Hypothesis testing as a decision rule

▶ Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

▶ Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

▶ We will use the following terminology:

- $T$ is the test statistic,
- $A$ is the acceptance region,
- $B$ is the rejection, or critical, region,
- Rejecting $H_0$ when $H_0$ is true is a type I error,
- Accepting $H_0$ when $H_0$ is false is a type II error,
- $\alpha = P(T \in B | H_0)$, the probability of a type I error, is the significance level of the test, also called the size of the type I error,
- $\beta = P(T \in A | H_A)$, the probability of a type II error, is called the size of the type II error,
- $1 - \beta = P(T \in B | H_A)$, the probability of rejecting $H_0$ when $H_0$ is false, is called the power of the test.
Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$. 

We will use the following terminology:

- $T$ is the test statistic,
- $A$ is the acceptance region,
- $B$ is the rejection, or critical, region,
- Rejecting $H_0$ when $H_0$ is true is a type I error,
- Accepting $H_0$ when $H_0$ is false is a type II error,
- $\alpha = P(T \in B | H_0)$ is the probability of a type I error, also called the significance level of the test,
- $\beta = P(T \in A | H_A)$ is the probability of a type II error, and
- $1 - \beta = P(T \in B | H_A)$ is the probability of rejecting $H_0$ when $H_0$ is false, called the power of the test.
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.
- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.
- We will use the following terminology:
Hypothesis testing as a decision rule

Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

We will use the following terminology:

- $T$ is the test statistic,
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1$, $X_2$, \ldots, $X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
  - $B$ is the rejection, or critical, region,
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
  - $B$ is the rejection, or critical, region,
  - Rejecting $H_0$ when $H_0$ is true is a type I error,
  - Accepting $H_0$ when $H_0$ is false is a type II error,
  - $\alpha = P(T \in B | H_0)$, the probability of a type I error, is the significance level of the test, also called the size of the type I error,
  - $\beta = P(T \in A | H_A)$, the probability of a type II error, is called the size of the type II error,
  - $1 - \beta = P(T \in B | H_A)$, the probability of rejecting $H_0$ when $H_0$ is false, is called the power of the test.
Hypothesis testing as a decision rule

- Suppose we wish to test \( H_0 \) versus \( H_A \) based on a random sample \( X_1, X_2, \ldots, X_n \), and we must make a decision in favor of \( H_0 \) or \( H_A \).

- Let \( T \) be a statistic and suppose we have selected subsets \( A \subset \mathbb{R} \) and \( B \subset \mathbb{R} \), where \( A \cup B = \mathbb{R} \) and \( A \cap B = \emptyset \), such that we will accept \( H_0 \) if \( T \in A \) and reject \( H_0 \) (that is, accept \( H_A \)) if \( T \in B \).

- We will use the following terminology:
  - \( T \) is the test statistic,
  - \( A \) is the acceptance region,
  - \( B \) is the rejection, or critical, region,
  - Rejecting \( H_0 \) when \( H_0 \) is true is a type I error,
  - Accepting \( H_0 \) when \( H_0 \) is false is a type II error,
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.

- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.

- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
  - $B$ is the rejection, or critical, region,
  - Rejecting $H_0$ when $H_0$ is true is a type I error,
  - Accepting $H_0$ when $H_0$ is false is a type II error,
  - $\alpha = P(T \in B \mid H_0)$, the probability of a type I error, is the significance level of the test, also called the size of the type I error,
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.
- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.
- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
  - $B$ is the rejection, or critical, region,
  - Rejecting $H_0$ when $H_0$ is true is a type I error,
  - Accepting $H_0$ when $H_0$ is false is a type II error,
  - $\alpha = P(T \in B \mid H_0)$, the probability of a type I error, is the significance level of the test, also called the size of the type I error,
  - $\beta = P(T \in A \mid H_A)$, the probability of a type II error, is called the size of the type II error, and

Dan Sloughter  (Furman University)  The Neyman-Pearson Paradigm  April 21, 2006  2 / 10
Hypothesis testing as a decision rule

- Suppose we wish to test $H_0$ versus $H_A$ based on a random sample $X_1, X_2, \ldots, X_n$, and we must make a decision in favor of $H_0$ or $H_A$.
- Let $T$ be a statistic and suppose we have selected subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, where $A \cup B = \mathbb{R}$ and $A \cap B = \emptyset$, such that we will accept $H_0$ if $T \in A$ and reject $H_0$ (that is, accept $H_A$) if $T \in B$.
- We will use the following terminology:
  - $T$ is the test statistic,
  - $A$ is the acceptance region,
  - $B$ is the rejection, or critical, region,
  - Rejecting $H_0$ when $H_0$ is true is a type I error,
  - Accepting $H_0$ when $H_0$ is false is a type II error,
  - $\alpha = P(T \in B \mid H_0)$, the probability of a type I error, is the significance level of the test, also called the size of the type I error,
  - $\beta = P(T \in A \mid H_A)$, the probability of a type II error, is called the size of the type II error, and
  - $1 - \beta = P(T \in B \mid H_A)$, the probability of rejecting $H_0$ when $H_0$ is false, is called the power of the test.
Controlling $\alpha$ and $\beta$

- Our object is to design a test so that $\alpha$ and $\beta$ are both small.
Controlling $\alpha$ and $\beta$

- Our object is to design a test so that $\alpha$ and $\beta$ are both small.
- For a given test and sample size, decreasing $\alpha$ will increase $\beta$. 
Controlling $\alpha$ and $\beta$

- Our object is to design a test so that $\alpha$ and $\beta$ are both small.
- For a given test and sample size, decreasing $\alpha$ will increase $\beta$.
- In the Neyman-Pearson testing paradigm, one typically fixes $\alpha$ as the acceptable rate of committing type I errors (usually $\alpha = 0.05$ or $\alpha = 0.01$), and then controls $\beta$ through the choice of the test statistic $T$ and the sample size.
Example

A machine is designed to fill boxes with 16 ounces of cereal. If $X$ is the weight of a filled box, suppose $X \sim N(\mu, 0.16)$. To test the hypotheses $H_0: \mu = 16$ $H_A: \mu \neq 16$, we decide to sample 10 boxes and reject $H_0$ if $|\bar{X} - 16| \geq 0.25$. That is, we take $\bar{X}$ as our test statistic, setting an acceptance region $A = (15.75, 16.25)$ and a critical region of $B = (-\infty, 15.75] \cup [16.25, \infty)$. 

Dan Sloughter  (Furman University)  The Neyman-Pearson Paradigm  April 21, 2006  4 / 10
Example

- A machine is designed to fill boxes with 16 ounces of cereal.
Example

- A machine is designed to fill boxes with 16 ounces of cereal.
- If $X$ is the weight of a filled box, suppose $X$ is $N(\mu, 0.16)$. 

To test the hypotheses

$H_0: \mu = 16$

$H_A: \mu \neq 16$,

we decide to sample 10 boxes and reject $H_0$ if $|\bar{X} - 16| \geq 0.25$.

That is, we take $\bar{X}$ as our test statistic, setting an acceptance region of $A = (15.75, 16.25)$ and a critical region of $B = (-\infty, 15.75] \cup [16.25, \infty)$. 

Dan Sloughter (Furman University)
Example

A machine is designed to fill boxes with 16 ounces of cereal. If $X$ is the weight of a filled box, suppose $X$ is $N(\mu, 0.16)$. To test the hypotheses

\[ H_0 : \mu = 16 \]
\[ H_A : \mu \neq 16, \]

we decide to sample 10 boxes and reject $H_0$ if $|\bar{X} - 16| \geq 0.25$. 
Example

- A machine is designed to fill boxes with 16 ounces of cereal.
- If $X$ is the weight of a filled box, suppose $X$ is $N(\mu, 0.16)$.
- To test the hypotheses

\[
H_0 : \mu = 16 \\
H_A : \mu \neq 16,
\]

we decide to sample 10 boxes and reject $H_0$ if $|\bar{X} - 16| \geq 0.25$.

- That is, we take $\bar{X}$ as our test statistic, setting an acceptance region of

\[
A = (15.75, 16.25)
\]

and a critical region of

\[
B = (-\infty, 15.75] \cup [16.25, \infty).
\]
Then the significance level of the test is
\[ \alpha = P\left( \bar{X} \in B \mid H_0 \right) = P\left( |\bar{X} - 16| \geq 0.25 \mid \mu = 16 \right) = 1 - P\left( -0.25 < \bar{X} - 16 < 0.25 \mid \mu = 16 \right) = 1 - \Phi(1.98) + \Phi(-1.98) = 1 - 0.9762 + 0.0238 = 0.0476. \]
Example (cont’d)

Then the significance level of the test is

\[ \alpha = P(\bar{X} \in B \mid H_0) \]
\[ = P(\mid \bar{X} - 16 \mid \geq 0.25 \mid \mu = 16) \]
\[ = 1 - P(-0.25 < \bar{X} - 16 < 0.25 \mid \mu = 16) \]
\[ = 1 - P \left( \frac{-0.25}{0.4 \sqrt{10}} < \frac{\bar{X} - 16}{0.4 \sqrt{10}} < \frac{0.25}{0.4 \sqrt{10}} \mid \mu = 16 \right) \]
\[ = 1 - (\Phi(1.98) - \Phi(-1.98)) \]
\[ = 1 - (0.9762 - 0.0238) \]
\[ = 0.0476. \]
Example (cont’d)

Note: we cannot compute $\beta = P(\bar{X} \in A | H_A)$ because $H_A$ is composite, and so does not uniquely specify a value of $\mu$.

However, we can evaluate the operating characteristic $\beta(\mu) = P(\bar{X} \in A | \mu)$, the probability of accepting $H_0$ when $\mu$ is the true mean.

And we can evaluate the power function, $\pi(\mu) = 1 - \beta(\mu) = P(\bar{X} \in B | \mu)$, the probability of rejecting $H_0$ when $\mu$ is the true mean.
Example (cont’d)

- Note: we cannot compute

\[ \beta = P(\bar{X} \in A \mid H_A) \]

because \( H_A \) is composite, and so does not uniquely specify a value of \( \mu \).
Example (cont’d)

- Note: we cannot compute

\[ \beta = P(\bar{X} \in A \mid H_A) \]

because \( H_A \) is composite, and so does not uniquely specify a value of \( \mu \).

- However, we can evaluate the operating characteristic

\[ \beta(\mu) = P(\bar{X} \in A \mid \mu), \]

the probability of accepting \( H_0 \) when \( \mu \) is the true mean.
Example (cont’d)

Note: we cannot compute

\[ \beta = P(\bar{X} \in A \mid H_A) \]

because \( H_A \) is composite, and so does not uniquely specify a value of \( \mu \).

However, we can evaluate the **operating characteristic**

\[ \beta(\mu) = P(\bar{X} \in A \mid \mu), \]

the probability of accepting \( H_0 \) when \( \mu \) is the true mean.

And we can evaluate the **power function**, 

\[ \pi(\mu) = 1 - \beta(\mu) = P(\bar{X} \in B \mid \mu), \]

the probability of rejecting \( H_0 \) when \( \mu \) is the true mean.
Example (cont’d)

Note: \( \pi(16) = \alpha \), the significance level of the test.

For example, \( \beta(16.5) = P(\bar{X} \in A | \mu = 16.5) = P(15.75 < \bar{X} < 16.25 | \mu = 16.5) = \Phi(-1.98) - \Phi(-5.93) = 0.0238 \), and \( \pi(16.5) = 1 - \beta(16.5) = 0.9762 \).

Dan Sloughter (Furman University)

The Neyman-Pearson Paradigm

April 21, 2006 7 / 10
Example (cont’d)

▶ Note: $\pi(16) = \alpha$, the significance level of the test.
Example (cont’d)

▶ Note: $\pi(16) = \alpha$, the significance level of the test.

▶ For example,

$$\beta(16.5) = P(\bar{X} \in A \mid \mu = 16.5)$$
$$= P(15.75 < \bar{X} < 16.25 \mid \mu = 16.5)$$
$$= P\left(\frac{15.75 - 16.5}{0.4/\sqrt{10}} < \frac{\bar{X} - 16.5}{0.4/\sqrt{10}} < \frac{16.25 - 16.5}{0.4/\sqrt{10}} \mid \mu = 16.5\right)$$
$$= \Phi(-1.98) - \Phi(-5.93)$$
$$= 0.0238,$$

and

$$\pi(16.5) = 1 - \beta(16.5) = 0.9762.$$
The following graph of the power function $\pi(\mu)$ was created using the R commands:

```r
> m <- seq(15,17,.01)
> b <- pnorm(16.25,mean=m,sd=.4/sqrt(10))
   - pnorm(15.75,mean=m,sd=.4/sqrt(10))
> p <- 1 - b
> plot(m,p,type="l",xlab="Mean",ylab="Power")
```
Example (cont’d)

What should the graph of an “ideal” power function look like?
Example (cont’d)

- Graph of the power function:
Example (cont’d)

- Graph of the power function:

What should the graph of an “ideal” power function look like?