

Contribuição Marginal

Econometria

Alexandre Gori Maia

Ementa:

- Definição;
- ANOVA para Contribuição Marginal;
- Correlação Parcial;

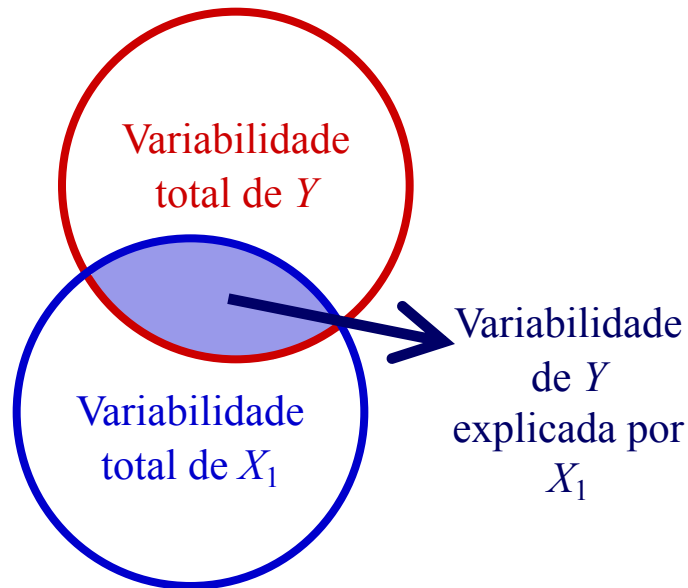
Bibliografia Básica:

Maia, Alexandre Gori (2017). *Econometria: Conceitos e Aplicações*. Cap. 9.

Contribuição Marginal

Seja o modelo de RLS:

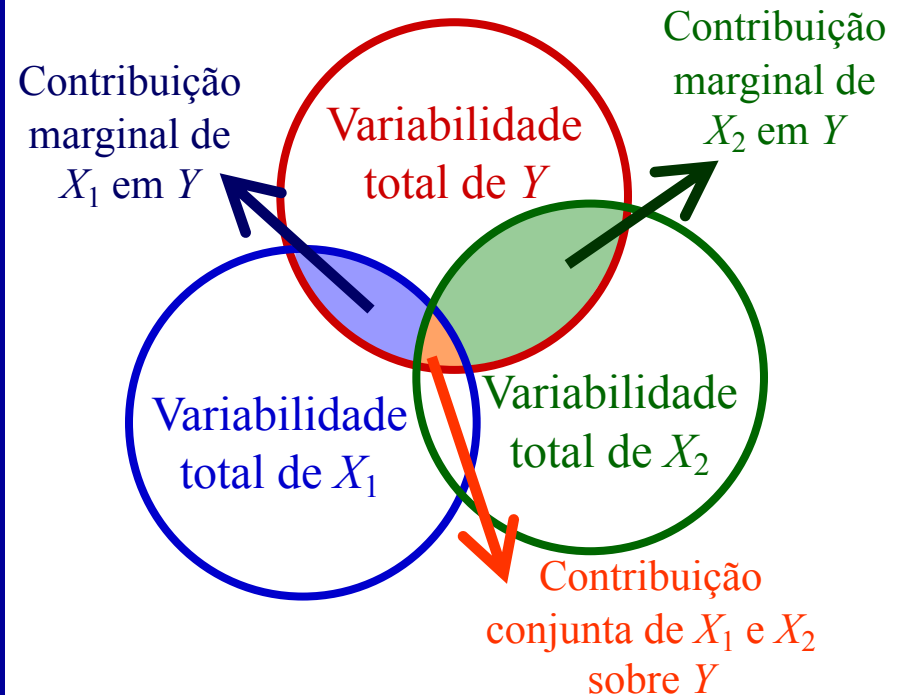
$$Y = \alpha + \beta_1 X_1 + e$$



O grau de relacionamento linear entre X_1 e Y determinará a parcela da variabilidade de Y explicada unicamente por X_1 .

Seja o modelo de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$



Qual a parcela da variabilidade de Y explicada isoladamente por X_2 ?

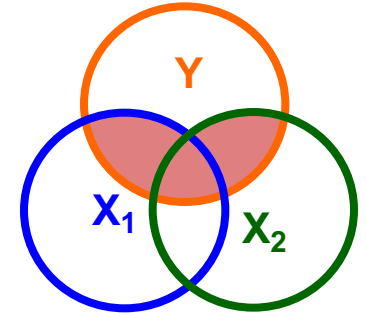
Em outras palavras, qual a contribuição marginal de X_2 , depois de considerada a contribuição de X_1 ?

Contribuição Marginal ($q=1$)

SQReg devido a X_1 e X_2 (Irrestrito):

$$SQReg(Y / X_1 \text{ e } X_2) \quad \text{ou} \quad SQReg_{ir}$$

Variabilidade de Y explicada pelo conjunto das variáveis X_1 e X_2 .

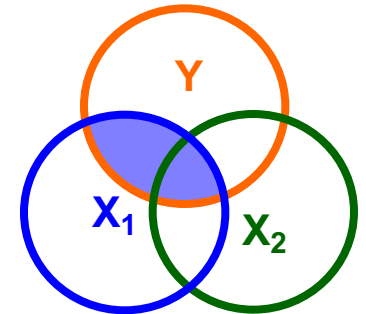


Graus de liberdade: 2 *coeficientes angulares do modelo*
 $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e.$

SQReg devido exclusivamente a X_1 (Restrito):

$$SQReg(Y / X_1) \quad \text{ou} \quad SQReg_r$$

Variabilidade de Y explicada exclusivamente por X_1 .

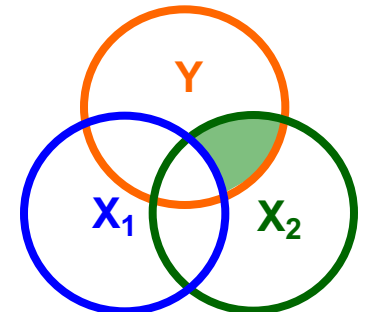


Graus de liberdade: 1 *coeficiente angular do modelo*
 $Y = \alpha + \beta_1 X_1 + e.$

SQReg devido ao acréscimo de X_2 :

$$\text{Contribuição } X_2 = SQReg_{ir} - SQReg_r$$

Variabilidade de Y explicada por X_2 após considerada a variabilidade explicada por X_1 .



Graus de liberdade: 1 *novo coeficiente angular incorporado no modelo (β_2).*

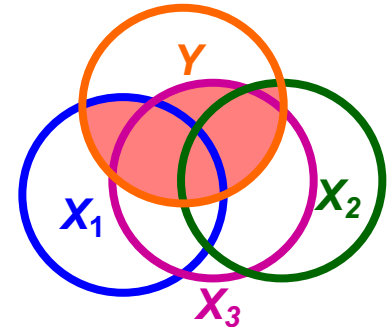
Contribuição Marginal ($q=2$)

SQReg devido a X_1 , X_2 e X_3 (Irrestrito):

$$SQReg(Y / X_1, X_2, X_3) \quad \text{ou} \quad SQReg_{ir}$$

Variabilidade de Y explicada pelo conjunto das variáveis X_1 , X_2 e X_3 :

Graus de liberdade: 3 *coeficientes angulares do modelo*
 $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$

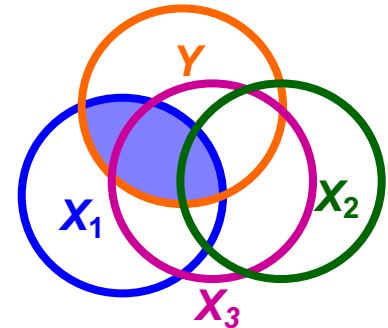


SQReg devido exclusivamente a X_1 (Restrito):

$$SQReg(Y / X_1) \quad \text{ou} \quad SQReg_r$$

Variabilidade de Y explicada exclusivamente por X_1

Graus de liberdade: 1 *coeficiente angular do modelo $Y = \alpha + \beta_1 X_1 + e$.*

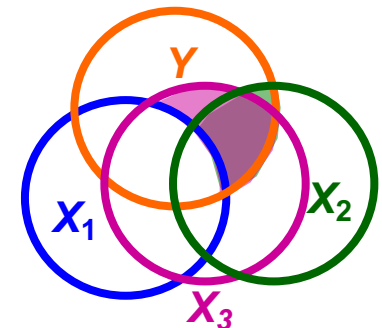


SQReg devido ao acréscimo de X_2 e X_3 :

$$\text{Contribuição de } X_2 \text{ e } X_3 = SQReg_{ir} - SQReg_r$$

Variabilidade de Y explicada por X_2 e X_3 após considerada a variabilidade explicada por X_1

Graus de liberdade: 2 *novos coeficientes angulares incorporados no modelo (β_2 e β_3).*



Teste de Restrição aos Parâmetros

Seja o modelo *irrestrito* de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

Para verificarmos se a contribuição de um grupo de q variáveis independentes é significativa no modelo irrestrito devemos, primeiro, elaborar um modelo com restrição aos parâmetros. Suponha que, por simplicidade, as q variáveis que desejamos testar são as últimas das k variáveis do modelo irrestrito (a ordem, obviamente, não faz importância). Nosso modelo *restrito* seria:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{k-q} X_{k-q} + e$$

Em outras palavras, estaríamos interessados em testar a hipótese nula de que os q coeficientes do modelo irrestrito são nulos:

$$H_0 : \beta_{k-q+1} = 0, \dots, \beta_k = 0 \quad \text{e} \quad H_1 : \text{pelo menos um } \beta_{k-q+j} \neq 0$$

O teste estatístico para restrição aos parâmetros consiste em verificar se a contribuição marginal dessas q variáveis é significativa. A estatística F será dada por:

$$F = \frac{(SQ\text{Re } g_{ir} - SQ\text{Re } g_r) / q}{SQ\text{Re } s_{ir} / (n - k - 1)} \quad \text{ou} \quad F = \frac{(SQ\text{Re } s_r - SQ\text{Re } s_{ir}) / q}{SQ\text{Re } s_{ir} / (n - k - 1)}$$

Onde $SQ\text{Reg}_{ir}$ e $SQ\text{Reg}_r$ são, respectivamente, a soma dos quadrados da regressão sem e com restrição nos parâmetros, $SQ\text{Res}_{ir}$ é a soma dos quadrados dos resíduos da regressão sem restrição.

ANOVA para Contribuição Marginal

Fonte	gl	Soma dos Quadrados	Quadrados Médios	F
Regressão Restrita	$k - q$	$SQRe g_r$	$SQRe g_r / (k - q)$	$F = \frac{QMReg_{contribuição}}{QMRes_{ir}}$
Contribuição	q	$SQRe g_{ir} - SQRe g_r$	$\frac{(SQRe g_{ir} - SQRe g_r)}{q}$	
Regressão Irrestrita	k	$SQRe g_{ir}$	$SQRe g_{ir} / k$	$F = \frac{QMReg_{ir}}{QMRes_{ir}}$
Resíduos	$n - (k + 1)$	$SQRe s_{ir}$	$SQRe s_{ir} / (n - k - 1)$	
Total	$n - 1$	STQ		

Teste F para Contribuição Marginal

Seja o modelo de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

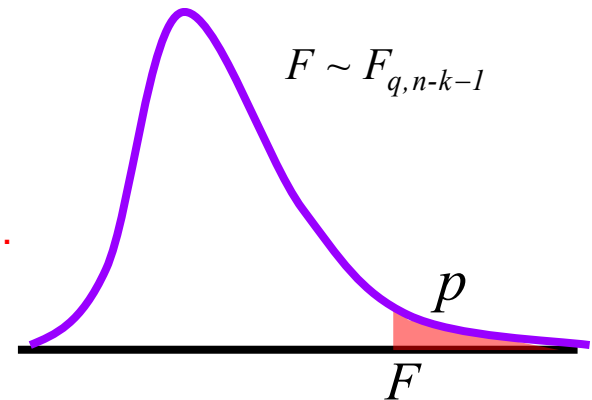
Para testarmos a contribuição marginal de um grupo de q variáveis, teremos as hipóteses:

$$\begin{cases} H_0: \beta_{k-q+1} = \dots = \beta_k = 0 \text{ (não contribui)} \\ H_1: \text{Pelo menos um } \beta_j \neq 0 \text{ (contribui)} \end{cases}$$

A estatística de teste será

$$F = \frac{(SQReg_{ir} - SQReg_r)/q}{SQRes_{ir}/(n - k - 1)}$$

Considerando H_0
verdadeiro, a
distribuição de F será...



Rejeitar H_0 significa afirmar que o grupo de q variáveis contribui para explicar Y , ou seja, há relação significativa entre pelo menos uma destas variáveis explicativas e a variável dependente.

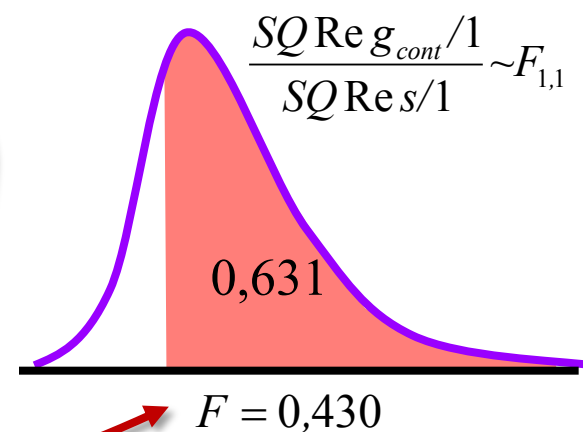
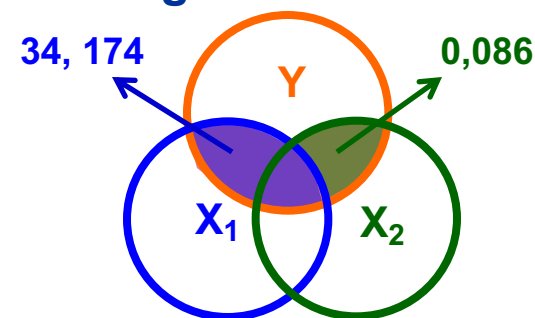
Contribuição Marginal - Exemplo

Na relação entre renda familiar (Y), anos de estudo (X_1) e idade (X_2) do responsável pela família, a contribuição marginal da idade é significativa?

Modelo Irrestrito: $Y_i = 1,9 + 1X_{1_i} + 0,06X_{2_i} + \hat{e}_i$

Modelo Restrito: $Y_i = 2,714 + 1,286X_{1_i} + \hat{e}_i$

Fonte	gl	Soma dos Quadrados	Quadrados Médios	F
Regressão Restrita	1	34.174	34.174	
Contribuição Marginal	1	0.086	0.086	0.430
Regressão Irrestrita	2	34.800	17.400	87.000
Resíduos	1	0.200	0.200	
Total	3	35.000		



Não há evidências para afirmar que a contribuição marginal da idade sobre a variabilidade da renda familiar seja significativa. A probabilidade de erro ao fazermos tal afirmação é muito alta, de aproximadamente 63%.

Correlação Parcial - Definição

Seja o modelo de RLM:

$$Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + e_i$$

Da mesma forma que análise de variância permite considerar a contribuição marginal de uma variável explanatória, podemos estender o conceito de correlação simples para ver em que medida a variável dependente e uma das variáveis independentes estão relacionadas, depois de isolados os efeitos das demais variáveis explanatórias do modelo:

$$r_{Y1.23\dots k}$$

O $r_{Y1.23\dots k}$ é a correlação parcial entre Y e X_1 , mantendo-se constantes os efeitos das $k-1$ variáveis independentes restantes. Seu quadrado, $r^2_{Y1.23\dots k}$ é o respectivo coeficiente de determinação parcial.

O respectivo coeficiente de determinação parcial, que mede a parcela da variabilidade de Y não associada aos demais regressores que é explicada exclusivamente por X_1 será representado por

$$r^2_{Y1.23\dots k}$$

Correlação Parcial - Conceito

Seja o modelo de RLM: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Para calcularmos $r_{Y1.2}$, poderíamos seguir os seguintes passos:

Passo 1) Estimar resíduos do ajuste: $Y = \hat{\delta}_0 + \hat{\delta}_1 X_2 + \hat{e}_{y2}$

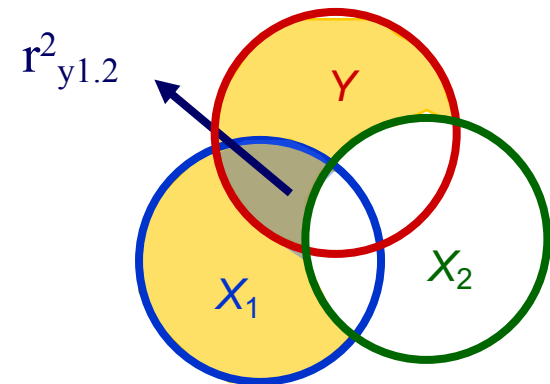
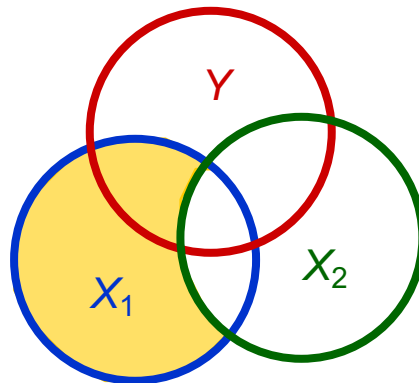
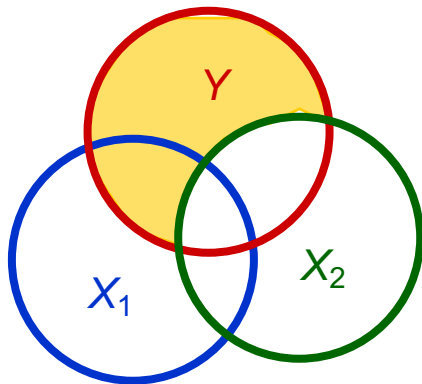
$\hat{e}_{y2}$ contém a parcela de Y não associada a X_2 .

Passo 2) Estimar resíduos da relação: $X_1 = \hat{\gamma}_0 + \hat{\gamma}_1 X_2 + \hat{e}_{12}$

$\hat{e}_{12}$ contém a parcela de X_1 não associada a X_2 .

Passo 3) Estimar correlação entre os resíduos dos dois modelos:

$$r_{Y1.2} = r_{\hat{e}_{y2}\hat{e}_{12}}$$



Correlação Parcial - Cálculo

Coefficiente de Correlação Parcial

No caso de duas variáveis independentes:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Então a correlação parcial poderá ser diretamente obtida por:

$$r_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{(1-r_{12}^2)(1-r_{Y2}^2)}} \quad \text{e} \quad r_{Y2.1} = \frac{r_{Y2} - r_{Y1}r_{12}}{\sqrt{(1-r_{12}^2)(1-r_{Y1}^2)}}$$

Analogamente, o coeficiente de determinação parcial será dado por:

$$r^2_{Y1.2} = \frac{R^2 - r_{Y2}^2}{1 - r_{Y2}^2}$$

Correlação Parcial - Exemplo

Vamos considerar a relação para a renda (Y) como função dos anos de estudo (X_1) e idade (X_2).

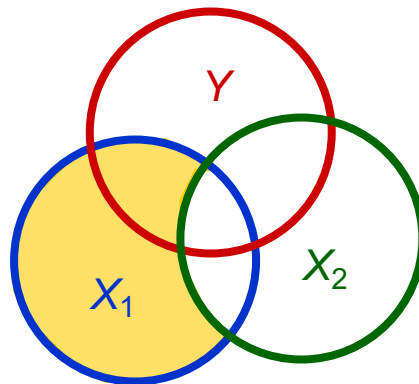
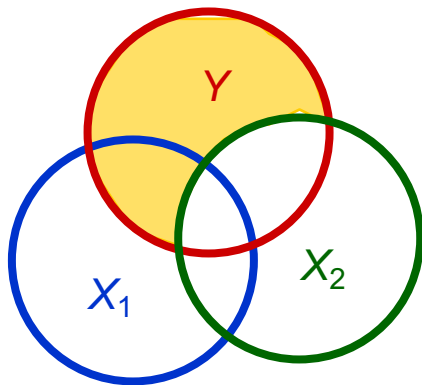
$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e \quad \Rightarrow \quad Y_i = 1,9 + 1X_{1_i} + 0,06X_{2_i} + \hat{e}_i$$

Para calcularmos a correlação parcial entre renda e anos de estudo ($r_{Y1.2}$), podemos seguir os seguintes passos:

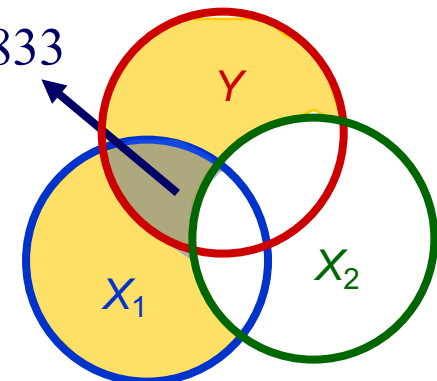
Passo 1) Estimar resíduos do ajuste: $Y = 2,714 + 1,286X_2 + \hat{e}_{y2}$

Passo 2) Estimar resíduos da relação: $X_1 = -2,5 + 0,2X_2 + \hat{e}_{12}$

Passo 3) Estimar correlação entre os resíduos dos dois modelos: $r_{Y1.2} = r_{\hat{e}_{y2}\hat{e}_{12}} = 0,913$



$$r_{y1.2}^2 = 0,833$$



Correlação Parcial - Exemplo

Vamos considerar a relação para a renda (Y) como função dos anos de estudo (X_1) e idade (X_2).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e \quad \Rightarrow \quad Y_i = 1,9 + 1X_{1_i} + 0,06X_{2_i} + \hat{e}_i$$

Podemos ainda calcular a correlação parcial entre renda e anos de estudo ($r_{Y1.2}$) diretamente por:

$$r_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{(1-r_{12}^2)(1-r_{Y2}^2)}} = \frac{(0,9959) - (0,9827)(-0,9757)}{\sqrt{[1 - (0,9759)^2][1 - (0,9827)^2]}} = 0,913$$

Onde:

$$r_{Y1} = \frac{\sum_{i=1}^n x_{1i} y_i}{\sqrt{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n y_i^2}} = \frac{27}{\sqrt{(21)(35)}} = 0,9959$$

$$r_{Y2} = \frac{\sum_{i=1}^n x_{2i} y_i}{\sqrt{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n y_i^2}} = \frac{130}{\sqrt{(500)(35)}} = 0,9827$$

$$r_{12} = \frac{\sum_{i=1}^n x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2}} = \frac{100}{\sqrt{(21)(500)}} = 0,9759$$