* Announcement

① Exam prep section : 2~3 PM

↳ Will show $Var(T_1)$, $Var(T_2)$, $Var(T_3)$ when $X_1, \dots, X_n \sim Unif(0, 2\theta)$ (Lecture 35 Page 3)

② Pre-final grade report (11/24)

③ Final logistics (11/23 ~ 11/24)

HW13 (11/23 ~ 11/30) — 2Qs, HW14 (11/30 ~ 12/7) — 3Qs

HW13 (11/30 ~ 12/7) — 5Qs

STAT 88: Lecture 36

Contents

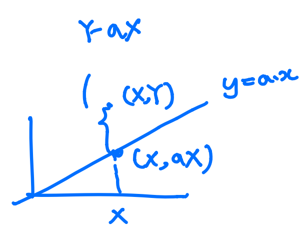
Warm up: (Exercise 11.6.3)

Sometimes data scientists want to fit a linear model that has no intercept term. For example, this might be the case when the data are from a scientific experiement in which the attribute X can have values near 0 and there is a physical reason why the response Y must be 0 when $X = 0$.

So let (X, Y) be a random pair and suppose you want to predict Y by an estimator of the form aX for some a. Find the least squares predictor $\widehat{Y}$ among all predictors of this form.

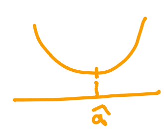
$Y \leftarrow aX$

$MSE(a) = E((Y - aX)^2)$. Find $\hat{a}$ that minimizes $MSE(a)$



$MSE(a) = E(Y^2 - 2aXY + a^2X^2)$

$= E(Y^2) - 2aE(XY) + a^2 \underbrace{E(X^2)}_{\geq 0}$ ↪



$\frac{dMSE(a)}{da} = 0 - 2E(XY) + 2aE(X^2)$.

$\hat{a}$ minimizes $MSE(a)$.

$\frac{dMSE(a)}{da}\Big|_{a=\hat{a}} = 0$

$\Rightarrow 0 - 2E(XY) + 2\hat{a}E(X^2) = 0$

$2\hat{a}E(X^2) = 2E(XY)$

$\Rightarrow \hat{a} = \frac{E(XY)}{E(X^2)}$

$\Rightarrow \hat{Y} = \hat{a} \cdot X$

Last time

Least squares regression

Let (X, Y) be a random pair. We write

- $E(X) = \mu_X$, $\mathrm{SD}(X) = \sigma_X$. kg
- $E(Y) = \mu_Y$, $\mathrm{SD}(Y) = \sigma_Y$. kg
- Correlation

unitless

$$\overset{\text{unitless}}{r} = \frac{\overbrace{E((\underbrace{X - \mu_X}_{= D_X})(\underbrace{Y - \mu_Y}_{= D_Y}))}^{\text{Covariance}}}{\sigma_X \sigma_Y}. \qquad \frac{\cancel{kg^2}}{\cancel{kg^2}}$$

We wish to find the best fitting line $\widehat{Y} = \widehat{a}X + \widehat{b}$, through the scatter plot at all (X, Y) pairs. We showed that

$$\widehat{a} = r\frac{\sigma_Y}{\sigma_X} \text{ and } \widehat{b} = \mu_Y - \widehat{a} \cdot \mu_X.$$

$\hat{a}, \hat{b}$ minimize $MSE(a,b) = E((Y - (aX+b))^2)$

11.4. Bounds on Correlation

linearity

For a random pair (X, Y), the correlation is defined as

$$r = r(X,Y) = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = E\left(\underbrace{\frac{X-\mu_X}{\sigma_X}}_{X^*} \cdot \underbrace{\frac{Y-\mu_Y}{\sigma_Y}}_{Y^*}\right) = E(X^\star Y^\star),$$

where $X^\star$ and $Y^\star$ are standardizations of X and Y respectively.

Our goal is to show that

$$-1 \le r \le 1.$$

As a preliminary, find $E(X^\star)$, $\mathrm{Var}(X^\star)$, and $E(X^{\star 2})$.

$E(X^\star) = 0$, $\mathrm{Var}(X^\star) = 1$, $E(X^{\star 2}) = Var(X^*) + (E(X^*))^2 = 1$

Lower Bound We will show that $r = E(X^\star Y^\star) \ge -1$.

we know $(X^* + Y^*)^2 \geq 0$

$\Rightarrow X^{*2} + Y^{*2} + 2X^*Y^* \geq 0$

$\Rightarrow \underbrace{E(X^{*2})}_{=1} + \underbrace{E(Y^{*2})}_{=1} + 2E(X^*Y^*) \geq 0$

$\Rightarrow 2E(X^*Y^*) \geq -2$

$\Rightarrow r = E(X^*Y^*) \geq -1$

Upper Bound Similarly,

$(X^* - Y^*)^2 \geq 0$

$\Rightarrow X^{*2} + Y^{*2} - 2X^*Y^* \geq 0$

$\Rightarrow \underbrace{E(X^{*2})}_{=1} + \underbrace{E(Y^{*2})}_{=1} - 2E(X^*Y^*) \geq 0$

$\Rightarrow 2 \geq 2E(X^*Y^*)$

$\Rightarrow r = E(X^*Y^*) \leq 1$

Other Properties We can show

(a) $r(X, Y) = r(Y, X)$.

(b) $r(aX+b, cY+d) = \begin{cases} r(X,Y) & \text{if } ac > 0 \\ -r(X,Y) & \text{if } ac < 0 \end{cases}$

(a) $r(X,Y) = E(X^*Y^*) = E(Y^*X^*) = r(Y,X)$

(b) $r(aX+b, cY+d) = E\left((aX+b)^*(cY+d)^*\right)$

$$\begin{cases} E(aX+b) = a\mu_X + b, & E(cY+d) = c\mu_Y + d \\ SD(aX+b) = |a| \cdot \sigma_X, & SD(cY+d) = |c| \cdot \sigma_Y \end{cases}$$

$$(aX+b)^* = \frac{aX + \cancel{b} - (a\mu_X + \cancel{b})}{|a| \cdot \sigma_X} = \frac{a(X-\mu_X)}{|a| \cdot \sigma_X} = \frac{a}{|a|} \cdot X^*$$

$$(cY+d)^* = \frac{c}{|c|} \cdot Y^*$$

$$\Rightarrow r(aX+b, cY+d) = E\left((aX+b)^*(cY+d)^*\right)$$

$$= E\left(\frac{a}{|a|}X^* \cdot \frac{c}{|c|} \cdot Y^*\right)$$

$$= \frac{ac}{|ac|} \cdot E(X^*Y^*)$$

$$= \frac{ac}{|ac|} \cdot r(X,Y)$$

$$= \begin{cases} r(X,Y) & \text{if } ac > 0 \\ -r(X,Y) & \text{if } ac < 0 \end{cases}$$

Example: (Exercise 11.6.7) Let (X, Y) be a random pair and let $r = r(X, Y)$. Let $X^\star$ be X in standard units and let $Y^\star$ be Y in standard units.

(a) Find $r(X^\star, Y^\star)$.

(b) Write the equation for $\widehat{Y}^\star$, the least squares linear predictor of $Y^\star$, based on $X^\star$.

(a) $X^* = \dfrac{X-\mu_X}{\sigma_X} = \underbrace{\dfrac{1}{\sigma_X}}_{=a} X - \underbrace{\dfrac{\mu_X}{\sigma_X}}_{=b} = aX + b$

$Y^* = \dfrac{Y-\mu_Y}{\sigma_Y} = \underbrace{\dfrac{1}{\sigma_Y}}_{=c} Y - \underbrace{\dfrac{\mu_Y}{\sigma_Y}}_{=d} = cY + d \rightarrow a \cdot c = \dfrac{1}{\sigma_X \sigma_Y} > 0$

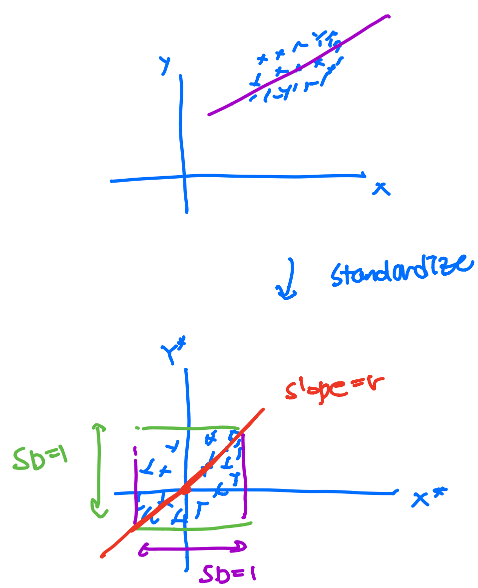
$r(X^*, Y^*) = r(X,Y)$ because $ac > 0$

(b) (X,Y) random pair $\rightarrow$ $\hat{Y} = \hat{a}X + \hat{b}$, $\hat{a} = r(X,Y) \cdot \dfrac{\sigma_Y}{\sigma_X}$, $\hat{b} = \mu_Y - \hat{a}\mu_X$

Now (X^*, Y^*) is our random pair.

$$\Rightarrow \begin{cases} \hat{a} = r(X^*, Y^*) \cdot \dfrac{\sigma_{Y^*}}{\sigma_{X^*}} = r \dfrac{1}{1} = r \\ \hat{b} = \underbrace{\mu_{Y^*}}_{=0} - \hat{a} \underbrace{\mu_{X^*}}_{=0} = 0 \end{cases}$$

$\hat{Y}^* = \hat{a}X^* + \hat{b} = rX^*$



11.5. The Error in Regression

The error in the regression estimate is called the residual and is defined as

$$D = Y - \widehat{Y}.$$

$= \widehat{a}X + \widehat{b}$

It is useful to write this in terms of the deviations $D_X = X - \mu_X$ and $D_Y = Y - \mu_Y$.

$$\widehat{Y} = \widehat{a}X + \widehat{b} = \widehat{a}X + \mu_Y - \widehat{a}\mu_X = \widehat{a}(X - \mu_X) + \mu_Y.$$

So,

$$\begin{aligned} D &= Y - \widehat{Y} \\ &= Y - [\widehat{a}(X - \mu_X) + \mu_Y] \\ &= \underbrace{Y - \mu_Y}_{D_Y} - \widehat{a}\underbrace{(X - \mu_X)}_{D_X} \\ &= D_Y - \widehat{a}D_X. \end{aligned}$$

$D_X = X - \mu_X$

$\Rightarrow E(D_X) = E(X - \mu_X)$
$= E(X) - \mu_X = 0$

$\widehat{a} = r\frac{\sigma_Y}{\sigma_X}$ constant

What is $E(D)$?

$$\begin{aligned} E(D) &= E(D_Y - \widehat{a}D_X) \\ &= E(D_Y) - \widehat{a}E(D_X) \\ &= 0 \end{aligned}$$

$= E((Y - (\widehat{a}X + \widehat{b}))^2)$

$MSE(\widehat{a}, \widehat{b})$

$\parallel$

Mean Squared Error of Regression

The mean squared error of regression is $E((Y - \widehat{Y})^2) = E(D^2)$. Since $E(D) = 0$, we have $\text{Var}(D) = E(D^2)$. Recall $\widehat{a} = r\frac{\sigma_Y}{\sigma_X}$ and $E(D_X D_Y) = r\sigma_X\sigma_Y$.

$= E(D^2) - (E(D))^2$

Let's find $\text{Var}(D)$:

$(D_Y - \widehat{a}D_X)^2$

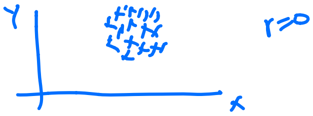
$= r\frac{\sigma_Y}{\sigma_X}$ $= r\sigma_X\sigma_Y$ $= E((X-\mu_X)^2) = \text{Var}(X) = \sigma_X^2$

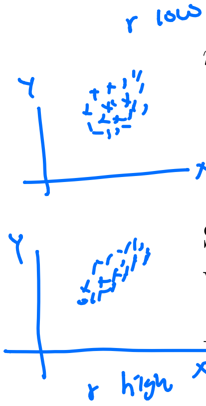
$$\begin{aligned} \text{Var}(D) &= E(D^2) \\ \sigma_Y^2 = \quad &= E(D_Y^2) - 2\widehat{a}E(D_X D_Y) + \widehat{a}^2 E(D_X^2) \\ &= \sigma_Y^2 - 2r\frac{\sigma_Y}{\sigma_X}r\sigma_X\sigma_Y + r^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 \\ &= \sigma_Y^2 - 2r^2\sigma_Y^2 + r^2\sigma_Y^2 \\ &= \sigma_Y^2 - r^2\sigma_Y^2 \\ &= (1 - r^2)\sigma_Y^2. \end{aligned}$$

So

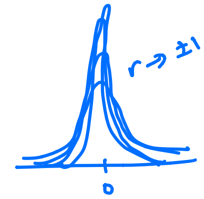
$$\text{SD}(D) = \sqrt{1 - r^2}\,\sigma_Y.$$

$\Rightarrow$ D has mean 0 and SD $\sqrt{1-r^2}\,\sigma_Y$.





r **As a Measure of Linear Association** Note that



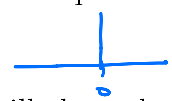
$$E(D) = 0 \text{ and } \mathrm{SD}(D) = \sqrt{1 - r^2}\sigma_Y.$$

$D = Y - \widehat{Y}$

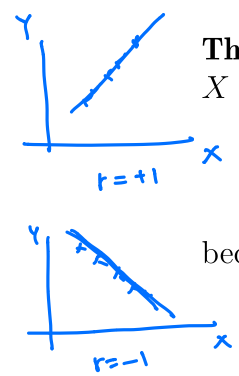
So if r is close to ± 1, $\mathrm{SD}(D)$ is close to 0, which implies that Y is close to $\widehat{Y}$. In other words, Y is close to being a linear function of X.

$= \hat{a}x + \hat{b}$

In the extreme case $r = \pm 1$, $\mathrm{SD}(D) = 0$ and Y is a perfectly linear function of X.



$Y = \widehat{Y} = \hat{a}x + \hat{b}$



The Residual is Uncorrelated with X We will show that the correlation between X and residual D is zero. Note that

$$r(D, X) = \frac{E((D - \mu_D)(X - \mu_X))}{\sigma_D \sigma_X} = \frac{1}{\sigma_D \sigma_X} E(DD_X),$$

because $\mu_D = 0$. We thus show $E(DD_X) = 0$:

$$\begin{aligned} E(DD_X) &= E((D_Y - \hat{a}D_X)D_X) \\ &= E(D_X D_Y) - \hat{a}E(D_X^2) \\ &= r\sigma_X\sigma_Y - r\frac{\sigma_Y}{\sigma_X}\sigma_X^2 \\ &= 0. \end{aligned}$$