YSU ASDS, Statistics, Fall 2019

Lecture 27

Michael Poghosyan

27 Nov 2019

Contents

Last Lecture ReCap

- What are the two Sample Z- or t-Test about?

Last Lecture ReCap

- What are the two Sample Z- or t-Test about?
- Describe the two Sample t-test.

From the last lecture

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$,

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2),\ Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2),$

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**.

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k;

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$;

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of D_k, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since X_k and Y_k can be dependent. But that's OK, we do not need it.

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size n is large. We jut need to use the CLT.

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of D_k, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since X_k and Y_k can be dependent. But that's OK, we do not need it.

Additional Assumption: We will assume that the Differences D_k are Normally Distributed[1].

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size n is large. We jut need to use the CLT.

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of D_k, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since X_k and Y_k can be dependent. But that's OK, we do not need it.

Additional Assumption: We will assume that the Differences D_k are Normally Distributed[1].

Null Hypothesis: $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

[1]The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size n is large. We jut need to use the CLT.

Paired t-Test for the Difference of two Normals Means

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_X, σ_Y **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

Note: Here we have the same number of X_k and Y_k; also, importantly, X_k and Y_k can be dependent!

Notation: $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of D_k, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since X_k and Y_k can be dependent. But that's OK, we do not need it.

Additional Assumption: We will assume that the Differences D_k are Normally Distributed[1].

Null Hypothesis: $\mathcal{H}_0 : \ \mu_X - \mu_Y = \mu_0$

Asymptotic Significance Level: $\alpha \in (0, 1)$;

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size n is large. We jut need to use the CLT.

Paired t-Test for the Difference of two Normals Means, Cont'd

Test Statistics: $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where S_D is the Sample Deviation of D.

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if D_k-s are not Normal, but n is large.

Paired t-Test for the Difference of two Normals Means, Cont'd

Test Statistics: $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where S_D is the Sample Deviation of D.

Distrib of the Test-Statistics Under $\mathcal{H}_0$:[2] $t \sim t(n-1)$;

[2]Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if D_k-s are not Normal, but n is large.

Paired t-Test for the Difference of two Normals Means, Cont'd

Test Statistics: $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where S_D is the Sample Deviation of D.

Distrib of the Test-Statistics Under $\mathcal{H}_0$:[2] $t \sim t(n-1)$;

Rejection Region:

$\mathcal{H}_1$ is	RR is
$\mu_X - \mu_Y \neq \mu_0$	$\lvert t \rvert > t_{n-1,1-\frac{\alpha}{2}}$
$\mu_X - \mu_Y > \mu_0$	$t > t_{n-1,1-\alpha}$
$\mu_X - \mu_Y < \mu_0$	$t < t_{n-1,\alpha}$

[2]Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if D_k-s are not Normal, but n is large.

Paired t-Test for the Difference of two Normals Means, Cont'd

Test Statistics: $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where S_D is the Sample Deviation of D.

Distrib of the Test-Statistics Under $\mathcal{H}_0$:[2] $t \sim t(n-1)$;

Rejection Region:

$\mathcal{H}_1$ is	RR is
$\mu_X - \mu_Y \neq \mu_0$	$\lvert t\rvert > t_{n-1,1-\frac{\alpha}{2}}$
$\mu_X - \mu_Y > \mu_0$	$t > t_{n-1,1-\alpha}$
$\mu_X - \mu_Y < \mu_0$	$t < t_{n-1,\alpha}$

Note: This Test is called the **Paired t-Test**

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if D_k-s are not Normal, but n is large.

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better.

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better. We are running our algorithms over n Test Cases, and mark by 1 if the Algorithm is giving correct answer, and 0 otherwise.

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better. We are running our algorithms over n Test Cases, and mark by 1 if the Algorithm is giving correct answer, and 0 otherwise. So, for each algorithm, we will have size n vector of 0-s and 1-s.

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better. We are running our algorithms over n Test Cases, and mark by 1 if the Algorithm is giving correct answer, and 0 otherwise. So, for each algorithm, we will have size n vector of 0-s and 1-s. So we will have something like:

$$Algo1 : 0, 1, 1, 0, 0, 1, 0$$

$$Algo2 : 1, 1, 1, 0, 0, 1, 1$$

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better. We are running our algorithms over n Test Cases, and mark by 1 if the Algorithm is giving correct answer, and 0 otherwise. So, for each algorithm, we will have size n vector of 0-s and 1-s. So we will have something like:

$$Algo1 : 0, 1, 1, 0, 0, 1, 0$$

$$Algo2 : 1, 1, 1, 0, 0, 1, 1$$

Here we have paired results, since, the first elements in each line, 0 and 1, are the results of running Algorithms on the same Test case.

Example

Example: Assume we have 2 Classification Algorithms, and we want to see which one is better. We are running our algorithms over n Test Cases, and mark by 1 if the Algorithm is giving correct answer, and 0 otherwise. So, for each algorithm, we will have size n vector of 0-s and 1-s. So we will have something like:

$$Algo1 : 0, 1, 1, 0, 0, 1, 0$$

$$Algo2 : 1, 1, 1, 0, 0, 1, 1$$

Here we have paired results, since, the first elements in each line, 0 and 1, are the results of running Algorithms on the same Test case. Now, we want to do a Test - paired t-Test, to see if the Probability of correct Classification is the same for each Algorithm.

Example, cont'd

Here is the code for some hypothetical results from two hypothetical Algorithms:

```
res.algo1 <- rbinom(300, size = 1, prob = 0.4)
res.algo2 <- rbinom(300, size = 1, prob = 0.45)
t.test(res.algo1,res.algo2,paired = T)
```

```
##
##  Paired t-test
##
## data:  res.algo1 and res.algo2
## t = -1.0952, df = 299, p-value = 0.2743
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -0.1212004  0.0345337
## sample estimates:
## mean of the differences
##             -0.04333333
```

Likelihood Ratio Test

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$;

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$; We assume the Parametric Family $\mathcal{F}_\theta$ is given by its PD(M)Fs $f(x|\theta)$.

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$; We assume the Parametric Family $\mathcal{F}_\theta$ is given by its PD(M)Fs $f(x|\theta)$.

Hypothesis: $\mathcal{H}_0 : \ \theta \in \Theta_0 \quad \text{vs} \quad \mathcal{H}_1 : \ \theta \in \Theta_0^c$

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$; We assume the Parametric Family $\mathcal{F}_\theta$ is given by its PD(M)Fs $f(x|\theta)$.

Hypothesis: $\mathcal{H}_0: \ \theta \in \Theta_0 \quad \text{vs} \quad \mathcal{H}_1: \ \theta \in \Theta_0^c$

Significance Level: $\alpha \in (0, 1)$;

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$; We assume the Parametric Family $\mathcal{F}_\theta$ is given by its PD(M)Fs $f(x|\theta)$.

Hypothesis: $\mathcal{H}_0 : \ \theta \in \Theta_0 \quad \text{vs} \quad \mathcal{H}_1 : \ \theta \in \Theta_0^c$

Significance Level: $\alpha \in (0, 1)$;

Rejection Region: we define the **Likelihood Ratio** by

$$LR = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}$$

Likelihood Ratio Test (LRT)

LRT is one of a general methods to Design a Test.

Model: $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$, $\theta \in \Theta$; We assume the Parametric Family $\mathcal{F}_\theta$ is given by its PD(M)Fs $f(x|\theta)$.

Hypothesis: $\mathcal{H}_0: \ \theta \in \Theta_0 \quad \text{vs} \quad \mathcal{H}_1: \ \theta \in \Theta_0^c$

Significance Level: $\alpha \in (0, 1)$;

Rejection Region: we define the **Likelihood Ratio** by

$$LR = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}$$

where $\mathcal{L}(\theta)$ is our GOF *Likelihood Function*:

$$\mathcal{L}(\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdot ... \cdot f(X_n|\theta).$$

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$.

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$. If $LR \approx 0$, then the Likelihood of our Data for $\theta \in \Theta_0$ is pretty small compared to the one with $\theta \notin \Theta_0$.

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$. If $LR \approx 0$, then the Likelihood of our Data for $\theta \in \Theta_0$ is pretty small compared to the one with $\theta \notin \Theta_0$. So we do not believe that $\theta \in \Theta_0$, and we reject the Null.

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$. If $LR \approx 0$, then the Likelihood of our Data for $\theta \in \Theta_0$ is pretty small compared to the one with $\theta \notin \Theta_0$. So we do not believe that $\theta \in \Theta_0$, and we reject the Null.

Note: If we have the MLE for θ, $\hat{\theta}^{MLE}$, then, clearly,

$$\sup_{\theta \in \Theta} \mathcal{L}(\theta) =$$

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$. If $LR \approx 0$, then the Likelihood of our Data for $\theta \in \Theta_0$ is pretty small compared to the one with $\theta \notin \Theta_0$. So we do not believe that $\theta \in \Theta_0$, and we reject the Null.

Note: If we have the MLE for θ, $\hat{\theta}^{MLE}$, then, clearly,

$$\sup_{\theta \in \Theta} \mathcal{L}(\theta) = \mathcal{L}\left(\hat{\theta}^{MLE}\right).$$

Likelihood Ratio Test (LRT), Cont'd

Now, our RR is of the form:

$$RR = \{LR \leq c\}$$

where we choose $c \in (0, 1)$ from the requirement to have a Significance Level α:

$$\sup_{\Theta_0} \mathbb{P}(\text{Type I Error}) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}(LR \leq c \mid \theta = \theta_0) \leq \alpha.$$

Tests, obtained by this method, are called LRTs.

Idea: Clearly, $0 \leq LR \leq 1$. If $LR \approx 0$, then the Likelihood of our Data for $\theta \in \Theta_0$ is pretty small compared to the one with $\theta \notin \Theta_0$. So we do not believe that $\theta \in \Theta_0$, and we reject the Null.

Note: If we have the MLE for θ, $\hat{\theta}^{MLE}$, then, clearly,

$$\sup_{\theta \in \Theta} \mathcal{L}(\theta) = \mathcal{L}\left(\hat{\theta}^{MLE}\right).$$

And similarly, $\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)$ can be thought as a *restricted MLE for* θ *over* Θ_0.

Neyman-Pearson Lemma

Assume $X_1, ..., X_n \sim \mathcal{F}_\theta$, and we want to Test, at α-level, a Simple Hypothesis vs Simple Hypothesis:

$$\mathcal{H}_0 : \ \theta = \theta_0 \qquad vs \qquad \mathcal{H}_1 : \ \theta = \theta_1.$$

Neyman-Pearson Lemma

Assume $X_1, ..., X_n \sim \mathcal{F}_\theta$, and we want to Test, at α-level, a Simple Hypothesis vs Simple Hypothesis:

$$\mathcal{H}_0 : \ \theta = \theta_0 \qquad vs \qquad \mathcal{H}_1 : \ \theta = \theta_1.$$

Let $\mathcal{L}(\theta)$ be the Likelihood Function for $X_1, ..., X_n$, and let

$$LR = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)}$$

be the Likelihood Ratio for our Test.

Neyman-Pearson Lemma

Assume $X_1, ..., X_n \sim \mathcal{F}_\theta$, and we want to Test, at α-level, a Simple Hypothesis vs Simple Hypothesis:

$$\mathcal{H}_0 : \ \theta = \theta_0 \qquad vs \qquad \mathcal{H}_1 : \ \theta = \theta_1.$$

Let $\mathcal{L}(\theta)$ be the Likelihood Function for $X_1, ..., X_n$, and let

$$LR = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)}$$

be the Likelihood Ratio for our Test. Assume there exists a number c such that for the Rejection Region

$$RR = \{LR \leq c\},$$

the significance level is α:

$$\mathbb{P}(LR \leq c \mid \mathcal{H}_0 \text{ is true }) = \alpha.$$

Neyman-Pearson Lemma

Assume $X_1, ..., X_n \sim \mathcal{F}_\theta$, and we want to Test, at α-level, a Simple Hypothesis vs Simple Hypothesis:

$$\mathcal{H}_0 : \ \theta = \theta_0 \qquad vs \qquad \mathcal{H}_1 : \ \theta = \theta_1.$$

Let $\mathcal{L}(\theta)$ be the Likelihood Function for $X_1, ..., X_n$, and let

$$LR = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)}$$

be the Likelihood Ratio for our Test. Assume there exists a number c such that for the Rejection Region

$$RR = \{LR \leq c\},$$

the significance level is α:

$$\mathbb{P}(LR \leq c \mid \mathcal{H}_0 \text{ is true }) = \alpha.$$

Then this Test has the highest Power among all other Tests of Significance Level α.

Example: LRT

Assume $X_1, X_2, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**.

Example: LRT

Assume $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**. We want to construct the LRT, for Significance Level α, for

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \ \mu \neq \mu_0.$$

Example: LRT

Assume $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**. We want to construct the LRT, for Significance Level α, for

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1: \ \mu \neq \mu_0.$$

Step 1: We calculate the Likelihood:

Example: LRT

Assume $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**. We want to construct the LRT, for Significance Level α, for

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1 : \ \mu \neq \mu_0.$$

Step 1: We calculate the Likelihood: the PDF is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}},$$

Example: LRT

Assume $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**. We want to construct the LRT, for Significance Level α, for

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1 : \ \mu \neq \mu_0.$$

Step 1: We calculate the Likelihood: the PDF is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}},$$

and the Likelihood Function is:

$$\mathcal{L}(\mu) = f(X_1|\mu, \sigma^2) \cdot f(X_2|\mu, \sigma^2) \cdot ... \cdot f(X_n|\mu, \sigma^2)$$

Example: LRT

Assume $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and σ^2 is **known**. We want to construct the LRT, for Significance Level α, for

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1 : \ \mu \neq \mu_0.$$

Step 1: We calculate the Likelihood: the PDF is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}},$$

and the Likelihood Function is:

$$\mathcal{L}(\mu) = f(X_1|\mu, \sigma^2) \cdot f(X_2|\mu, \sigma^2) \cdot ... \cdot f(X_n|\mu, \sigma^2)$$

i.e.

$$\mathcal{L}(\mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum_{k=1}^{n} (X_k - \mu)^2}$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} =$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\left(\hat{\mu}^{MLE}\right),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\Big(\hat{\mu}^{MLE}\Big),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.And, of course, we know that in this model

$$\hat{\mu}^{MLE} =$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\left(\hat{\mu}^{MLE}\right),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.And, of course, we know that in this model

$$\hat{\mu}^{MLE} = \overline{X}.$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\left(\hat{\mu}^{MLE}\right),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.And, of course, we know that in this model

$$\hat{\mu}^{MLE} = \overline{X}.$$

Plugging the obtained into the LR formula, we will obtain:

$$LR = \frac{\mathcal{L}(\mu_0)}{\mathcal{L}(\overline{X})} =$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\Big(\hat{\mu}^{MLE}\Big),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.And, of course, we know that in this model

$$\hat{\mu}^{MLE} = \overline{X}.$$

Plugging the obtained into the LR formula, we will obtain:

$$LR = \frac{\mathcal{L}(\mu_0)}{\mathcal{L}(\overline{X})} = \frac{e^{-\frac{1}{2\sigma^2}\cdot\sum_{k=1}^{n}(X_k-\mu_0)^2}}{e^{-\frac{1}{2\sigma^2}\cdot\sum_{k=1}^{n}(X_k-\overline{X})^2}} =$$

Example: LRT, Cont'd

Step 2: Now, we calculate the LR:

$$LR = \frac{\sup_{\mu=\mu_0} \mathcal{L}(\mu)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)} = \frac{\mathcal{L}(\mu_0)}{\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu)}$$

We can write

$$\sup_{\mu\in\mathbb{R}} \mathcal{L}(\mu) = \mathcal{L}\Big(\hat{\mu}^{MLE}\Big),$$

since, by the very definition, $\hat{\mu}^{MLE}$ is the global max point of $\mathcal{L}(\mu)$.And, of course, we know that in this model

$$\hat{\mu}^{MLE} = \overline{X}.$$

Plugging the obtained into the LR formula, we will obtain:

$$LR = \frac{\mathcal{L}(\mu_0)}{\mathcal{L}(\overline{X})} = \frac{e^{-\frac{1}{2\sigma^2}\cdot\sum_{k=1}^{n}(X_k-\mu_0)^2}}{e^{-\frac{1}{2\sigma^2}\cdot\sum_{k=1}^{n}(X_k-\overline{X})^2}} = e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2}$$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} =$$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2} \leq c\}$$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2} \cdot (\overline{X} - \mu_0)^2} \leq c\}$$

which is equivalent to

$$-\frac{n}{2\sigma^2} \cdot (\overline{X} - \mu_0)^2 \leq \ln c$$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2} \leq c\}$$

which is equivalent to

$$-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2 \leq \ln c$$

This is equivalent to

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq \sqrt{-2\ln c} =: C$$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2} \leq c\}$$

which is equivalent to

$$-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2 \leq \ln c$$

This is equivalent to

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq \sqrt{-2\ln c} =: C$$

But this is exactly

$$|Z| \geq C$$

where Z is our Z-Test Statistics. So we have arrived at our GOF Z-Test!

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2} \leq c\}$$

which is equivalent to

$$-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2 \leq \ln c$$

This is equivalent to

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq \sqrt{-2\ln c} =: C$$

But this is exactly

$$|Z| \geq C$$

where Z is our Z-Test Statistics. So we have arrived at our GOF Z-Test! And, in particular, we know that value of C: $C =$

Example: LRT, Cont'd

Step 3: Our RR is (c will be calculated later)

$$RR = \{LR \leq c\} = \{e^{-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2} \leq c\}$$

which is equivalent to

$$-\frac{n}{2\sigma^2}\cdot(\overline{X}-\mu_0)^2 \leq \ln c$$

This is equivalent to

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq \sqrt{-2\ln c} =: C$$

But this is exactly

$$|Z| \geq C$$

where Z is our Z-Test Statistics. So we have arrived at our GOF Z-Test! And, in particular, we know that value of C: $C = z_{1-\alpha/2}$.

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

The first reason is that it is easy to “translate” the Rejection Region into the one using the Log-Likelihood Ratio:

$$LR \leq c \quad \text{is equivalent to} \quad -2 \cdot \ln LR \geq -2 \ln c = c'.$$

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

The first reason is that it is easy to “translate” the Rejection Region into the one using the Log-Likelihood Ratio:

$$LR \leq c \quad \text{is equivalent to} \quad -2 \cdot \ln LR \geq -2 \ln c = c'.$$

And, it is much easier to deal with $\ln LR$, since LR is, in some sense, a product of n terms, so the logarithm will have a form of a sum.

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

The first reason is that it is easy to "translate" the Rejection Region into the one using the Log-Likelihood Ratio:

$$LR \leq c \quad \text{is equivalent to} \quad -2 \cdot \ln LR \geq -2 \ln c = c'.$$

And, it is much easier to deal with $\ln LR$, since LR is, in some sense, a product of n terms, so the logarithm will have a form of a sum.

The other important reason (and the reason to have -2 in front) is that, under some conditions on the PD(M)F, the Asymptotic Distribution of $-2 \cdot \ln LR$ under the Null Hypothesis is

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

The first reason is that it is easy to "translate" the Rejection Region into the one using the Log-Likelihood Ratio:

$$LR \leq c \quad \text{is equivalent to} \quad -2 \cdot \ln LR \geq -2 \ln c = c'.$$

And, it is much easier to deal with $\ln LR$, since LR is, in some sense, a product of n terms, so the logarithm will have a form of a sum.

The other important reason (and the reason to have -2 in front) is that, under some conditions on the PD(M)F, the Asymptotic Distribution of $-2 \cdot \ln LR$ under the Null Hypothesis is $\chi^2(\nu)$, where

Log-Likelihood Ratio, its Distribution

Sometimes Statisticians use the Log-Laikelihood Ratio to obtain the LRT: the quantity

$$-2 \cdot \ln LR = -2 \cdot \Big(\ln \big(\sup_{\theta \in \Theta_0} \mathcal{L}(\theta) \big) - \ln \big(\sup_{\theta \in \Theta} \mathcal{L}(\theta) \big) \Big).$$

The first reason is that it is easy to "translate" the Rejection Region into the one using the Log-Likelihood Ratio:

$$LR \leq c \quad \text{is equivalent to} \quad -2 \cdot \ln LR \geq -2 \ln c = c'.$$

And, it is much easier to deal with $\ln LR$, since LR is, in some sense, a product of n terms, so the logarithm will have a form of a sum.

The other important reason (and the reason to have -2 in front) is that, under some conditions on the PD(M)F, the Asymptotic Distribution of $-2 \cdot \ln LR$ under the Null Hypothesis is $\chi^2(\nu)$, where

$$\nu = \dim \Theta - \dim \Theta_0,$$

and $\dim \Theta$ and $\dim \Theta_0$ are the numbers of Free Parameters under Θ and Θ_0, respectively.

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \ \mu \neq \mu_0$$

and σ is given,

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$).

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1 : \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$). So $\dim \Theta_0 = 0$.

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$). So $\dim \Theta_0 = 0$. Under $\mathcal{H}_1$, μ is our only Free Parameter, so $\dim \Theta = 1$.

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$). So $\dim \Theta_0 = 0$. Under $\mathcal{H}_1$, μ is our only Free Parameter, so $\dim \Theta = 1$. Thus,

$$df = \nu = \dim \Theta - \dim \Theta_0 = 1,$$

and

$$-2 \cdot \ln LR \approx \chi^2(1).$$

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0 : \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1 : \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$). So $\dim \Theta_0 = 0$. Under $\mathcal{H}_1$, μ is our only Free Parameter, so $\dim \Theta = 1$. Thus,

$$df = \nu = \dim \Theta - \dim \Theta_0 = 1,$$

and

$$-2 \cdot \ln LR \approx \chi^2(1).$$

In fact, it is easy to calculate in this case, for the Normal Model with given σ, that exactly

$$-2 \cdot \ln LR \sim \chi^2(1).$$

Log-Likelihood Ratio, its Distribution

For example, if we are Testing for the $\mathcal{N}(\mu, \sigma^2)$ Model

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu \neq \mu_0$$

and σ is given, then under $\mathcal{H}_0$ we do not have any Free Parameters (the only Parameter is μ, but under $\mathcal{H}_0$ the value of μ is fixed, $\mu = \mu_0$). So $\dim \Theta_0 = 0$. Under $\mathcal{H}_1$, μ is our only Free Parameter, so $\dim \Theta = 1$. Thus,

$$df = \nu = \dim \Theta - \dim \Theta_0 = 1,$$

and

$$-2 \cdot \ln LR \approx \chi^2(1).$$

In fact, it is easy to calculate in this case, for the Normal Model with given σ, that exactly

$$-2 \cdot \ln LR \sim \chi^2(1).$$

And having this Distribution, we can find the Rejection Region using

$$\mathbb{P}(\text{Reject } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ is True}) = \mathbb{P}(-2 \ln LR \geq c') = \alpha.$$

p-Values

3 Methods of Decision Making in Testing

When Testing a Hypothesis, at a Significance Level α, we can make the Decision in 3 ways:

3 Methods of Decision Making in Testing

When Testing a Hypothesis, at a Significance Level α, we can make the Decision in 3 ways:

- Based on the Test Statistics TS and the Rejection Region RR: if $TS \in RR$, then we Reject Null Hypothesis, othewise, we Fail to Reject Null;

3 Methods of Decision Making in Testing

When Testing a Hypothesis, at a Significance Level α, we can make the Decision in 3 ways:

- Based on the Test Statistics TS and the Rejection Region RR: if $TS \in RR$, then we Reject Null Hypothesis, othewise, we Fail to Reject Null;
- Based on the Confidence Interval CI for the Parameter: if θ is our Parameter, (L, U) is a CI of $(1-\alpha)$-level for θ, and our Null is $\mathcal{H}_0: \ \theta = \theta_0$, then we Reject Null if and only if $\theta_0 \notin (L, U)$;

3 Methods of Decision Making in Testing

When Testing a Hypothesis, at a Significance Level α, we can make the Decision in 3 ways:

- Based on the Test Statistics TS and the Rejection Region RR: if $TS \in RR$, then we Reject Null Hypothesis, othewise, we Fail to Reject Null;
- Based on the Confidence Interval CI for the Parameter: if θ is our Parameter, (L, U) is a CI of $(1-\alpha)$-level for θ, and our Null is $\mathcal{H}_0: \ \theta = \theta_0$, then we Reject Null if and only if $\theta_0 \notin (L, U)$;
- Based on the p-Value: if $p < \alpha$, Reject Null, otherwise, Fail to Reject.

3 Methods of Decision Making in Testing

When Testing a Hypothesis, at a Significance Level α, we can make the Decision in 3 ways:

- Based on the Test Statistics TS and the Rejection Region RR: if $TS \in RR$, then we Reject Null Hypothesis, othewise, we Fail to Reject Null;
- Based on the Confidence Interval CI for the Parameter: if θ is our Parameter, (L, U) is a CI of $(1-\alpha)$-level for θ, and our Null is $\mathcal{H}_0: \ \theta = \theta_0$, then we Reject Null if and only if $\theta_0 \notin (L, U)$;
- Based on the p-Value: if $p < \alpha$, Reject Null, otherwise, Fail to Reject.

Note: When doing Tests, say, with `t.test`, **R** is calculating the p-Value, and sometimes also the CI. So, to decide whether to Reject Null or Not, using **R**, you can use the 2nd and 3rd Methods.

p-Values

Now, about the p-Value.

p-Values

Now, about the p-Value. Assume we are Testing a Hypothesis

$$\mathcal{H}_0 \qquad vs \qquad \mathcal{H}_1$$

based on the Test Statistics TS.

p-Values

Now, about the p-Value. Assume we are Testing a Hypothesis

$$\mathcal{H}_0 \qquad vs \qquad \mathcal{H}_1$$

based on the Test Statistics TS. Assume we already have Observations, and we calculate the value of TS, let us denote that by TS_{obs} (this is just a number).

p-Values

Now, about the p-Value. Assume we are Testing a Hypothesis

$$\mathcal{H}_0 \qquad vs \qquad \mathcal{H}_1$$

based on the Test Statistics TS. Assume we already have Observations, and we calculate the value of TS, let us denote that by TS_{obs} (this is just a number). We know that, for a given Significance Level α, we will Reject $\mathcal{H}_0$, iff TS_{obs} will be in the RR.

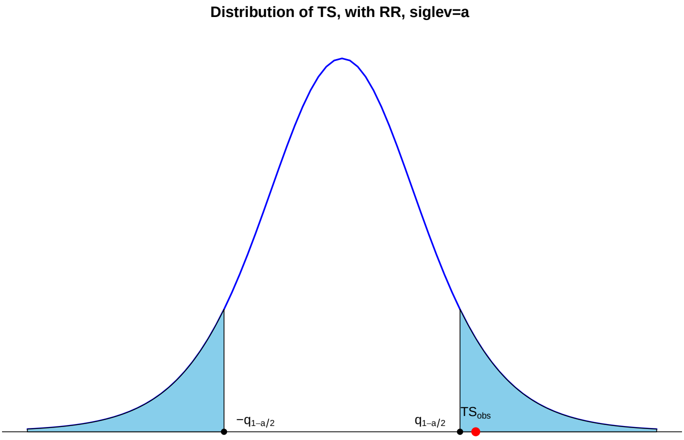
Now, assume the Distribution of TS, our Test Statistics, **under** $\mathcal{H}_0$, is given like this (I am drawing for Z- or t-Statistics, for Two Tailed Test, the other cases can be considered in a similar way):

p-Values

Let us take a Significance Level $a \in (0, 1)$:

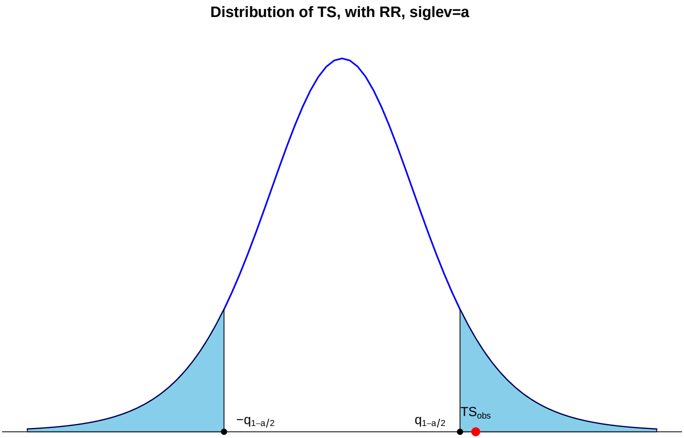
p-Values

Let us take a Significance Level $a \in (0, 1)$:



p-Values

Let us take a Significance Level $a \in (0, 1)$:



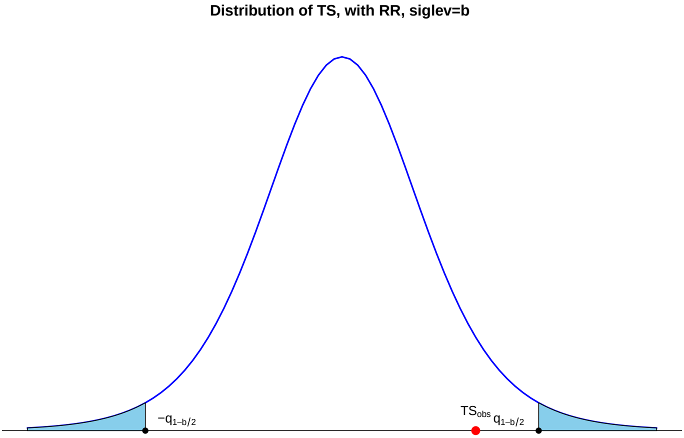
We Reject $\mathcal{H}_0$ at the level a

p-Values

Now, let us change our Significance Level to $b < a$:

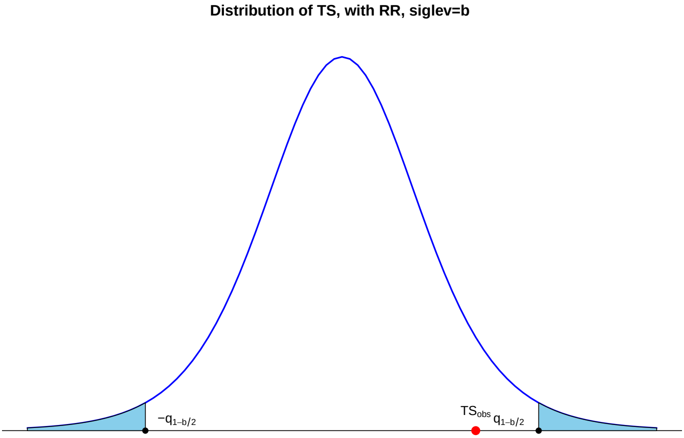
p-Values

Now, let us change our Significance Level to $b < a$:



p-Values

Now, let us change our Significance Level to $b < a$:



We Do Not Reject $\mathcal{H}_0$ at the level b

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$.

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

Then we will have a point $\alpha^* \in (0, 1)$ such that

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

Then we will have a point $\alpha^* \in (0, 1)$ such that

- We Reject $\mathcal{H}_0$ for any $\alpha > \alpha^*$

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

Then we will have a point $\alpha^* \in (0, 1)$ such that

- We Reject $\mathcal{H}_0$ for any $\alpha > \alpha^*$
- We Fail to Reject $\mathcal{H}_0$ for any $\alpha < \alpha^*$

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

Then we will have a point $\alpha^* \in (0, 1)$ such that

- We Reject $\mathcal{H}_0$ for any $\alpha > \alpha^*$
- We Fail to Reject $\mathcal{H}_0$ for any $\alpha < \alpha^*$

Note: Give here the real line with picture, MP!

p-Values

Now, it is clear that, having our Observed Test Statistics TS_{obs}, we will Reject our Null for some large values of α, but Fail to Reject for very small values of α.

Also, clearly, if we are Rejecting Null at the level α, then we will Reject also at any level $\alpha' \geq \alpha$. And, similarly, if we Fail to Reject at the level β, then we also will Fail to Reject at any level $\beta' \leq \beta$.

Then we will have a point $\alpha^* \in (0, 1)$ such that

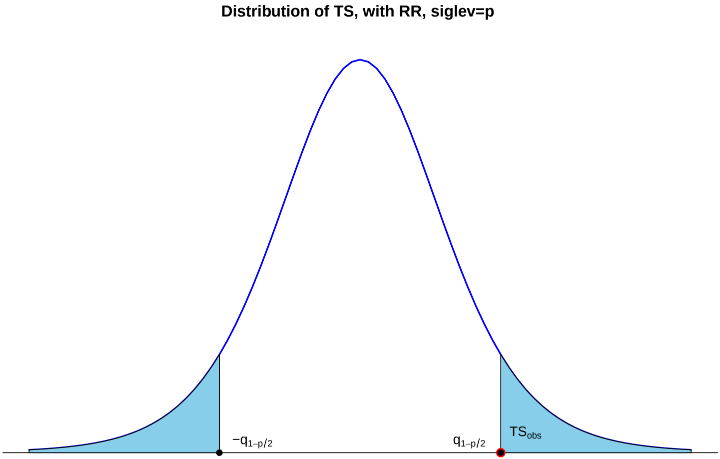
- We Reject $\mathcal{H}_0$ for any $\alpha > \alpha^*$
- We Fail to Reject $\mathcal{H}_0$ for any $\alpha < \alpha^*$

Note: Give here the real line with picture, MP!

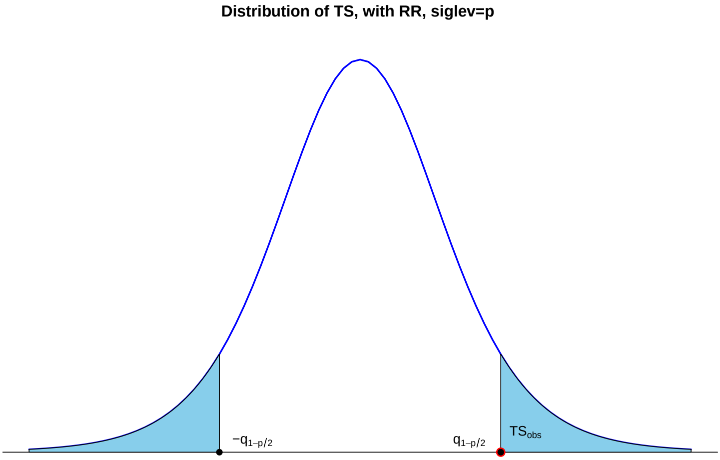
Now, we denote $p = \alpha^*$ and call it the *p*-**Value of the Test:**

$$p\text{-Value} = p = \inf\{\alpha : \text{ we Reject } \mathcal{H}_0 \text{ at level } \alpha\}.$$

p-Values



p-Values



p-Value, the inf value of α at which we Reject $\mathcal{H}_0$

p-Values

It is clear from the Figure above that

$$p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True}).$$

p-Values

It is clear from the Figure above that

$$p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True}).$$

In words, p-Value is the Probability, under the Null Hypothesis, that we will have a value of Test Statistics TS as extreme as the Observed one.

p-Values

It is clear from the Figure above that

$$p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True}).$$

In words, p-Value is the Probability, under the Null Hypothesis, that we will have a value of Test Statistics TS as extreme as the Observed one. I.e., p-Value is measuring somehow *how extreme is our Observed Test Statistics* TS_{obs} under $\mathcal{H}_0$.

p-Values

It is clear from the Figure above that

$$p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True}).$$

In words, p-Value is the Probability, under the Null Hypothesis, that we will have a value of Test Statistics TS as extreme as the Observed one. I.e., p-Value is measuring somehow *how extreme is our Observed Test Statistics* TS_{obs} under $\mathcal{H}_0$.

And, if p-Value is small, then TS_{obs} is very unprobable, very unbelievable, under $\mathcal{H}_0$, so we safely Reject $\mathcal{H}_0$.

p-Values

It is clear from the Figure above that

$$p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True}).$$

In words, p-Value is the Probability, under the Null Hypothesis, that we will have a value of Test Statistics TS as extreme as the Observed one. I.e., p-Value is measuring somehow *how extreme is our Observed Test Statistics* TS_{obs} under $\mathcal{H}_0$.

And, if p-Value is small, then TS_{obs} is very unprobable, very unbelievable, under $\mathcal{H}_0$, so we safely Reject $\mathcal{H}_0$.

To Remember:

- If p-Value$< \alpha$, then we Reject $\mathcal{H}_0$
- If p-Value$\geq \alpha$, then we Fail to Reject $\mathcal{H}_0$

R Code for the Graphics

```
df <- 8;
x <- seq(-4,4,0.1); y <- dt(x, df = df)
plot.new()
plot.window(xlim = c(-4, 4), ylim = c(-0.05,0.4))
plot(x,y, type="l",col="blue",lwd=2,xaxt="n",yaxt="n",
     bty="n",xlab="",ylab="")
abline(h=0)
title("Distribution of TS, with RR, siglev=a ")
qpoint <- 1.5; tspoint <- 1.7
cord.x <- c(qpoint,seq(qpoint,4,0.01),4)
cord.y <- c(0,dt(seq(qpoint,4,0.01), df=df),0)
polygon(cord.x,cord.y,col='skyblue')
points(c(qpoint), c(0), pch=20, cex=1.4)
text(c(qpoint-0.38),c(0.01),labels=expression("q"[1-a/2]))
cord.x1 <- c(-4,seq(-4,-qpoint,0.01),-qpoint)
cord.y1 <- c(0,dt(seq(-4,-qpoint,0.01), df=df),0)
polygon(cord.x1,cord.y1,col='skyblue')
points(c(-qpoint), c(0), pch=20, cex=1.4)
text(c(-qpoint+0.4),c(0.01),labels=expression("-q"[1-a/2]))
points(c(tspoint), c(0), col="red", pch=19, cex=1.4)
text(c(tspoint), c(0.02), labels = expression("TS"[obs]))
```

Example

Example: Assume we have an observation from $\mathcal{N}(\mu, 3^2)$, and we want to Test

$$\mathcal{H}_0 : \ \mu = 1.2 \qquad \textit{vs} \qquad \mathcal{H}_1 : \ \mu \neq 1.2.$$

Example

Example: Assume we have an observation from $\mathcal{N}(\mu, 3^2)$, and we want to Test

$$\mathcal{H}_0: \ \mu = 1.2 \qquad \textit{vs} \qquad \mathcal{H}_1: \ \mu \neq 1.2.$$

So we will do a Z-Test.

Example

Example: Assume we have an observation from $\mathcal{N}(\mu, 3^2)$, and we want to Test

$$\mathcal{H}_0 : \ \mu = 1.2 \qquad \textit{vs} \qquad \mathcal{H}_1 : \ \mu \neq 1.2.$$

So we will do a Z-Test. Assume that after plugging our observations into the Z-Statistics, we get

$$Z_{obs} = 1.72.$$

Example

Example: Assume we have an observation from $\mathcal{N}(\mu, 3^2)$, and we want to Test

$$\mathcal{H}_0: \ \mu = 1.2 \qquad vs \qquad \mathcal{H}_1: \ \mu \neq 1.2.$$

So we will do a Z-Test. Assume that after plugging our observations into the Z-Statistics, we get

$$Z_{obs} = 1.72.$$

Then, the p-Value of the Test is

$$p\text{-Value} = \mathbb{P}(|Z| > Z_{obs} \mid \text{Null is True}) = \mathbb{P}\Big(|Z| > 1.72 \mid Z \sim \mathcal{N}(0,1)\Big)$$

Example

Example: Assume we have an observation from $\mathcal{N}(\mu, 3^2)$, and we want to Test

$$\mathcal{H}_0: \ \mu = 1.2 \qquad vs \qquad \mathcal{H}_1: \ \mu \neq 1.2.$$

So we will do a Z-Test. Assume that after plugging our observations into the Z-Statistics, we get

$$Z_{obs} = 1.72.$$

Then, the p-Value of the Test is

$$p\text{-Value} = \mathbb{P}(|Z| > Z_{obs} \mid \text{Null is True}) = \mathbb{P}\Big(|Z| > 1.72 \mid Z \sim \mathcal{N}(0,1)\Big)$$

and the value of p-Value is ⌣̈

```
1-(pnorm(1.72,mean=0,sd=1)-pnorm(-1.72,mean=0,sd=1))
```

```
## [1] 0.08543244
```