Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu
chenle02@gmail.com

Emory University
Atlanta, GA

Last updated on Spring 2021
Last compiled on January 15, 2023

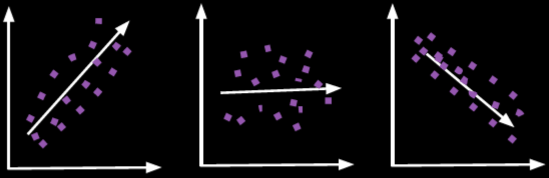
2021 Spring

Chapter 11. Regression

Plan

Chapter 11. Regression

CORRELATION



Positive Correlation

Zero Correlation

Negative Correlation

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}$$

Covarianced normalized by Standard Deviation

Correlation between X and Y

Standard deviation of X

Standard deviation of Y

Notation: $Corr(X, Y) = \rho(X, Y) = \rho_{XY}$

Computing: $\mathrm{Var}(X) = \sigma_X^2$, $\mathrm{Var}(Y) = \sigma_Y^2$, $\mathrm{Cov}(X, Y) = \sigma_{XY}$

$$\Downarrow$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\mathrm{Corr}(X,Y) = \underbrace{\frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}}_{\text{Covarianced normalized by Standard Deviation}}$$

Correlation between X and Y

Standard deviation of X

Standard deviation of Y

Notation: $Corr(X, Y) = \rho(X, Y) = \rho_{XY}$

Computing: $\mathrm{Var}(X) = \sigma_X^2$, $\mathrm{Var}(Y) = \sigma_Y^2$, $\mathrm{Cov}(X, Y) = \sigma_{XY}$

$$\Downarrow$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}$$

Covarianced normalized by Standard Deviation

Correlation between X and Y

Standard deviation of X

Standard deviation of Y

Notation: $Corr(X, Y) = \rho(X, Y) = \rho_{XY}$

Computing: $\text{Var}(X) = \sigma_X^2$, $\text{Var}(Y) = \sigma_Y^2$, $\text{Cov}(X, Y) = \sigma_{XY}$

$$\Downarrow$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Thm. For any two random variables X and Y,

a. $|\rho(X, Y)| \leq 1$

b. $\rho(X, Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $b \in \mathbb{R}$;
$\rho(X, Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$ and $b \in \mathbb{R}$.

Proof. (a)

$$|\rho(X, Y)| \leq 1$$

$$\Updownarrow$$

$$\begin{aligned} |\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)| &\leq \sqrt{\text{Var}(X)\text{Var}(Y)} \\ &= \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}\sqrt{\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)} \end{aligned}$$

which is nothing but the Cauchy-Schwartz inequality.

Thm. For any two random variables X and Y,

a. $|\rho(X, Y)| \le 1$

b. $\rho(X, Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $b \in \mathbb{R}$; $\rho(X, Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$ and $b \in \mathbb{R}$.

Proof. (a)

$$|\rho(X, Y)| \le 1$$

$$\Updownarrow$$

$$\begin{aligned} |\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)| &\le \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \\ &= \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}\sqrt{\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)} \end{aligned}$$

which is nothing but the Cauchy-Schwartz inequality.

Thm. For any two random variables X and Y,

a. $|\rho(X, Y)| \leq 1$

b. $\rho(X, Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $b \in \mathbb{R}$;
$\rho(X, Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$ and $b \in \mathbb{R}$.

Proof. (a)

$$|\rho(X, Y)| \leq 1$$

$$\Updownarrow$$

$$\begin{aligned} |\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)| &\leq \sqrt{\text{Var}(X)\text{Var}(Y)} \\ &= \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}\sqrt{\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)} \end{aligned}$$

which is nothing but the Cauchy-Schwartz inequality.

Thm. For any two random variables X and Y,

a. $|\rho(X, Y)| \leq 1$

b. $\rho(X, Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $b \in \mathbb{R}$;
$\rho(X, Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$ and $b \in \mathbb{R}$.

Proof. (a)

$$|\rho(X, Y)| \leq 1$$

$$\Updownarrow$$

$$\begin{aligned}|\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)| &\leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \\ &= \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}\sqrt{\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)}\end{aligned}$$

which is nothing but the Cauchy-Schwartz inequality.

Thm. For any two random variables X and Y,

a. $|\rho(X,Y)| \leq 1$

b. $\rho(X,Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $b \in \mathbb{R}$;
$\rho(X,Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$ and $b \in \mathbb{R}$.

Proof. (a)

$$|\rho(X,Y)| \leq 1$$

$$\Updownarrow$$

$$\begin{aligned} |\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right)| &\leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \\ &= \sqrt{\mathbb{E}\left((X-\mathbb{E}(X))^2\right)}\sqrt{\mathbb{E}\left((Y-\mathbb{E}(Y))^2\right)} \end{aligned}$$

which is nothing but the Cauchy-Schwartz inequality.

(b) In the Cauchy-Schwartz inequality, the equality holds if and only if for some $a \neq 0$,

$$X - \mathbb{E}(X) = a[Y - E(Y)]$$

namely,

$$X = aY + b, \qquad \text{with} \quad b = \mathbb{E}(X) - a\mathbb{E}(Y).$$

In particular, $a > 0$ corresponds to the case $\rho(X, Y) = 1$ and $a < 0$ to $\rho(X, Y) = -1$. $\square$

(b) In the Cauchy-Schwartz inequality, the equality holds if and only if for some $a \neq 0$,

$$X - \mathbb{E}(X) = a[Y - E(Y)]$$

namely,

$$X = aY + b, \qquad \text{with} \quad b = \mathbb{E}(X) - a\mathbb{E}(Y).$$

In particular, $a > 0$ corresponds to the case $\rho(X, Y) = 1$ and $a < 0$ to $\rho(X, Y) = -1$. $\square$

Estimating $\rho(X, Y)$ – Sample correlation coefficient

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

$$= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}\sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2}}$$

$$\Downarrow$$

$$R = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2}}$$

Pearson product-moment correlation coefficient

or

Sample correlation coefficient

Estimating $\rho(X, Y)$ – Sample correlation coefficient

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

$$= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}\sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2}}$$

$$\Downarrow$$

$$R = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2}}$$

Pearson product-moment correlation coefficient

or

Sample correlation coefficient

Estimating $\rho(X, Y)$ – Sample correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

$$= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}\sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2}}$$

$$\Downarrow$$

$$R = \frac{n\sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n Y_i\right)}{\sqrt{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2}\sqrt{n\sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i\right)^2}}$$

Pearson product-moment correlation coefficient

or

Sample correlation coefficient

Estimating $\rho(X, Y)$ – Sample correlation coefficient

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

$$= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}\sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2}}$$

$$\Downarrow$$

$$R = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2}}$$

Pearson product-moment correlation coefficient

or

Sample correlation coefficient

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i\right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Thm.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSTR}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2, \quad \widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2, \quad \text{and} \quad SSTR = SST - SSE.$$

Remark SSE: sum of square errors $\sim$ the variation in y_i's not explained by L.M.

SST: Total sum of squares $\sim$ total variability.

SSTR: Treatment sum of sqrs. $\sim$ the variation in y_i's explained by L.M.

R^2 (or r^2 when X_i and Y_i are replaced by x_i and y_i) $\sim$ proportion of total variation in the y_i's that can be attributed to L.M.

Coefficient of determination or simply *R squared*

Proof

□

Adjusted R-squared

Def. The adjusted R-squareed:

$$R^2_{adj} := 1 - \frac{MSE}{MST}$$

where

$$MSE = \frac{SSE}{n-q} \qquad \text{and} \qquad MST = \frac{SST}{n-1}$$

and q is number of parameters in the model.

Relation:

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n-1}{n-q}$$

MSE: Mean squared error.

MST: Mean squared total.

MSR = MSTR: Mean square for treatment (or regresssion).

$$MSR = MSTR = \frac{SSTR}{q-1}$$

Adjusted R-squared

Def. The adjusted R-squareed:

$$R^2_{adj} := 1 - \frac{MSE}{MST}$$

where

$$MSE = \frac{SSE}{n-q} \qquad \text{and} \qquad MST = \frac{SST}{n-1}$$

and q is number of parameters in the model.

Relation:

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-q}$$

MSE: Mean squared error.

MST: Mean squared total.

MSR = MSTR: Mean square for treatment (or regresssion).

$$MSR = MSTR = \frac{SSTR}{q-1}$$

Adjusted R-squared

Def. The adjusted R-squareed:

$$R^2_{adj} := 1 - \frac{MSE}{MST}$$

where

$$MSE = \frac{SSE}{n-q} \quad \text{and} \quad MST = \frac{SST}{n-1}$$

and q is number of parameters in the model.

Relation:

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n-1}{n-q}$$

MSE: Mean squared error.

MST: Mean squared total.

MSR = MSTR: Mean square for treatment (or regresssion).

$$MSR = MSTR = \frac{SSTR}{q-1}$$

Adjusted R-squared

Def. The adjusted R-squareed:

$$R^2_{adj} := 1 - \frac{MSE}{MST}$$

where

$$MSE = \frac{SSE}{n-q} \qquad \text{and} \qquad MST = \frac{SST}{n-1}$$

and q is number of parameters in the model.

Relation:

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-q}$$

MSE: Mean squared error.

MST: Mean squared total.

MSR = MSTR: Mean square for treatment (or regresssion).

$$MSR = MSTR = \frac{SSTR}{q-1}$$

Adjusted R-squared

Def. The adjusted R-squareed:

$$R^2_{adj} := 1 - \frac{MSE}{MST}$$

where

$$MSE = \frac{SSE}{n-q} \qquad \text{and} \qquad MST = \frac{SST}{n-1}$$

and q is number of parameters in the model.

Relation:

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-q}$$

MSE: Mean squared error.

MST: Mean squared total.

MSR = MSTR: Mean square for treatment (or regresssion).

$$MSR = MSTR = \frac{SSTR}{q-1}$$