Answer 1

The objective function (or loss function since minimizing) is:

$$SE = \sum_{i=1}^{m} (\widehat{w}x_i + \hat{b} - y_i)^2$$

SE = Squared Error

This represents the ordinary least squares problem of regression.

We have to find $\widehat{w}$ and $\hat{b}$ which minimizes the loss function (minimizes the squared error to regression line).

So, we take the partial derivative of the loss function w.r.t. $\widehat{w}$ and $\hat{b}$ and set it equal to 0.

$$\frac{\partial SE}{\partial \widehat{w}} = 0 \qquad \text{and} \qquad \frac{\partial SE}{\partial \hat{b}} = 0$$

$$\frac{\partial SE}{\partial \widehat{w}} = 0$$

or,

$$\frac{\partial \sum_{i=1}^{m} (\widehat{w}x_i + \hat{b} - y_i)^2}{\partial \widehat{w}} = 0$$

or,

$$2\sum_{i=1}^{m} \left(\widehat{w}x_i + \hat{b} - y_i\right) x_i = 0$$

or,

$$\frac{\sum_{i=1}^{m} (\widehat{w}{x_i}^2 + \hat{b}x_i - y_i x_i)}{m} = 0$$

or,

$$\widehat{w}\frac{1}{m}\sum_{i=1}^{m} {x_i}^2 + \hat{b}\frac{1}{m}\sum_{i=1}^{m} x_i - \frac{1}{m}\sum_{i=1}^{m} x_i y_i = 0$$

The yellow highlighted values are the mean values or the expected values, so we can rewrite as:

$$\widehat{w}\mathbb{E}\,[x^2] + \hat{b}\mathbb{E}\,[x] - \mathbb{E}\,[xy] = 0$$

or,

$$\widehat{w}\mathbb{E}\,[x^2] + \hat{b}\mathbb{E}\,[x] = \mathbb{E}\,[xy] \qquad (1)$$

$$\frac{\partial SE}{\partial \hat{b}} = 0$$

or,

$$\frac{\partial \sum_{i=1}^{m} (\widehat{w}x_i + \hat{b} - y_i)^2}{\partial \hat{b}} = 0$$

or,

$$2\sum_{i=1}^{m} \left(\widehat{w}x_i + \hat{b} - y_i\right) = 0$$

or,

$$\frac{\sum_{i=1}^{m} (\widehat{w}x_i + \hat{b} - y_i)}{m} = 0$$

or,

$$\hat{w}\frac{1}{m}\sum_{i=1}^{m} x_i + \hat{b}\frac{1}{m}\sum_{i=1}^{m} 1 - \frac{1}{m}\sum_{i=1}^{m} y_i = 0$$

The yellow highlighted values are the mean values or the expected values, so we can rewrite as:

$$\hat{w}\mathbb{E}\,[x] + \hat{b} - \mathbb{E}\,[y] = 0$$

or,

$$\hat{w}\mathbb{E}\,[x] + \hat{b} = \mathbb{E}\,[y] \qquad (2)$$

Now we just need the solve the two simultaneous equations (1) and (2) to find $\hat{w}$ and $\hat{b}$

$$\hat{w}\mathbb{E}\,[x^2] + \hat{b}\mathbb{E}\,[x] = \mathbb{E}\,[xy]$$
$$\hat{w}\mathbb{E}\,[x] + \hat{b} = \mathbb{E}\,[y]$$

Rewriting equation (2)

$$\hat{w}\mathbb{E}\,[x] + \hat{b} = \mathbb{E}\,[y]$$

or,

$$\hat{b} = \mathbb{E}\,[y] - \hat{w}\mathbb{E}\,[x] \qquad (3)$$

or equivalently,

$$\hat{b} = \overline{y} - \hat{w}\overline{x} \qquad (3)$$

Rewriting equation (1)

$$\hat{w}\mathbb{E}\,[x^2] + \hat{b}\mathbb{E}\,[x] = \mathbb{E}\,[xy]$$

Substituting $\hat{b}$ in the above,

$$\hat{w}\mathbb{E}\,[x^2] + (\mathbb{E}\,[y] - \hat{w}\mathbb{E}\,[x])\mathbb{E}\,[x] = \mathbb{E}\,[xy]$$

or,

$$\hat{w}\mathbb{E}\,[x^2] + \mathbb{E}\,[y]\mathbb{E}\,[x] - \hat{w}(\mathbb{E}\,[x])^2 = \mathbb{E}\,[xy]$$

or,

$$\hat{w}\left(\mathbb{E}\,[x^2] - (\mathbb{E}\,[x])^2\right) = \mathbb{E}\,[xy] - \mathbb{E}\,[x]\mathbb{E}\,[y]$$

or,

$$\hat{w} = \frac{\mathbb{E}\,[xy] - \mathbb{E}\,[x]\mathbb{E}\,[y]}{\mathbb{E}\,[x^2] - (\mathbb{E}\,[x])^2} \qquad (4)$$

$$Cov(x, y) = \mathbb{E}\,[(x - \mathbb{E}[\boldsymbol{x}])(y - \mathbb{E}[y])]$$

or,

$$Cov(x, y) = \mathbb{E}\,[xy - x\mathbb{E}[y] - \mathbb{E}[x]y + \mathbb{E}[x]\mathbb{E}[y]]$$

or,

$$Cov(x, y) = \mathbb{E}\,[xy] - \mathbb{E}[x][y] - \mathbb{E}[x][y] + \mathbb{E}[x][y]$$

or,

$$Cov(x, y) = \mathbb{E}\,[xy] - \mathbb{E}[x][y]$$

Using the above, rewriting equation (4)

$$\hat{w} = \frac{Cov(x, y)}{Var(x)}$$

Substituting $\hat{w}$ from above in (3)

$$\hat{b} = \mathbb{E}\,[y] - \left(\frac{Cov(x, y)}{Var(x)}\right)\mathbb{E}\,[x]$$

or equivalently,

$$\hat{b} = \overline{y} - \left(\frac{Cov(x, y)}{Var(x)}\right)\overline{x}$$

or equivalently,

$$\hat{b} = \overline{y} - \left(\frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}\right)\overline{x}$$

Answer 2

From (4)

$$\hat{w} = \frac{\mathbb{E}\,[xy] - \mathbb{E}\,[x]\mathbb{E}\,[y]}{\mathbb{E}\,[x^2] - (\mathbb{E}\,[x])^2}$$

Previously, we divided by m to get mean, let us undo that (i.e. multiply numerator and denominator by m), then:

$$\hat{w} = \frac{\sum_{i=1}^{m} y_i x_i - m\mathbb{E}[x]\mathbb{E}\,[y]}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2} \tag{5}$$

Given true linear model

$$y_i = wx_i + b + \epsilon_i$$

Substituting the above y_i in (5),

$$\hat{w} = \frac{\sum_{i=1}^{m}(wx_i + b + \epsilon_i)x_i - m\mathbb{E}[x]\mathbb{E}(wx + b + \epsilon)}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}(wx_i + b + \epsilon_i)x_i - m\mathbb{E}[x](w\mathbb{E}[x] + b + \mathbb{E}[\epsilon])}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

or,

$$\hat{w} = \frac{w\sum_{i=1}^{m} x_i{}^2 + b\sum_{i=1}^{m} x_i + \sum_{i=1}^{m} \epsilon_i x_i - m\mathbb{E}[x]w\mathbb{E}[x] - m\mathbb{E}[x]b - m\mathbb{E}[x]\mathbb{E}[\epsilon]}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

or,

$$\hat{w} = \frac{w\sum_{i=1}^{m} x_i{}^2 + bm\mathbb{E}[x] + \sum_{i=1}^{m} \epsilon_i x_i - mw(\mathbb{E}[x])^2 - bm\mathbb{E}[x] - m\mathbb{E}[x]\mathbb{E}[\epsilon]}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

or,

$$\hat{w} = \frac{w(\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}[x])^2) + \sum_{i=1}^{m} \epsilon_i x_i - m\mathbb{E}[x]\mathbb{E}[\epsilon]}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

or,

$$\hat{w} = w + \frac{\sum_{i=1}^{m} \epsilon_i x_i - m\mathbb{E}[x]\mathbb{E}[\epsilon]}{\sum_{i=1}^{m} x_i{}^2 - m(\mathbb{E}\,[x])^2}$$

But the mean of errors/noise is 0(given), therefore $\mathbb{E}[\epsilon] = 0$
and $\mathbb{E}[\sum_{i=1}^{m} \epsilon_i x_i] = 0$ because $\mathbb{E}[Cov(x, \epsilon)] = 0$

Therefore,

$$\mathbb{E}[\hat{w}] = w$$

This means that $\hat{w}$ is an unbiased estimate of w i.e. it is correct on average.

Now we know

$$y_i = wx_i + b + \epsilon_i$$

or,

$$\mathbb{E}\,[y] = \mathbb{E}\,[wx_i + b + \epsilon_i]$$

or,

$$\mathbb{E}\,[y] = w\mathbb{E}\,[x] + \mathbb{E}[b] + \mathbb{E}[\epsilon]$$

or,

$$\mathbb{E}\,[y] = w\mathbb{E}\,[x] + b + 0 \qquad (6)$$

From (3),

$$\hat{b} = \mathbb{E}\,[y] - \hat{w}\mathbb{E}\,[x]$$

or,

$$\mathbb{E}[\hat{b}] = \mathbb{E}\big[\mathbb{E}\,[y] - \hat{w}\mathbb{E}\,[x]\big]$$

or,

$$\mathbb{E}[\hat{b}] = \mathbb{E}\big[\mathbb{E}\,[y]\big] - \mathbb{E}\big[\hat{w}\mathbb{E}\,[x]\big]$$

or,

$$\mathbb{E}[\hat{b}] = \mathbb{E}\,[y] - w\mathbb{E}\,[x]$$

since $\mathbb{E}[\hat{w}] = w$ and expected value of an expected value is just that.

Substituting $\mathbb{E}\,[y]$ from (6)

$$\mathbb{E}[\hat{b}] = w\mathbb{E}\,[x] + b - w\mathbb{E}\,[x]$$

or,

$$\mathbb{E}[\hat{b}] = b$$

Alternatively, in Answer 1, we proved the following:

$$\hat{w} = \frac{Cov(x,y)}{Var(x)}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(wx_i + b + \epsilon_i - w\bar{x} - b - \bar{\epsilon})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(wx_i + \epsilon_i - w\bar{x} - \bar{\epsilon})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})\big(w(x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon})\big)}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = \frac{\sum_{i=1}^{m}\big(w(x_i - \bar{x})^2 + (x_i - \bar{x})(\epsilon_i - \bar{\epsilon})\big)}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = \frac{w\sum_{i=1}^{m}(x_i - \bar{x})^2}{\sum_{i=1}^{m}(x_i - \bar{x})^2} + \frac{\sum_{i=1}^{m}(x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$\hat{w} = w + \frac{\sum_{i=1}^{m}(x_i - \bar{x})(\epsilon_i)}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

But $\mathbb{E}[\epsilon_i] = 0$, therefore,

$$\mathbb{E}[\hat{w}] = w$$

This means that $\hat{w}$ is an unbiased estimate of w i.e. it is correct on average.

From (3),

$$\hat{b} = \bar{y} - \hat{w}\bar{x}$$

or,

$$\mathbb{E}[\hat{b}] = \mathbb{E}[\bar{y} - \hat{w}\bar{x}]$$

or,

$$\mathbb{E}[\hat{b}] = w\bar{x} + b - \mathbb{E}[\hat{w}]\bar{x}$$

or (since $\mathbb{E}[\hat{w}] = w$),

$$\mathbb{E}[\hat{b}] = w\bar{x} + b - w\bar{x}$$

or,

$$\mathbb{E}[\hat{b}] = b$$

The variances of these estimators are asked to be proven in the next question, so please refer to Answer 3

Answer 3

Finding Variance of $\hat{w}$:

$$Var(\hat{w}) = Var\left(\frac{\sum_{i=1}^{m}(x_i - \bar{x})(\epsilon_i)}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right)$$

or,

$$Var(\hat{w}) = \frac{1}{\left(\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^2} * Var\left(\sum_{i=1}^{m}(x_i - \bar{x})(\epsilon_i)\right)$$

or (given $Var(\epsilon_i) = \sigma^2$),

$$Var(\hat{w}) = \frac{1}{\left(\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^2} * \sum_{i=1}^{m}(x_i - \bar{x})^2\sigma^2$$

or,

$$Var(\hat{w}) = \frac{1}{\left(\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^2} * \sigma^2\sum_{i=1}^{m}(x_i - \bar{x})^2$$

or,

$$Var(\hat{w}) = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{\left(\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^2} * \sigma^2$$

or,

$$Var(\hat{w}) = \sigma^2 * \frac{1}{\sum_{i=1}^{m}(x_i - \bar{x})^2} \tag{7}$$

or (m-1 for sample variance; for population it would have been m),

$$Var(\hat{w}) = \frac{\sigma^2}{(m-1)} * \frac{(m-1)}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$Var(\hat{w}) = \frac{\sigma^2}{(m-1)} * \frac{1}{Var(x)}$$

or (for large m, m-1=m),

$$Var(\widehat{w}) \approx \frac{\sigma^2}{m} * \frac{1}{Var(x)}$$

Finding Variance of $\widehat{b}$:

$$Var(\widehat{b}) = Var(\bar{y} - \widehat{w}\bar{x})$$

or [using Var(A + B) = Var(A) + Var(B) + 2Cov(A, B)],

$$Var(\widehat{b}) = Var(\bar{y}) + Var(-\widehat{w}\bar{x}) + 2 * Cov(\bar{y}, -\widehat{w}\bar{x})$$

or,

$$Var(\widehat{b}) = Var(\bar{y}) + \bar{x}^2 * Var(\widehat{w}) - 2 * \bar{x} * Cov(\bar{y}, \widehat{w})$$

or,

$$Var(\widehat{b}) = \frac{\sigma^2}{m} + \bar{x}^2 * \left(\frac{\sigma^2}{m} * \frac{1}{Var(x)}\right) - 2 * \bar{x} * Cov(\bar{y}, \widehat{w})$$

Evaluating the above yellow highlighted term

$$Cov(\bar{y}, \widehat{w}) = Cov\left(\sum_{i=1}^{m} \frac{1}{m}(y_i), \sum_{j=1}^{m} \frac{x_j - \bar{x}}{m * Var(x)}(y_j)\right)$$

or,

$$Cov(\bar{y}, \widehat{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{x_j - \bar{x}}{m^2 * Var(x)} Cov(y_i, y_j)$$

or,

$$Cov(\bar{y}, \widehat{w}) = \sum_{i=1}^{m} \frac{x_j - \bar{x}}{m^2 * Var(x)} \sigma^2 + 0$$

or,

$$Cov(\bar{y}, \widehat{w}) = 0$$

Therefore,

$$Var(\widehat{b}) = \frac{\sigma^2}{m} + \bar{x}^2 * \left(\frac{\sigma^2}{m} * \frac{1}{Var(x)}\right) - 0$$

or,

$$Var(\widehat{b}) = \frac{\sigma^2}{m}\left(1 + \frac{\bar{x}^2}{Var(x)}\right)$$

or,

$$Var(\widehat{b}) \approx \frac{\sigma^2}{m}\left(\frac{\mathbb{E}[x^2]}{Var(x)}\right)$$

Answer 4

After centering, new $x_i' = x_i - \mu$

μ is the mean of x values, hence we are performing mean centering as discussed in the lecture.

$$\mu = \bar{x}$$

Let the new error on $\widehat{w}$ be $\widehat{w}'$ and $\widehat{b}$ be $\widehat{b}'$ (after centering)

From (7)

$$Var(\widehat{w}) = \sigma^2 * \frac{1}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$Var(\widehat{w}') = \sigma^2 * \frac{1}{\sum_{i=1}^{m}(x_i' - \bar{x}')^2}$$

or,

$$Var(\widehat{w}') = \sigma^2 * \frac{1}{\sum_{i=1}^{m}\big((x_i - \mu) - (\bar{x} - \mu)\big)^2}$$

or,

$$Var(\widehat{w}') = \sigma^2 * \frac{1}{\sum_{i=1}^{m}(x_i - \mu - \bar{x} + \mu)^2}$$

or,

$$Var(\widehat{w}') = \sigma^2 * \frac{1}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

or,

$$Var(\widehat{w}') = Var(\widehat{w})$$

Hence the new error (after centering) is same as the previous error.

From the result in Answer 3, we have

$$Var(\hat{b}) \approx \frac{\sigma^2}{m}\left(\frac{\mathbb{E}[x^2]}{Var(x)}\right)$$

or,

$$Var(\hat{b}) \approx \sigma^2 * \left(\frac{\mathbb{E}[x^2]}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right)$$

or,

$$Var(\hat{b}') \approx \sigma^2 * \left(\frac{\mathbb{E}[x'^2]}{\sum_{i=1}^{m}\big((x_i - \mu) - (\bar{x} - \mu)\big)^2}\right)$$

or,

$$Var(\hat{b}') \approx \sigma^2 * \left(\frac{\mathbb{E}[(x - \mu)^2]}{\sum_{i=1}^{m}(x_i - \mu - \bar{x} + \mu)^2}\right)$$

or,

$$Var(\hat{b}') \approx \sigma^2 * \left(\frac{\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right)$$

But $\bar{x} = \mu = \mathbb{E}[x]$, therefore,

$$Var(\hat{b}') \approx \sigma^2 * \left(\frac{\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2}{\sum_{i=1}^{m}(x_i - \mu)^2}\right)$$

or,

$$Var(\hat{b}') \approx \frac{\sigma^2}{m}$$

Proof that new error on $\hat{b}$ i.e. $\hat{b}$ is minimized

$$\frac{Var(\hat{b}')}{Var(\hat{b})} \approx \frac{\frac{\sigma^2}{m}}{\frac{\sigma^2}{m}\left(\frac{\mathbb{E}[x^2]}{Var(x)}\right)}$$

or,

$$\frac{Var(\hat{b}')}{Var(\hat{b})} \approx \frac{Var(x)}{\mathbb{E}[x^2]}$$

or,

$$\frac{Var(\hat{b}')}{Var(\hat{b})} \approx \frac{\mathbb{E}[x^2] - (\mathbb{E}[x])^2}{\mathbb{E}[x^2]}$$

Since $(\mathbb{E}[x])^2$ is a square hence it is a positive quantity, therefore the numerator is smaller than the denominator. Hence, $Var(\hat{b}') \leq Var(\hat{b})$.

Answer 5

Code:

Importing Libs

```
import numpy as np
```

Functions to Generate Xs and Y

```
# gen X (m datapoints (1 Dim))
def gen_X(m):
#https://numpy.org/doc/stable/reference/random/generated/numpy.random.unifo
rm.html
    X=np.random.uniform(low=100,high=102,size=m)
    return X
```

```
# gen Y/target/labels/output
def gen_Y(X,w,b,variance):
    Y=np.empty((np.shape(X)[0]))

    for i in range(np.shape(Y)[0]):
#https://numpy.org/doc/stable/reference/random/generated/numpy.random.norma
l.html
        #scale takes std. dev. hence we have to sqrt the variance
        Y[i]=X[i]*w + b + np.random.normal(loc=0, scale=pow(variance,0.5))

    return Y
```

```
# gen new X i.e. X' (m datapoints (1 Dim))
def gen_X_dash(X):
    X_dash=np.empty((np.shape(X)[0]))

    for i in range(np.shape(X_dash)[0]):
        X_dash[i] = X[i] - 101

    return X_dash
```

Helper functions

```
def compute_w_and_b(X, Y):
    #https://numpy.org/doc/stable/reference/generated/numpy.cov.html
    #https://stackoverflow.com/questions/15317822/calculating-covariance-
with-python-and-numpy

    #w_hat=Cov(x,y)/Var(x) ... proven in Answer 1
    w=(np.cov(X, Y, bias=True)[0][1])/np.var(X)

    #b_hat=y_mean-w_hat*x_mean ... equation (3)
```

```
    b=np.mean(Y)-w*np.mean(X)

    return w, b
```

```
def sim(m, w, b, variance, iters=1000):
    #storing sim results in lists
    w_hat_list = []
    b_hat_list = []
    w_hat_dash_list = []
    b_hat_dash_list = []
    for i in range(iters):
        X=gen_X(m)
        Y=gen_Y(X,w,b,variance)
        X_dash=gen_X_dash(X)

        w_hat, b_hat = compute_w_and_b(X, Y)
        w_hat_list.append(w_hat)
        b_hat_list.append(b_hat)

        w_hat_dash, b_hat_dash = compute_w_and_b(X_dash, Y)
        w_hat_dash_list.append(w_hat_dash)
        b_hat_dash_list.append(b_hat_dash)

    print("Expected value of w_hat: ",np.mean(w_hat_list))
    print("Expected value of w_hat_dash: ",np.mean(w_hat_dash_list))
    print("Expected value of b_hat: ",np.mean(b_hat_list))
    print("Expected value of b_hat_dash: ",np.mean(b_hat_dash_list))

    print("Variance of w_hat: ",np.var(w_hat_list))
    print("Variance of w_hat_dash: ",np.var(w_hat_dash_list))
    print("Variance of b_hat: ",np.var(b_hat_list))
    print("Variance of b_hat_dash: ",np.var(b_hat_dash_list))
```

Running the sim

```
m=200
w=1
b=5
variance=0.1
sim(m, w, b, variance, 1000)
```

Output:

```
Expected value of w_hat:  0.9993276553126749
Expected value of w_hat_dash:  0.999327655312675
Expected value of b_hat:  5.067982014234828
Expected value of b_hat_dash:  106.00007520081499
Variance of w_hat:  0.0014741891806137291
Variance of w_hat_dash:  0.0014741891806137278
Variance of b_hat:  15.034748195851964
Variance of b_hat_dash:  0.0004470032298959514
```

The results agree.
First, we are generating one dimensional data (**X**) uniformly distributed between 100 and 102.
Then, we are generating the regression line (**Y**) based on **X**.
Then, we are generating **X' (X_dash)** i.e. X centered around the mean of X.

As, we have shown in Answer 4, that centering the data around the mean, produces the same error on the weights but reduces/minimizes the error on the bias. This is confirmed from the above simulation results. The variance of weights on non-centered data and centered data is almost the same (difference only in higher precision). The variance of bias on non-centered data and centered data is different. The error on bias is much reduced on the centered data.

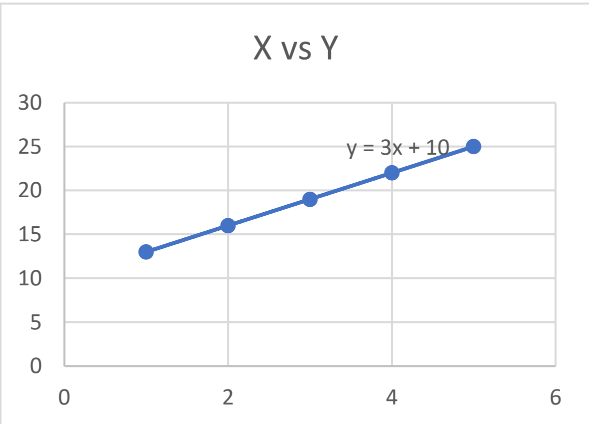
Answer 6

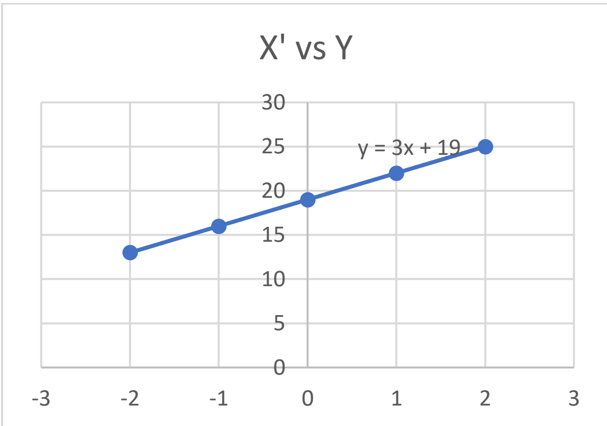
Consider the following dataset (different w and b values from question 5):

	A	B
1	**w**	3
2	**b**	10
3		
4	**X**	**Y**
5	1	13
6	2	16
7	3	19
8	4	22
9	5	25
10		
11	**X'**	**Y**
12	-2	13
13	-1	16
14	0	19
15	1	22
16	2	25

	A	B
1	**w**	3
2	**b**	10
3		
4	**X**	**Y**
5	1	=A5*B$1+B$2
6	2	=A6*B$1+B$2
7	3	=A7*B$1+B$2
8	4	=A8*B$1+B$2
9	5	=A9*B$1+B$2
10		
11	**X'**	**Y**
12	=A5-AVERAGE(A$5:A$9)	=B5
13	=A6-AVERAGE(A$5:A$9)	=B6
14	=A7-AVERAGE(A$5:A$9)	=B7
15	=A8-AVERAGE(A$5:A$9)	=B8
16	=A9-AVERAGE(A$5:A$9)	=B9

And the resultant plots with trendline:





After generating the trendline in excel, we see that the coefficient of **x** in the equation of the line remains same even after centering the data around the mean.
This is because all the points are just re positioned by the same value. However, the bias (y-intercept) changes. Here there is no error term but even if there was, then on average the slope (w) would remain the same. Hence the estimate or mean of the slope has no change when the data is shifted.

We can intuitively say that we are just grabbing the dataset and repositioning the whole thing to our wish (w will remain the same if centered around a value other than the mean as well). Hence the slope doesn't change.

This is true for higher dimensional data as well. The structure remains the same, only the position of the whole thing is changed.

Answer 7

After augmenting (going from one dimension to two dimensions), the new X would be

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

Consequently,

$$X^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix}$$

Now,

$$\Sigma = X^T X$$

or,

$$\Sigma = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

or,

$$\Sigma = \begin{bmatrix} 1*1+1*1+\cdots+1*1 & 1*x_1+1*x_2+\cdots 1*x_m \\ x_1*1+x_2*1+\cdots+x_m*1 & x_1*x_1+x_2*x_2+\cdots x_m*x_m \end{bmatrix}$$

or,

$$\Sigma = \begin{bmatrix} m & m\mathbb{E}[x] \\ m\mathbb{E}[x] & m\mathbb{E}[x^2] \end{bmatrix}$$

or,

$$\Sigma = m \begin{bmatrix} 1 & \mathbb{E}[x] \\ \mathbb{E}[x] & \mathbb{E}[x^2] \end{bmatrix}$$

To find the condition number of Σ i.e. κ(Σ), we need to find the find the eigenvalues of ∑ (largest and smallest eigenvalue). *Later we will find $\kappa(\Sigma')$ to compare both.*

$$\det(\Sigma - \lambda I) = 0$$

or,

$$\det\left(\begin{bmatrix} 1-\lambda & \mathbb{E}[x] \\ \mathbb{E}[x] & \mathbb{E}[x^2]-\lambda \end{bmatrix}\right) = 0$$

or,

$$(1-\lambda)(\mathbb{E}[x^2]-\lambda) - (\mathbb{E}[x])^2 = 0$$

or,

$$(\mathbb{E}[x^2] - \lambda - \lambda\mathbb{E}[x^2] + \lambda^2) - (\mathbb{E}[x])^2 = 0$$

or,

$$\lambda^2 - \lambda(1+\mathbb{E}[x^2]) + (\mathbb{E}[x^2] - (\mathbb{E}[x])^2) = 0$$

We have a quadratic equation in terms of λ, solving which will give us two eigenvalues of the 2*2 matrix Σ.

$$\lambda = \frac{(1+\mathbb{E}[x^2]) \pm \sqrt{(1+\mathbb{E}[x^2])^2 - 4*1*(\mathbb{E}[x^2]-(\mathbb{E}[x])^2)}}{2*1}$$

We will consider the '+' term as the largest eigenvalue and the '-' term as the smallest eigenvalue.

$$\lambda_{largest} = \frac{(1+\mathbb{E}[x^2]) + \sqrt{(1+\mathbb{E}[x^2])^2 - 4(\mathbb{E}[x^2]-(\mathbb{E}[x])^2)}}{2*1}$$

$$\lambda_{smallest} = \frac{(1+\mathbb{E}[x^2]) - \sqrt{(1+\mathbb{E}[x^2])^2 - 4(\mathbb{E}[x^2]-(\mathbb{E}[x])^2)}}{2*1}$$

$$\kappa(\Sigma) = \frac{\lambda_{largest}}{\lambda_{smallest}}$$

or,

$$\kappa(\Sigma) = \frac{(1+\mathbb{E}[x^2]) + \sqrt{(1+\mathbb{E}[x^2])^2 - 4(\mathbb{E}[x^2] - (\mathbb{E}[x])^2)}}{(1+\mathbb{E}[x^2]) - \sqrt{(1+\mathbb{E}[x^2])^2 - 4(\mathbb{E}[x^2] - (\mathbb{E}[x])^2)}}$$

or,

$$\kappa(\Sigma) = \frac{(1+\mathbb{E}[x^2]) + \sqrt{(1+\mathbb{E}[x^2])^2 - 4\mathbb{E}[x^2] + 4(\mathbb{E}[x])^2}}{(1+\mathbb{E}[x^2]) - \sqrt{(1+\mathbb{E}[x^2])^2 - 4\mathbb{E}[x^2] + 4(\mathbb{E}[x])^2}}$$

or [using $(a-b)^2 = (a+b)^2 - 4ab$],

$$\kappa(\Sigma) = \frac{(1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}}{(1+\mathbb{E}[x^2]) - \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}}$$

or,

or (multiplying both numerator and denominator by the numerator),

$$\kappa(\Sigma) = \frac{\left((1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}\right)^2}{(1+\mathbb{E}[x^2])^2 - ((1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2)}$$

or,

$$\kappa(\Sigma) = \frac{\left((1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}\right)^2}{(1+\mathbb{E}[x^2])^2 - (1-\mathbb{E}[x^2])^2 - 4(\mathbb{E}[x])^2}$$

or,

$$\kappa(\Sigma) = \frac{\left((1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}\right)^2}{4\mathbb{E}[x^2] - 4(\mathbb{E}[x])^2}$$

or,

$$\kappa(\Sigma) = \frac{\frac{1}{4}\left((1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}\right)^2}{Var(X)}$$

Let us compute Σ'

$$X' = \begin{bmatrix} 1 & x_1 - \mathbb{E}[x] \\ 1 & x_2 - \mathbb{E}[x] \\ \vdots & \vdots \\ 1 & x_m - \mathbb{E}[x] \end{bmatrix}$$

Consequently,

$$X^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \mathbb{E}[x] & x_2 - \mathbb{E}[x] & \cdots & x_m - \mathbb{E}[x] \end{bmatrix}$$

Now,

$$\Sigma' = (X')^T(X')$$

or,

$$\Sigma' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \mathbb{E}[x] & x_2 - \mathbb{E}[x] & \cdots & x_m - \mathbb{E}[x] \end{bmatrix} \begin{bmatrix} 1 & x_1 - \mathbb{E}[x] \\ 1 & x_2 - \mathbb{E}[x] \\ \vdots & \vdots \\ 1 & x_m - \mathbb{E}[x] \end{bmatrix}$$

or,

$$\Sigma' = \begin{bmatrix} m & \sum_{i=1}^{m}(x_i - \mathbb{E}[x]) \\ \sum_{i=1}^{m}(x_i - \mathbb{E}[x]) & \sum_{i=1}^{m}(x_i - \mathbb{E}[x])^2 \end{bmatrix}$$

But, $\sum_{i=1}^{m}(x_i - \mathbb{E}[x]) = 0$, therefore,

$$\Sigma' \approx \begin{bmatrix} m & 0 \\ 0 & mVar(X) \end{bmatrix}$$

or,

$$\Sigma' \approx m \begin{bmatrix} 1 & 0 \\ 0 & Var(X) \end{bmatrix}$$

To find the condition number of Σ' i.e. $\kappa(\Sigma')$, we need to find the find the eigenvalues of $\sum'$ (largest and smallest eigenvalue).

$$\det(\Sigma' - \lambda I) = 0$$

or,

$$\det\left(\begin{bmatrix} 1-\lambda' & 0 \\ 0 & Var(X) - \lambda' \end{bmatrix}\right) = 0$$

or,

$$(1-\lambda')(Var(X) - \lambda') = 0$$

or,

$$Var(X) - \lambda' - \lambda' Var(X) + \lambda^2 = 0$$

or,

$$\lambda'^2 - \lambda'\big(1 + Var(X)\big) + Var(X) = 0$$

We have a quadratic equation in terms of λ, solving which will give us two eigenvalues of the 2*2 matrix Σ.

$$\lambda' = \frac{\big(1+Var(X)\big) \pm \sqrt{\big(1+Var(X)\big)^2 - 4 * 1 * Var(X)}}{2 * 1}$$

We will consider the '+' term as the largest eigenvalue and the '-' term as the smallest eigenvalue.

$$\lambda'_{largest} = \frac{\big(1+Var(X)\big) + \sqrt{\big(1+Var(X)\big)^2 - 4 * 1 * Var(X)}}{2 * 1}$$

$$\lambda'_{smallest} = \frac{\big(1+Var(X)\big) - \sqrt{\big(1+Var(X)\big)^2 - 4 * 1 * Var(X)}}{2 * 1}$$

$$\kappa(\Sigma') = \frac{\lambda'_{largest}}{\lambda'_{smallest}}$$

or,

$$\kappa(\Sigma') = \frac{\big(1+Var(X)\big) + \sqrt{\big(1+Var(X)\big)^2 - 4 * 1 * Var(X)}}{\big(1+Var(X)\big) - \sqrt{\big(1+Var(X)\big)^2 - 4 * 1 * Var(X)}}$$

or [using $(a-b)^2 = (a+b)^2 - 4ab$],

$$\kappa(\Sigma') = \frac{\big(1+Var(X)\big)+\big(1-Var(X)\big)}{\big(1+Var(X)\big)-\big(1-Var(X)\big)}$$

or,

$$\kappa(\Sigma') = \frac{2}{2Var(X)}$$

or,

$$\kappa(\Sigma') = \frac{1}{Var(X)}$$

Now, we know that (proven in this answer),

$$\kappa(\Sigma) = \frac{\frac{1}{4}\left((1+\mathbb{E}[x^2]) + \sqrt{(1-\mathbb{E}[x^2])^2 + 4(\mathbb{E}[x])^2}\right)^2}{Var(X)}$$

It is always true that $\mathbb{E}[x^2] \geq 0$ and $(\mathbb{E}[x])^2 \geq 0$, so if we replace them by 0, we get,

$$\kappa(\Sigma) \geq \frac{\frac{1}{4}\left((1+0) + \sqrt{(1-0)^2 + 4(0)^2}\right)^2}{Var(X)}$$

or,

$$\kappa(\Sigma) \geq \frac{\frac{1}{4}(1+1)^2}{Var(X)}$$

or,

$$\kappa(\Sigma) \geq \frac{1}{Var(X)}$$

Hence,

$$\kappa(\Sigma) \geq \kappa(\Sigma')$$

Or in other words, $\kappa(\Sigma')$ is minimized when centering around the mean.

References

1. Lecture Notes and Videos
2. https://www.khanacademy.org/math/statistics-probability/describing-relationships-quantitative-data/more-on-regression/v/proof-part-4-minimizing-squared-error-to-regression-line
3. https://en.wikipedia.org/wiki/Covariance
4. https://math.stackexchange.com/questions/2947788/what-is-varx-y
5. http://www.unm.edu/~lspear/geog525/24linreg2.pdf
6. http://home.iitk.ac.in/~shalab/econometrics/Chapter2-Econometrics-SimpleLinearRegressionAnalysis.pdf

***