Advanced statistics and modelling

5. week

Parameter inference in statistics
Hypothesis testing, MLE, Bootstrap

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
  - We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
  - Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
  - We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
  - Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
  - We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
  - Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
  - We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
  - Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
- We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
- Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

The basic statistical inference problem

- Recall: The basic statistical inference problem was the following:
  - We have some observed data: $X_1, X_2, \ldots, X_n \sim F$.
  - Based on the observations we would like to **infer** (or estimate or learn) some parameters (e.g. p in a Binomial).
- In data science: based on the observations validate an assumption (hypothesis) on some parameters.
- Statistical decision: calculate the probability, that the observations are consistent with the hypothesis.
- Calculate some parameters (3 methods):
  - assume normal distribution,
  - assume the observation is the most probable realization
  - assume all possible values to be observed

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter
- We collect data $X \in \mathcal{X}$, and we want to make a decision on the hypothesis.

Formal definition

- The **parameter space** Θ is partitioned:
  Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter
- We collect data $X \in \mathcal{X}$, and we want to make a decision on the hypothesis.
- The **rejection region** is $R \subset \mathcal{X}$ if
  $X \in R \Rightarrow \text{reject } H_0$
  $X \notin R \Rightarrow \text{retain}(\text{not reject})H_0$

Formal definition

- The **parameter space** Θ is partitioned:
  Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter
- We collect data $X \in \mathcal{X}$, and we want to make a decision on the hypothesis.
- The **rejection region** is $R \subset \mathcal{X}$ if
  $X \in R \Rightarrow \text{reject } H_0$
  $X \notin R \Rightarrow \text{retain(not reject)} H_0$
- A function $T : \mathcal{X} \to \mathbb{R}$ is a **test statistics**

Formal definition

- The **parameter space** Θ is partitioned: Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter
- We collect data $X \in \mathcal{X}$, and we want to make a decision on the hypothesis.
- The **rejection region** is $R \subset \mathcal{X}$ if
  $X \in R \Rightarrow \text{reject } H_0$
  $X \notin R \Rightarrow \text{retain(not reject)} H_0$
- A function $T : \mathcal{X} \to \mathbb{R}$ is a **test statistics**
- The **critical value** is c if the rejection region can be expressed as $R = \{x \in \mathcal{X} \mid T(x) > c\}$

Formal definition

- The **parameter space** Θ is partitioned:
  Θ_0 and Θ_1 ($\Theta_0 \cap \Theta_1 = \emptyset$).
- The **null hypothesis** is $H_0 : \theta \in \Theta_0$
- The **alternative hypothesis** is $H_1 : \theta \in \Theta_1$
- Note: θ is a fixed parameter
- We collect data $X \in \mathcal{X}$, and we want to make a decision on the hypothesis.
- The **rejection region** is $R \subset \mathcal{X}$ if
  $X \in R \Rightarrow \text{reject } H_0$
  $X \notin R \Rightarrow \text{retain(not reject)} H_0$
- A function $T : \mathcal{X} \to \mathbb{R}$ is a **test statistics**
- The **critical value** is c if the rejection region can be expressed as
  $R = \{x \in \mathcal{X} \mid T(x) > c\}$
- The **hypothesis test** is to find T and c, which leads the least harmful decision.
  (e.g. Do I have cancer? or Do I got the most points for home work?)

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

- The **power function of the test** is a $\Theta \to \mathbb{R}$

$$\beta(\theta = \mu) = P_\theta(X \in R)$$

where R is the rejection region and X are some collected data.

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

- The **power function of the test** is a $\Theta \to \mathbb{R}$

$$\beta(\theta = \mu) = P_\theta(X \in R)$$

where R is the rejection region and X are some collected data.

- The test is **two sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

- The **power function of the test** is a $\Theta \to \mathbb{R}$

$$\beta(\theta = \mu) = P_\theta(X \in R)$$

where R is the rejection region and X are some collected data.

- The test is **two sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$
- The test is **one sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$

  or $H_0 : \theta = \theta_0$ and $H_1 : \theta < \theta_0$

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

- The **power function of the test** is a $\Theta \to \mathbb{R}$

$$\beta(\theta = \mu) = P_\theta(X \in R)$$

  where R is the rejection region and X are some collected data.

- The test is **two sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$
- The test is **one sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$

  or $H_0 : \theta = \theta_0$ and $H_1 : \theta < \theta_0$

- The **level** of the test is $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

  the maximum probability for rejecting the observations, when the parameter is in the retained set.

Formal definition

- Possible outcomes:

	Retain H_0	Reject H_0
H_0 true	OK	type I error
H_1 true	type II error	OK

- The **power function of the test** is a $\Theta \to \mathbb{R}$ $\beta(\theta = \mu) = P_\theta(X \in R)$ where R is the rejection region and X are some collected data.
- The test is **two sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$
- The test is **one sided** if $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$

  or $H_0 : \theta = \theta_0$ and $H_1 : \theta < \theta_0$
- The **level** of the test is $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

  the maximum probability for rejecting the observations, when the parameter is in the retained set.

Note some alternatives in the literature:

- $H_0 \leq \theta_0$ for one sided tests (later we see problems with this)
- $1 - \beta(\theta) = P_\theta(X \in R)$ (though the **power** is the same!)

Example:
We observe $X_1, X_2, \ldots, X_n$ standard normally distributed, independent variables, and we want to test:
Do these values tend to be negative or positive?
.

Example:
We observe $X_1, X_2, \ldots, X_n$ standard normally distributed, independent variables, and we want to test:
Do these values tend to be negative or positive?
.
Possible solution: the variables are from a $N(\mu = 0, \sigma)$ distribution and we test for $\mu \leq 0$.

Example:
We observe $X_1, X_2, \ldots, X_n$ standard normally distributed, independent variables, and we want to test:
Do these values tend to be negative or positive?
.
Possible solution: the variables are from a $N(\mu = 0, \sigma)$ distribution and we test for $\mu \leq 0$.
Parameter space: $\Theta = (-\infty, \infty)$,
partition: $\Theta_0 = (-\infty, 0]$, $\Theta_1 = (0, \infty)$,

Example:
We observe $X_1, X_2, \ldots, X_n$ standard normally distributed, independent variables, and we want to test:
Do these values tend to be negative or positive?
.
Possible solution: the variables are from a $N(\mu = 0, \sigma)$ distribution and we test for $\mu \leq 0$.
Parameter space: $\Theta = (-\infty, \infty)$,
partition: $\Theta_0 = (-\infty, 0]$, $\Theta_1 = (0, \infty)$,
Null hypothesis: $H_0 : \mu = 0$, $H_1 : \mu > 0$
(Note the different relations $H_0 \leftrightarrow \Theta_0$ and $H_1 \leftrightarrow \Theta_1$!)

Example:
We observe $X_1, X_2, \dots, X_n$ standard normally distributed, independent variables, and we want to test:
Do these values tend to be negative or positive?
.
Possible solution: the variables are from a $N(\mu = 0, \sigma)$ distribution and we test for $\mu \leq 0$.
Parameter space: $\Theta = (-\infty, \infty)$,
partition: $\Theta_0 = (-\infty, 0]$, $\Theta_1 = (0, \infty)$,
Null hypothesis: $H_0 : \mu = 0$, $H_1 : \mu > 0$
(Note the different relations $H_0 \leftrightarrow \Theta_0$ and $H_1 \leftrightarrow \Theta_1$!)
Let us define $T = \overline{X}$ (average of observed values), and find c which rejects H_0 if $T > c$.
Rejection region: $R = \{x : T(x) > c\}$.

Example

The power function:

$$\begin{aligned}\beta(\theta=\mu) &= P_\mu(\overline{X}>c)\\ &= P_\mu\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n}>\sqrt{n}\frac{c-\mu}{\sigma}\right).\end{aligned}$$

Example

The power function:

$$\begin{aligned} \beta(\theta = \mu) &= P_\mu(\overline{X} > c) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \sqrt{n}\frac{c - \mu}{\sigma}\right). \end{aligned}$$

Because $\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is the z-transformed (normalized) value of the average,

Example

The power function:

$$\begin{aligned} \beta(\theta = \mu) &= P_\mu(\overline{X} > c) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \sqrt{n}\frac{c - \mu}{\sigma}\right). \end{aligned}$$

Because $\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is the z-transformed (normalized) value of the average, it is $Z \sim N(0, 1)$ distributed:

$$\begin{aligned} \beta(\theta) = \beta(\mu) &= P_\mu\left(Z > \sqrt{n}\frac{c - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\sqrt{n}\frac{c - \mu}{\sigma}\right) \end{aligned}$$

Example

The power function:

$$\begin{aligned} \beta(\theta = \mu) &= P_\mu(\overline{X} > c) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \sqrt{n}\frac{c - \mu}{\sigma}\right). \end{aligned}$$

Because $\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is the z-transformed (normalized) value of the average, it is $Z \sim N(0, 1)$ distributed:
$\beta(\theta) = \beta(\mu) = P_\mu\left(Z > \sqrt{n}\frac{c-\mu}{\sigma}\right) = 1 - \Phi\left(\sqrt{n}\frac{c-\mu}{\sigma}\right)$
For level (α), note: $\beta(\mu)$ is increasing and $\Theta_0 = (-\infty, 0]$,

Example

The power function:

$$\begin{aligned}\beta(\theta = \mu) &= P_\mu(\overline{X} > c)\\ &= P_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \sqrt{n}\frac{c - \mu}{\sigma}\right).\end{aligned}$$

Because $\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is the z-transformed (normalized) value of the average, it is $Z \sim N(0, 1)$ distributed:
$\beta(\theta) = \beta(\mu) = P_\mu\left(Z > \sqrt{n}\frac{c-\mu}{\sigma}\right) = 1 - \Phi\left(\sqrt{n}\frac{c-\mu}{\sigma}\right)$
For level (α), note: $\beta(\mu)$ is increasing and $\Theta_0 = (-\infty, 0]$,
so $\sup \beta(\mu)$ for $\mu \in \Theta_0$ is at $\mu = 0$.

Example

The power function:

$$\begin{aligned}\beta(\theta=\mu) &= P_\mu(\overline{X}>c)\\ &= P_\mu\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n}>\sqrt{n}\frac{c-\mu}{\sigma}\right).\end{aligned}$$

Because $\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is the z-transformed (normalized) value of the average, it is $Z\sim N(0,1)$ distributed:
$\beta(\theta)=\beta(\mu)=P_\mu\left(Z>\sqrt{n}\frac{c-\mu}{\sigma}\right)=1-\Phi\left(\sqrt{n}\frac{c-\mu}{\sigma}\right)$
For level (α), note: $\beta(\mu)$ is increasing and $\Theta_0=(-\infty,0]$, so $\sup\beta(\mu)$ for $\mu\in\Theta_0$ is at $\mu=0$.
Find the level of the test:

$$\begin{aligned}\alpha &= \sup_{\mu\in[-\infty,0]}\beta(\mu)=\sup_{\mu\in[-\infty,0]}\left\{1-\Phi\left(\sqrt{n}\frac{c-\mu}{\sigma}\right)\right\}\\ &= 1-\Phi\left(\sqrt{n}\frac{c}{\sigma}\right)\\ \textit{or vica}-\textit{verse}: &\quad \text{at given level find the critical value}\\ c &= \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}\end{aligned}$$

Example

We reject H_0

if $T = \overline{X} > c = \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}$

Example

We reject H_0

if $T = \overline{X} > c = \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}$

Or

$$\frac{\sqrt{n}(\overline{X}-0)}{\sigma} > z_\alpha$$

where z_α is the tabulated critical value of the standard normal distribution.

.

Example

We reject H_0

if $T = \overline{X} > c = \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}$

Or

$$\frac{\sqrt{n}(\overline{X}-0)}{\sigma} > z_\alpha$$

where z_α is the tabulated critical value of the standard normal distribution.

.

Note the substituted value of $\mu = 0$, which is the H_0 null hypothesis expressed as an equation, instead of an inequality.

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$
Test statistics: reject H_0 if

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$
Test statistics: reject H_0 if

$$\left| \frac{(\hat{\theta} - \theta)}{\hat{se}} \right| > z_{\alpha/2}$$

in case of a two sided test

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$
Test statistics: reject H_0 if

$$\frac{(\hat{\theta} - \theta)}{\hat{se}} > z_\alpha$$

in case of a one sided test

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$
Test statistics: reject H_0 if

$$\frac{(\hat{\theta} - \theta)}{\hat{se}} < z_\alpha$$

in an other case of a one sided test

Standard test I.: Wald test

This test is valid for asymptotically normally distributed variables.

- θ denotes the real value of the parameter,
- $\hat{\theta}$ an estimated value of the parameter,
- $\hat{se}$ estimated standard error of $\hat{\theta}$

The formal test: $H_0 : \theta = \hat{\theta}$ and $H_1 : \theta \neq \hat{\theta}$
Test statistics: reject H_0 if

$$\left| \frac{(\hat{\theta} - \theta)}{\hat{se}} \right| > z_{\alpha/2}$$

$$\frac{(\hat{\theta} - \theta)}{\hat{se}} > z_\alpha \quad or \quad \frac{(\hat{\theta} - \theta)}{\hat{se}} < z_\alpha$$

Note the difference between the one sided and the two sided test:
one sided compares with: z_α and no $|\ .\ |$
two sided compares with: $z_{\alpha/2}$ and takes absolute value

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$P(X \in R) = 1 - P(X \notin R)$$

where

$$P(X \notin R) = P(\theta_\star \text{ close to } \hat{\theta}) = \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} + z_{\alpha/2}\right) - \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} - z_{\alpha/2}\right)$$

so

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$P(X \in R) = 1 - P(X \notin R)$$

where

$$P(X \notin R) = P(\theta_\star \text{ close to } \hat{\theta}) = \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} + z_{\alpha/2}\right) - \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} - z_{\alpha/2}\right)$$

so

$$\beta = 1 - \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} - z_{\alpha/2}\right)$$

or with the true value

$$\beta = 1 - \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} - z_{\alpha/2}\right)$$

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$\beta = 1 - \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{\hat{\theta} - \theta_\star}{\hat{se}} - z_{\alpha/2}\right)$$

or with the true value

$$\beta = 1 - \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} - z_{\alpha/2}\right)$$

which is the probability correctly rejecting a false H_0 under assuming $\theta_\star$ as the parameter instead θ.

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$\beta = 1 - \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} - z_{\alpha/2}\right)$$

which is the probability correctly rejecting a false H_0 under assuming $\theta_\star$ as the parameter instead θ.

The power increases if:

- difference between $\theta_\star$ and θ increases (note: we change only $\theta_\star$)

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$\beta = 1 - \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} - z_{\alpha/2}\right)$$

which is the probability correctly rejecting a false H_0 under assuming $\theta_\star$ as the parameter instead θ.

The power increases if:

- difference between $\theta_\star$ and θ increases (note: we change only $\theta_\star$)
- sample variance decreases ($\hat{se} \sim 1/\sqrt{n}$ decreases as n grows)

Some properties of Wald test

When the true value would be different what we expect as real parameter:

$$\theta_\star \neq \theta,$$

the power of the test ($\beta = P(X \in R)$) will be:

$$\beta = 1 - \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{(\hat{\theta} - \theta) - (\theta_\star - \theta)}{\hat{se}} - z_{\alpha/2}\right)$$

which is the probability correctly rejecting a false H_0 under assuming $\theta_\star$ as the parameter instead θ.
The power increases if:

- difference between $\theta_\star$ and θ increases (note: we change only $\theta_\star$)
- sample variance decreases ($\hat{se} \sim 1/\sqrt{n}$ decreases as n grows)

Wald test is equivalent with a confidence interval:

$$C = (\hat{\theta} - z_{\alpha/2}\hat{se}, \hat{\theta} + z_{\alpha/2}\hat{se},)$$

$H_0 : \hat{\theta} = \theta_\star$ rejected at level α if and only if $\theta_\star \notin C$

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1 - p_1)/n)$
(p_2 similar)
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1 - p_1)/n)$
(p_2 similar)
If the two samples are independent, variances are additive.
and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1 - p_1)/n)$
(p_2 similar)
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1 - p_1)/n)$
(p_2 similar)
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

$$\beta = P(z_{sample} > z_{crit})$$

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1-p_1)/n)$
(p_2 similar)
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

$$\beta = P(z_{sample} > z_{crit})$$

Conclusion: power increases as $\hat{p_1}$ is farther from $\hat{p_2}$
or sample sizes (n and m) increase

Standard test II: Comparing two predictions

$W \sim \mathrm{Binomial}(p_1, n)$ and $Y \sim \mathrm{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1-p_1)/n)$
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

$$\beta = P(z_{sample} > z_{crit})$$

Conclusion: power increases as $\hat{p_1}$ is farther from $\hat{p_2}$
or sample sizes (n and m) increase
Note: the minimum of $\min(n, m)$ dominates!

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1 - p_1)/n)$
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

$$\beta = P(z_{sample} > z_{crit})$$

Conclusion: power increases as $\hat{p_1}$ is farther from $\hat{p_2}$ or sample sizes (n and m) increase
Note: the minimum of $\min(n, m)$ dominates!
Note2: X and Y independent, two different samples

Standard test II: Comparing two predictions

$W \sim \text{Binomial}(p_1, n)$ and $Y \sim \text{Binomial}(p_2, m)$ are the number of incorrect predictions of two methods in two different samples.
Question: is $p_1 = p_2$?
Test: $H_0 : \delta = p_1 - p_2 = 0$ and $H_1 : \delta \neq 0$
Power of the test: $\beta = P(X \in R)$
For large samples: Binomial $\sim$ Gauss distribution.
$\hat{p}_1 = \frac{W}{n}$ is approx. Normal distr.: $N(\mu, \sigma^2) = N(p_1, p_1(1-p_1)/n)$
If the two samples are independent, variances are additive.
Results to:

$$z_{sample} = \frac{\delta - 0}{\hat{se}} = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$$

and

$$\beta = P(z_{sample} > z_{crit})$$

Conclusion: power increases as $\hat{p_1}$ is farther from $\hat{p_2}$
or sample sizes (n and m) increase
Note: the minimum of $\min(n, m)$ dominates!
Note2: X and Y independent, two different samples
Note3: if X and Y measured together (**paired samples**), then $\delta = \overline{D}$
and $\hat{se}^2 = \overline{(D - \overline{D})^2} / \sqrt{n}$, where $D_i = X_i - Y_i$ for each case i

Yet another definition: p-value

- If a test rejects at level α, then it rejects for all $\alpha' > \alpha$.
- The rejection region R depends on the α level (smaller α results smaller R).
- The **p-value** of a test is the smallest α where the test rejects.

$$\mathrm{p-value} = \inf \{\alpha \mid \exists X \in R_\alpha\}$$

Recall: X is a random variable taking values from the observed values $\mathcal{X}$
Note: the p-value is a measure **against** H_0, and different from $\mathbb{P}(H_0|data)$ (which is the probability of H_0 being true with the condition of observed data.) The latter will be discussed under Bayesian inference later.

Properties of p-value

- The smaller the p-value, the stronger evidence against H_0

Properties of p-value

- The smaller the p-value, the stronger evidence against H_0
- Large p-value can result when:

- H_0 is true

Properties of p-value

- The smaller the p-value, the stronger evidence against H_0
- Large p-value can result when:
  - H_0 is true
  - H_1 is true and the test has low power

Properties of p-value

- The smaller the p-value, the stronger evidence against H_0
- Large p-value can result when:
  - H_0 is true
  - H_1 is true and the test has low power

Let's calculate the p-value:

$$\text{reject } H_0 : T(X) > c_\alpha \Leftrightarrow \text{p--value} = \sup_{\theta \in \Theta_0, \xi \in \mathcal{X}} \mathbb{P}_\theta(T(\xi) \geq T(X))$$

which is the probability of observing a test statistics as extreme or more extreme that was actually observed.

Properties of p-value

- Why prefer the null hypothesis $H_0 : \theta = \theta_0$ against $H_0 : \theta \leq \theta_0$?
  - In practice, the probability is calculated at a given value of the parameter θ, which is set according to the null hypothesis.
  - If the distribution of $T(X)$ is continuous, and $H_0 : \theta = \theta_0$, then the p-value has a uniform distribution on [0,1]

The test is **statistically significant** if $\alpha >$p-value.

Properties of p-value

- Why prefer the null hypothesis $H_0 : \theta = \theta_0$ against $H_0 : \theta \leq \theta_0$?
  - In practice, the probability is calculated at a given value of the parameter θ, which is set according to the null hypothesis.
  - If the distribution of $T(X)$ is continuous, and $H_0 : \theta = \theta_0$, then the p-value has a uniform distribution on [0,1]

The test is **statistically significant** if $\alpha >$p-value.
Note: a test can be statistically significant, but practically not significant, if the confidence interval is very small.