Significance of regression

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 31

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Simple linear regression: a quick review

Last time, we introduced **simple linear regression**. More specifically we talked about the **simple two-dimensional case** of predicting one dependent variable y using one independent variable x:

- $(x_i, y_i), i = 1, \ldots, n$: a series of n data points.
- Main idea: connect all points through line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - β_0: intercept;
  - β_1: slope;
  - ϵ_i: "noise" associated with point i $\sim \mathcal{N}(0, \sigma^2)$
- To find the "best" β_0, β_1, we optimize the least squares function:

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Simple linear regression: a quick review

Last time, we introduced **simple linear regression**. More specifically we talked about the **simple two-dimensional case** of predicting one dependent variable y using one independent variable x:

- (x_i, y_i), $i = 1, \ldots, n$: a series of n data points.
- Main idea: connect all points through line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - β_0: intercept;
  - β_1: slope;
  - ϵ_i: "noise" associated with point i $\sim \mathcal{N}(0, \sigma^2)$
- To find the "best" β_0, β_1, we optimize the least squares function:

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Simple linear regression: a quick review

Last time, we introduced **simple linear regression**. More specifically we talked about the **simple two-dimensional case** of predicting one dependent variable y using one independent variable x:

- (x_i, y_i), $i = 1, \ldots, n$: a series of n data points.
- Main idea: connect all points through line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - β_0: intercept;
  - β_1: slope;
  - ϵ_i: "noise" associated with point i $\sim \mathcal{N}(0, \sigma^2)$
- To find the "best" β_0, β_1, we optimize the least squares function:

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

ILLINOIS

Simple linear regression: a quick review

Last time, we introduced **simple linear regression**. More specifically we talked about the **simple two-dimensional case** of predicting one dependent variable y using one independent variable x:

- (x_i, y_i), $i = 1, \ldots, n$: a series of n data points.
- Main idea: connect all points through line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - β_0: intercept;
  - β_1: slope;
  - ϵ_i: "noise" associated with point i $\sim \mathcal{N}\left(0, \sigma^2\right)$
- To find the "best" β_0, β_1, we optimize the least squares function:

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

ILLINOIS

Simple linear regression: a quick review

Last time, we introduced **simple linear regression**. More specifically we talked about the **simple two-dimensional case** of predicting one dependent variable y using one independent variable x:

- (x_i, y_i), $i = 1, \ldots, n$: a series of n data points.
- Main idea: connect all points through line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - β_0: intercept;
  - β_1: slope;
  - ϵ_i: “noise” associated with point i $\sim \mathcal{N}\left(0, \sigma^2\right)$
- To find the “best” β_0, β_1, we optimize the least squares function:

$$L = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

ILLINOIS

Simple linear regression: a quick review

Taking the derivatives and setting to zero, we obtain:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

So, we now have:

1. the *observed values*: y_i for given x_i;
2. the *fitted values*: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$;
3. the *residuals/errors*: $e_i = y_i - \hat{y}_i$;
4. the *sum of squares of errors*: $SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Simple linear regression: a quick review

Taking the derivatives and setting to zero, we obtain:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

So, we now have:

1. the *observed values*: y_i for given x_i;
2. the *fitted values*: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$;
3. the *residuals/errors*: $e_i = y_i - \hat{y}_i$;
4. the *sum of squares of errors*: $SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Simple linear regression: a quick review

Taking the derivatives and setting to zero, we obtain:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

So, we now have:

1. the *observed values*: y_i for given x_i;
2. the *fitted values*: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$;
3. the *residuals/errors*: $e_i = y_i - \hat{y}_i$;
4. the *sum of squares of errors*: $SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Simple linear regression: a quick review

Taking the derivatives and setting to zero, we obtain:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

So, we now have:

1. the *observed values*: y_i for given x_i;
2. the *fitted values*: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$;
3. the *residuals/errors*: $e_i = y_i - \hat{y}_i$;
4. the *sum of squares of errors*: $SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Simple linear regression: a quick review

Taking the derivatives and setting to zero, we obtain:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

So, we now have:

1. the *observed values*: y_i for given x_i;
2. the *fitted values*: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$;
3. the *residuals/errors*: $e_i = y_i - \hat{y}_i$;
4. the *sum of squares of errors*: $SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Mean square error

- We note here that the sum of squares of errors has $n-2$ degrees of freedom.
  - You may view the "2" in the degrees of freedom as the number of parameters (β_0, β_1) we are estimating in simple linear regression.
- We then define the **mean square error** as:

$$MS_E = \frac{SS_E}{n-2}.$$

- Remember that noise is normally distributed with mean 0 and variance σ^2: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Mean square error

- We note here that the sum of squares of errors has $n-2$ degrees of freedom.
  - You may view the "2" in the degrees of freedom as the number of parameters (β_0, β_1) we are estimating in simple linear regression.
- We then define the **mean square error** as:

$$MS_E = \frac{SS_E}{n-2}.$$

- Remember that noise is normally distributed with mean 0 and variance σ^2: $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$

$$E\left[MS_E\right] = E\left[\frac{SS_E}{n-2}\right] = \sigma^2.$$

And hence: $\boxed{\hat{\sigma}^2 = MS_E}$

Mean square error

- We note here that the sum of squares of errors has $n - 2$ degrees of freedom.
  - You may view the “2” in the degrees of freedom as the number of parameters (β_0, β_1) we are estimating in simple linear regression.
- We then define the **mean square error** as:

$$MS_E = \frac{SS_E}{n-2}.$$

- Remember that noise is normally distributed with mean 0 and variance σ^2: $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$

$$E\left[MS_E\right] = E\left[\frac{SS_E}{n-2}\right] = \sigma^2.$$

And hence: $\boxed{\hat{\sigma}^2 = MS_E}$

Mean square error

- We note here that the sum of squares of errors has $n-2$ degrees of freedom.
  - You may view the "2" in the degrees of freedom as the number of parameters (β_0, β_1) we are estimating in simple linear regression.
- We then define the **mean square error** as:

$$MS_E = \frac{SS_E}{n-2}.$$

- Remember that noise is normally distributed with mean 0 and variance σ^2: $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$

$$E\left[MS_E\right] = E\left[\frac{SS_E}{n-2}\right] = \sigma^2.$$

And hence: $\boxed{\hat{\sigma}^2 = MS_E}$

ILLINOIS

Slope redefined

We can also write that

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

Define:

- $S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$
- $S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$

to get that

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

ILLINOIS

Slope redefined

We can also write that

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

Define:

- $S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$
- $S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$

to get that

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

ILLINOIS

Significance of regression

- So, is our regression *significant*?
- Let us define what this means:
  - Is there enough evidence to suggest that x and y are related?
  - Or is it a “random” phenomenon?
- Hypothesis testing!

When are x and y unrelated? When $\beta_1 = 0$.

Significance of regression

- So, is our regression *significant*?
- Let us define what this means:
  - Is there enough evidence to suggest that x and y are related?
  - Or is it a "random" phenomenon?
- Hypothesis testing!

When are x and y unrelated? When $\beta_1 = 0$.

Significance of regression

- So, is our regression *significant*?
- Let us define what this means:
  - Is there enough evidence to suggest that x and y are related?
  - Or is it a “random” phenomenon?
- Hypothesis testing!

When are x and y unrelated? When $\beta_1 = 0$.

Significance of regression

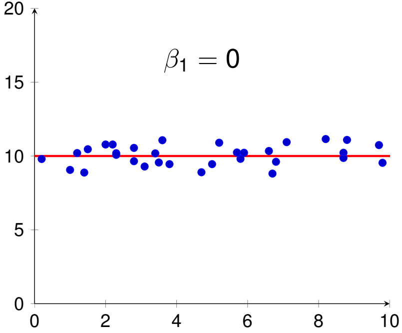
- So, is our regression *significant*?
- Let us define what this means:
  - Is there enough evidence to suggest that x and y are related?
  - Or is it a “random” phenomenon?
- Hypothesis testing!

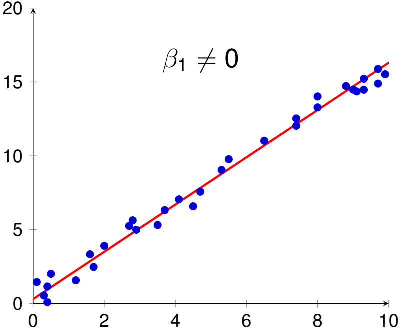
When are x and y unrelated? When $\beta_1 = 0$.

Significance of regression

- So, is our regression *significant*?
- Let us define what this means:
  - Is there enough evidence to suggest that x and y are related?
  - Or is it a “random” phenomenon?
- Hypothesis testing!

When are x and y unrelated? When $\beta_1 = 0$.





Hypothesis testing for significance of regression

We have:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0.$$

But, how is $\hat{\beta}_1$ distributed?

- Recall that $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
- We then have that:

$$\beta_1 : \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \to \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 is known: z test statistic.
- If σ^2 is unknown, then use an estimator for σ^2: t test statistic.

Which estimator can we use for σ?

Recall that $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2}$.

ILLINOIS

Hypothesis testing for significance of regression

We have:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0.$$

But, how is $\hat{\beta}_1$ distributed?

- Recall that $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
- We then have that:

$$\beta_1 : \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \rightarrow \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 is known: z test statistic.
- If σ^2 is unknown, then use an estimator for σ^2: t test statistic.

Which estimator can we use for σ?

Recall that $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2}$.

ILLINOIS

Hypothesis testing for significance of regression

We have:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0.$$

But, how is $\hat{\beta}_1$ distributed?

- Recall that $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$.
- We then have that:

$$\beta_1\text{:}\ \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right) \to \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 is known: z test statistic.
- If σ^2 is unknown, then use an estimator for σ^2: t test statistic.

Which estimator can we use for σ?

Recall that $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2}$.

Hypothesis testing for significance of regression

We have:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0.$$

But, how is $\hat{\beta}_1$ distributed?

- Recall that $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$.
- We then have that:

$$\beta_1\text{:} \quad \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) \rightarrow \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 is known: z test statistic.
- If σ^2 is unknown, then use an estimator for σ^2: t test statistic.

Which estimator can we use for σ?

Recall that $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2}$.

ILLINOIS

Hypothesis testing for significance of regression

We have:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0.$$

But, how is $\hat{\beta}_1$ distributed?

- Recall that $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$.
- We then have that:

$$\beta_1\text{:}\ \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) \to \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 is known: z test statistic.
- If σ^2 is unknown, then use an estimator for σ^2: t test statistic.

Which estimator can we use for σ?

Recall that $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2}$.

ILLINOIS

Procedure for significance testing

Null hypothesis:

$$H_0 : \beta_1 = 0.$$

Test statistic:

$$T_0 = \frac{\hat{\beta}_1}{\sqrt{MSE/S_{xx}}}.$$

Distribution under H_0:

$$T_0 \sim T_{n-2}.$$

H_1	Rejection region	CI region
$\beta_1 \neq 0$	$\lvert T_0 \rvert > t_{\alpha/2,n-2}$	$\left[\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}}, \hat{\beta}_1 + t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}}\right]$

Reject if:

- T_0 falls in the rejection region, or
- 0 falls outside the CI region.