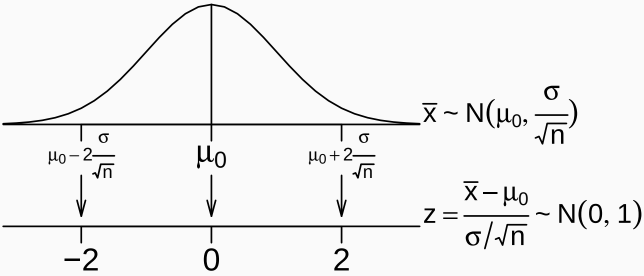
To gauge how unusual the observed sample mean $\bar{x} = 9.7$ is relative to its the hypothesized sampling distribution above, the *test statistic* we used is the z-statistic, which is the z-score of the sample mean relative to the distribution above

$$z\text{-statistic} = \frac{\bar{x} - \mu_0}{SE} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{9.7 - 8}{7/\sqrt{106}} \approx 2.5$$

$$\sim N(0, 1) \quad \text{under } H_0 : \ \mu = \mu_0 = 8$$

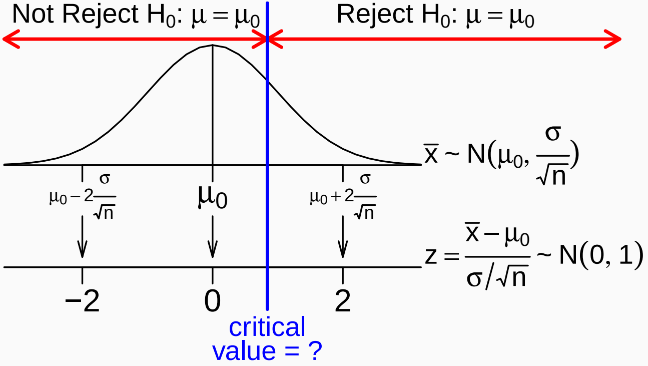
Rejection Region, Critical Value, and Type 1 Error Rate

To test H$_0$: $\mu = \mu_0$ against H$_a$: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H$_a$ and only in such cases should H$_0$ be rejected.



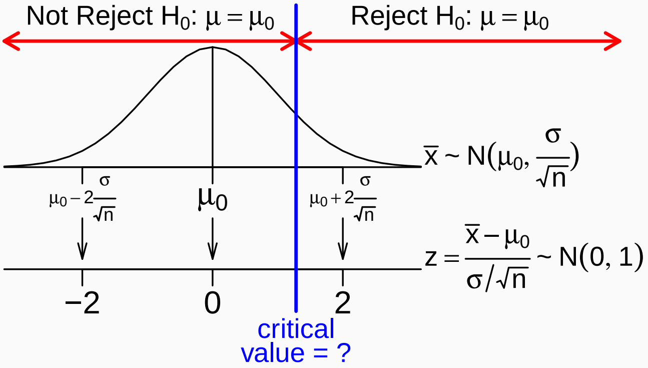
Rejection Region, Critical Value, and Type 1 Error Rate

To test H_0: $\mu = \mu_0$ against H_a: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H_a and only in such cases should H_0 be rejected.



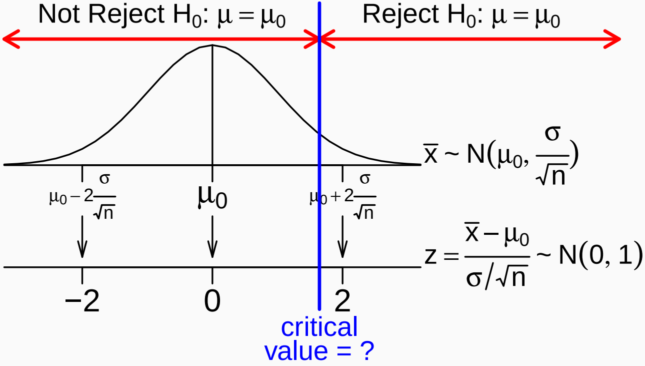
Rejection Region, Critical Value, and Type 1 Error Rate

To test H$_0$: $\mu = \mu_0$ against H$_a$: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H$_a$ and only in such cases should H$_0$ be rejected.



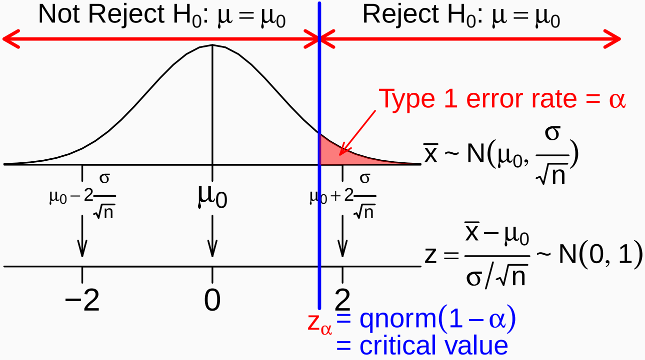
Rejection Region, Critical Value, and Type 1 Error Rate

To test H$_0$: $\mu = \mu_0$ against H$_a$: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H$_a$ and only in such cases should H$_0$ be rejected.



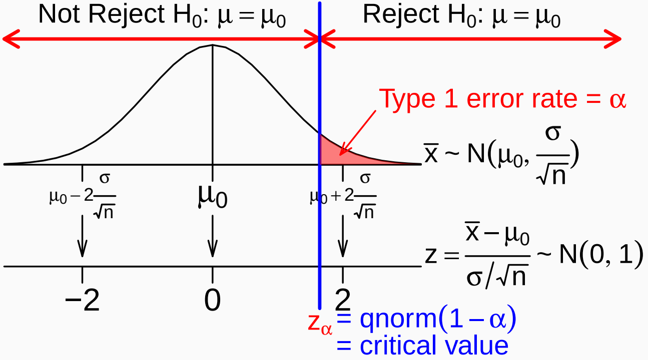
Rejection Region, Critical Value, and Type 1 Error Rate

To test H_0: $\mu = \mu_0$ against H_a: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H_a and only in such cases should H_0 be rejected.



Rejection Region, Critical Value, and Type 1 Error Rate

To test H_0: $\mu = \mu_0$ against H_a: $\mu > \mu_0$, only a sample mean $\bar{x}$ far above μ_0 is evidence for H_a and only in such cases should H_0 be rejected.

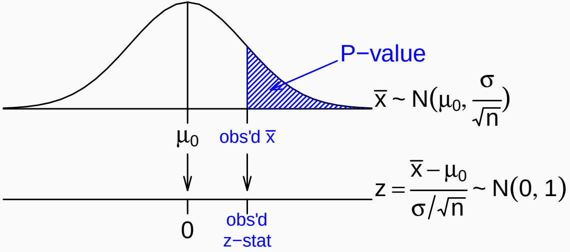


To control Type 1 error rate $= P(\text{rejecting } H_0 | H_0 \text{ is true})$ at the significance level α, we should reject H_0 only when

$$\text{the } z\text{-statistic} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha = \texttt{qnorm}(1 - \alpha).$$

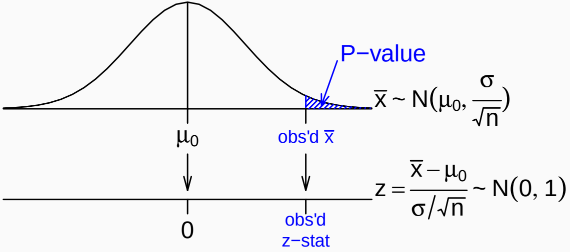
P-value

To test H$_0$: $\mu = \mu_0$ against H$_a$: $\mu > \mu_0$, the P-value is
$P(\bar{x} >$ observed value of $\bar{x} \mid \mu = \mu_0)$ or the blue shaded region below.



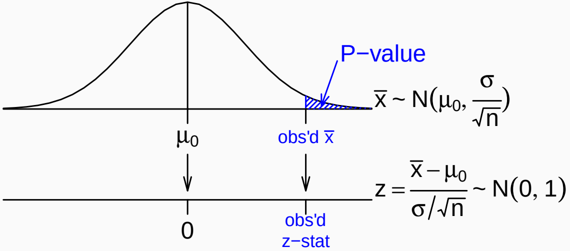
P-value

To test H_0: $\mu = \mu_0$ against H_a: $\mu > \mu_0$, the P-value is $P(\bar{x} > \text{observed value of } \bar{x} \mid \mu = \mu_0)$ or the blue shaded region below.



P-value

To test H$_0$: $\mu = \mu_0$ against H$_a$: $\mu > \mu_0$, the P-value is $P(\bar{x} >$ observed value of $\bar{x} \mid \mu = \mu_0)$ or the blue shaded region below.



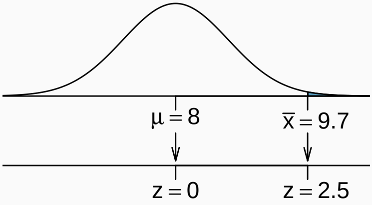
For the College Applications example, the P-value for testing H$_0$: $\mu = 8$ against H$_a$: $\mu > 8$ is

$$P\text{-value} = P(\bar{x} > 9.7 \mid \mu = 8)$$

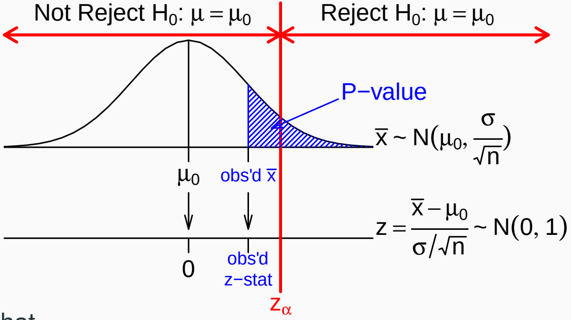
$$= P\left(Z > \frac{9.7 - 8}{7/\sqrt{106}} \approx 2.500\right)$$

$$= 1 - \texttt{pnorm(2.500)}$$

$$\approx 0.0062$$



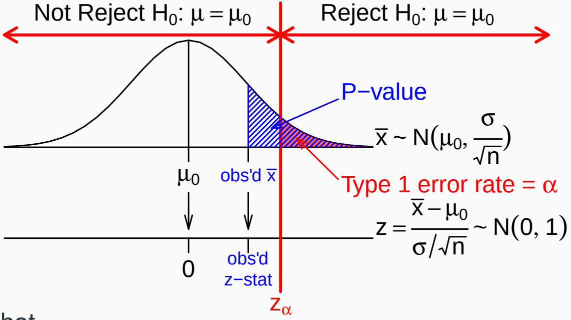
P-value and Critical Value Approaches for Hypotheses Tests



Observed that

$$\text{if } z\text{-statistic} < z_\alpha \quad \text{then} \quad P\text{-value} > \alpha$$

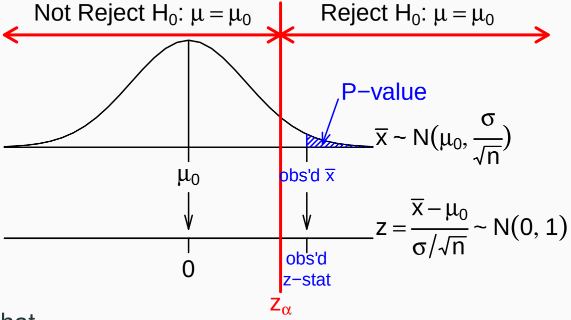
P-value and Critical Value Approaches for Hypotheses Tests



Observed that

if z-statistic $< z_\alpha$ then P-value $> \alpha$

P-value and Critical Value Approaches for Hypotheses Tests

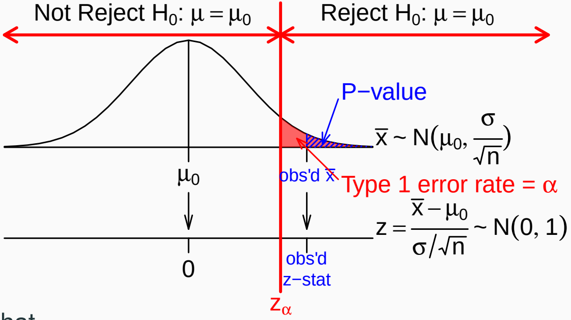


Observed that

if z-statistic $< z_\alpha$ then P-value $> \alpha$

if z-statistic $> z_\alpha$ then P-value $< \alpha$

P-value and Critical Value Approaches for Hypotheses Tests

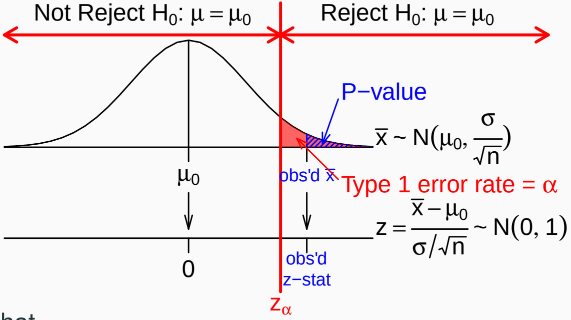


Observed that

if z-statistic $< z_\alpha$ then P-value $> \alpha$

if z-statistic $> z_\alpha$ then P-value $< \alpha$

P-value and Critical Value Approaches for Hypotheses Tests



Observed that

$$\begin{array}{lll} \text{if } z\text{-statistic} < z_\alpha & \text{then} & P\text{-value} > \alpha \\ \text{if } z\text{-statistic} > z_\alpha & \text{then} & P\text{-value} < \alpha \end{array}$$

There are two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_a: $\mu > \mu_0$ and control the Type 1 error rate at the significance level α.

- *Critical value approach*: one can compute the z-statistic = $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and the critical value $z_\alpha = \texttt{qnorm}(1-\alpha)$, and reject H_0 if the z-statistic > z_α.
- *P-value approach*: one can compute the P-value from the z-statistic and reject H_0 when the P-value < α

Two-Sided Hypothesis Tests

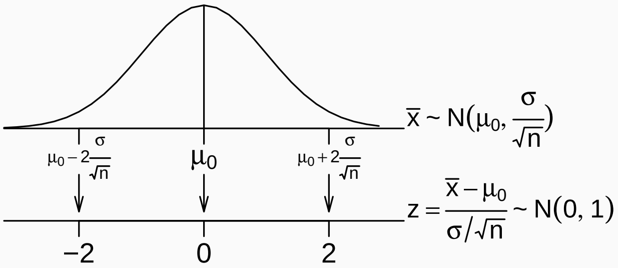
If the dean wanted to know whether the data provide convincing evidence that the average number of colleges applied is *different* than the recommended 8 schools, the alternative hypothesis would be different.

$$H_a : \mu \neq 8$$

In this case, a sample mean $\bar{x}$ far below 8 would also be evidence in favor of H$_a$.

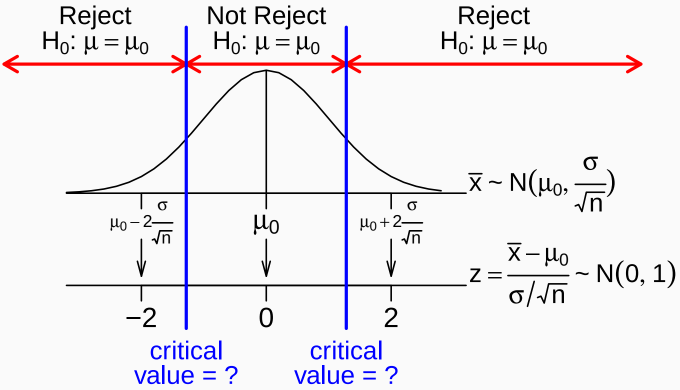
Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{H_a\text{: } \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H_a and hence H_0 should be rejected when $|\bar{x} - \mu_0|$ is large.



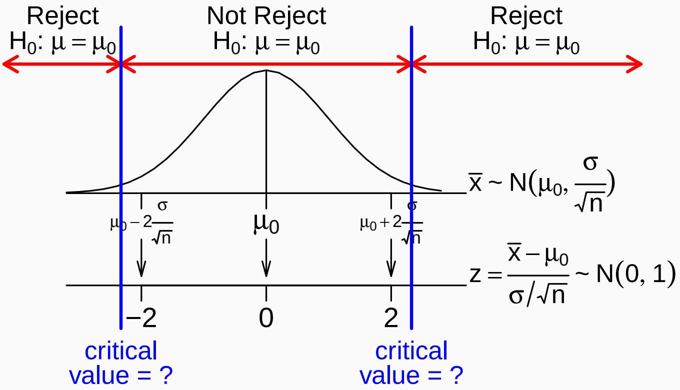
Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{H_a\colon \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H_a and hence H_0 should be rejected when $|\bar{x} - \mu_0|$ is large.



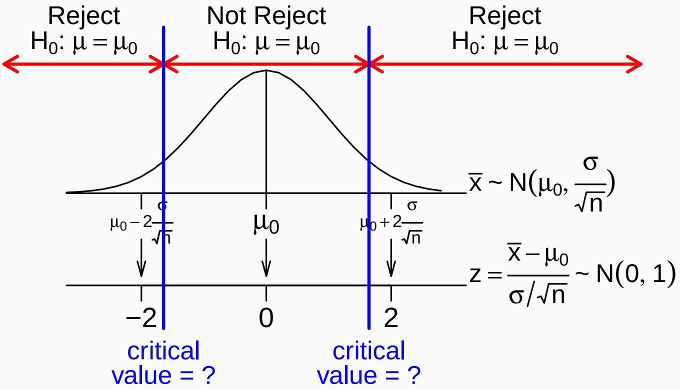
Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{\mathrm{H}_a\colon \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H_a and hence H_0 should be rejected when $|\bar{x} - \mu_0|$ is large.



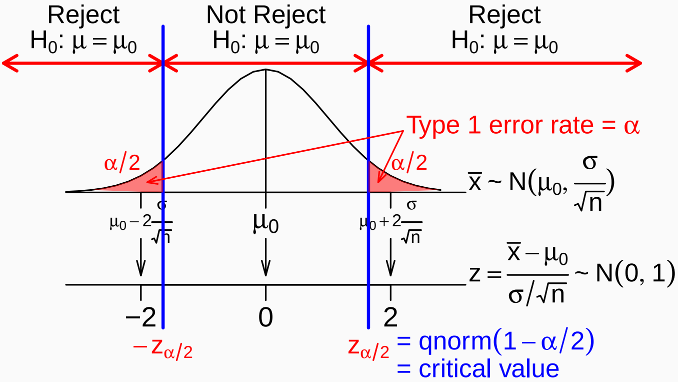
Two-Sided Hypothesis Tests

To test H$_0$: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{\text{H}_a\colon \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H$_a$ and hence H$_0$ should be rejected when $|\bar{x} - \mu_0|$ is large.



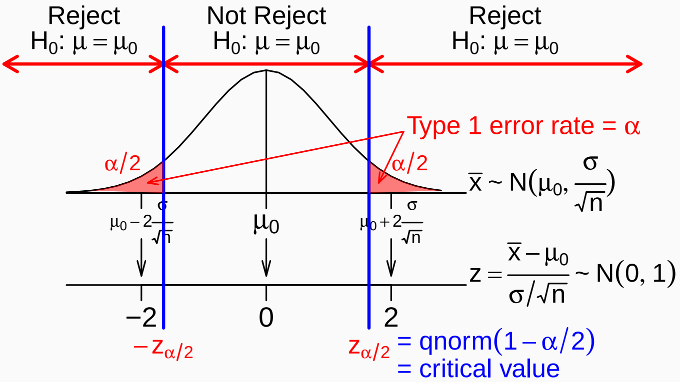
Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{H_a\colon \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H_a and hence H_0 should be rejected when $|\bar{x} - \mu_0|$ is large.



Two-Sided Hypothesis Tests

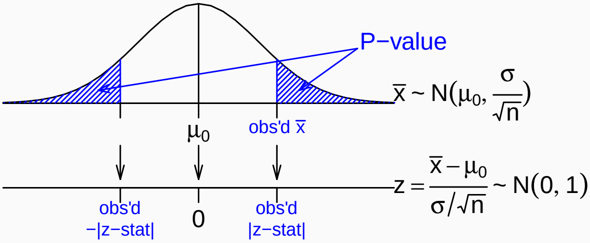
To test H_0: $\mu = \mu_0$ against the **two-sided alternative** $\boxed{\mathrm{H}_a: \mu \neq \mu_0}$, both $\bar{x}$ far above or below μ_0 are evidence for H_a and hence H_0 should be rejected when $|\bar{x} - \mu_0|$ is large.



To control Type 1 error rate at the significance level α, we should reject H_0 only when the $|z\text{-statistic}| = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} = \texttt{qnorm}(1-\alpha/2)$.

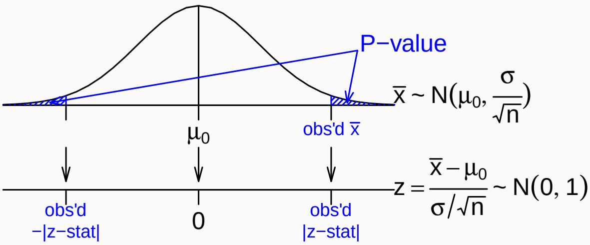
P-values for Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against **two-sided alternative** $\boxed{\mathrm{H}_a\text{: } \mu \neq \mu_0}$, the P-value is the two tail probability (the blue shaded region) below.



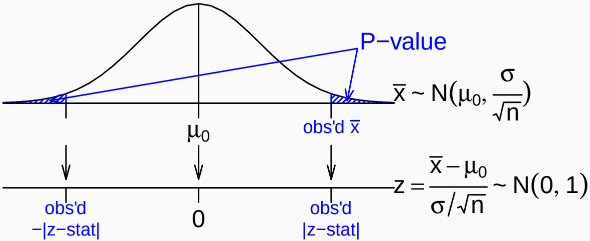
P-values for Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against **two-sided alternative** $\boxed{\mathrm{H}_a\text{: } \mu \neq \mu_0}$, the P-value is the two tail probability (the blue shaded region) below.

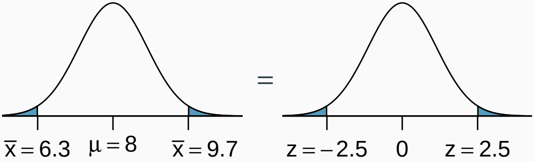


P-values for Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against **two-sided alternative** $\boxed{\mathrm{H}_a\text{: } \mu \neq \mu_0}$, the P-value is the two tail probability (the blue shaded region) below.

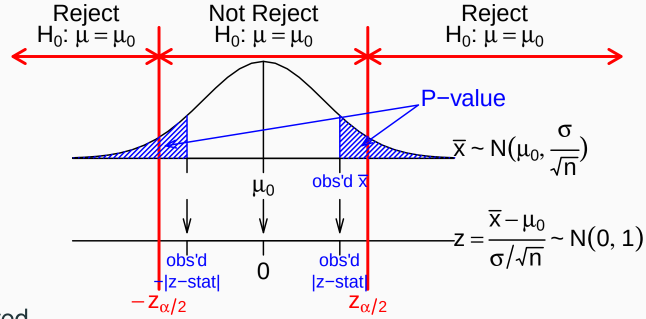


For the College Applications example, the P-value for testing H_0: $\mu = 8$ against H_a: $\mu \neq 8$ is



p-value
$= 0.0062 \times 2$
$= 0.0124$

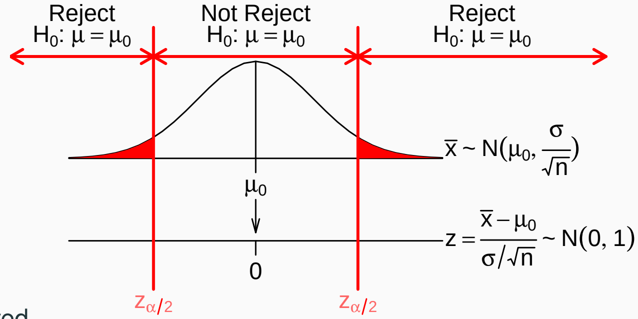
P-value and Critical Value Approaches for Two-Sided Tests



Observed

if $|z\text{-statistic}| < z_{\alpha/2}$ then 2-sided P-value $> \alpha$

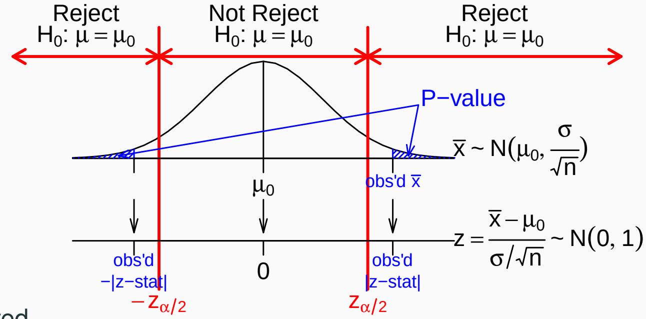
P-value and Critical Value Approaches for Two-Sided Tests



Observed

if $|z$-statistic$| < z_{\alpha/2}$ then 2-sided P-value $> \alpha$

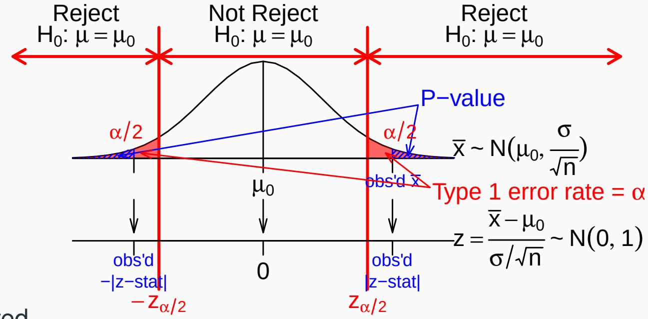
P-value and Critical Value Approaches for Two-Sided Tests



Observed

if $\lvert z$-statistic$\rvert < z_{\alpha/2}$	then	2-sided P-value $> \alpha$
if $\lvert z$-statistic$\rvert > z_{\alpha/2}$	then	2-sided P-value $< \alpha$

P-value and Critical Value Approaches for Two-Sided Tests

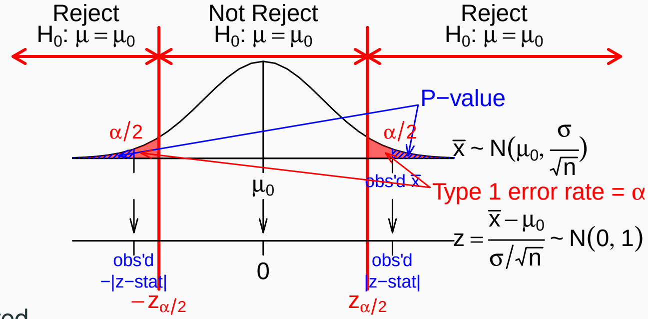


Observed

if $|z$-statistic$| < z_{\alpha/2}$ then 2-sided P-value $> \alpha$

if $|z$-statistic$| > z_{\alpha/2}$ then 2-sided P-value $< \alpha$

P-value and Critical Value Approaches for Two-Sided Tests



Observed

if $\lvert z\text{-statistic}\rvert < z_{\alpha/2}$	then	2-sided P-value $> \alpha$
if $\lvert z\text{-statistic}\rvert > z_{\alpha/2}$	then	2-sided P-value $< \alpha$

There are also two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_a: $\mu \neq \mu_0$ and control the Type 1 error rate at the significance level α.

- *Critical value approach*: reject H_0 if the absolute value of the z-statistic $= \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2} = \texttt{qnorm}(1-\alpha/2)$
- *P-value approach*: one can compute the 2-sided P-value from the z-statistic and reject H_0 when the P-value $< \alpha$

Lower One-Sided Hypothesis Test

If the dean wanted to know whether the data provide convincing evidence that the average number of colleges applied is *lower* than the recommended 8 schools, the alternative hypothesis would be

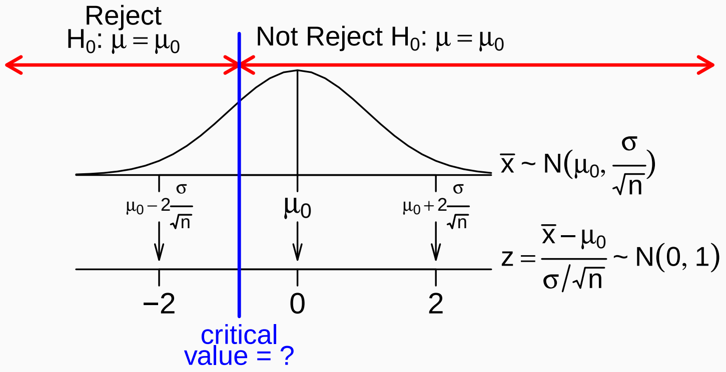
$$H_a : \mu < 8$$

Three types of alternative hypotheses:

- Upper one-sided: H$_a$: $\mu > 8$
- Lower one-sided: H$_a$: $\mu < 8$
- Two-sided: H$_a$: $\mu \neq 8$

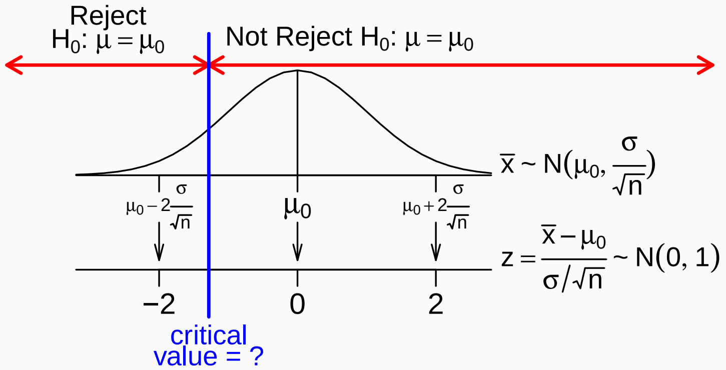
Lower One-Sided Tests

To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{\mathrm{H}_a\text{: } \mu < \mu_0}$, only a sample mean $\bar{x}$ far below μ_0 is evidence for H_a and H_0 should be rejected only in such cases.



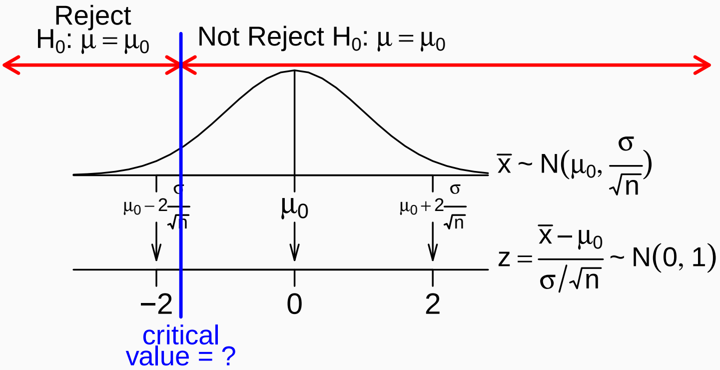
Lower One-Sided Tests

To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{H_a\colon \mu < \mu_0}$, only a sample mean $\bar{x}$ far below μ_0 is evidence for H_a and H_0 should be rejected only in such cases.



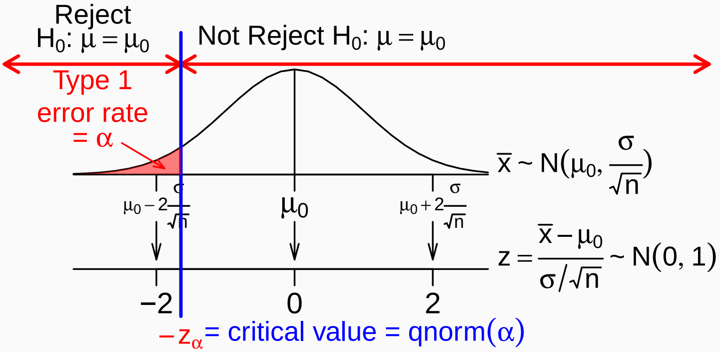
Lower One-Sided Tests

To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{H_a\colon \mu < \mu_0}$, only a sample mean $\bar{x}$ far below μ_0 is evidence for H_a and H_0 should be rejected only in such cases.



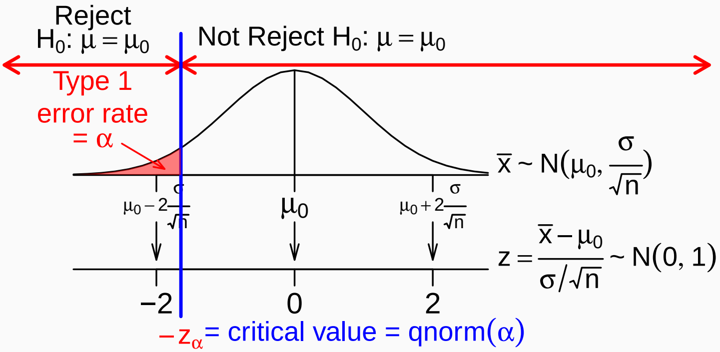
Lower One-Sided Tests

To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{\mathrm{H}_a: \mu < \mu_0}$, only a sample mean $\bar{x}$ far below μ_0 is evidence for H_a and H_0 should be rejected only in such cases.



Lower One-Sided Tests

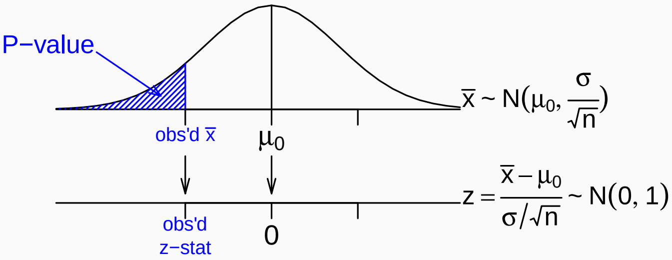
To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{\mathrm{H}_a\colon \mu < \mu_0}$, only a sample mean $\bar{x}$ far below μ_0 is evidence for H_a and H_0 should be rejected only in such cases.



To control Type 1 error rate at the significance level α, we should reject H_0 only when the z-statistic $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha = \mathtt{qnorm}(\alpha)$.

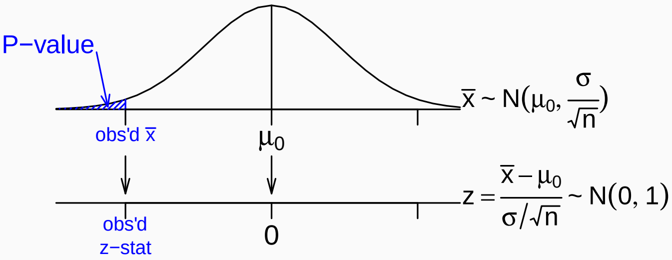
P-values for Lower One-Sided Hypothesis Tests

To test H$_0$: $\mu = \mu_0$ v.s. **lower one-sided alternative** $\boxed{\text{H}_a\text{: } \mu < \mu_0}$,
the P-value is the lower tail probability (blue shaded region) below.



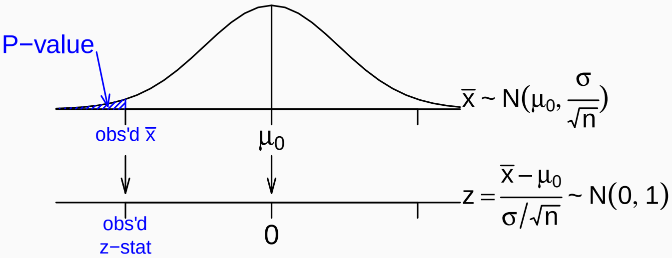
P-values for Lower One-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ v.s. **lower one-sided alternative** $\boxed{H_a: \mu < \mu_0}$, the P-value is the lower tail probability (blue shaded region) below.



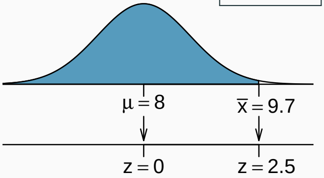
P-values for Lower One-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ v.s. **lower one-sided alternative** $\boxed{H_a\colon \mu < \mu_0}$,
the P-value is the lower tail probability (blue shaded region) below.

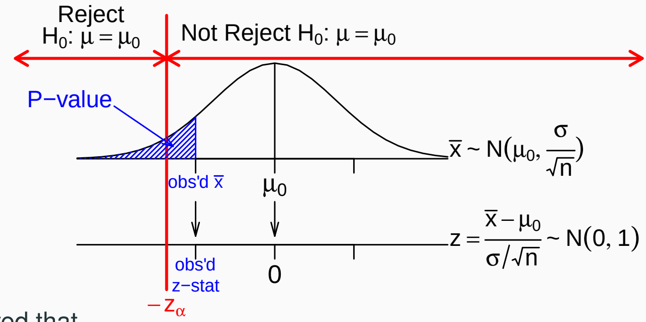


For the College Applications example, the P-value for testing H_0: $\mu = 8$ against H_a: $\mu < 8$ is the lower tail area $1 - 0.0062 = \boxed{0.9938}$,

which makes sense since $\bar{x} = 9.7 > 8$.
Hence H_a: $\mu < 8$ is less plausible than H_0: $\mu = 8$. no reason to reject H_0.



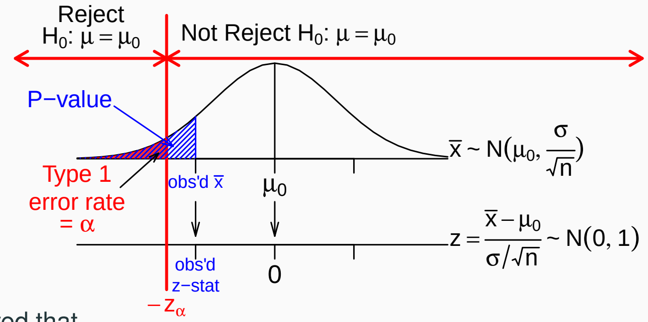
P-value & Critical Value Approaches for Lower One-Sided Tests



Observed that

if z-statistic $> -z_\alpha$ then P-value $> \alpha$

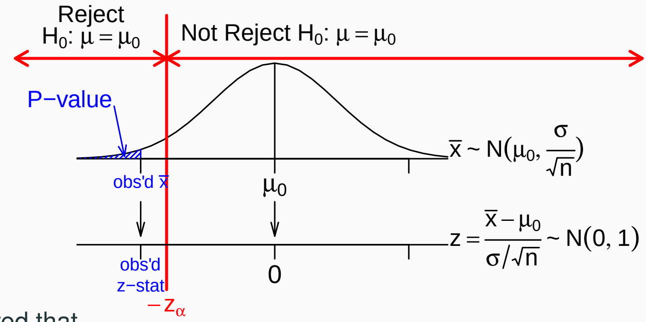
P-value & Critical Value Approaches for Lower One-Sided Tests



Observed that

if z-statistic $> -z_\alpha$ then P-value $> \alpha$

P-value & Critical Value Approaches for Lower One-Sided Tests

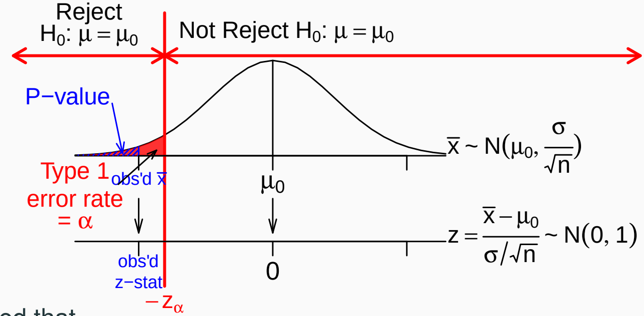


Observed that

if z-statistic $> -z_\alpha$ then P-value $> \alpha$

if z-statistic $< -z_\alpha$ then P-value $< \alpha$

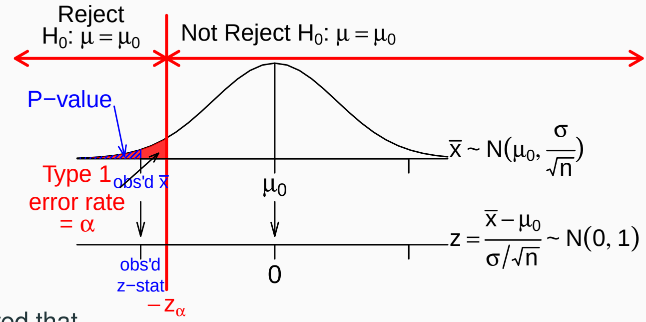
P-value & Critical Value Approaches for Lower One-Sided Tests



Observed that

$$\text{if } z\text{-statistic} > -z_\alpha \quad \text{then} \quad P\text{-value} > \alpha$$
$$\text{if } z\text{-statistic} < -z_\alpha \quad \text{then} \quad P\text{-value} < \alpha$$

P-value & Critical Value Approaches for Lower One-Sided Tests



Observed that

if z-statistic $> -z_\alpha$ then P-value $> \alpha$

if z-statistic $< -z_\alpha$ then P-value $< \alpha$

There are also two equivalent approaches to test H$_0$: $\mu = \mu_0$ v.s. H$_a$: $\mu < \mu_0$ and control the Type 1 error rate at the significance level α.

- *Critical value approach*: reject H$_0$ if the z-statistic $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha =$ `qnorm`(α)
- *P-value approach*: one can compute the lower one-sided P-value from the z-statistic and reject H$_0$ when the P-value $< \alpha$

P-value Approach or Critical Approach?

We have introduced both the critical value approach and the P-value approach for hypothesis testing. They are equivalent but we generally *recommend the P-value approach*, for two reasons.

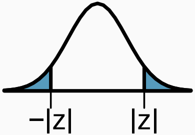
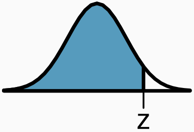
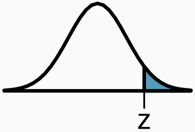
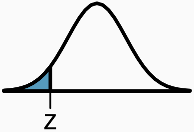
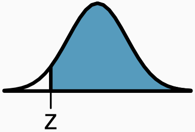
- The rejection rule is simpler, just compare the P-value with the significance level α
- More importantly, we can simply report the P-value and let people choose their own significance level α (the Type 1 error rate) and decide whether to reject or not to reject the H_0

From now on, we will just stick with the P-value approach.

Recap: How to Compute One-Sided & Two Sided P-values

The z-statistic for testing $\text{H}_0 : \mu = \mu_0$ is $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$.

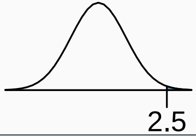
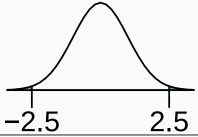
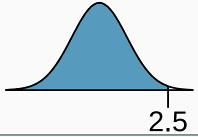
The p-value depends on H_a.

<table>
<tr><th></th><th>Two-sided test</th><th colspan="2">One-sided test</th></tr>
<tr><td>H_a</td><td>$\mu \neq \mu_0$</td><td>$\mu < \mu_0$</td><td>$\mu > \mu_0$</td></tr>
<tr><td rowspan="2">P-value</td><td rowspan="2">
</td><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>R</td><td><code>2*pnorm(|z|,lower.tail=F)</code></td><td><code>pnorm(z)</code></td><td><code>pnorm(z,lower.tail=F)</code></td></tr>
</table>

Then we reject H_0 when P-value $< \alpha$.

The bell curve above is the standard normal curve.

Back to the College Applications Example

H_a	$\mu > 8$	$\mu \neq 8$	$\mu < 8$
P-value	0.0062 	2×0.0062 	$1 - 0.0062$ 

For H_a: $\mu > 8$ and H_a: $\mu \neq 8$, we *reject H_0* since p-value is *low* (under 5%)

- The data provide convincing evidence that freshmen in this university have applied to more than (different from) 8 schools on average.
- The diff. betw. the null value of 8 schools and observed sample mean of 9.7 schools is beyond sampling variability.

For the H_a: $\mu < 8$, there is no reason to reject H_0: $\mu = 8$ since the observed sample mean $9.7 > \mu = 8$. The alternative H_a: $\mu < 8$ is even less plausible than H_0: $\mu = 8$.

Conclusion when the P-value is Low

When the P-value is lower than the significance level, we say

- The H_0 is rejected
- There is strong evidence that freshmen in this university had applied to over 8 schools on average (H_a is true)
- The mean number of schools freshmen in this university had applied is *significantly* over 8

We don't say

- The H_a is accepted
- We fail to reject H_a

Conclusion when the P-value is Not Low

When the P-value exceeds the significance level, we say

- We fail to reject H_0
- No strong evidence that freshmen in this university had applied to over 8 schools on average (H_a is true)
- The mean number of schools freshmen in this university had applied is *not significantly* over 8

We don't say

- the H_0 is accepted
- we fail to accept H_a
- there is strong evidence that H_0 is true — because we might have made a Type 2 error, and the chance of making a Type 2 error is not controlled, which can be quite big

More Incorrect Statements of Hypotheses

Please note that the terms: *significant(ly)* and *reject*, are only used to state the conclusions of the hypotheses tests. Do NOT use them in the hypotheses. It's incorrect to state the hypotheses as

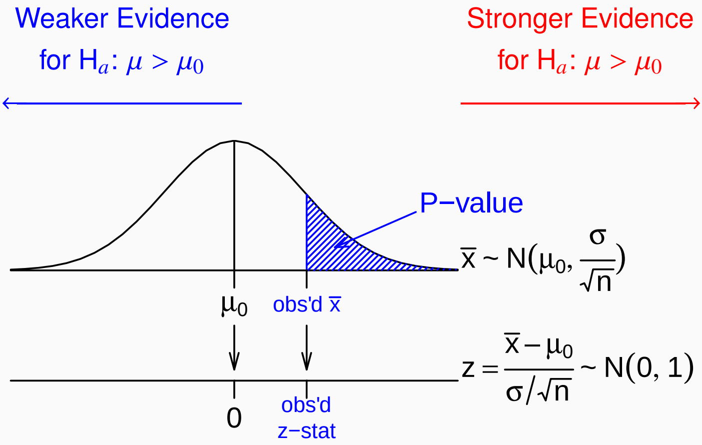
- H_0: The mean number of schools students have applied is *not significantly* over 8
- H_a: The mean number of schools students have applied is *significantly* over 8

or

- H_0: We don't reject that the mean number of schools students have applied is 8
- H_a: We reject that the mean number of schools students applied is 8

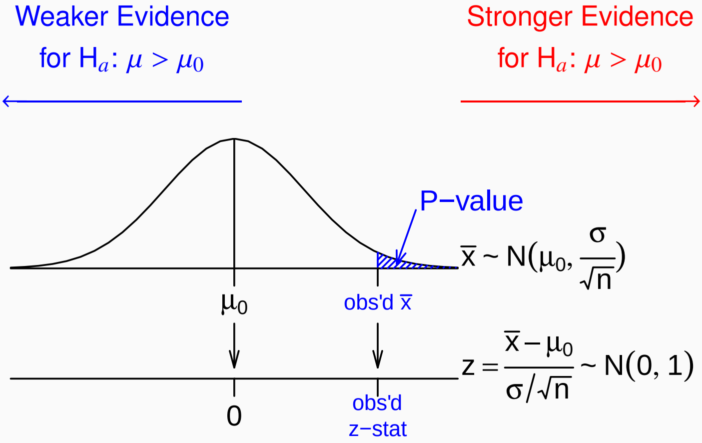
Interpretation of P-Values — Upper One-Sided Tests

The P**-value** is *the probability of getting data such that the evidence for the H_a is* ***at least as strong as*** *our observed data, if in fact H_0: $\mu = \mu_0$ were true*.



Interpretation of P-Values — Upper One-Sided Tests

The P-**value** is *the probability of getting data such that the evidence for the H_a is* **at least as strong as** *our observed data, if in fact H_0: $\mu = \mu_0$ were true*.

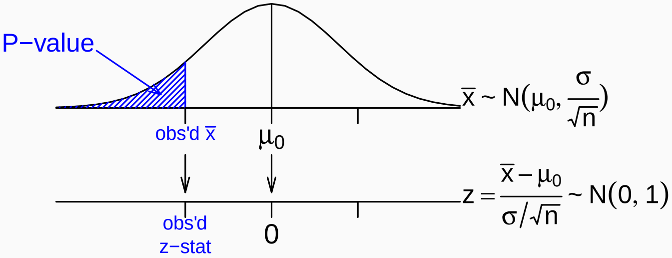


Interpretation of P-Values — Lower One-Sided Tests

The P**-value** is *the probability of getting data such that the evidence for the H_a is* **at least as strong as** *our observed data, if in fact H_0: $\mu = \mu_0$ were true*.

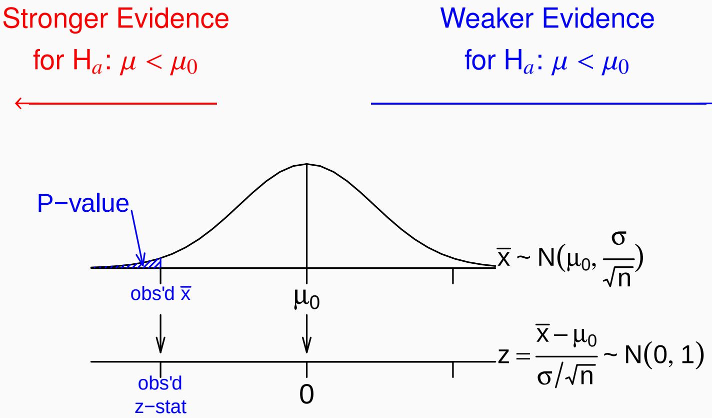
Stronger Evidence for H$_a$: $\mu < \mu_0$ $\longleftarrow$

Weaker Evidence for H$_a$: $\mu < \mu_0$ $\longrightarrow$



Interpretation of P-Values — Lower One-Sided Tests

The P**-value** is *the probability of getting data such that the evidence for the* H_a *is* **at least as strong as** *our observed data, if in fact* H_0*:* $\mu = \mu_0$ *were true*.



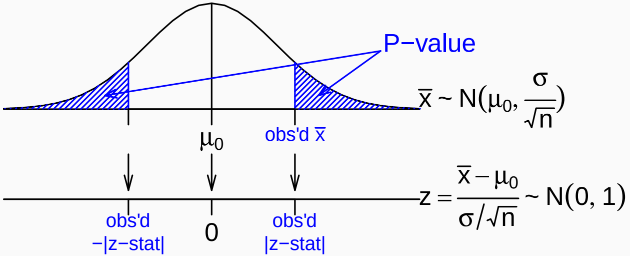
Interpretation of P-Values — Two-Sided Tests

The P**-value** is *the probability of getting data such that the evidence for the H_a is* **at least as strong as** *our observed data, if in fact H_0: $\mu = \mu_0$ were true*.

Stronger evidence for H$_a$: $\mu \neq \mu_0$

Weak evidence for H$_a$: $\mu \neq \mu_0$

Stronger evidence for H$_a$: $\mu \neq \mu_0$



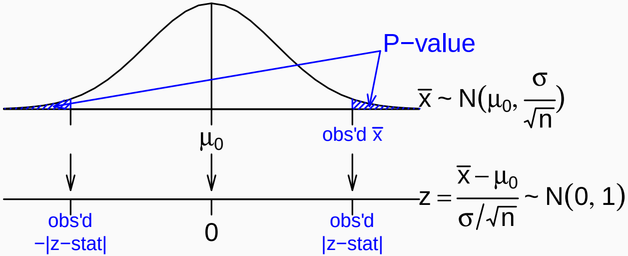
Interpretation of P-Values — Two-Sided Tests

The P**-value** is *the probability of getting data such that the evidence for the* H_a *is* **at least as strong as** *our observed data, if in fact* H_0*:* $\mu = \mu_0$ *were true*.

Stronger evidence for H_a: $\mu \neq \mu_0$ ⟵

Weak evidence for H_a: $\mu \neq \mu_0$ ⟶ ⟵

Stronger evidence for H_a: $\mu \neq \mu_0$ ⟶



Example: Number of College Applications – Conditions

As CLT is used in the hypothesis test above, we need to check the same conditions as we construct confidence intervals for the population mean.

- Observations must be *independent*
    - Use your knowledge to judge if the data might be dependent
- The population distribution of the number of colleges students apply to should not be extremely skewed.
- In the z-statistic $= \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$, if the unknown population SD σ is replaced with the sample SD s, we need to further check that
    - sample size cannot be too small (at least 30)
    - no outliers & not too skewed ⇒ Check the histogram of data!

Recap: Hypothesis Testing for a Population Mean

1. Set the hypotheses
    - $H_0 : \mu = \mu_0$
    - $H_A : \mu <$ or $>$ or $\neq \mu_0$
2. Check assumptions and conditions
    - Independence
    - Normality: nearly normal population or $n \geq 30$, no extreme skew – or use the t distribution (Section 5.1)
3. Calculate a *test statistic* and a p*-value* (draw a picture!)

$$Z = \frac{\bar{x} - \mu_0}{SE}, \ where \ SE = \frac{\sigma}{\sqrt{n}}$$

4. (Optional) Make a decision
    - If p-value $< \alpha$, reject H_0
    - If p-value $> \alpha$, do not reject H_0

Relationship Between Confidence Intervals and Two-Sided Hypothesis Tests

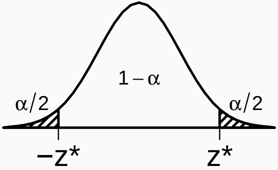
Confidence Intervals and Two-Sided Hypothesis Tests

For a two-sided test:

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu \neq \mu_0$$

the following are equivalent:

- p-value $> \alpha$ (and hence $\mathrm{H}_0 : \mu = \mu_0$ is not rejected at level α)
- $|z\text{-statistic}| = |(\bar{x} - \mu_0)/SE| < z^*$, where z^* is a value such that



- μ_0 is in the $100(1 - \alpha)\%$ confidence interval for μ

$$\bar{x} - z^* SE < \mu_0 < \bar{x} + z^* SE$$

Example

Suppose in a study,

- 90% CI for μ is $(4.81, 11.39)$;
- 95% CI for μ: $(4.18, 12.02)$;
- 99% CI for μ: $(2.95, 13.25)$.

Then

- $H_0 : \mu = 4$ is rejected at 5% level but not at 1% level
  (2-sided p-value is between 1% and 5%)
  because 4 is in the 99% CI but not in the 95% CI
- $H_0 : \mu = 4.5$ is rejected at 10% level but not at 5% level
  because 4.5 is in the 95% CI but not in the 90% CI