4.2 Coeficiente de determinação R^2

Vejamos primeiro um resultado que nos permite entender a importância de escolher modelos lineares. Encontremos a função $f(X_1, \cdots, X_p)$, preditora que maximiza a correlação entre Y e f, considerando as variáveis preditoras $X_1, \cdots, X_p$ aleatórias. Entendemos aqui por função preditora a qualquer função que satisfaça $\mathrm{E}(Y) = f(X_1, \cdots, X_p)$.

Teorema 4.6. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2 \mathrm{I})$. Seja $f(\underset{\sim}{X})$ uma função qualquer preditora de Y, considerando o vetor variáveis $\underset{\sim}{X} = (X_1, \cdots, X_p)$ como aleatório. Então, o coeficiente de correlação entre Y e $\mathrm{E}(Y|\underset{\sim}{X})$ é não-negativo e satisfaz que*

$$\rho[Y, \mathrm{E}(Y|\underset{\sim}{X})] \geq |\rho(Y, f)|\cdot$$

Demonstração. Para qualquer função f temos que

$$\begin{aligned} \mathrm{Cov}(Y, f) &= \mathrm{E}\{[f - \mathrm{E}(f)][Y - \mathrm{E}(Y)]\} = \mathrm{E}\{[f - \mathrm{E}(f)]Y\} \\ &= \mathrm{E}\{[f - \mathrm{E}(f)]\,\mathrm{E}(Y|\underset{\sim}{X})\} = \mathrm{Cov}[\,\mathrm{E}(Y|\underset{\sim}{X}), f]\cdot \end{aligned}$$

(Ver exercício 1-(b) em Exercícios da Seção 4.2). Quando $f = \mathrm{E}(Y|\underset{\sim}{X})$, $\mathrm{Cov}[Y, \mathrm{E}(Y|\underset{\sim}{X})] = \mathrm{Cov}[\,\mathrm{E}(Y|\underset{\sim}{X}), \mathrm{E}(Y|\underset{\sim}{X})]$, pelo resultado anterior. Então $\rho[Y, \mathrm{E}(Y|\underset{\sim}{X})] = \sqrt{\mathrm{Var}[\,\mathrm{E}(Y|\underset{\sim}{X})]/\,\mathrm{Var}(Y)} \geq 0$.

Agora

$$\rho^2(Y, f) = \frac{\mathrm{Cov}^2(Y, f)}{\mathrm{Var}(Y)\,\mathrm{Var}(f)} = \frac{\mathrm{Cov}^2(f, \mathrm{E}(Y|\underset{\sim}{X}))}{\mathrm{Var}(f)\,\mathrm{Var}[\,\mathrm{E}(Y|\underset{\sim}{X})]} \times \frac{\mathrm{Var}[\,\mathrm{E}(Y|\underset{\sim}{X})]}{\mathrm{Var}(Y)},$$

do qual obtemos que

$$\rho^2(Y, f) = \rho^2[f, \mathrm{E}(Y|\underset{\sim}{X})]\rho^2[Y, \mathrm{E}(Y|\underset{\sim}{X})] \leq \rho^2[Y, \mathrm{E}(Y|\underset{\sim}{X})]\cdot$$

□

O coeficiente de Pearson é apropriado principalmente para indicar associações lineares.

Obtemos da demonstração deste teorema que o limite superior de $\rho(Y, f)$ é atingido quando

$$\rho(f, \mathrm{E}(Y|\underset{\sim}{X})) = 1,$$

o qual implica que f seja uma função linear de $\mathrm{E}(Y|\underset{\sim}{X})$. Novamente o modelo linear é a resposta.

Faz sentido então utilizar o coeficiente de correlação entre Y e $\mathrm{E}(Y|\underset{\sim}{X})$ como medida de qualidade de um modelo e claro, este resultado nos permite entender melhor a definição a seguir do coeficiente de determinação.

Sabemos que uma medida eficaz de calcular a relação linear entre duas variáveis aleatórias é o coeficiente de correlação (ver Definição 1.5) e o coeficiente de determinação é justamente a máxima correlação ao quadrado entre a resposta e a melhor combinação linear de X. Esta definição deve-se a Anderson (1984).

Definição 4.2. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2 \mathrm{I})$. Define-se o coeficiente de determinação paramétrico, denotando-se por ϕ^2, como*

$$\phi^2 = \max_{\beta} \rho^2(Y, X\beta), \tag{4.6}$$

onde ρ representa o coeficiente de correlação entre Y e $X\beta$.

Uma medida eficaz de calcular a relação entre duas variáveis aleatórias é o coeficiente de correlação e o coeficiente de determinação é justamente a correlação ao quadrado entre as observações e os valores preditos pelo modelo. Especificamente, o coeficiente de determinação pode ser definido como a máxima correlação ao quadrado entre Y e a melhor combinação linear de X. Esta definição deve-se a Anderson (1984).

Observemos que a normalidade não é utilizada, meramente a existência dos segundos momentos. Um outro detalhe interessante é que esta definição teve origem nos estudos de estatística multivariada nos quais considera-se a matriz X como aleatória. No caso de modelos de regressão as variáveis regressoras são consideradas fixas, Helland (1987) provou que também podemos utilizar a definição acima na situação dos modelos lineares.

Tal como para outros modelos, dado um conjunto de dados, é importante ser capaz de verificar a qualidade do ajuste do modelo de regressão. Com este objetivo utilizamos o coeficiente de determinação paramétrico, denotado como ϕ^2. O estimador de ϕ^2 será chamado simplesmente de coeficiente de determinação e denotar-se-á R^2. Este coeficiente é uma medida da bondade do ajuste do modelo selecionado e também uma medida da precisão na predição, tanto de novas observações quanto da média de novas observações do modelo de regressão linear ajustado.

Existiram muitas formas diferentes de definir este coeficiente na sua longa história de existência. Considera-se que desde os primeiros desenvolvimentos de modelos de regressão no século XIX existem referências indiretas a este coeficiente. Percebemos que o coeficiente de determinação no modelo de regressão linear é justamente a máxima correlação ao quadrado entre as observações da variável resposta e os valores preditos pelo modelo. Durante muitos anos, esta definição do coeficiente de determinação foi a definição de R^2 melhor aceita na comunidade científica. Atualmente com o desenvolvimento de modelos de regressão mais complexos, como os modelos lineares generalizados e outros, esta definição do ϕ^2 ficou restrita e novas definições foram propostas. O leitor interessado em algumas destas novas definições pode consultar, por exemplo, o artigo de van der Linde & Tutz (2008).

Vejamos agora como estimar o coeficiente de determinação. Uma primeira observação na Definição 4.2 nos permite perceber que para estimar o ϕ^2 devemos estimar o coeficiente de correlação; desde que o estimador do coeficiente de correlação ao quadrado seja um estimador adequado para o coeficiente de determinação. Isto será resolvido no próximo resultado apresentado em Zehna (1966).

Teorema 4.7 (Princípio de Invariância)**.** *Sejam $\widehat{\theta}$ o estimador de máxima verossimilhança de $\theta \in \Theta$ e $h : \theta \to \mathbb{R}$, uma função injetora de Θ nos reais. Então $h(\widehat{\theta})$ é o estimador de máxima verossimilhança de $h(\theta)$.*

Teorema 4.8. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2 \mathrm{I})$. Seja $f(x_1, \cdots, x_p)$ uma função qualquer das variáveis explicativas de Y. Então, o coeficiente de correlação $\rho(Y, X\beta)$ é não-negativo e satisfaz que*

$$\rho(Y, X\beta) \geq |\rho(Y, f)|,$$

para qualquer função f.

Demonstração. Para cada $\lambda \in \mathbb{R}$, definamos

$$\Theta_\lambda = \{\theta : \theta \in \Theta, h(\theta) = \lambda\}$$

e

$$M(\lambda) = \sup_{\theta \in \Theta_\lambda} \ell(\theta)\cdot$$

Então, M definida nos reais é a função de verossimilhança induzida por h. Se $\widehat{\theta}$ é um estimador de máxima verossimilhança de θ, então $\widehat{\theta}$ pertence a um, e somente um, conjunto $\Theta_{\widehat{\lambda}}$. Dado que $\widehat{\theta} \in \Theta_{\widehat{\lambda}}$, $\widehat{\lambda} = h(\widehat{\theta})$. Agora

$$M(\widehat{\lambda}) = \sup_{\theta \in \Theta_{\widehat{\lambda}}} \ell(\theta) \geq \ell(\widehat{\theta})$$

e $\widehat{\lambda}$ maximiza M, de modo que

$$M(\widehat{\lambda}) \leq \sup_{\theta \in \mathbb{R}} M(\lambda) = \sup_{\theta \in \Theta} \ell(\theta) = \ell(\widehat{\theta}),$$

do qual obtemos que $M(\widehat{\lambda}) = \sup_{\theta \in \mathbb{R}} M(\lambda)$. Segue então que $\widehat{\lambda}$ é o estimador de máxima verossimilhança de $h(\theta)$, onde $\widehat{\lambda} = h(\widehat{\theta})$. □

Nosso objetivo é propor um estimador adequado para ϕ^2 e posteriormente investigar suas propriedades como medida da qualidade de um modelo. Prosseguindo nosso desenvolvimento, partimos do conhecimento prévio de que o coeficiente de correlação amostral ou o estimador de máxima verossimilhança do coeficiente de correlação entre duas variáveis aleatórias (Sen & Singer, 1993) é

$$\rho(Y, X) = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\mathrm{Var}(Y)\,\mathrm{Var}(X)}}.$$

Utilizando o Princípio da Invariância (Teorema 4.7) sugere-se como estimador de ϕ^2 ao quadrado de $\rho(Y, X\widehat{\beta})$. Isto justifica-se pelo fato da função quadrática ser injetora no intervalo $(0, 1)$, intervalo no qual está definido $\rho(Y, X\widehat{\beta})$, segundo demonstrado no Teorema 4.6. Esta é a ideia da demonstração do teorema a seguir, o qual fornece-nos a expressão mais conhecida do coeficiente de determinação amostral.

Ainda devemos esclarecer que, segundo o Teorema 4.6, se Y satisfaz um modelo de regressão então

$$\max_{\beta} \rho(Y, X\beta) = \rho(Y, X\widehat{\beta}).$$

O objetivo é selecionar um modelo que representa o máximo de variação em Y como é prático. Uma vez que o símbolo R2 não pode diminuir à medida que as variáveis independentes são adicionados ao modelo, o modelo que proporciona a máxima R2 será necessariamente o modelo que contém todas as variáveis independentes. O enredo típico de R2 em relação ao número de variáveis no modelo começa como uma curva acentuadamente ascendente, então os níveis de fora perto do R2 máximo uma vez que as variáveis mais importantes foram incluídos. Assim, a utilização do critério de R2 para a construção de modelos requer uma decisão quanto ao facto de o aumento de R2 a partir de variáveis adicionais, justifica o aumento da complexidade do modelo. O tamanho é escolhido subconjunto perto da curva onde a curva tende a aplanar.

Teorema 4.9. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2\,\mathrm{I})$. O estimador do coeficiente de determinação paramétrico ϕ^2, denotado por R^2 e conhecido como coeficiente de determinação amostral ou simplesmente coeficiente de determinação, é dado por*

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \widehat{\mu}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}, \tag{4.7}$$

onde $\widehat{\mu}_i$ representa os valores preditos pelo modelo de regressão linear normal

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i, \tag{4.8}$$

e $\overline{Y}$ *o correspondente valor predito pelo modelo*

$$Y_i = \beta_0 + \epsilon_i \cdot \tag{4.9}$$

Demonstração. O estimador do coeficiente de correlação entre Y e $X\widehat{\beta}$ é

$$\begin{aligned}\rho(Y, X\widehat{\beta}) &= \frac{\sum_{i=1}^n (Y_i - \overline{Y})(\widehat{\mu}_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}} = \frac{\sum_{i=1}^n (Y_i - \overline{Y} \pm \overline{Y})(\widehat{\mu}_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}} \\ &= \frac{\sum_{i=1}^n \left[(Y_i - \widehat{\mu}_i)(\widehat{\mu}_i - \overline{Y}) + (\widehat{\mu}_i - \overline{Y})^2\right]}{\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}} = \sqrt{\frac{\sum_{i=1}^n (\widehat{\mu}_i - \overline{Y})^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}}.\end{aligned}$$

Então

$$R^2 = \rho^2(Y, X\widehat{\beta}) = \frac{\sum_{i=1}^n (\widehat{\mu}_i - \overline{Y})^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2},$$

utilizando a relação em (2.28) provamos o resultado em (4.7). □

Lembremos que $\mathrm{E}(Y_i) = \mu_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$ para o modelo (4.8) com estimador $\widehat{\mu}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \ldots + \widehat{\beta}_p x_{ip}$ e no caso do modelo (4.9), $\mathrm{E}(Y_i) = \mu = \beta_0$ e $\widehat{\mu} = \widehat{\beta}_0 = \overline{Y}$.

Em 1885, Sir Francis Galton definiu o primeiro modelo de regressão e completou a teoria de correlação bivariada. Uma década mais tarde, Karl Pearson[3] desenvolveu o índice que ainda usamos para medir a correlação ρ, também chamado de coeficiente de correlação de Pearson.

Existem diversas formas de escrever R^2, por exemplo

$$R^2 = \frac{\sum_{i=1}^n (\widehat{\mu}_i - \overline{Y})^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}, \tag{4.10}$$

a qual foi obtida na demonstração do Teorema 4.9 que pode ser escrito como

$$R^2 = \frac{\frac{1}{n}\sum_{i=1}^n (\widehat{\mu}_i - \overline{Y})^2}{\mathrm{Var}(Y)}$$

ou

$$R^2 = \frac{SQReg}{SQT}.$$

Assim, o R^2 está relacionado com a decomposição da variância total de Y mostrando que este coeficiente é o quociente entre a soma de quadrados da regressão (SQReg) e a soma de quadrados total (SQT).

Com a introdução de generalizadas modelos de regressão linear, a atenção foi ainda mais para o facto de que um modelo de regressão pode especificar não só o E média condicional (Y | x), mas também a densidade todo condicional p (y | x): Em particular, no parâmetro multi-famílias exponenciais, como a família normal com média e variância tanto dependendo covariáveis??, a proporção de variância entre a variância total não leva em conta se a variância é adequadamente modelado. R2 apenas força a média condicional para estar perto das observações, mesmo que sob um modelo correto que não deve ser esperado para determinadas condições experimentais. A razão de variância, assim, parece ser uma medida questionável particularmente no exemplo clássico de um modelo de regressão gaussiana com variâncias heterogêneas onde causou muita discussão [3]. Esta experiência novamente motiva tendo em conta a amostragem todo condicional densidade e medindo a dependência entre X e Y de modo mais geral.

[3]Karl Pearson (1857 - 1936) foi um grande contribuidor para o desenvolvimento da Estatística como uma disciplina científica séria e independente. Ele foi o fundador do Departamento de Estatística Aplicada (Department of Applied Statistics) na University College London em 1911; foi o primeiro departamento universitário dedicado à estatística em todo o mundo.

4.2.1 Propriedades do coeficiente de determinação

Podemos afirmar que o R^2 é uma medida da proporção que a soma de quadrados dos desvios de cada Y_i em relação a $\overline{Y}$ pode ser explicada pelas covariáveis $X_1, \ldots, X_p$. Então, o R^2 é uma medida da bondade do ajuste do modelo (4.8), incluindo as covariáveis, em relação ao modelo (4.9), no qual nenhuma das covariáveis é considerada. Assim, para conjuntos de dados com variáveis dependentes, valores de R^2 perto de 1 refletem a capacidade de predição do modelo de regressão.

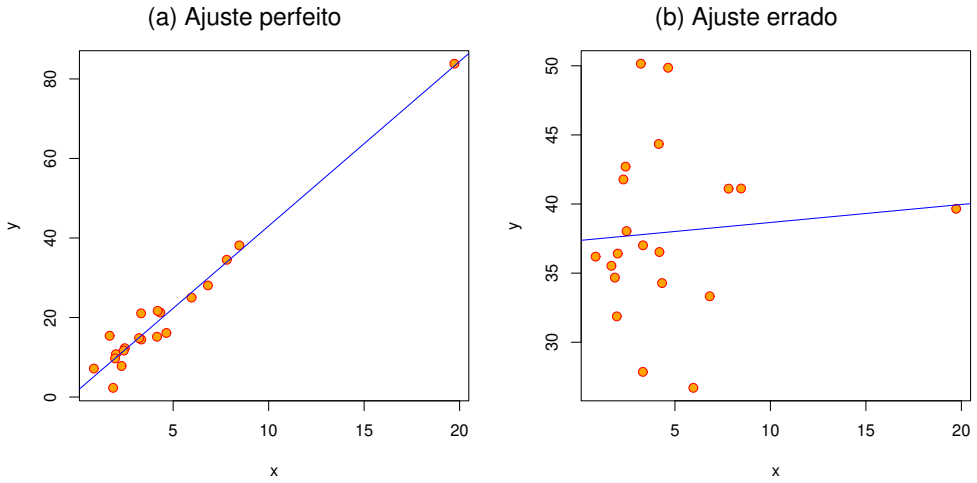

Figura 4.3: Diferentes modelos de regressão linear simples, (a) ajuste perfeito, (b) modelo errado, nesta situação não faz sentido um modelo de regressão.

Por exemplo, os dados mostrados na Figura 4.3 (a) foram gerados pelo modelo $Y_i = 2+4X_i+\epsilon_i$ satisfazendo que $\text{Var}(Y_i) = 9$ e cada $\epsilon_i \sim N(0,1)$, $i = 1, \ldots, 20$. As estimativas do modelo ajustado são quase perfeitas, o modelo estimado para os dados mostrados na Figura 4.3 (a) foi $\widehat{\mu} = 1.6358+4.1386X$ e o valor do coeficiente de determinação foi de $R^2 = 0.9655$. Na outra situação considerada, agora na Figura 4.3 (b), em casos onde não faz sentido um modelo de regressão o R^2 reflete isso, o valor deste coeficiente é $R^2 = 0.007684$, indicando independência entre a variável explicativa e a variável resposta. O modelo de regressão estimado para os dados mostrados na Figura **??** (b) foi $\widehat{\mu} = 37.3627 + 0.1302X$.

O coeficiente de determinação satisfaz algumas propriedades interessantes as quais demonstraremos nos resultados a continuação. A primeira destas propriedades nos permite melhor interpretar o R^2, esta propriedade nos diz que $0 \leq R^2 \leq 1$.

Teorema 4.10. *O coeficiente de determinação estimado R^2 é não negativo e limitado superiormente por 1.*

Demonstração. Primeiro provemos que $R^2 \geq 0$. A expressão alternativa do coeficiente de determinação amostral

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{\mu}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}, \tag{4.11}$$

permite-nos perceber que $R^2 \geq 0$. Tanto o numerador quanto o denominador são somas de quadrado, logo são quantidades positivas ou nulas. O numerador pode, eventualmente, ser zero e o denominador nunca será nulo numa amostra aleatória. Podemos perceber que $R^2 = 0$ somente quando $\text{E}(Y) = \overline{Y}$, como é o caso do modelo (4.9); nessa situação $\text{E}(Y_i) = \mu = \beta_0$, $\widehat{\mu} = \widehat{\beta}_0 = \overline{Y}$,

logo $\sum_{i=1}^{n}(Y_i - \widehat{\mu}_i)^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$ e

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 0\cdot$$

Para provar que o coeficiente R^2 é limitado superiormente por 1 observaremos primeiro que podemos escrever matricialmente a expressão em (4.11) como

$$R^2 = \frac{\widehat{\beta}^\top X^\top Y - n\overline{Y}^2}{Y^\top Y - n\overline{Y}^2}\cdot$$

Assim, para provar que $R^2 \leq 1$ devemos provar que $\widehat{\beta}^\top X^\top Y \leq Y^\top Y$. Utilizando as expressões dos estimadores do modelo linear, temos que

$$\widehat{\beta}^\top X^\top Y = Y^\top X(X^\top X)^{-1}X^\top Y = Y^\top HY,$$

onde $H = X(X^\top X)^{-1}X^\top$ é a matriz de predição, a qual sabemos é simétrica e idempotente. Observemos que

$$Y^\top Y - \widehat{\beta}^\top X^\top Y = Y^\top IY - Y^\top HY = Y^\top(I - H)Y,$$

o objetivo agora é demonstrar que $Y^\top(I-H)Y$ é uma forma quadrática positiva. Para isso devemos provar que $I - H$ é uma matriz definida não negativa. Em Rao (1973, pg.72) afirma-se que toda matriz simétrica e idempotente é definida não negativa. Portanto

$$Y^\top(I - H)Y \geq 0,$$

logo $Y^\top Y \geq \widehat{\beta}^\top X^\top Y$, concluindo-se que $R^2 \leq 1$. □

Interpretamos então que se $R^2 \approx 0$ o modelo não é apropriado para explicar a variável resposta através das variáveis explicativas selecionadas, significando que o R^2 é uma medida da utilidade dos outros termos além do β_0 no modelo. Um modelo cujo ajuste seja perfeito implicaria que $\widehat{\mu}_i = Y_i$, portanto $\sum_{i=1}^{n}(Y_i - \widehat{\mu}_i)^2 = 0$ e consequentemente $R^2 = 1$. Significando que quanto mais próximo de 1 estiver o valor do coeficiente de determinação melhor o ajuste aos dados do modelo proposto.

É importante notar que altos valores de R^2 não necessariamente implicam que o modelo de regressão está bem ajustado. Podemos visualizar isto através dos exemplos em Anscombe (1973) reproduzidos na Tabela 4.2.

Nesse trabalho o autor apresentou quatro conjuntos de dados com as mesmas médias, variâncias e correlação entre as variáveis resposta e explicativa. Algumas estatísticas descritivas importantes destes dados, como média, variancia, correlação entre x e y e outras, assumem os mesmos valores e portanto, as retas de regressão também coincidem. Outras estatísticas descritivas que não influenciam na estimação da reta de regressão não coincidem, como é o caso da mediana e os valores extremos.

O modelo de regressão estimado é $Y_1 = 3.0001 + 0.5001X_1$ o qual é comum a todos os outros modelos, ou seja, as estimativas dos modelos de regressão relacionando os pares de variáveis (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) e (X_4, Y_4) coincidem, sendo que o $R^2 = 0.6665$ e o desvio padrão dos resíduos é 1.2370. Também são comuns os resultados da análise de variância da regressão.

No entanto, observando as Figura 4.4 e Figura 4.5 fica claro que simplesmente com o valor do R^2 não seria possível perceber que nos conjuntos de dados No.2 e No.4 um modelo de regressão linear não faz sentido, no conjunto No.2 devemos transformar a variável explicativa para obtermos um

	No.1		No.2		No.3		No.4	
	X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
	10	8.04	10	9.14	10	7.46	8	6.58
	8	6.95	8	8.14	8	6.77	8	5.76
	13	7.58	13	8.74	13	12.74	8	7.71
	9	8.81	9	8.77	9	7.11	8	8.84
	11	8.33	11	9.26	11	7.81	8	8.47
	14	9.96	14	8.10	14	8.84	8	7.04
	6	7.24	6	6.13	6	6.08	8	5.25
	4	4.26	4	3.10	4	5.39	19	12.50
	12	10.84	12	9.13	12	8.15	8	5.56
	7	4.82	7	7.26	7	6.42	8	7.91
	5	5.68	5	4.74	5	5.73	8	6.89
média	9	7.5	9	7.5	9	7.5	9	7.5
variância	11	4.12	11	4.12	11	4.12	11	4.12
correlação	0.82		0.82		0.82		0.82	

Tabela 4.2: Quatro conjuntos de dados de Anscombe e medidas descritivas assim como a correlação entre X e Y em cada caso.

Coeficientes	Estimativa	Desvio padrão	t_{obs}	$P(>\|t_{obs}\|)$
Intercepto		1.1247	2.6670	0.0257
X_1		0.1179	4.2410	0.0021

Tabela 4.3: Tabela de análise de regressão.

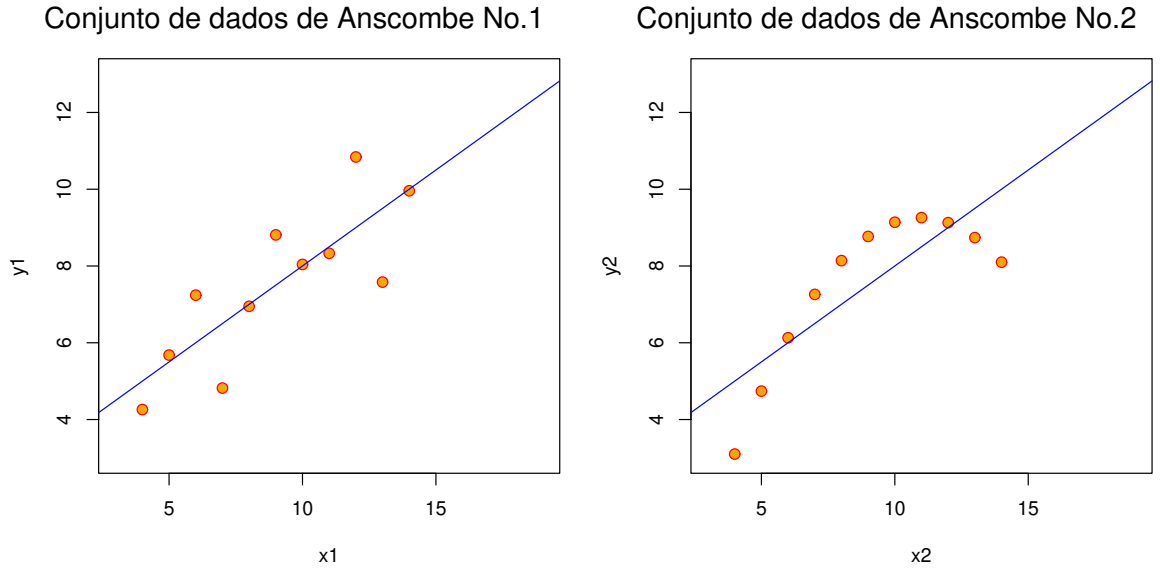

Figura 4.4: Gráficos de dispersão e reta de regressão estimada para os dois primeiros conjuntos de dados apresentados na Tabela **??**.

melhor ajustar e no conjunto de dados No.4 não existe relação alguma entre a variável explicativa e resposta.

No caso do conjunto de dados No.3 existe uma observação muito diferente das outras que atrapalha completamente e somente no conjunto de dados No.1 existe uma relação linear entre a covariável e a resposta. Neste exemplo, por ser uma regressão simples, é mais fácil perceber em

quais situações um modelo de regressão não é o mais adequado. Em situações mais complexas devemos realizar uma análise minuciosa dos resíduos para poder confiar na qualidade do R^2 obtido.

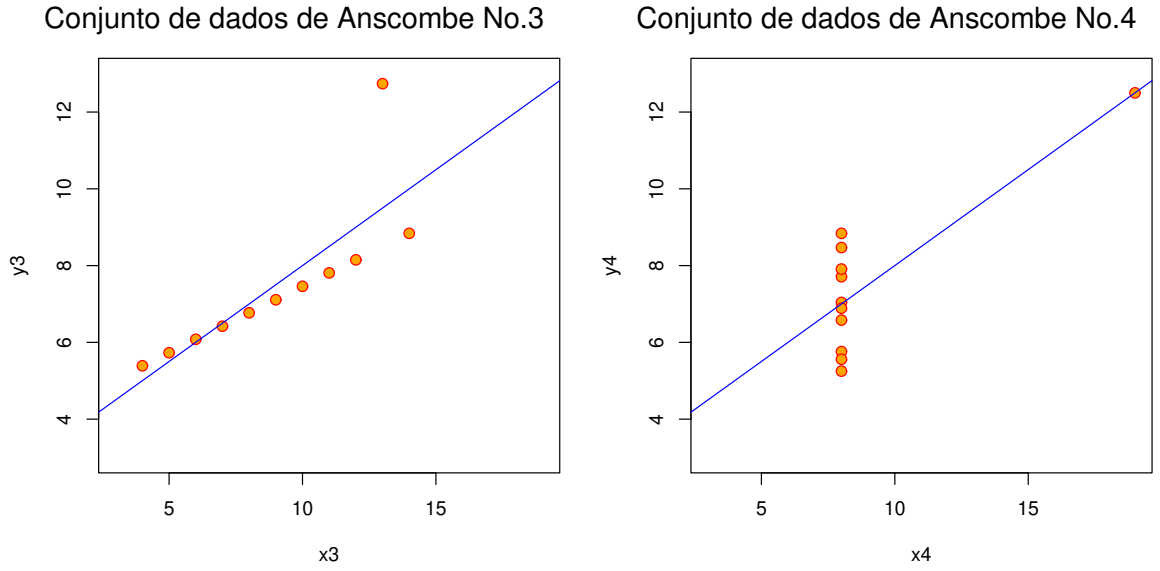

Figura 4.5: Gráficos de dispersão e reta de regressão estimada para os dois últimos conjuntos de dados apresentados na Tabela 4.2.

No entanto, observando as Figuras 4.4 (a) e (b) fica claro que simplesmente com o valor do R^2 não seria possível perceber que nos conjuntos de dados No.2 e No.4 um modelo de regressão linear não faz sentido, no conjunto No.2 devemos transformar a variável explicativa para obtermos um melhor ajustar e no conjunto de dados No.4 não existe relação alguma entre a variável explicativa e resposta. No caso do conjunto de dados No.3 existe uma observação muito diferente das outras que atrapalha completamente e somente no conjunto de dados No.1 existe uma relação linear entre a covariável e a resposta. Neste exemplo, por ser uma regressão simples, é mais fácil perceber em quais situações um modelo de regressão não é o mais adequado. Em situações mais complexas devemos realizar uma análise minuciosa dos resíduos para poder confiar na qualidade do R^2 obtido.

Uma outra propriedade do R^2 é ele ser crescente conforme aumenta o número de variáveis explicativas, mesmo que as novas variáveis acrescidas nada tenham a ver com a resposta. Adicionando variáveis regressoras podemos incrementar o valor de R^2 sem importar se as novas variáveis regressoras contribuem de fato ao modelo. Então é possível que alguns modelos tenham grandes valores de R^2 e sua qualidade seja ruim para estimação ou predição de novas observações.

Teorema 4.11. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X_p\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2\,\mathrm{I})$, X_p uma matriz $n \times p$ e coeficiente de determinação R_p^2. Seja agora $X_{p+1} = \begin{pmatrix} X_p & Z \end{pmatrix}$ uma nova matriz de constantes conhecidas, de dimensão $n \times p+1$, definindo um novo modelo de regressão $Y = X_{p+1}\beta + \epsilon$, com coeficiente de determinação R_{p+1}^2. Então*

$$R_{p+1}^2 > R_p^2. \tag{4.12}$$

Demonstração. Podemos escrever o R_{p+1}^2, para qualquer número de covariáveis, em função das somas de quadrados da Análise de Variância da regressão, como

$$R_{p+1}^2 = \frac{SQT - SQRes_{p+1}}{SQT},$$

e, desta expressão, vamos provar que aumentado o número de variáveis explicativas diminuímos a soma de quadrados do resíduo $SQRes_{p+1}$; isso logicamente implica no acréscimo do R^2_{p+1}. Segundo o Lema 3.8, se $X_{p+1} = \begin{pmatrix} X_p & Z \end{pmatrix}$ então $H_{p+1} = H_p + H_Z$, do qual temos que

$$SQRes_{p+1} = Y^\top (I - H_{p+1})Y = Y^\top (I - H_p)Y - Y^\top H_Z Y.$$

Obtemos que

$$R^2_{p+1} = \frac{SQT - Y^\top (I - H_p)Y + Y^\top H_Z Y}{SQT} = R^2_p + \frac{Y^\top H_Z Y}{SQT}.$$

Da expressão em (2.33) podemos perceber que $Y^\top H_Z Y > 0$, logo $R^2_{p+1} > R^2_p$. □

Devemos observar que, no Teorema 4.11, Z representa um vetor de dimensão $n \times 1$. Isto significa que acrescentamos somente uma nova variável explicativa ao modelo $Y = X_p\beta + \epsilon$. Evidente que podemos acrescer um número qualquer de variáveis explicativas.

Acontece que se aumentamos em um a dimensão deste espaço, pela inclusão de uma variável explicativa ao modelo, a estimação no espaço de dimensão p seria uma minimização restrita no espaço de dimensão $p+1$, e sabemos por cálculo que mínimos restritos são maiores do que mínimos absolutos. Logo o R^2 obtido no modelo de ordem p é ligeiramente menor do que o R^2 obtido no modelo de ordem $p + 1$.

Esta propriedade nos diz que adicionando infinitas covariáveis ao modelo, mesmo que não tenham nada a ver com o problema em questão, podemos artificialmente melhorar o coeficiente de determinação. Por esse motivo o R^2 serve para medir a qualidade do ajuste mas não é o mais apropriado para comparar modelos.

Para comparar modelos de regressão desenvolveu-se o coeficiente de determinação ajustado, do qual trata-se na sub-seção 4.2.2. Melhor ainda, no Capítulo 5 trata-se de indicadores e procedimentos mais adequados à escolha e modelos.

Uma questão mais difícil é encontrar a esperança de R^2 e intervalos de confiança para ϕ^2. Desde muito cedo foi percebido que a distribuição amostral de R^2 é complexa. Trabalhos clássicos de Fisher (1928), Ezekiel (1929), Gurland (1968) e outros foram dedicados à obtenção da função de densidade de R^2 e, a partir desta, desenvolver as propriedades procuradas. Aqui seguiremos os desenvolvimentos em Muirhead (1982) e Anderson (1984) para encontrarmos tanto a função de densidade quanto a esperança do coeficiente de determinação.

Teorema 4.12. *O coeficiente de determinação estimado R^2 é uma variável aleatória com função de densidade*

$$\begin{aligned} f(R^2) &= \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}p]\Gamma[\frac{1}{2}(n-p-1)]}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}(1-\phi^2)^{\frac{n-1}{2}} \times \\ &\quad \times {}_2F_1\left[\tfrac{1}{2}(n-1),\, \tfrac{1}{2}(n-1);\, \tfrac{1}{2}p;\, \phi^2 R^2\right], \end{aligned}$$

para $0 < R^2 < 1$.

Neste Teorema ${}_2F_1$ representa a função hipergeométrica definida como

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)} \frac{z^n}{n!}.$$

Demonstração. Seja (Y, X) um vetor aleatório com distribuição $N_{p+1}(\mu, \Sigma)$, sendo Σ a matriz de variâncias e covariâncias. Seja $\widehat{\Sigma}$ o estimador de máxima verossimilhança de Σ com base nas n

observações. Por sua vez, seja $\widehat{A} = n\widehat{\Sigma}$, então $\widehat{A} \sim W_{p+1}(n-1, \Sigma)$[4], distribuição Wishart. Se particionamos a matriz $\widehat{A}$ como

$$\widehat{A} = \begin{pmatrix} \widehat{a}_{11} & \widehat{\mathbf{a}}_{12}^\top \\ \widehat{\mathbf{a}}_{12} & \widehat{A}_{22} \end{pmatrix},$$

o coeficiente de determinação amostral pode ser escrito como

$$R^2 = \frac{\widehat{\mathbf{a}}_{12}^\top \widehat{A}_{22}^{-1} \widehat{\mathbf{a}}_{12}}{\widehat{a}_{11}}, \tag{4.13}$$

portanto

$$\frac{R^2}{1-R^2} = \frac{\widehat{\mathbf{a}}_{12}^\top \widehat{A}_{22}^{-1} \widehat{\mathbf{a}}_{12}}{\widehat{a}_{11.2}}, \tag{4.14}$$

onde $\widehat{a}_{11.2} = \widehat{a}_{11} - \widehat{\mathbf{a}}_{12}^\top \widehat{A}_{22}^{-1} \widehat{\mathbf{a}}_{12}$. Pelo Teorema 2.17 sabemos que o numerador e denominador à direita em (4.14) são independentes. Assim

$$\widehat{a}_{11.2}/\sigma_{11.2} \sim \chi^2_{n-m+1},$$

sendo $\sigma_{11.2} = \sigma_{11} - \sigma_{12}^\top \Sigma_{22}^{-1} \sigma_{12}$ e

$$\widehat{\mathbf{a}}_{12} | \widehat{A}_{22} \sim N(\widehat{A}_{22} \Sigma_{22}^{-1} \sigma_{12}, \sigma_{11.2} \widehat{A}_{22}) \cdot$$

Portanto, □

No artigo de Fisher (1928) encontramos uma das primeiras demonstrações deste resultado, nesse artigo se obteve que

$$\begin{aligned} f(R^2) &= \frac{\Gamma[\frac{1}{2}n]}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(n-p-1)]} (R^2)^{\frac{p-2}{2}} (1-R^2)^{\frac{n-p-3}{2}} (1-\phi^2)^{\frac{n-1}{2}} \times \\ &\quad \times \frac{1}{\pi} \int_{\psi=0}^{\pi} \int_{\beta=-\infty}^{\infty} \frac{\operatorname{sen}^{p-2}(\psi)}{\{\cosh(\beta) - \phi^2 R \cos(\psi)\}^n} \, \mathrm{d}\beta \, \mathrm{d}\psi \cdot \end{aligned}$$

Este resultado Fisher o encontrou através de raciocínio geométrico. Pela expansão de $\cos(\psi)$ em série de potências e integrando termo a termo, finalmente, se obtém a expressão no teorema. Maiores detalhes desta demonstração podem ser encontrados no referido artigo de Fisher, no artigo de Gurland (1968) e no livro de Muirhead (1982), dentre outros.

Observemos que se $I_x(\alpha, \beta)$ denota a função beta incompleta

$$I_x(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} \, \mathrm{d}t, \tag{4.15}$$

para $0 < x < 1$ e $\alpha, \beta > 0$, a função de distribuição F-Fisher $F(n_1, n_2)$ e a função beta incompleta estão relacionadas pela identidade

$$P\left[F(n_1, n_2) \leq x\right] = I_z(\frac{1}{2}n_1, \frac{1}{2}n),$$

[4]A distribuição Wishart é fundamental na estatística multivariada e é definida como

$$f(\widehat{A}) = \frac{|\widehat{A}|^{\frac{1}{2}(n-p-1)} \exp\left[-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\widehat{A})\right]}{2^{\frac{1}{2}np} \pi^{\frac{1}{4}p(p-1)} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]} \cdot$$

onde $z = n_1x/(n_2 + n_1x)$.

Uma questão mais difícil é encontrar a esperança de R^2 e encontrar intervalos de confiança para ϕ^2. Desde muito cedo foi percebido que $\mathrm{E}(R^2) \neq \phi^2$, por exemplo, nos trabalhos de Fisher (1928), Ezekiel (1929), Gurland (1968) e outros foi percebido que a distribuição amostral de R^2 é complexa.

Teorema 4.13. *A função de distribuição de R^2 pode ser expressa como*

$$P(R^2 \leq x) = \sum_{k=0}^{\infty} c_k P\left(F_{m-1+2k;n-m+1} \leq \frac{n-m+1}{m-1+2k}\frac{x}{1-x}\right),$$

onde c_k é da forma

$$c_k = (-1)^k \binom{-\frac{1}{2}n}{k} (1-\phi^2)^{n/2}(\phi^2)^k.$$

Demonstração. Encontremos primeiro a função de densidade de R^2. Então a função de densidade é

$$f(R^2) = \frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2}(p-1))\Gamma(\frac{1}{2}(n-p-2))}(R^2)^{(p-3)/2}(1-R^2)^{(n-p-3)/2}(1-\phi^2)^{\frac{n-1}{2}},$$

para $0 < R^2 < 1$. Ver em http://www.jstor.org/stable/2984506 o artigo de Gurland (1968)

Ver em http://www.doc88.com/p-5486116501056.html o livro de Muirhead (1982), Capítulo 5, Seção 5.2.4, pg.198 Teorema 5.2.4

Em Olkin & Pratt (1958) os autores provaram que, para $h(R^2)$ ser uma função não viciada de $\mathrm{E}(R^2)$, esta deve ser da forma

$$h(R^2) = 1 - \frac{n-3}{n-1-p}(1-R^2)F[1,1;(n-p+1)/2;1-R^2],$$

onde F é a função hipergeométrica

$$F(\alpha,\beta;\gamma;x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)\Gamma(\beta+k)\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma+k)}\frac{x^k}{k!}.$$

□

Observemos que $h(1) = 1$, no entanto, $h(0) = -p/(n-p-3)$. Ainda percebemos que a função $h(R^2)$ é estritamente crescente em R^2 e difere de R^2 somente por um termo de ordem $1/(n-1)$.

Teorema 4.14. *Os momentos do coeficiente de determinação são dados por*

$$\mathrm{E}[(1-R^2)^k] = \frac{[\frac{1}{2}(n-m+1)]_k}{(\frac{1}{2}n)_k}(1-\phi^2)^k\ {}_2F_1(k,k;\tfrac{1}{2}n+k;\phi^2).$$

Demonstração. Os momentos de R^2 serão obtidos utilizando a representação da distribuição do coeficiente de determinação R^2 obtida no Teorema 4.13. Nesse teorema foi demonstrado que a distribuição de R^2 é uma mistura de distribuições beta. Utilizando o fato de que o k-ésimo momento da distribuição beta com parâmetros α e β é

$$\frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$$

encontramos que

$$\mathrm{E}[(1-R^2)^k] = \sum_{h=0}^{\infty} c_h \frac{\Gamma[\frac{1}{2}(n-m+1)+k]}{\Gamma[\frac{1}{2}(n-m+1)]} \frac{\Gamma(\frac{1}{2}n+h)}{\Gamma(\frac{1}{2}n+k+h)},$$

onde c_h é dado por

$$c_h = (-1)^h \binom{-\frac{1}{2}n}{h} (1-\phi^2)^{\frac{n}{2}} (\phi^2)^h = \frac{(\frac{1}{2}n)_h}{h!} (1-\phi^2)^{\frac{n}{2}} (\phi^2)^h.$$

Portanto podemos escrever

$$\begin{aligned} \mathrm{E}[(1-R^2)^k] &= \frac{[\frac{1}{2}(n-m+1)]_k}{(\frac{1}{2}n)_k} (1-\phi^2)^{\frac{n}{2}} \sum_{h=0}^{\infty} \frac{(\frac{1}{2}n)_h (\frac{1}{2}n)_h}{(\frac{1}{2}n+k)_h} \frac{(\phi^2)^h}{h!} \\ &= \frac{[\frac{1}{2}(n-m+1)]_k}{(\frac{1}{2}n)_k} (1-\phi^2)^{\frac{n}{2}} \, {}_2F_1(\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}n+k; \phi^2). \end{aligned}$$

Utilizando a relação de Euler[5] chegamos a que

$$\mathrm{E}[(1-R^2)^k] = \frac{[\frac{1}{2}(n-m+1)]_k}{(\frac{1}{2}n)_k} (1-\phi^2)^k \, {}_2F_1(k, k; \tfrac{1}{2}n+k; \phi^2).$$

□

Uma consequência deste teorema é podermos encontrar a média e variância do R^2. Assim podemos perceber que este estimador, embora muito útil, é viciado e isso provocou a construção dos chamados coeficientes de determinação ajustados ou R^2_{adj}.

Teorema 4.15. *O coeficiente de determinação estimado R^2 é viciado para ϕ^2.*

Demonstração. Pelo Teorema 4.14, temos que a média de R^2 é

$$\begin{aligned} \mathrm{E}(R^2) &= 1 - \left(\frac{n-m+1}{n}\right) (1-\phi^2) \, {}_2F_1\left(1, 2; \tfrac{1}{2}n+1; \phi^2\right) \\ &= \phi^2 + \frac{m-1}{n}(1-\phi^2) + \frac{2}{n+2}\phi^2(1-\phi^2) + O(n^{-2}), \end{aligned} \tag{4.16}$$

o qual implica que o R^2 é viciado para ϕ^2. □

Como consequência deste teorema também percebemos que

$$\begin{aligned} \mathrm{Var}(R^2) &= \mathrm{E}(R^4) - \mathrm{E}^2(R^2) \\ &= \mathrm{E}[(1-R^2)^2] - \mathrm{E}^2(1-R^2) \\ &= \frac{[\frac{1}{2}(n-p)]_2}{(\frac{1}{2}n)_2} (1-\phi^2)^2 \, {}_2F_1\left(2, 2; \tfrac{1}{2}n+2; \phi^2\right) \\ &\quad - \left[\left(\frac{n-p}{n}\right) (1-\phi^2)^2 \, {}_2F_1\left(1, 1; \tfrac{1}{2}n+1; \phi^2\right)\right]^2 \end{aligned}$$

[5] A relação de Euler estabelece que

$${}_2F_1(a, b; c; z) = (1-z)^{c-a-b} \, {}_2F_1(c-a, c-b; c; z).$$

Consultar Erdélyi (1953) e Rainville (1960)

e utilizando propriedades da função hipergeométrica obtemos que

$$\mathrm{Var}(R^2) = \frac{n-p}{n^2(n+2)}(1-\phi^2)^2\left\{2p + 4\phi^2\frac{4p+n(n-p)}{n+4} + O(n^{-2})\right\}. \tag{4.17}$$

Observe que as ordens de magnitude de $\mathrm{Var}(R^2)$ dependem se $\phi = 0$ ou se $\phi \neq 0$. Assim, para $\phi \neq 0$ temos que

$$\begin{aligned}\mathrm{Var}(R^2) &= \frac{4\phi^2(1-\phi^2)^2(n-p)^2}{n(n+2)(n+4)} + O(n^{-2}) \\ &= \frac{4\phi^2(1-\phi^2)^2}{n} + O(n^{-2}).\end{aligned}$$

Enquanto para $\phi = 0$ temos que

$$\mathrm{Var}(R^2) = \frac{2p(n-p)}{(n+2)n^2},$$

a qual é uma expressão exata, válida somente no caso nulo.

Uma questão importante tem sido encontrar um estimador não viciado para ϕ^2, mas esta questão não tem resposta simples. Algumas respostas as apresentamos na Seção 4.2.2 e, dentre estas, encontra-se o conhecido como coeficiente de determinação ajustado, o qual não é um e sim diversos. Tecnicamente, todos os resultados que apresentaremos na próxima seção podem ser considerados como coeficientes de determinação ajustados.

4.2.2 Coeficientes de determinação ajustados

Há muito se sabe que, mesmo para dados normais, a estimativa do coeficiente de correlação tem um pequeno viés (Fisher, 1915). Esse viés nos dados normais é conservador, levando à subestimação do verdadeiro coeficiente de correlação absoluta. É importante ressaltar que esse viés de dados normal é uma preocupação apenas para amostras pequenas; o viés absoluto se torna desprezível (menor que 0,01) para um tamanho de amostra maior que 20. Para corrigir esse viés, Fisher (1915) e outros (por exemplo, Olkin e Pratt, 1958) recomendaram vários ajustes aproximados. O problema, porém, é que cada um desses ajustes assume uma normalidade bivariada e poderia causar mais danos do que benefícios quando a anormalidade está presente.

Em uma revisão da literatura, foram identificados diversas propostas para corrigir o vício de R^2. Apresentamos nesta seção diversas expressões analíticas para corrigi-lo, a final, a expressão da esperança de R^2 é aproximada e diversos trabalhos tiveram como objetivo investigar o comportamento destas propostas modificando o tamanho da amostra e o número de parâmetros da regressão.

Denotaremos por R^2_{adj} ao coeficiente de determinação corrigido, chamado de coeficiente de determinação ajustado, de maneira que $\mathrm{E}(R^2_{adj}) \approx \phi^2 + O(n^{-1})$. Isto significa que a média do coeficiente de determinação ajustado é aproximadamente igual a ϕ^2 mais termos que decrescem conforme o tamanho da amostra aumenta.

A primeira proposta conhecida na literatura foi apresentada por Ezekiel (1929). Então, definimos o coeficiente de determinação ajustado

$$R^2_{adj} = 1 - \frac{n}{n-p}(1-R^2). \tag{4.18}$$

No artigo mencionado o autor menciona que BB Smith desenvolveu recentemente uma fórmula de correção a aplicar-se o coeficiente de correlação múltipla nos casos em que o número de variáveis é significativo em comparação com o número de observações, de modo a reduzir o valor da correlação

observada para a correlação, que provavelmente se obtém no o universo do qual a amostra foi colhida. RA Fisher chega exatamente a mesma correção por uma fórmula que mostra o quanto o erro padrão de estimativa para os casos incluídos na amostra tem de ser aumentada para indicar o erro padrão que é susceptível de ser obtido quando a mesma equação de regressão é usado para estimar valores da variável dependente para novas observações extraídas do mesmo universo.

Definição 4.3. *Seja Y um vetor aleatório satisfazendo o modelo de regressão $Y = X\beta + \epsilon$, onde $\epsilon \sim N_n(0, \sigma^2\,\mathrm{I})$ e coeficiente de determinação R^2. Definimos o coeficiente de determinação ajustado como*

$$R^2_{adj} = 1 - \frac{n}{n-p-1}(1-R^2)\cdot$$

shadecolor

Esta fórmula está a ser implementado por pacotes estatísticos populares para calcular o R_2 ajustado em procedimentos de regressão múltipla. A expressão

Uma outra transformação foi proposta por Wherry (1931).

No artigo de Olkin & Pratt (1958), os autores propuseram ..

A primeira proposta conhecida na literatura foi apresentada por Ezekiel (1929), nesse artigo propõe-se que

$$R^2_{adj} = 1 - \frac{(n-3)(1-R^2)}{n-p-1}\left[1 + \frac{2(1-R^2)}{n-p+1}\right]\cdot \tag{4.19}$$

como coeficiente ajustado These three formulas are basically the same equation in different algebraic forms, and they are all approximations of Olkin and Pratt's (1958) unbiased estimate of the squared multiple correlation .

No artigo de

De acordo com Anderson- Sprecher (1994), "o coeficiente de determinação múltipla , R2, é uma medida muitos estatísticos ama odiar . Esta hostilidade existe principalmente porque o uso generalizado de R2 inevitavelmente conduz a pelo menos mau uso ocasional "(Anderson , Sprecher , 1994, p . 113) . Enquanto a controvérsia sobre R2 tem a sua origem na literatura estatísticas (Kavalseth , 1985; Helland , 1987; Willett e Singer , 1988; Lavergne , 1996; Korn e Simon , 1991; Scott and Wild , 1991; McGuirk e Driscoll , 1995), o debate R2 é importante para todas as áreas do conhecimento que utilizam modelos de regressão linear . McGregor (1993) argumenta que "não é de admirar que o modelo de regressão atingiu o seu status preferencial nas ciências sociais "(McGregor , 1993 p. 801/802) 4 . Além disso, ele afirma que "o modelo de regressão domina trabalho empírico em ciência política . Evidência áspera isso pode ser encontrado em uma revisão da American Political Science Review : quase todos os artigos que exibia resultados em forma de tabela utilizada alguma forma de análise de regressão "(. McGregor , 1993 p 801) 5 .

De fato, a atratividade do modelo de regressão pode ser parcialmente explicado pela sua capacidade de resumir a relação entre variáveis diferentes em uma abordagem sistemática e parcimonioso. Portanto, uma vez que o uso de modelos de regressão têm vindo a aumentar nas ciências sociais em geral e ciência política em particular, é importante entender o papel controverso de R^2 e os estudiosos significado substantivo pode tirar dele .

Pelo fato de esta novo estimador considerar o tamanho da amostra n e a quantidade de variáveis explicativas q recomenda-se utilizar o R^2_{adj} para comparar modelos e, desta forma, escolher o modelo mais adequado. Vejamos isto através do seguinte exemplo.

Exemplo 4.3. *Seja .....*

```
R2  R2.Smith R2.Wherry2 R2.Olkin.Pratt  R2.Pratt R2.Claudy
```