AUA CS 108, Statistics, Fall 2019

Lecture 42

Michael Poghosyan
YSU, AUA
michael@ysu.am, mpoghosyan@aua.am

02 Dec 2019

Contents

Last Lecture ReCap

Last Lecture ReCap

- Give the definition of the p-Value of a Test.

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

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

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- p-Value $= p = \inf\{\alpha :$ we Reject $\mathcal{H}_0$ at level $\alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- p-Value $= p = \inf\{\alpha :$ we Reject $\mathcal{H}_0$ at level $\alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;
- p-Value $= p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0$ is True $)$;

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- p-Value $= p = \inf\{\alpha : \text{ we Reject } \mathcal{H}_0 \text{ at level } \alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;
- p-Value $= p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True })$; i.e., 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

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- p-Value $= p = \inf\{\alpha : \text{ we Reject } \mathcal{H}_0 \text{ at level } \alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;
- p-Value $= p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True })$; i.e., 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. In other words, p-Value is measuring somehow *how extreme is our Observed Test Statistics* TS_{obs} under $\mathcal{H}_0$.

p-Values

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- p-Value $= p = \inf\{\alpha : \text{ we Reject } \mathcal{H}_0 \text{ at level } \alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;
- p-Value $= p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True })$; i.e., 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. In other words, 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

In our last lecture we gave two equivalent definitions of the p-Value of a Test:

- $p\text{-Value} = p = \inf\{\alpha : \text{ we Reject } \mathcal{H}_0 \text{ at level } \alpha\}$; i.e., given observations, p-Value is the smallest level of α at which we can Reject Null;
- $p\text{-Value} = p = \mathbb{P}(|TS| > TS_{obs} \mid \mathcal{H}_0 \text{ is True })$; i.e., 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. In other words, 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 $p\text{-Value} < \alpha$, then we Reject $\mathcal{H}_0$
- If $p\text{-Value} \geq \alpha$, then we Fail to Reject $\mathcal{H}_0$

Example

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

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

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

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

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 $\ddot\smile$

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

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

Linear Regression

Intro to Linear Regression

Note: From this slide on, **LR** will mean Linear Regression, not the Likelihood Ratio $\ddot\smile$

Intro to Linear Regression

Note: From this slide on, **LR** will mean Linear Regression, not the Likelihood Ratio $\ddot\smile$

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

Intro to Linear Regression

Note: From this slide on, **LR** will mean Linear Regression, not the Likelihood Ratio $\ddot\smile$

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;

Intro to Linear Regression

Note: From this slide on, **LR** will mean Linear Regression, not the Likelihood Ratio $\ddot\smile$

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;
- to find a CI for θ of given CL;

Intro to Linear Regression

Note: From this slide on, **LR** will mean Linear Regression, not the Likelihood Ratio $\ddot\smile$

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;
- to find a CI for θ of given CL;
- to Test a Hypothesis about θ, say, is it likely that $\theta = 3.1415$ or not.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets. That was defined for two **Numerical Dataset**, without any assumptions behind the Process generating that Datasets.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets. That was defined for two **Numerical Dataset**, without any assumptions behind the Process generating that Datasets. Now, if we assume that that Datasets are coming from some Distribution, we are at the stage of doing a Statistical Inference, Statistical Analysis.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Note: The coordinates of $\mathbf{X}$ are called **Explanatory, Predictor, Independent Variables or Features**,

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Note: The coordinates of $\mathbf{X}$ are called **Explanatory, Predictor, Independent Variables or Features**, and Y is called the **Dependent, Response Variable or the Label**.

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS.

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.

Examples

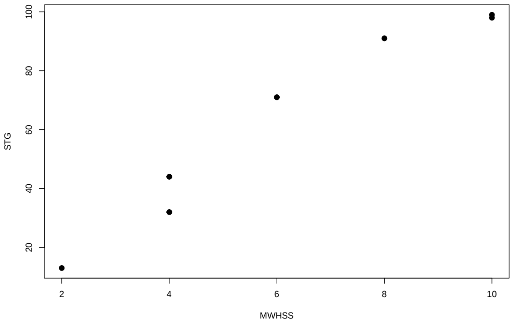
Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.To estimate the relationship, we need some data, say, we can have:

$$(10, 98);\ (6, 71);\ (10, 99);\ (4, 44);\ (2, 13);\ (8; 91);\ (4, 32).$$

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.To estimate the relationship, we need some data, say, we can have:

$$(10, 98);\ (6, 71);\ (10, 99);\ (4, 44);\ (2, 13);\ (8; 91);\ (4, 32).$$



Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S).

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), \dots$$

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), \ldots$$

Example: (from Kaggle, Link) We want to predict the House Price by having some Features of teh House (say, Floor, Zone, YearBuilt, RoofStyle etc). So we want to model the relationship between the House Price and all other variables.

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), \dots$$

Example: (from Kaggle, Link) We want to predict the House Price by having some Features of teh House (say, Floor, Zone, YearBuilt, RoofStyle etc). So we want to model the relationship between the House Price and all other variables. The Number of explanatory variables is 79. So here $\mathbf{X}$ is 79-D.