STAT 24400 Lecture 18
P-values
Tests & Confidence Intervals for Normal Distributions

Yibi Huang
Department of Statistics
University of Chicago

P-value

P-values

The P*-value* of a test is the **probability of obtaining a test statistic such that the evidence for the alternative hypothesis H_1 is at least as strong as our observed data, assuming the H_0 is true**.

P-values

The *P-value* of a test is the **probability of obtaining a test statistic such that the evidence for the alternative hypothesis $\mathbf{H}_1$ is at least as strong as our observed data, assuming the $\mathbf{H}_0$ is true**.

The definition is mouthful. Here are some key points

- The P-value is a probability, and thus it's between 0 and 1

P-values

The *P-value* of a test is the **probability of obtaining a test statistic such that the evidence for the alternative hypothesis H_1 is at least as strong as our observed data, assuming the H_0 is true**.

The definition is mouthful. Here are some key points

- The P-value is a probability, and thus it's between 0 and 1
- This probability is calculated assuming the H_0 is true.

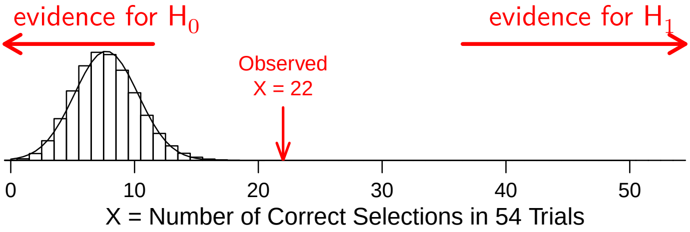
P-values

The *P-value* of a test is the **probability of obtaining a test statistic such that the evidence for the alternative hypothesis H_1 is at least as strong as our observed data, assuming the H_0 is true**.

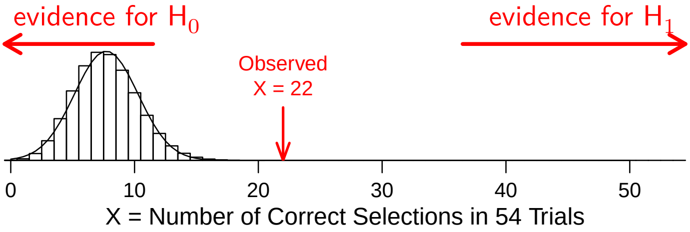
The definition is mouthful. Here are some key points

- The P-value is a probability, and thus it's between 0 and 1
- This probability is calculated assuming the H_0 is true.
- To determine the P-value, we must first decide which values of the test statistic are the evidence for H_1 to be stronger than or as as the value obtained from our sample

P-Value — Dogs-Smell-Cancer Study



P-Value — Dogs-Smell-Cancer Study



- Observed $X = 22$
- Evidence for H_1 is stronger than or as strong as the observed $X = 22$ if $X \geq 22$
- Under H_0, $X \sim Bin(n = 54, p = 1/7)$

$$P\text{-value} = \mathrm{P}(X \geq 22 \mid H_0) = \sum_{k=22}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{54-k} \approx 1.86 \times 10^{-6}$$

- Note P-value is NOT $\mathrm{P}(X = 22 \mid H_0)$

Test Procedure Based on the P-value

As an alternative to test procedures based on rejection regions, one can use test procedures based on P-values

1. Select a significance level $\alpha =$ the desired P(Type I error).
2. Then
   - reject H_0 if the P-value $\leq \alpha$
   - do not reject H_0 if the P-value $> \alpha$

Test Procedure Based on the P-value

As an alternative to test procedures based on rejection regions, one can use test procedures based on P-values

1. Select a significance level $\alpha =$ the desired P(Type I error).
2. Then
   - reject H_0 if the P-value $\leq \alpha$
   - do not reject H_0 if the P-value $> \alpha$

Remark: Hypothesis tests using a "Rejection Region" and those using the "P-value" are equivalent. In fact,

- the test statistic is in the rejection region with significance level α if and only if the P-value $<$ the significance level α

Test Procedure Based on the P-value

As an alternative to test procedures based on rejection regions, one can use test procedures based on P-values

1. Select a significance level $\alpha =$ the desired P(Type I error).
2. Then
   - reject H_0 if the P-value $\leq \alpha$
   - do not reject H_0 if the P-value $> \alpha$

Remark: Hypothesis tests using a "Rejection Region" and those using the "P-value" are equivalent. In fact,

- the test statistic is in the rejection region with significance level α if and only if the P-value $<$ the significance level α

In the rest of L18, we will outline the test procedures for 6 major tests about the normal distribution, using both the critical-value and the P-value approach.

Six Tests for Normal Distributions

One sample: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

1. One sample test for mean, with known σ^2
   H$_0$: $\mu = \mu_0$ v.s. H$_1$: $\mu \neq \mu_0$ (or $\mu > \mu_0$, $\mu < \mu_0$)
2. One sample test for mean, with unknown σ^2
   H$_0$: $\mu = \mu_0$ v.s. H$_1$: $\mu \neq \mu_0$ (or $\mu > \mu_0$, $\mu < \mu_0$)
3. One sample test for variance, with unknown μ
   H$_0$: $\sigma^2 = \sigma_0^2$ v.s. H$_1$: $\sigma^2 \neq \sigma_0^2$ $\sigma^2 > \sigma_0^2$, $\sigma^2 < \sigma_0^2$)

Two indep samples: $X_{11}, \ldots, X_{1n_1} \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$, and $X_{21}, \ldots, X_{2n_2} \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$

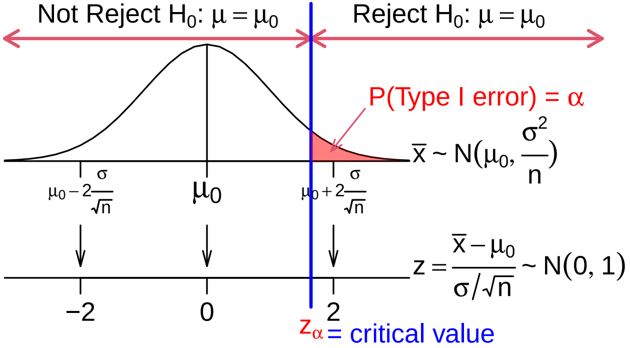
4. Two sample tests for mean, assuming $\sigma_1^2 = \sigma_2^2$
   H$_0$: $\mu_1 = \mu_2$ v.s. H$_1$: $\mu_1 \neq \mu_2$ (or $\mu_1 > \mu_2$, $\mu_1 < \mu_2$)
5. Two sample tests for mean, NOT assuming $\sigma_1^2 = \sigma_2^2$
   H$_0$: $\mu_1 = \mu_2$ v.s. H$_1$: $\mu_1 \neq \mu_2$ (or $\mu_1 > \mu_2$, $\mu_1 < \mu_2$)
6. Two sample tests for variance, μ_1 and μ_2 unknown
   H$_0$: $\sigma_1^2 = \sigma_2^2$ v.s. H$_1$: $\sigma_1^2 \neq \sigma_2^2$ (or $\sigma_1^2 > \sigma_2^2$, $\sigma_1^2 < \sigma_2^2$)

One Sample Tests for Mean, Known σ^2

Upper One-Sided One Sample Tests for Mean, Known σ^2

The test statistic for testing H_0: $\mu = \mu_0$ against H_1: $\mu > \mu_0$ is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1), \quad \text{under } H_0: \ \mu = \mu_0.$$



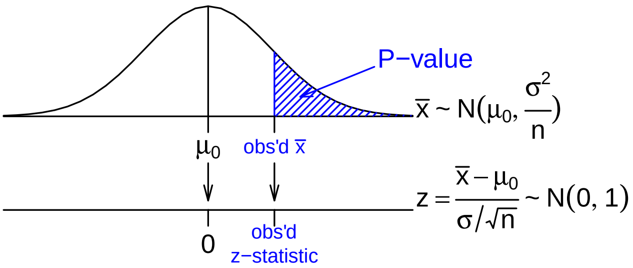
To control P(Type I error) $= \mathrm{P}($rejecting $H_0 \mid H_0$ is true$)$ at the significance level α, we reject H_0 when

P-value for Upper One-Sided Test

Let $\overline{x}$ be the observed value of $\overline{X}$. The P-value for testing H_0: $\mu = \mu_0$ against H_1: $\mu > \mu_0$ is

$$\mathrm{P}(Z > z) = 1 - \Phi(z), \quad \text{where } z = \text{obs'd z-stat} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

or the blue shaded region below.

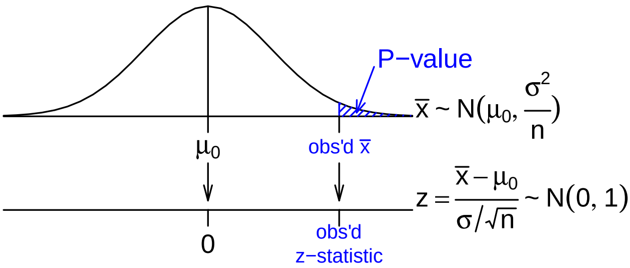


P-value for Upper One-Sided Test

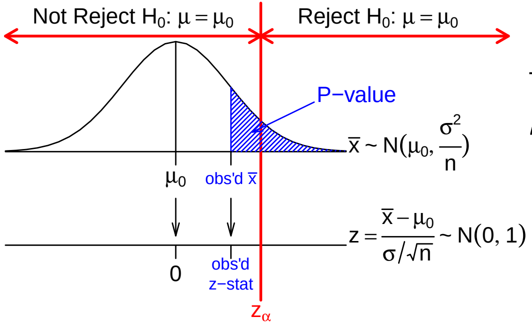
Let $\overline{x}$ be the observed value of $\overline{X}$. The P-value for testing H_0: $\mu = \mu_0$ against H_1: $\mu > \mu_0$ is

$$\mathrm{P}(Z > z) = 1 - \Phi(z), \quad \text{where } z = \text{obs'd z-stat} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

or the blue shaded region below.



P-value v.s. Critical Value



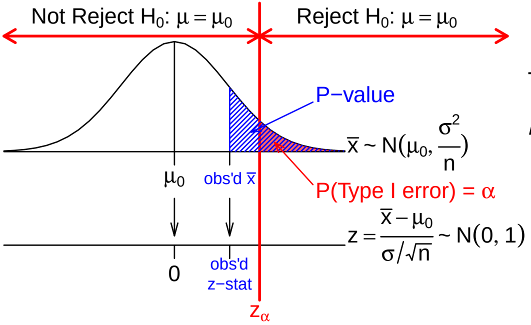
Observed that

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

Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu > \mu_0$ and control P(Type I error) at a significance level α:

- *Critical value approach*: compute the z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and the critical value z_α, and reject H_0 if the z-stat $> z_\alpha$.
- *P-value approach*: compute the P-value from the z-stat and reject H_0 when P-value $< \alpha$

P-value v.s. Critical Value



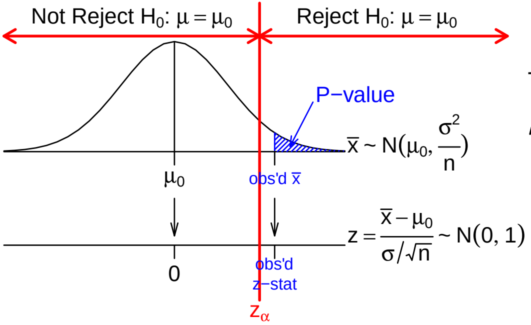
Observed that

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

Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu > \mu_0$ and control P(Type I error) at a significance level α:

- *Critical value approach*: compute the z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and the critical value z_α, and reject H_0 if the z-stat $> z_\alpha$.
- *P-value approach*: compute the P-value from the z-stat and reject H_0 when P-value $< \alpha$

P-value v.s. Critical Value



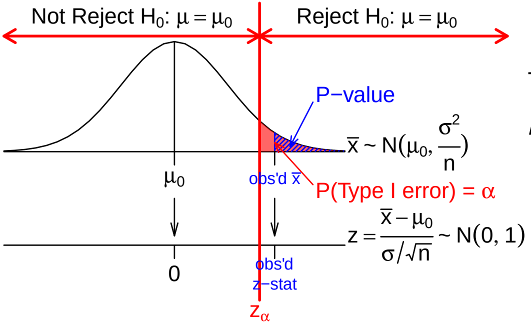
Observed that

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

Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu > \mu_0$ and control P(Type I error) at a significance level α:

- *Critical value approach*: compute the z-stat $= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ and the critical value z_α, and reject H_0 if the z-stat $> z_\alpha$.
- *P-value approach*: compute the P-value from the z-stat and reject H_0 when P-value $< \alpha$

P-value v.s. Critical Value



Observed that

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

Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu > \mu_0$ and control P(Type I error) at a significance level α:

- *Critical value approach*: compute the z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and the critical value z_α, and reject H_0 if the z-stat $> z_\alpha$.
- *P-value approach*: compute the P-value from the z-stat and reject H_0 when P-value $< \alpha$

P-value is the Smallest Significance Level to Reject H_0

The P-value is the **smallest significance level** α **at which the** H_0 **can be rejected**.

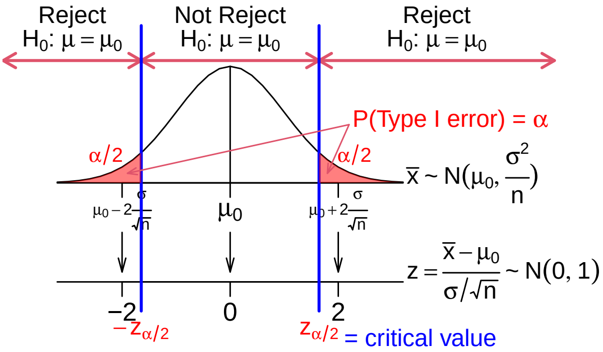
- e.g., the P-value for the dog study is 1.86×10^{-6}.
  The H_0 won't be rejected unless the significance level is as small as 1.86×10^{-6}

Because of this, the P-value is alternatively referred to as the *observed significance level* for the data.

Two-Sided One Sample Tests for Mean, Known σ^2,

For a two-sided test of H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$, the test statistic remains to be

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1), \quad \text{under } H_0: \ \mu = \mu_0.$$



To control P(Type I error) at the significance level α, reject H_0 when

$$|z\text{-stat}| = \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2},$$

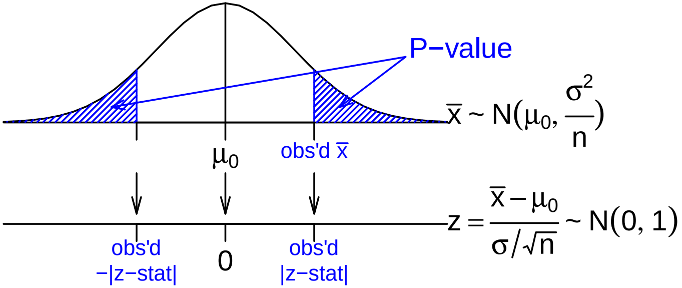
where $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$.

P-values for Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against **two-sided alternative** $\boxed{H_1\colon \mu \neq \mu_0}$, the P-value is the two-tail probability

$$P(|Z| > |z|) = 2(1 - \Phi(|z|)), \quad \text{where } z = \text{obs'd z-stat} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

(the blue shaded region below).

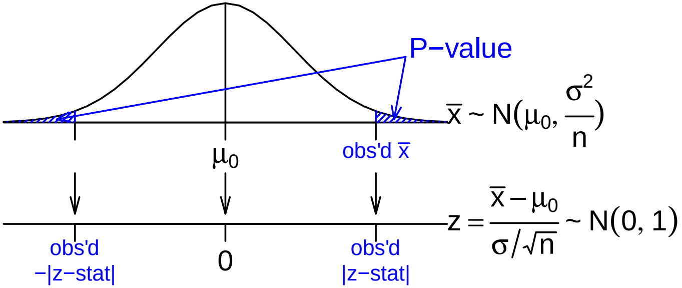


P-values for Two-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ against **two-sided alternative** $\boxed{\mathrm{H}_1\text{: } \mu \neq \mu_0}$, the P-value is the two-tail probability

$$\mathrm{P}(|Z| > |z|) = 2(1 - \Phi(|z|)), \quad \text{where } z = \text{obs'd z-stat} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

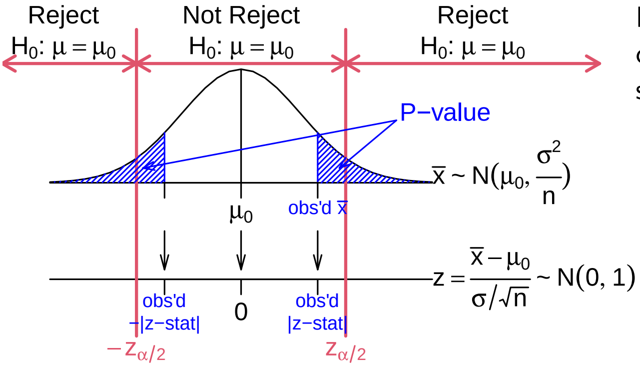
(the blue shaded region below).



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

Observed

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



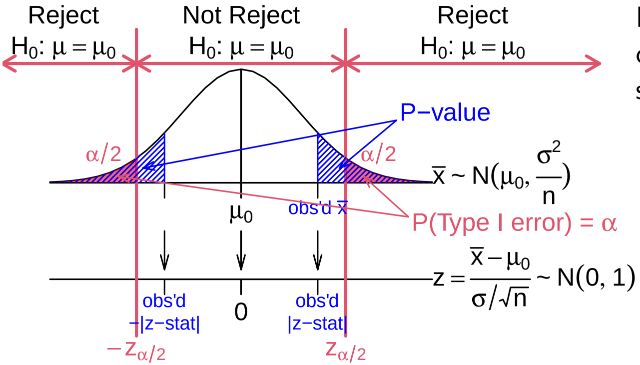
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu \neq \mu_0$ and control P(Type I error) at the significance level α.

- *Critical value approach*: reject H_0 if $|z\text{-stat}| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
- *P-value approach*: compute the 2-sided P-value from the z-statistic and reject H_0 when the P-value $< \alpha$

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

Observed

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



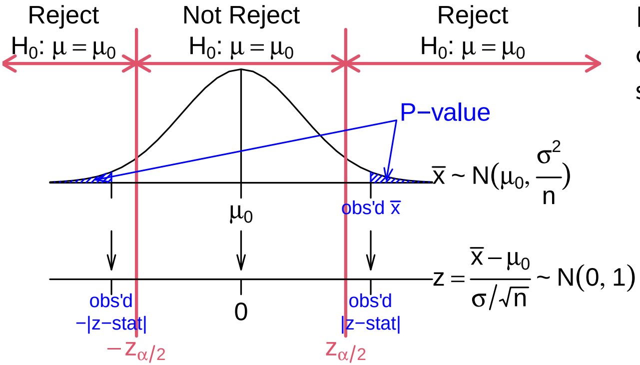
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu \neq \mu_0$ and control P(Type I error) at the significance level α.

- *Critical value approach*: reject H_0 if $|z\text{-stat}| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
- *P-value approach*: compute the 2-sided P-value from the z-statistic and reject H_0 when the P-value $< \alpha$

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

Observed

$$\begin{array}{lll} \text{if } |z\text{-stat}| < z_{\alpha/2} & \text{then} & \text{2-sided } P\text{-value} > \alpha \\ \text{if } |z\text{-stat}| > z_{\alpha/2} & \text{then} & \text{2-sided } P\text{-value} < \alpha \end{array}$$



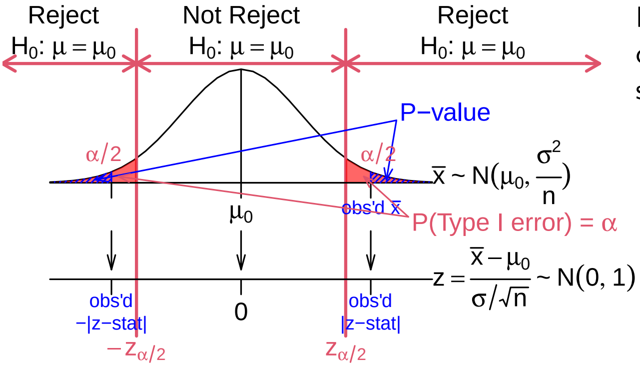
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu \neq \mu_0$ and control P(Type I error) at the significance level α.

- *Critical value approach*: reject H_0 if $|z\text{-stat}| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
- *P-value approach*: compute the 2-sided P-value from the z-statistic and reject H_0 when the P-value $< \alpha$

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

Observed

$$\begin{array}{lll} \text{if } |z\text{-stat}| < z_{\alpha/2} & \text{then} & \text{2-sided } P\text{-value} > \alpha \\ \text{if } |z\text{-stat}| > z_{\alpha/2} & \text{then} & \text{2-sided } P\text{-value} < \alpha \end{array}$$



Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu \neq \mu_0$ and control P(Type I error) at the significance level α.

- *Critical value approach*: reject H_0 if $|z\text{-stat}| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
- *P-value approach*: compute the 2-sided P-value from the z-statistic and reject H_0 when the P-value $< \alpha$

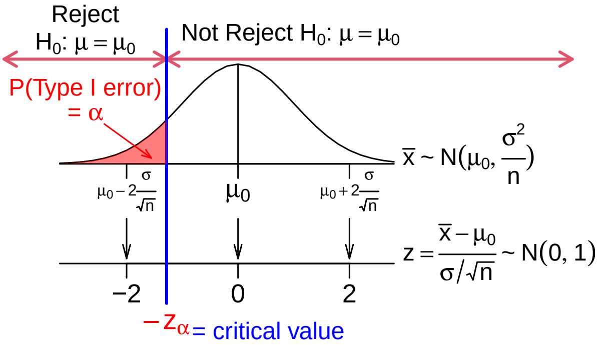
Three Types of Alternative Hypotheses:

- Upper one-sided: H_1: $\mu > \mu_0$
- Lower one-sided: H_1: $\mu < \mu_0$
- Two-sided: H_1: $\mu \neq \mu_0$

Lower One-Sided Tests

To test H_0: $\mu = \mu_0$ against the **lower one-sided** alternative $\boxed{H_1\colon\ \mu < \mu_0}$, the test statistic remains to be

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1), \quad \text{under } H_0:\ \mu = \mu_0.$$



To control P(Type I error) at the significance level α, we reject H_0 when

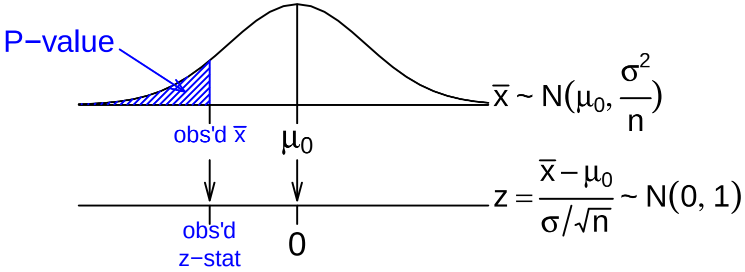
$$z\text{-stat} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha.$$

P-values for Lower One-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ v.s. **lower one-sided alternative** $\boxed{\mathrm{H}_1\colon\ \mu < \mu_0}$, the P-value is the lower tail probability

$$\mathrm{P}(Z < z) = \Phi(z), \quad \text{where } z = \text{obs'd z-stat} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

(blue shaded region below).

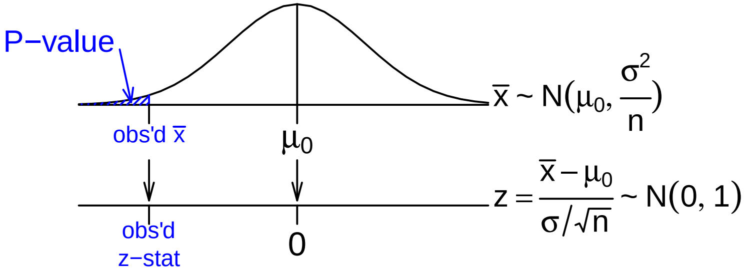


P-values for Lower One-Sided Hypothesis Tests

To test H_0: $\mu = \mu_0$ v.s. **lower one-sided alternative** $\boxed{\mathrm{H}_1\text{: } \mu < \mu_0}$, the P-value is the lower tail probability

$$\mathrm{P}(Z < z) = \Phi(z), \quad \text{where } z = \text{obs'd z-stat} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

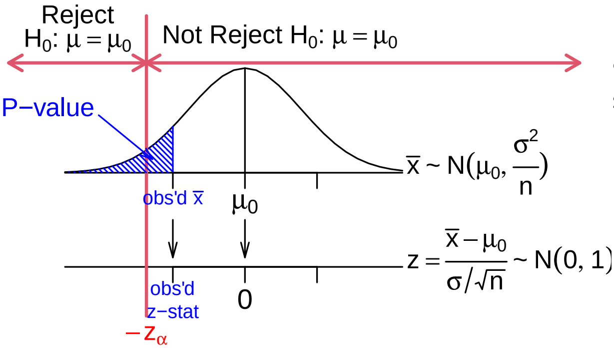
(blue shaded region below).



P-value v.s. Critical Value for Lower One-Sided Tests

Observed that

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



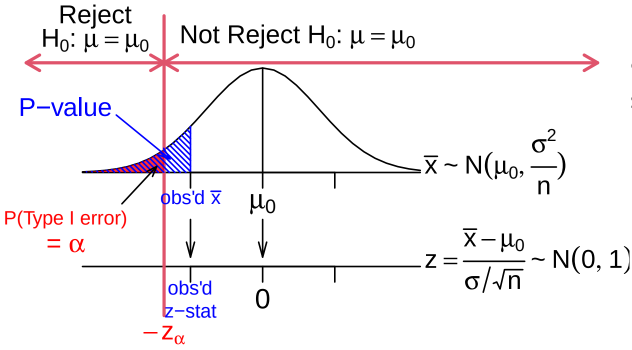
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu < \mu_0$ control P(Type I error) at the significance level α:

- *Critical value approach*: reject H_0 if z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$
- *P-value approach*: compute the lower one-sided P-value from the z-stat and reject H_0 when the P-value $< \alpha$

P-value v.s. Critical Value for Lower One-Sided Tests

Observed that

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



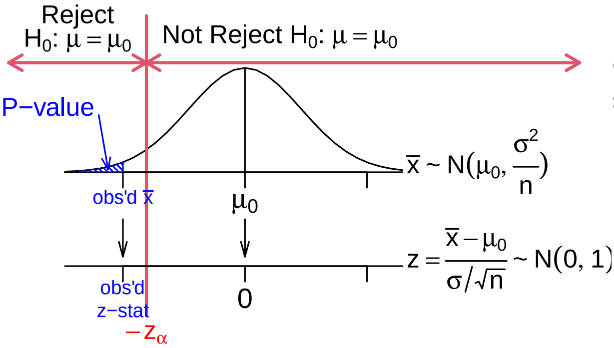
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu < \mu_0$ control P(Type I error) at the significance level α:

- *Critical value approach*: reject H_0 if z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$
- *P-value approach*: compute the lower one-sided P-value from the z-stat and reject H_0 when the P-value $< \alpha$

P-value v.s. Critical Value for Lower One-Sided Tests

Observed that

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



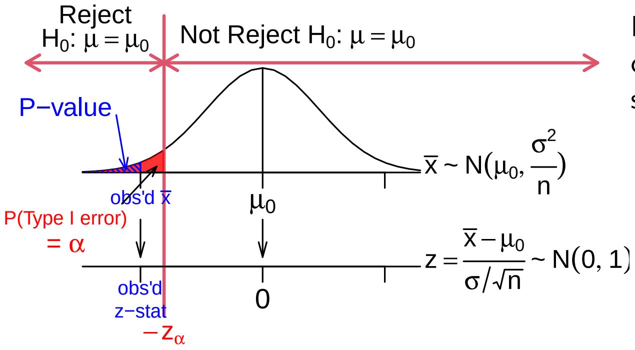
Two equivalent approaches to test H_0: $\mu = \mu_0$ v.s. H_1: $\mu < \mu_0$ control P(Type I error) at the significance level α:

- *Critical value approach*: reject H_0 if z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$
- *P-value approach*: compute the lower one-sided P-value from the z-stat and reject H_0 when the P-value $< \alpha$

P-value v.s. Critical Value for Lower One-Sided Tests

Observed that

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



Two equivalent approaches to test $H_0\colon \mu = \mu_0$ v.s. $H_1\colon \mu < \mu_0$ control P(Type I error) at the significance level α:

- *Critical value approach*: reject H_0 if z-stat $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$
- *P-value approach*: compute the lower one-sided P-value from the z-stat and reject H_0 when the P-value $< \alpha$

P-value Approach or Critical Value Approach?

We 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.

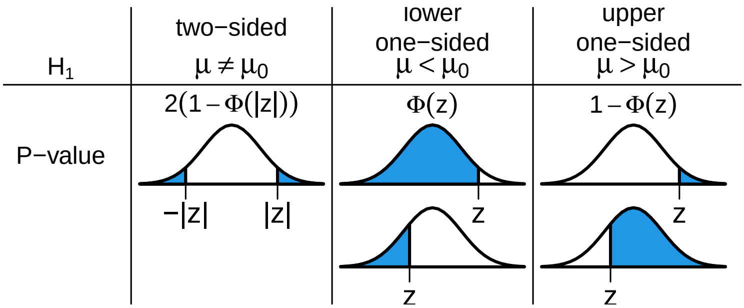
- The rejection rule is simpler, just compare the P-value with the significance level α
- More importantly, we can simply report the P-value and let people choose their own significance level $\alpha =$ P(Type I error) and decide whether to reject or not to reject the H_0

Recap: 1- & 2-Sided Rejection Regions & P-values

For H_0: $\mu = \mu_0$, z-stat $= \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$, reject H_0 at level α if

- z-stat $> z_\alpha$ for H_1: $\mu > \mu_0$
- z-stat $< -z_\alpha$ for H_1: $\mu < \mu_0$
- $|z\text{-stat}| > z_{\alpha/2}$ for H_1: $\mu \neq \mu_0$

The P-values are as follows, where the bell-shape curve is the standard normal curve

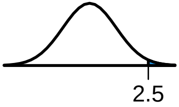
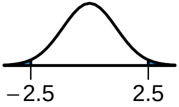
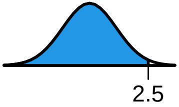


Example w/ Data

Data: $X_1, \dots, X_{100} \overset{\text{iid}}{\sim} N(\mu, \sigma^2 = 6^2)$, w/ sample mean $\overline{x} = 9.5$.
For H_0: $\mu = 8$,

$$z\text{-stat} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{9.5 - 8}{6/\sqrt{100}} = \frac{1.5}{0.6} = 2.5,$$

$$P\text{-value} = \begin{cases} 1 - \Phi(2.5) \approx 0.0062 & \text{if H}_1\text{: } \mu > 8 \\ 2(1 - \Phi(2.5)) \approx 0.0124 & \text{if H}_1\text{: } \mu \neq 8 \\ \Phi(2.5) \approx 1 - 0.0062 = 0.9938 & \text{if H}_1\text{: } \mu < 8 \end{cases}$$

H_1	$\mu > 8$	$\mu \neq 8$	$\mu < 8$
P−value	0.0062 	2×0.0062 	$1 - 0.0062$ 

For H_1: $\mu > 8$ or $\mu \neq 8$, we <u>reject H_0</u> since P-value $< 5\%$.

For H_1: $\mu < 8$, no reason to reject H_0: $\mu = 8$ since H_1: $\mu < 8$ is less plausible than

One Sample Tests for Mean, Unknown σ^2

One Sample Tests for Mean (Unknown σ^2) — Rejection Regions

Data: $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

The test statistic for testing H_0: $\mu = \mu_0$ with unknown σ^2 is

$$T = \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}}, \quad \text{where} \quad S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}.$$

Under H_0: $\mu = \mu_0$, $T \sim t_{n-1}$, we reject H_0 at level α if

- t-stat $> t_{n-1,\alpha}$ for H_1: $\mu > \mu_0$
- t-stat $< -t_{n-1,\alpha}$ for H_1: $\mu < \mu_0$
- $|t\text{-stat}| > t_{n-1,\alpha/2}$ for H_1: $\mu \neq \mu_0$

where t-stat is the observed value of T

$$t\text{-stat} = \frac{\overline{x} - \mu_0}{\sqrt{s^2/n}}, \quad \text{in which} \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}.$$

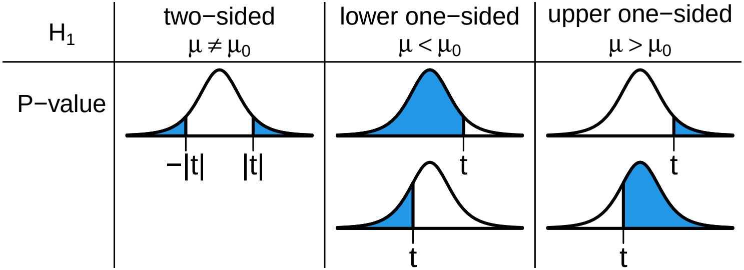
and $t_{n-1,\alpha}$ satisfies

$$\mathrm{P}(T > t_{n-1,\alpha}) = \alpha \quad \text{for } T \sim t_{n-1}.$$

One Sample Tests for Mean (Unknown σ^2) — P-values

The P-values for testing H_0: $\mu = \mu_0$ with unknown σ^2 is

$$P\text{-value} = \begin{cases} P(T > t\text{-stat}) & \text{if } H_1\text{: } \mu > \mu_0 \\ P(|T| > |t\text{-stat}|) = 2P(T > |t\text{-stat}|) & \text{if } H_1\text{: } \mu \neq \mu_0 \\ P(T < t\text{-stat}) & \text{if } H_1\text{: } \mu < \mu_0 \end{cases}$$



The bell-shape curve above is the t-curve with df $= n - 1$, not the normal curve. We reject H_0 when P-value $< \alpha$.

One Sample Test for Variance

One Sample Test for Variance — Test Statistic

Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$
The test statistic for testing H_0: $\sigma^2 = \sigma_0^2$ with unknown μ is

$$V = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}.$$

- What's the distribution of V under H_0: $\sigma^2 = \sigma_0^2$?

- $V \geq 0$
- Large V far above 1 is evidence for H_1: $\sigma^2 > \sigma_0^2$
- V far below 1 is evidence for H_1: $\sigma^2 < \sigma_0^2$
- V being far from 1 is evidence for H_1: $\sigma^2 \neq \sigma_0^2$

One Sample Test for Variance — Test Statistic

Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$
The test statistic for testing H_0: $\sigma^2 = \sigma_0^2$ with unknown μ is

$$V = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}.$$

- What's the distribution of V under H_0: $\sigma^2 = \sigma_0^2$?
  $V \sim \chi^2_{n-1}$, a **chi-squared** distribution w/ $n-1$ degrees of freedom
- $V \geq 0$
- Large V far above 1 is evidence for H_1: $\sigma^2 > \sigma_0^2$
- V far below 1 is evidence for H_1: $\sigma^2 < \sigma_0^2$
- V being far from 1 is evidence for H_1: $\sigma^2 \neq \sigma_0^2$

One Sample Test of Equal Variance — Rejection Region

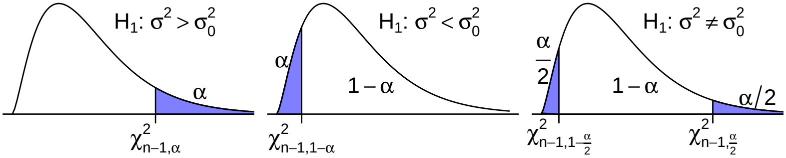
We reject H_0 at level α if

- v-stat $> \chi^2_{n-1,\alpha}$ for H_1: $\sigma^2 > \sigma_0^2$
- v-stat $< \chi^2_{n-1,1-\alpha}$ for H_1: $\sigma^2 < \sigma_0^2$
- v-stat $> \chi^2_{n-1,\alpha/2}$ or v-stat $< \chi^2_{n-1,1-\alpha/2}$ or for H_1: $\sigma^2 \neq \sigma_0^2$

where v-stat is the observed value of V

$$v\text{-stat} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sigma_0^2}.$$

and $\chi^2_{n-1,\alpha}$ satisfies $\mathrm{P}(V > \chi^2_{n-1,\alpha}) = \alpha$ for $V \sim \chi^2_{n-1}$.

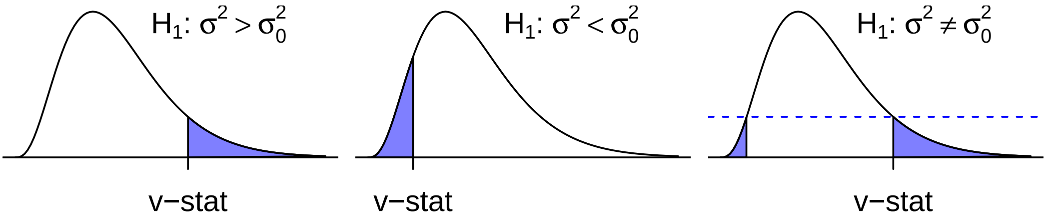


One Sample Test for Variance — P-value

The P-values for testing $\mathsf{H}_0\colon\ \sigma^2 = \sigma_0^2$ with unknown μ is

$$P\text{-value} = \begin{cases} \mathrm{P}(V > v\text{-stat}) & \text{if } \mathsf{H}_1\colon\ \sigma^2 > \sigma_0^2 \\ \mathrm{P}(V < v\text{-stat}) & \text{if } \mathsf{H}_1\colon\ \sigma^2 < \sigma_0^2 \end{cases}$$

What's the two-sided P-value?



Two Sample Tests for Mean (Equal Variance)

Two Sample Test for Mean (Equal Variance) — Test Statistic

Consider two normal random samples of size n_1 and n_2 respectively

$$\left.\begin{array}{ll} X_{11}, X_{12}, \dots, X_{1n_1} & \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma^2) \\ X_{21}, X_{22}, \dots\dots, X_{2n_2} & \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma^2) \end{array}\right\} \to \text{indep., same } \sigma^2.$$

For testing H_0: $\mu_1 = \mu_2$, the two-sample T-statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S^2}}, \text{ where } S^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2}{n_1 + n_2 - 2}$$

Under H_0: $\mu_1 = \mu_2$, $\boxed{T \sim t_{n_1+n_2-2}}$.

Two Sample Test for Mean (Equal Variance) — Rejection Region

We reject H_0: $\mu_1 = \mu_2$ at level α if

- t-stat $> t_{n_1+n_2-2,\alpha}$ for H_1: $\mu_1 > \mu_2$
- t-stat $< -t_{n_1+n_2-2,\alpha}$ for H_1: $\mu_1 < \mu_2$
- $|t\text{-stat}| > t_{n_1+n_2-2,\alpha/2}$ for H_1: $\mu_1 \neq \mu_2$

where t-stat is the observed value of T

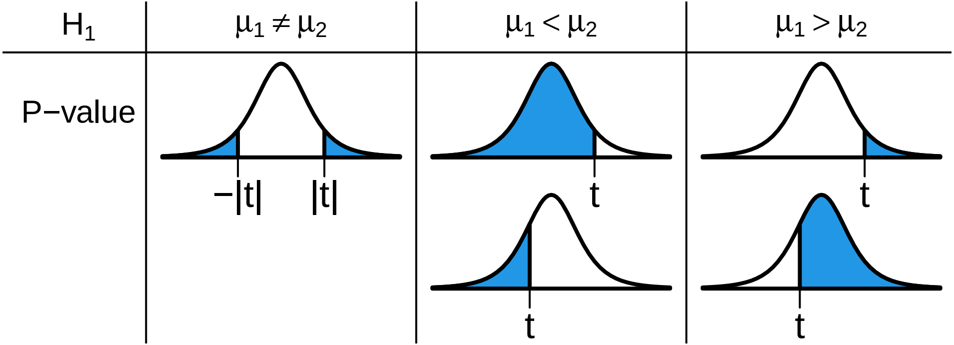
$$t\text{-stat} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})s^2}}, \text{ in which } s^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \overline{x}_1)^2 + \sum_{j=1}^{n_2}(x_{2j} - \overline{x}_2)^2}{n_1 + n_2 - 2}.$$

and $t_{n_1+n_2-2,\alpha}$ satisfies

$$\mathrm{P}(T > t_{n_1+n_2-2,\alpha}) = \alpha \quad \text{for } T \sim t_{n_1+n_2-2}.$$

In L17, we show that a two-sided two-sample test for mean is equivalent to the GLR test.

Two Sample Test for Mean (Equal Variance) — P-Value



The bell curve above is the t-curve with $n_1 + n_2 - 2$ degrees of freedom.

Two Sample Tests for Mean (Unequal Variance)

Two Sample Test for Mean (Unequal Variance)

Without the equal variance assumption, by the indep of the two samples, we know

$$\mathrm{Var}(\overline{X}_1 - \overline{X}_2) = \mathrm{Var}(\overline{X}_1) + \mathrm{Var}(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

The two-sample T-statistic without the equal variance assumption is

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{where} \quad \begin{aligned} S_1^2 &= \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2}{n_1 - 1} \\ S_2^2 &= \frac{\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2}{n_2 - 1} \end{aligned}$$

- Unfortunately, the T-statistic above does NOT have a t-distribution, even under H_0: $\mu_1 - \mu_2$
- Fortunately, it can be approximated by a t-distribution with a certain degrees of freedom.

See the next slide for the approximation

Approximate Distribution of the Two-Sample t-Statistic

Under H_0: $\mu_1 - \mu_2$, the two-sample t-statistic has an **approximate** t_k **distribution**, with the degrees of freedom k as follows

$$k = \frac{(w_1 + w_2)^2}{w_1^2/(n_1 - 1) + w_2^2/(n_2 - 1)}, \quad \text{where} \quad \begin{matrix} w_1 = s_1^2/n_1, \\ w_2 = s_2^2/n_2. \end{matrix}$$

The rejection regions and the calculation of the P-value are similar to the equal variance case, except for the degrees of freedom and thus is not repeated here.

Two Sample Tests of Equal Variance

Two Sample Tests of Equal Variance

Consider two normal random samples of size n_1 and n_2 respectively

$$\left.\begin{array}{ll} X_{11}, X_{12}, \dots, X_{1n_1} & \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2) \\ X_{21}, X_{22}, \dots\dots, X_{2n_2} & \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma_2^2) \end{array}\right\} \rightarrow \text{indep.}$$

For testing H_0: $\sigma_1^2 = \sigma_2^2$, the test-statistic is

$$F = \frac{S_1^2}{S_2^2} \quad \text{where } S_k^2 = \frac{\sum_{i=1}^{n_k}(X_{ki} - \overline{X}_k)^2}{n_k - 1}, \; k = 1, 2.$$

- What's the distribution of F under H_0: $\sigma_1^2 = \sigma_2^2$?
  - $(n_1 - 1)S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ and $(n_2 - 1)S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$ are indep
  - So $F \sim F_{n_1-1, n_2-1}$ has an F-distribution w/ $n_1 - 1$ and $n_2 - 1$ degrees of freedom under H_0: $\sigma_1^2 = \sigma_2^2$
- $F \geq 0$
- F far above 1 is evidence for H_1: $\sigma_1^2 > \sigma_2^2$
- F far below 1 is evidence for H_1: $\sigma_1^2 < \sigma_2^2$
- F being far away from 1 is evidence for H_1: $\sigma^2 \neq \sigma_0^2$

Two-Sample Test for Variance — Rejection Region

We reject H_0 at level α if

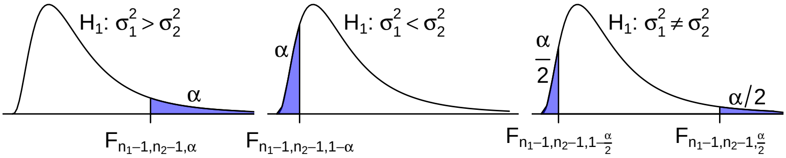
- f-stat $> F_{n_1-1,n_2-2,\alpha}$ for H_1: $\sigma_1^2 > \sigma_2^2$
- f-stat $< F_{n_1-1,n_2-2,1-\alpha}$ for H_1: $\sigma_1^2 < \sigma_2^2$
- f-stat $> F_{n_1-1,n_2-2,\alpha/2}$ or f-stat $< F_{n_1-1,n_2-1,1-\alpha/2}$ for H_1: $\sigma_1^2 \neq \sigma_2^2$

where f-stat is the observed value of V

$$f\text{-stat} = \frac{s_1^2}{s_2^2}.$$

and $F_{n_1-1,n_2-1,\alpha}$ satisfies

$$\mathrm{P}(F > F_{n_1-1,n_2-1,\alpha}) = \alpha \quad \text{for } F \sim F_{n_1-1,n_2-1}.$$

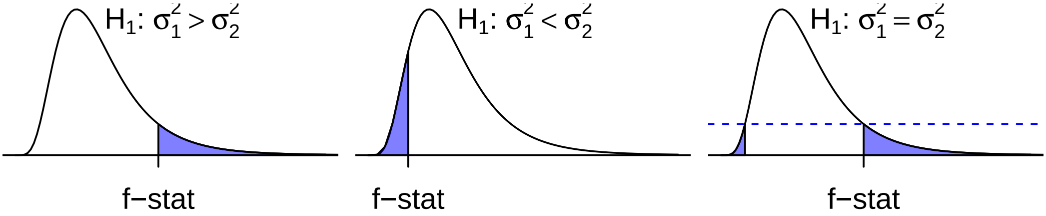


Two-Sample Test for Equal Variance — P-value

The P-values for testing H_0: $\sigma_1^2 = \sigma_2^2$ with unknown μ is

$$P\text{-value} = \begin{cases} \mathrm{P}(F > f\text{-stat}) & \text{if } \text{H}_1\text{: } \sigma_1^2 > \sigma_2^2 \\ \mathrm{P}(F < f\text{-stat}) & \text{if } \text{H}_1\text{: } \sigma_1^2 < \sigma_2^2 \end{cases}$$

What's the two-sided P-value?



Robustness to Non-Normality

If the data are NOT normally distributed,

- the 4 tests about the mean are approx. valid if the sample size is large by CLT.
- the 2 tests about the variance are no longer valid