Slope and intercept defined.

We know y = mx + c
what is m here define ?

In the equation $y = mx + c$, m represents the **slope** of the line. The slope indicates the steepness and direction of the line on a graph. It is calculated as the change in the y-value divided by the change in the x-value between two points on the line. Mathematically, the slope m is given by:

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

where (x_1, y_1) and (x_2, y_2) are two distinct points on the line.

In simpler terms, m tells you how much y changes for a unit change in x. For example:

- If $m > 0$, the line rises as x increases (positive slope).
- If $m < 0$, the line falls as x increases (negative slope).
- If $m = 0$, the line is horizontal and there is no change in y as x changes.

The other term, c, is the **y-intercept**, which is the point where the line crosses the y-axis.

OK . Then what is the below expression
For example, if you are using values of X to predict values of Y, guessing
too high might be better (or worse) than guessing too low. In that case you
might want to compute some cost function, cost(#i), and minimize total cost.
However, computing a least squares fit is quick, easy and often good
enough, so here's how:
1. Compute the sample means, ¯ x and ¯ y, the variance of X, and the covariance
of X and Y.
2. The estimated slope is
ˆb
=
Cov(X,Y)
Var(X)
3. And the intercept is
ˆa
= ¯ y ˆb ¯

then what is beta relation with m

In the context of your provided expression, you are describing the process of computing a linear regression line to predict Y from X. This process involves calculating the sample means, variances, and covariances to estimate the slope and intercept of the regression line.

Let's break down the steps and relate them to the general form $y = mx + c$:

1. **Compute the sample means** $\bar{x}$ and $\bar{y}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

   where n is the number of data points.

2. **Compute the variance of** X:

$$\mathrm{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

3. **Compute the covariance of** X **and** Y:

$$\mathrm{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

4. **Estimated slope (**$\hat{b}$**)**:

$$\hat{b} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}$$

   This $\hat{b}$ is the estimated slope of the regression line.

5. **Estimated intercept (**$\hat{a}$**)**:

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

   This $\hat{a}$ is the estimated intercept of the regression line.

Relation of β (beta) with m

In the context of linear regression, β often represents the slope of the regression line, which is analogous to m in the equation $y = mx + c$. Therefore, in your expression:

- $\hat{b}$ (estimated slope) is equivalent to m.

Thus, the relationship is:

$$m = \beta = \hat{b}$$

Where:

- m is the slope of the line in the equation $y = mx + c$.
- β (or $\hat{b}$) is the estimated slope in the linear regression context, derived from the sample data using the formulas provided.

But beta is telling that the amount of covariance change per unit change in var (X) , it is not telling about the y change per x .

I see your point. Let's clarify the interpretation of β in the context of your statement and how it relates to the change in y per unit change in x.

Interpreting β and Its Relation to m

In the context of linear regression:

1. **Slope (m) in the equation $y = mx + c$:**
   - This is the coefficient m which directly tells you the change in y for a unit change in x.
   - It is calculated using:

$$m = \frac{\Delta y}{\Delta x}$$

2. **β (or $\hat{b}$) in the regression context:**
   - This is the estimated slope of the regression line, calculated as:

$$\hat{b} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

   - Here, $\text{Cov}(X, Y)$ is the covariance between X and Y, and $\text{Var}(X)$ is the variance of X.

Understanding Covariance and Variance in Slope Estimation

- **Covariance ($\text{Cov}(X, Y)$):**
  - Measures the joint variability of X and Y. If X and Y tend to increase together, the covariance is positive; if one tends to increase when the other decreases, the covariance is negative.
- **Variance ($\text{Var}(X)$):**

- Measures how much X varies around its mean $\bar{x}$.

The slope $\hat{b}$ (or β) is calculated by normalizing the covariance by the variance of X, which means:

$$\hat{b} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

This ratio effectively tells you how much Y changes for a unit change in X.

Connecting β with m

Although β is expressed in terms of covariance and variance, its fundamental interpretation remains the same as that of m:

- **β tells you the expected change in Y for a unit change in X.**
- This is analogous to saying β represents the slope m in the linear equation $y = mx + c$.

Conclusion

Therefore, while the expression $\hat{b} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$ uses statistical measures (covariance and variance), the result β or $\hat{b}$ still directly indicates the change in y per unit change in x. So, β is indeed the slope m in the context of linear regression, representing the same concept: the change in y for a unit change in x.

How come coefficeint of determination has the relation with peason correlation

The coefficient of determination, denoted as R^2, is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). When there is only one independent variable, the relationship between the coefficient of determination (R^2) and the Pearson correlation coefficient (r) is straightforward.

Pearson Correlation Coefficient (r)

The Pearson correlation coefficient (r) measures the linear relationship between two variables, X and Y. It ranges from -1 to 1, where:

- $r = 1$: Perfect positive linear relationship
- $r = -1$: Perfect negative linear relationship
- $r = 0$: No linear relationship

Mathematically, it is defined as:

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

Coefficient of Determination (R^2)

In the context of simple linear regression with one independent variable, R^2 is the square of the Pearson correlation coefficient (r):

$$R^2 = r^2$$

Explanation of the Relationship

1. **Variance Explanation**:
   - R^2 indicates how much of the variance in the dependent variable (Y) is explained by the independent variable (X).
   - For example, if $R^2 = 0.64$, it means 64% of the variance in Y can be explained by X.

2. **Square of Correlation**:
   - The Pearson correlation coefficient (r) measures the strength and direction of a linear relationship between X and Y.
   - When you square the correlation coefficient (r), you obtain R^2, which expresses the proportion of variance in Y explained by X.

Mathematical Derivation

If we consider the formula for r:

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

When we square r:

$$r^2 = \left(\frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}\right)^2$$

This simplifies to:

$$r^2 = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X) \cdot \text{Var}(Y)}$$

In the context of simple linear regression, this r^2 is equivalent to R^2.

Summary

The coefficient of determination R^2 and the Pearson correlation coefficient r are closely related in the context of simple linear regression:

- $R^2 = r^2$
- $r = \sqrt{R^2}$

This relationship shows that R^2 provides the proportion of the variance in Y explained by X, while r provides the direction and strength of the linear relationship between X and Y.

when you say equivalent , do you mean r
2 == R
2

When I say R^2 is equivalent to the square of the Pearson correlation coefficient r in the context of simple linear regression, I mean that:

$$R^2 = r^2$$

This equation indicates that the coefficient of determination R^2 is numerically equal to the square of the Pearson correlation coefficient r.

To clarify further:

- **Pearson correlation coefficient** r: Measures the strength and direction of the linear relationship between two variables X and Y. It ranges from -1 to 1.
- **Coefficient of determination** R^2: Represents the proportion of the variance in the dependent variable Y that is predictable from the independent variable X. It ranges from 0 to 1.

In simple linear regression (with one independent variable X and one dependent variable Y):

- r indicates how strongly X and Y are linearly related.
- R^2 indicates how much of the variability in Y can be explained by X.

Therefore, the relationship between R^2 and r in simple linear regression is given by:

$$R^2 = (r)^2$$

In summary, yes, in the context of simple linear regression, the coefficient of determination R^2 is equal to the square of the Pearson correlation coefficient r:

$$R^2 = r^2$$

Range for both are different , then how would you say that ?

You are correct that the ranges for r and R^2 are different, and that is an important distinction to make. Here's a more precise explanation:

- **Pearson correlation coefficient** r: Measures the strength and direction of the linear relationship between two variables. Its range is from -1 to 1.
  - $r = 1$: Perfect positive linear relationship.
  - $r = -1$: Perfect negative linear relationship.
  - $r = 0$: No linear relationship.
- **Coefficient of determination** R^2: Represents the proportion of the variance in the dependent variable that is predictable from the independent variable. Its range is from 0 to 1.
  - $R^2 = 1$: The model explains all the variability of the response data around its mean.
  - $R^2 = 0$: The model explains none of the variability of the response data around its mean.

The relationship between R^2 and r in the context of simple linear regression (one independent variable) is:

$$R^2 = r^2$$

Explanation

1. **Different Ranges**:
   - r ranges from -1 to 1.
   - R^2 ranges from 0 to 1.
2. **Mathematical Relationship**:
   - Given that $R^2 = r^2$, R^2 will always be non-negative because squaring any real number (whether positive or negative) results in a non-negative number.
   - This means if $r = 0.8$, then $R^2 = 0.64$.
   - If $r = -0.8$, then $R^2 = 0.64$ as well.

Key Points

- While r can be negative or positive, indicating the direction of the relationship between X and Y, R^2 only measures the strength of the relationship, without regard to direction.
- R^2 is the square of r, meaning it represents the proportion of variance explained, while r indicates both the strength and direction of the linear relationship.

Summary

The statement that R^2 is equivalent to the square of r (i.e., $R^2 = r^2$) should be understood in terms of their mathematical relationship in simple linear regression. Although their ranges are different and they convey related but distinct information (proportion of variance explained vs. strength and direction of the relationship), R^2 is indeed derived from squaring r.

can you proof mathematically ?

Sure, let's prove mathematically that $R^2 = r^2$ in the context of simple linear regression.

Definitions

1. **Pearson Correlation Coefficient r**:

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

2. **Coefficient of Determination R^2**:
   In the context of simple linear regression, R^2 is defined as the square of the correlation coefficient between the observed and predicted values of the dependent variable.

Linear Regression

The linear regression line is given by:

$$\hat{y} = b_0 + b_1 x$$

where:

- $\hat{y}$ is the predicted value of y.
- b_0 is the y-intercept.
- b_1 is the slope (which we previously denoted as $\hat{b}$).

Steps to Prove $R^2 = r^2$

1. **Calculate the Total Sum of Squares (SST)**:

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

   where $\bar{y}$ is the mean of the observed values.

2. **Calculate the Regression Sum of Squares (SSR)**:

$$\text{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

3. **Calculate the Error Sum of Squares (SSE)**:

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

4. **Relationship between SST, SSR, and SSE**:

$$\text{SST} = \text{SSR} + \text{SSE}$$

5. **Define R^2 as**:

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

Expressing R^2 in Terms of r

We start by using the fact that in linear regression, $\hat{y}_i = b_0 + b_1 x_i$.

Recall that the sample covariance $\text{Cov}(X, Y)$ and variances $\text{Var}(X)$ and $\text{Var}(Y)$ are defined as:

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\text{Var}(Y) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The slope b_1 of the regression line can be written as:

$$b_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

The predicted values $\hat{y}_i$ are:

$$\hat{y}_i = \bar{y} + b_1(x_i - \bar{x})$$

Now, the regression sum of squares (SSR) is:

$$\text{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} (b_1(x_i - \bar{x}))^2 = b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = b_1^2 \cdot \text{Var}(X) \cdot n$$

The total sum of squares (SST) is:

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{Var}(Y) \cdot n$$

So, R^2 can be expressed as:

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{b_1^2 \cdot \text{Var}(X) \cdot n}{\text{Var}(Y) \cdot n} = \frac{b_1^2 \cdot \text{Var}(X)}{\text{Var}(Y)}$$

Substituting $b_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$, we get:

$$R^2 = \frac{\left(\frac{\text{Cov}(X,Y)}{\text{Var}(X)}\right)^2 \cdot \text{Var}(X)}{\text{Var}(Y)} = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X) \cdot \text{Var}(Y)}$$

This is exactly the square of the Pearson correlation coefficient:

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

Therefore:

$$R^2 = r^2$$

Conclusion

This proof shows mathematically that the coefficient of determination R^2 in simple linear regression is indeed the square of the Pearson correlation coefficient r.