Revisão de Econometria

Mixtape - Cap. 2 - S. Cunningham

Pedro Hemsley e Romero Rocha

October 22, 2024

Introdução ao Modelo Populacional

- Estamos interessados em saber como a variável Y varia com X.
- Para isso, usamos um *modelo de regressão linear bivariada*.

Modelo de Regressão Linear Bivariada

Equação

$$Y = \beta_0 + \beta_1 X + u$$

Onde:

- Y: variável dependente.
- X: variável independente ou explicativa.
- u: termo de erro (fatores não observados que afetam Y).
- β_0: intercepto.
- β_1: coeficiente angular (slope).

Objetivo

- Estimar os valores de β_0 e β_1 utilizando dados.

Hipótese sobre o Erro

Suposição:

$$E(u) = 0$$

Isso implica que a média do erro na população é zero.

Interpretação do Intercepto

Caso $E(u) = \alpha_0$

Ajustamos o intercepto como:

$$Y = (\beta_0 + \alpha_0) + \beta_1 X + (u - \alpha_0)$$

Esse ajuste não afeta β_1.

Hipótese de Independência na Média

Independência na média:

$$E(u \mid X) = E(u)$$

Ou seja, para qualquer valor de X, o valor esperado de u é constante. X não muda, em média, quando u é alterado

Exemplo: Escolaridade e Salário

- Suponha que estamos estimando o efeito da escolaridade sobre os salários.
- O erro u inclui a habilidade não observada de uma pessoa.
- A independência na média requer que a habilidade média seja a mesma entre pessoas com diferentes níveis de escolaridade:

$$E(\text{habilidade} \mid \text{escolaridade} = 8) = E(\text{habilidade} \mid \text{escolaridade} = 12)$$

Hipótese de Independência Condicional

Juntando as hipóteses:

Ao combinar a Hipótese de independência da média com a normalização $E(u) = 0$, obtemos a seguinte condição chave:

$$E(u \mid X) = 0 \text{ para todos os valores de } X$$

- Isso é conhecido como a *Hipótese de média condicional zero*.
- Essa suposição é essencial para a identificação causal no modelo de regressão.
- Essa hipótese nos permite reescrever nossa equação básica...

Função de Regressão Populacional

Função de Expectativa Condicional (CEF)

$$E(Y \mid X) = \beta_0 + \beta_1 X$$

- A função de regressão populacional é uma função linear de X.
- Essa relação linear simplifica a interpretação dos parâmetros β_0 e β_1.

Mínimos Quadrados Ordinários (OLS)

- Queremos estimar os parâmetros da população, β_0 e β_1, dados os pares de observações (x_i, y_i), com $i = 1, 2, \ldots, n$, onde n é o tamanho da amostra.
- Para isso, usamos a equação:

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

- y_i e x_i são observados, mas u_i (o termo de erro) não é. Apenas sabemos que u_i está presente.

Mínimos Quadrados Ordinários (OLS)

- Usamos duas restrições populacionais para derivar as equações para estimação de β_0 e β_1:

$$E(u) = 0$$

$$E(u \mid x) = 0$$

- A segunda condição significa que o valor médio do termo de erro u não muda com diferentes valores de x.
- Essa suposição de independência implica que:

$$E(xu) = 0$$

Usando as Restrições

- A independência entre x e u implica também que $C(x, u) = 0$ (a covariância entre x e u é zero).
- Substituímos u por $y - \beta_0 - \beta_1 x$ e obtemos as equações:

$$E(y - \beta_0 - \beta_1 x) = 0$$

$$E(x[y - \beta_0 - \beta_1 x]) = 0$$

- Essas equações são usadas para derivar os valores estimados de β_0 e β_1.

Contraparte amostral

- Precisamos usar as contrapartes amostrais dessas equações: β_0 e β_1:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

- $\hat{\beta}_0$ e $\hat{\beta}_1$ são as estimativas obtidas a partir dos dados.
- Temos então duas equações e duas variáveis: $\hat{\beta}_0$ e $\hat{\beta}_1$

Desenvolvendo a Primeira Equação

- Aplicando as propriedades do somatório na primeira equação:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0 - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 x_i$$

- Como $\hat{\beta}_0$ e $\hat{\beta}_1$ são constantes, podemos reescrever essa expressão:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

- Como o lado esquerdo é igual a zero:

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

Substituindo $\hat{\beta}_0$ na Segunda Equação

- Substituímos $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ na segunda equação:

$$\sum_{i=1}^{n} x_i \left[y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \right] = 0$$

- Simplificando:

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Solução para $\hat{\beta}_1$

- Finalmente, a solução para $\hat{\beta}_1$ é:

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

- Essa é a fórmula do coeficiente angular, expressa como a covariância amostral entre x_i e y_i, dividida pela variância amostral de x_i.
- Podemos simplificar um pouco mais observando que $\sum_{i=1}^{n}(x_i - \bar{x})\,\bar{y} = \bar{y}\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. Dessa forma:

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

Estimativa de $\hat{\beta}_1$ (Coeficiente Angular)

- A fórmula anterior para $\hat{\beta}_1$ é importante porque nos mostra como usar os dados para calcular a estimativa da inclinação.
- A estimativa $\hat{\beta}_1$ é comumente referida como a *estimativa da inclinação de mínimos quadrados ordinários (OLS)*.
- Ela pode ser calculada sempre que a variância amostral de x_i não for zero. Em outras palavras, x_i precisa variar entre as observações.

Intuição sobre $\hat{\beta}_1$

- A variação de x nos permite identificar seu impacto em y.
- Isso também significa que não podemos determinar a inclinação de uma relação se todos os valores de x_i forem iguais, como se todos tivessem os mesmos anos de escolaridade, por exemplo.

Estimativa de $\hat{\beta}_0$ (Intercepto)

- Uma vez que calculamos $\hat{\beta}_1$, podemos calcular o valor do intercepto $\hat{\beta}_0$, como:

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

- Esta é a estimativa de intercepto OLS, pois é calculada usando médias amostrais.
- A fórmula é direta, pois $\hat{\beta}_0$ é linear em $\hat{\beta}_1$.

Uso de Computadores para Cálculos OLS

- Embora a fórmula de $\hat{\beta}_0$ seja simples, o uso de computadores e softwares estatísticos é fundamental.
- Mesmo quando n (tamanho da amostra) é pequeno, esses cálculos podem ser tediosos, e usamos computadores para realizá-los.

Valores Ajustados e Resíduos

- Para qualquer par de estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$, definimos o valor ajustado para cada observação i como:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Esse é o valor que prevemos para y_i, dado que $x = x_i$.
- No entanto, há erro de predição, já que $y \neq y_i$. Chamamos essa diferença de *resíduo*, denotado como $\hat{u}_i$.

Cálculo dos Resíduos

- O resíduo para cada observação i é calculado como:

$$\hat{u}_i = y_i - \hat{y}_i$$

- Substituindo $\hat{y}_i$:

$$\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

- Os resíduos medem o erro de predição para cada observação, ou seja, a diferença entre o valor observado e o valor ajustado.

Diferença entre Resíduo e Termo de Erro

- Tanto o resíduo quanto o termo de erro são representados por u, mas é importante entender a diferença.
- O *resíduo* é o erro de predição com base no valor ajustado $\hat{y}$ e no valor real y. Ele pode ser calculado facilmente com qualquer amostra de dados.
- O *termo de erro* u (sem o chapéu) é não observado pelo pesquisador e representa os determinantes não capturados pelo modelo.
- Essa distinção é crucial: o resíduo aparece nos dados após a regressão, enquanto o termo de erro nunca aparece na amostra.

Medindo o Tamanho do Erro

- Para medir o tamanho do erro para cada observação i, elevamos o resíduo ao quadrado. Isso elimina os valores negativos e torna todas as diferenças positivas.
- Isso é útil, pois somar os erros diretamente pode resultar em cancelamentos entre valores positivos e negativos.
- A soma dos resíduos quadráticos é expressa como:

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

Soma dos Quadrados dos Resíduos

- A equação acima é chamada de *soma dos quadrados dos resíduos*, pois $\hat{u}_i = y_i - \hat{y}_i$.
- No entanto, o resíduo é baseado nas estimativas da inclinação $\hat{\beta}_1$ e do intercepto $\hat{\beta}_0$.
- O objetivo do método de mínimos quadrados ordinários (OLS) é *minimizar a soma dos quadrados dos resíduos* escolhendo os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$.
- Usando cálculo, pode-se mostrar que a solução para esse problema de minimização fornece os mesmos parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$.

Equação da Regressão OLS

- Uma vez que temos os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ para um conjunto de dados, podemos escrever a equação da reta de regressão OLS:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Essa é a equação de predição que utilizamos para estimar o valor de y com base em qualquer valor de x.
- Podemos estudar variações em y a partir de mudanças em x a partir dessa equação:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x$$

Propriedades Algébricas dos Mínimos Quadrados Ordinários (OLS)

- As estimativas de $\hat{\beta}_0$ e $\hat{\beta}_1$ possuem algumas propriedades algébricas importantes.
- Essas propriedades garantem que os resíduos e as variáveis explicativas possuam relações específicas.

Soma dos Resíduos

Propriedade 1

A soma dos resíduos é sempre igual a zero:

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

- Essa propriedade é garantida pela maneira como os coeficientes $\hat{\beta}_0$ e $\hat{\beta}_1$ são escolhidos.

Resíduos e Variáveis Explicativas

Propriedade 2

A soma dos resíduos ponderada pelas variáveis explicativas também é zero:

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0$$

- Isso significa que não há correlação entre os resíduos e X, o que é uma consequência da minimização dos resíduos na OLS.

Resíduos e Valores Ajustados

Propriedade 3

Os resíduos são não correlacionados com os valores ajustados $\hat{Y}_i$.

$$\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$$

- Isso é uma consequência direta da forma como a OLS ajusta a linha de regressão aos dados.

Ponto Médio na Linha de Regressão

Propriedade 4

O ponto médio dos dados está na linha de regressão:

$$(\bar{X}, \bar{Y}) \text{ está na linha } Y = \hat{\beta}_0 + \hat{\beta}_1 X$$

- A linha de regressão passa sempre pelo ponto médio da amostra.

Soma dos Quadrados

Definições

- Soma Total dos Quadrados (SST):

$$SST = \sum (Y_i - \bar{Y})^2$$

- Soma dos Quadrados Explicada (SSE):

$$SSE = \sum (\hat{Y}_i - \bar{Y})^2$$

- Soma dos Quadrados dos Resíduos (SSR):

$$SSR = \sum (Y_i - \hat{Y}_i)^2$$

R-quadrado

Definição de R-quadrado

$$R^2 = \frac{SSE}{SST}$$

- R^2 mede a fração da variação total de Y explicada pela regressão.
- Varia entre 0 (nenhuma variação explicada) e 1 (toda a variação explicada).

Interpretação do R-quadrado

- Um R^2 próximo de 1 indica que a regressão explica a maior parte da variação em Y.
- No entanto, um alto R^2 não implica causalidade. O objetivo principal da regressão é a estimativa do efeito causal de X sobre Y, não apenas explicar a variação.

Goodness-of-fit (Qualidade de Ajuste)

- Para avaliar a qualidade de ajuste da regressão, comparamos a variação explicada pelos valores ajustados $\hat{Y}_i$ com a variação total de Y.
- Usamos as somas dos quadrados para medir isso.

Componentes da Variação

Somas dos Quadrados

- Soma Total dos Quadrados (SST): $\sum(Y_i - \bar{Y})^2$
- Soma dos Quadrados Explicada (SSE): $\sum(\hat{Y}_i - \bar{Y})^2$
- Soma dos Quadrados dos Resíduos (SSR): $\sum(Y_i - \hat{Y}_i)^2$

- $SST = SSR + SSE$, ou seja, a variação total é a soma da variação explicada e da variação residual.
- Essa é uma aplicação de Pitágoras

Interpretação do R-quadrado

- Um R^2 alto indica que a regressão ajusta bem os dados.
- No entanto, um R^2 alto não implica causalidade; ele apenas indica o grau de ajuste da regressão.
- A qualidade de ajuste é útil, mas não é o foco principal quando buscamos estimar efeitos causais.

Cuidado com o R^2 em Pesquisa Causal

- O R^2 não deve ser o foco em projetos de pesquisa que visam estimar efeitos causais.
- Embora seja uma medida útil da qualidade do ajuste, ele não nos diz nada sobre se o coeficiente β_1 representa um efeito causal verdadeiro.
- Em inferência causal, estamos mais preocupados com a validade dos nossos estimadores e suposições.

Valor Esperado dos Estimadores OLS

- Nosso objetivo é estudar as propriedades estatísticas dos estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$, em particular, se eles são viesados ou não.
- Em outras palavras, queremos saber se os estimadores OLS fornecem, em média, o verdadeiro valor dos parâmetros β_0 e β_1 - especialmente esse último.
- Isso envolve calcular o *valor esperado* do estimador $\hat{\beta}_1$.
- Queremos saber se $E(\hat{\beta}_1) = \beta_1$

Hipóteses para ausência de viés

- O modelo é linear nos parâmetros (ou: é corretamente especificado)
- Temos uma amostra aleatória, de tamanho n, obtida a partir do modelo populacional
- A variável explicativa não é constante
- *Média condicional igual a zero*: essa é a principal hipótese para inferência causal

Cálculo do Valor Esperado

- Para determinar se o estimador OLS é não viesado, precisamos calcular $E(\hat{\beta}_1)$.
- Vamos começar escrevendo $\hat{\beta}_1$ de uma forma alternativa

Reescrevendo o Estimador $\hat{\beta}_1$

- A fórmula para o estimador $\hat{\beta}_1$ é:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- Substituímos Y_i pelo modelo populacional $Y_i = \beta_0 + \beta_1 X_i + u_i$, onde u_i é o termo de erro:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})[(\beta_0 + \beta_1 X_i + u_i - \bar{Y})]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- Expanda $\bar{Y}$ usando a fórmula $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$, onde $\bar{u}$ é a média dos erros:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})[\beta_0 + \beta_1 X_i + u_i - \beta_0 - \beta_1 \bar{X} - \bar{u}]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Reescrevendo o Estimador $\hat{\beta}_1$

- Simplificando:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})[\beta_1(X_i - \bar{X}) + u_i - \bar{u}]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \beta_1 \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sum_{i=1}^{n}(X_i - \bar{X})[u_i - \bar{u}]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Reescrevendo o Estimador $\hat{\beta}_1$

- Podemos simplificar mais um pouco, notando que:

$$\sum_{i=1}^{n}(X_i - \bar{X})\bar{u} = \bar{u}\sum_{i=1}^{n}(X_i - \bar{X}) = 0$$

- Portanto:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- Observe que essa é uma função linear em u_i

Encontrando $E(\hat{\beta}_1)$

- Agora tomamos o valor esperado de ambos os lados da equação:

$$E(\hat{\beta}_1|X) = \beta_1 + E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}|X\right)$$

- Sabemos que $E(u_i \mid X_i) = 0$, ou seja, o termo de erro não está correlacionado com X.
- Portanto, a expectativa do termo com u_i é zero:

$$E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}u_i|X\right) = \left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)E(u_i|X) = 0$$

- Concluímos que:

$$E(\hat{\beta}_1) = \beta_1$$

- Logo, $\hat{\beta}_1$ é um estimador *não viesado* de β_1.

Análise de $E(\hat{\beta}_0)$

- A fórmula para $\hat{\beta}_0$ é:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

- Substituímos $\bar{Y}$ e $\hat{\beta}_1$:

$$\hat{\beta}_0 = \beta_0 + \beta_1 \bar{X} + \bar{u} - \hat{\beta}_1 \bar{X}$$

- Tomando o valor esperado de ambos os lados:

$$E(\hat{\beta}_0) = \beta_0 + \beta_1 \bar{X} + E(\bar{u}) - E(\hat{\beta}_1 \bar{X})$$

- Sabemos que $E(\bar{u}) = 0$ e $E(\hat{\beta}_1) = \beta_1$, então:

$$E(\hat{\beta}_0) = \beta_0$$

- Logo, $\hat{\beta}_0$ é um estimador *não viesado* de β_0.

Conclusão: Estimadores OLS Não Viesados

- Concluímos que, sob as suposições do modelo de regressão linear clássico, os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ são *não viesados*:

$$E(\hat{\beta}_0) = \beta_0 \quad \text{e} \quad E(\hat{\beta}_1) = \beta_1$$

- Ou seja, em média, os estimadores OLS fornecem os valores corretos dos parâmetros populacionais.
- Isso garante que o método dos mínimos quadrados ordinários (OLS) é um método confiável para estimar os parâmetros de uma regressão linear, desde que as suposições do modelo sejam válidas.

Lei das Expectativas Iteradas

- A *Lei das Expectativas Iteradas* é uma propriedade chave no cálculo da esperança condicional.
- Ela afirma que a expectativa total de uma variável aleatória pode ser obtida tomando a expectativa condicional em duas etapas.

Definição Formal

Lei das Expectativas Iteradas

$$E(Y) = E_X(E_Y(Y \mid X))$$

- A expectativa de Y é igual à expectativa da expectativa condicional de Y, dado X.
- Isso significa que, ao calcular a expectativa de uma variável, podemos primeiro calcular sua expectativa condicional em relação a outra variável, e depois tomar a expectativa disso.

Exemplo

- Suponha que você tenha uma população dividida em grupos com base em uma variável X.
- A expectativa total de uma variável Y pode ser obtida calculando a expectativa dentro de cada grupo de X e, em seguida, calculando a média ponderada dessas expectativas.
- Isso é exatamente o que a Lei das Expectativas Iteradas afirma.

Aplicação no Modelo de Regressão

- No contexto da regressão, essa lei nos ajuda a entender o comportamento da variável dependente Y, dado X.
- A expectativa condicional de Y em relação a X é a função de regressão:

$$E(Y \mid X) = \beta_0 + \beta_1 X$$

Conclusão

- A Lei das Expectativas Iteradas simplifica o cálculo da expectativa de uma variável ao permitir que seja dividida em expectativas condicionais.
- Essa propriedade é amplamente usada em inferência estatística e no desenvolvimento de modelos econométricos.

Propriedade de Decomposição da Função de Expectativa Condicional (CEF)

- A função de expectativa condicional (CEF) pode ser decomposta de maneira útil para análise.
- A relação entre a variável dependente Y e a variável explicativa X pode ser expressa como:

$$Y = E(Y \mid X) + u$$

Interpretação

- A parte $E(Y \mid X)$ é a expectativa condicional de Y dada X, ou seja, a parte explicada.
- O termo u é o erro ou parte não explicada, que tem expectativa zero.

Propriedade de Predição da Função de Expectativa Condicional (CEF)

- A função de expectativa condicional também tem uma propriedade de predição útil:

$$E(Y \mid X) \text{ minimiza } E((Y - g(X))^2)$$

Interpretação

- A função de expectativa condicional $E(Y \mid X)$ é a melhor previsão de Y, dado X, no sentido de minimizar o erro quadrático médio.
- Isso justifica o uso da CEF como preditor em modelos de regressão.

Teoria ANOVA

- A Análise de Variância (ANOVA) decompõe a variação total de Y em componentes atribuíveis a diferentes fontes.
- No contexto da regressão, a variação total de Y pode ser dividida em:

$$SST = SSE + SSR$$

Onde:

  - SST é a soma total dos quadrados.
  - SSE é a soma dos quadrados explicada pela regressão.
  - SSR é a soma dos quadrados dos resíduos (erro).

Interpretação

- A ANOVA nos permite quantificar quanta variação em Y é explicada pelo modelo de regressão em comparação à variação não explicada.
- Essa decomposição é útil para entender a qualidade de ajuste do modelo.
- Na prática... nunca usamos ANOVA em inferência causal.

Teorema da Função de Expectativa Condicional Linear (CEF)

- O *Teorema da CEF Linear* afirma que, se a relação entre Y e X for linear, a CEF também será linear:

$$E(Y \mid X) = \beta_0 + \beta_1 X$$

Implicações

- Isso significa que a melhor predição de Y, dado X, será uma função linear de X.
- Esse teorema fundamenta o uso de modelos de regressão linear para descrever relações causais entre variáveis.

Teorema do Melhor Previsor Linear (Best Linear Predictor)

- O *Teorema do Melhor Previsor Linear* afirma que, entre todos os preditores lineares de Y, o modelo OLS é o melhor, no sentido de minimizar o erro quadrático médio:

$$E((Y - \hat{Y})^2)$$

Implicações

- O OLS é o melhor previsor linear, independentemente da distribuição de Y e X.
- Isso justifica o uso do OLS como método padrão para estimar relações lineares.

Teorema da Regressão como Função de Expectativa Condicional (CEF)

- O *Teorema da Regressão CEF* afirma que a regressão OLS é uma aproximação da função de expectativa condicional:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Implicações

- A regressão linear estimada por OLS fornece a melhor aproximação possível para $E(Y \mid X)$, sob a hipótese de linearidade.
- OLS funciona bem mesmo se a CEF não for linear.

Teorema da Anatomia da Regressão

- O *Teorema da Anatomia da Regressão* permite decompor a variação em Y de forma que cada variável explicativa $X_1, X_2, \ldots, X_k$ possa ser analisada individualmente.
- Podemos utilizar uma *regressão auxiliar* para examinar a contribuição de uma variável específica à variação total de Y.

Regressão Auxiliar

- A *regressão auxiliar* consiste em regredir uma variável explicativa de interesse, digamos X_1, nas outras variáveis explicativas $X_2, X_3, \ldots, X_k$.
- A equação da regressão auxiliar é:

$$X_1 = \gamma_0 + \gamma_2 X_2 + \cdots + \gamma_k X_k + v$$

- O resíduo v da regressão auxiliar é a parte de X_1 que não é explicada pelas outras variáveis.

Uso do Resíduo da Regressão Auxiliar

- Uma vez que obtemos o resíduo v da regressão auxiliar, podemos regredir Y em v para entender o efeito puramente linear de X_1 sobre Y, eliminando a colinearidade com as outras variáveis $X_2, X_3, \ldots, X_k$.
- Essa segunda regressão isola a contribuição única de X_1 na explicação de Y.

Exemplo: Tamanho da Família e Oferta de Trabalho

- Considere o exemplo da relação entre o *tamanho da família* (X_1) e a *oferta de trabalho* (Y).
- Podemos estar interessados no efeito do tamanho da família, controlando por outras variáveis, como idade, escolaridade e localização.
- Regredimos o tamanho da família (X_1) nas variáveis de controle ($X_2, X_3, \ldots$) usando uma regressão auxiliar para obter o resíduo v, que é a parte de X_1 que não é explicada por essas variáveis.

Aplicação do Exemplo

- Depois, regredimos a oferta de trabalho Y no resíduo v. Isso nos permite ver o efeito direto do tamanho da família sobre a oferta de trabalho, sem a influência das outras variáveis.
- O resíduo v representa a variação "limpa" de X_1 (tamanho da família) que não está correlacionada com as outras variáveis explicativas.
- Podemos escrever

$$\beta_1 = Cov(Y, v)/Var(v)$$

- Na prática, para duas variáveis: 1- regredimos Y em X_1 e computamos os resíduos; 2- regredimos X_2 em X_1 e computamos os resíduos; 3- regredimos os resíduos do primeiro passo no segundo passo, e obtemos o mesmo coeficiente da regressão original.

Implicações do Teorema da Anatomia da Regressão

- O uso da regressão auxiliar e dos resíduos nos permite isolar o impacto de uma variável específica (X_1) na variável dependente (Y).
- Isso é particularmente útil em casos em que há colinearidade entre variáveis explicativas.
- A decomposição da variância nos ajuda a entender o quanto da variação de Y é explicada por cada variável, e o quanto não é explicada pelo modelo.

Variância dos Estimadores OLS

- A *variância dos estimadores OLS* $\hat{\beta}_0$ e $\hat{\beta}_1$ é crucial para determinar a precisão das estimativas de regressão.
- A variância de $\hat{\beta}_1$ mede a dispersão das estimativas de β_1 em diferentes amostras, refletindo a incerteza na estimativa do coeficiente.

Homocedasticidade e Variância de $\hat{\beta}_1$

- A *homocedasticidade* ocorre quando a variância do erro u é constante para todos os valores de X.
- Formalmente, sob a hipótese de homocedasticidade:

$$E(u^2 \mid X) = \sigma^2$$

- Embora a homocedasticidade *não seja necessária para a ausência de viés* nos estimadores OLS, ela facilita o cálculo da variância de $\hat{\beta}_1$.

Variância de Y Condicional a X

Variância Condicional

Sob homocedasticidade, a variância de Y condicional a X é constante e dada por:

$$V(Y \mid X) = \sigma^2$$

- Isso implica que a variância de Y não é afetada por X quando o erro u é homocedástico.
- O erro tem uma variância constante para todos os valores de X, o que simplifica os cálculos e interpretações.

Fórmula para a Variância de $\hat{\beta}_1$

Fórmula Geral

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2} = \frac{\sigma^2}{SST_X}$$

- Onde $SST_X = \sum(X_i - \bar{X})^2$ é a soma dos quadrados de X.
- A variância de $\hat{\beta}_1$ é inversamente proporcional à variabilidade de X.
- A maior variabilidade em X leva a uma menor variância de $\hat{\beta}_1$, ou seja, uma estimativa mais precisa.

Efeito do Tamanho da Amostra na Variância de $\hat{\beta}_1$

- O tamanho da amostra, n, afeta diretamente a variância de $\hat{\beta}_1$.
- Podemos relacionar SST_X com a variância amostral de X, pois:
$$\frac{SST_X}{n} = s_X^2$$
- Quanto maior for n, menor será a variância de $\hat{\beta}_1$, pois a variância amostral s_X^2 torna-se uma estimativa mais precisa da verdadeira variância de X na população.

Estimativa de σ^2 e Ajuste de Graus de Liberdade

- Para estimar σ^2, podemos usar a soma dos quadrados dos resíduos (SSR):
$$SSR = \sum \hat{u}_i^2$$
- Dividindo por n, obtemos:
$$\frac{SSR}{n}$$
- No entanto, essa é uma *estimativa viesada* de σ^2, pois não leva em conta o número de parâmetros estimados no modelo.

Ajuste nos Graus de Liberdade

- Para corrigir o viés, fazemos um ajuste nos graus de liberdade, dividindo o SSR por 2:

$$\hat{\sigma}^2 = \frac{SSR}{n-2}$$

- Esse ajuste produz uma estimativa não viesada de σ^2, que é utilizada para calcular a variância de $\hat{\beta}_1$ e $\hat{\beta}_0$.

Fatores que Afetam a Variância de $\hat{\beta}_1$

- A variância de $\hat{\beta}_1$ depende de:
  1. *Variância dos erros* (σ^2): Se os erros têm grande variabilidade, isso aumenta a variância de $\hat{\beta}_1$.
  2. *Número de observações* (n): Mais observações reduzem a variância de $\hat{\beta}_1$.
  3. *Dispersão de X* (SST_X): Quanto maior a variação de X, menor a variância de $\hat{\beta}_1$.

Fórmula para a Variância de $\hat{\beta}_0$

Fórmula

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{SST_X} \right)$$

- A variância de $\hat{\beta}_0$ depende da variância de X e do tamanho da amostra.
- Quanto maior a variação de X e o número de observações n, menor será a variância de $\hat{\beta}_0$.

Erros-Padrão

- A *raiz quadrada da variância* dos estimadores OLS nos dá os *erros-padrão*:
$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$
- Os erros-padrão são usados para construir intervalos de confiança e realizar testes de hipóteses sobre os coeficientes.
- Erros-padrão menores indicam maior precisão nas estimativas.

Resumo: Importância da Variância dos Estimadores

- As variâncias dos estimadores OLS $\hat{\beta}_0$ e $\hat{\beta}_1$ são fundamentais para avaliar a precisão das estimativas.
- A variabilidade de X, o tamanho da amostra e a variabilidade dos erros afetam diretamente a variância e, portanto, a precisão das estimativas.
- Variâncias menores resultam em estimativas mais confiáveis, o que é essencial para inferências robustas em modelos de regressão.

Erros-Padrão Robustos

- Quando as suposições clássicas do OLS são violadas, como heterocedasticidade, os erros-padrão estimados podem ser incorretos.
- *Erros-padrão robustos* ajustam essa inconsistência:

Implicações dos Erros-Padrão Robustos

- Erros-padrão robustos são usados para corrigir problemas de heterocedasticidade, sem precisar ajustar o modelo inteiro.
- Permitem que as inferências sejam válidas, mesmo que as variâncias dos erros não sejam constantes.

Erros-Padrão Robustos para Cluster

- Em muitos casos, as observações podem estar agrupadas em clusters (grupos) que compartilham características, como regiões ou empresas.
- Os *erros-padrão cluster-robustos* ajustam para dependência dentro dos clusters.

Implicações dos Erros-Padrão Cluster-Robustos

- Esses erros-padrão levam em conta a correlação dentro de clusters, permitindo que as inferências sejam válidas mesmo que as observações dentro de um cluster não sejam independentes.
- São amplamente utilizados em estudos empíricos, especialmente quando há agrupamentos naturais nas observações.