Decomposing variance

Pearson correlation

The population Pearson correlation coefficient of two jointly distributed random variables X and Y is

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

It is estimated by

$$\hat{\rho}_{XY} = \frac{\widehat{\mathrm{cov}}(X,Y)}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \cdot \sum_i (Y_i - \bar{Y})^2}} = \frac{(X - \bar{X})'(Y - \bar{Y})}{\|X - \bar{X}\| \cdot \|Y - \bar{Y}\|}.$$

By the Cauchy-Schwartz inequality,

$$\begin{array}{rcccl} -1 & \leq & \rho & \leq & 1 \\ -1 & \leq & \widehat{\rho} & \leq & 1. \end{array}$$

The sample correlation coefficient is slightly biased for the population correlation coefficient, but the bias is small unless $|\rho|$ is close to 1.

For the simple linear regression model

$$Y = \alpha + \beta X + \epsilon,$$

if we view X as a random variable that is uncorrelated with ϵ, then

$$\text{cov}(X, Y) = \beta\sigma_X^2$$

and the correlation is

$$\rho_{XY} \equiv \text{cor}(X, Y) = \frac{\beta}{\sqrt{\beta^2 + \sigma^2/\sigma_X^2}}.$$

The sample correlation coefficient is related to the least squares slope estimate:

$$\hat{\beta} = \frac{\widehat{\text{cov}}(X, Y)}{\hat{\sigma}_X^2} = \hat{\rho}_{XY}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}.$$

Orthogonality between fitted values and residuals

Recall that the fitted values are

$$\hat{Y} = X\hat{\beta} = PY$$

and the residuals are

$$R = Y - \hat{Y} = (I - P)Y.$$

Since $P(I - P) = 0$ it follows that $\hat{Y}'R = 0$.

Since $\bar{R} = 0$, it is equivalent to state that the sample correlation between R and $\hat{Y}$ is zero.

Coefficient of determination

A descriptive summary of the explanatory power of X for Y is given by the "coefficient of determination," also known as the "proportion of explained variance," or "multiple R^2." This is the quantity

$$R^2 \equiv 1 - \frac{\|Y - \hat{Y}\|^2}{\|Y - \bar{Y}\|^2} = \frac{\|\hat{Y} - \bar{Y}\|^2}{\|Y - \bar{Y}\|^2} = \frac{\widehat{\mathrm{var}}(\hat{Y})}{\widehat{\mathrm{var}}(Y)}.$$

The equivalence between the two expressions follows from the identity

$$\begin{aligned} \|Y - \bar{Y}\|^2 &= \|Y - \hat{Y} + \hat{Y} - \bar{Y}\|^2 \\ &= \|Y - \hat{Y}\|^2 + \|\hat{Y} - \bar{Y}\|^2 + 2(Y - \hat{Y})'(\hat{Y} - \bar{Y}) \\ &= \|Y - \hat{Y}\|^2 + \|\hat{Y} - \bar{Y}\|^2, \end{aligned}$$

It should be clear that $R^2 = 0$ iff $\hat{Y} = \bar{Y}$ and $R^2 = 1$ iff $\hat{Y} = Y$.

The coefficient of determination is equal to

$$\widehat{\text{cor}}(\hat{Y}, Y)^2.$$

To see this, note that

$$\begin{aligned}
\widehat{\text{cor}}(\hat{Y}, Y) &= \frac{(\hat{Y} - \bar{Y})'(Y - \bar{Y})}{\|\hat{Y} - \bar{Y}\| \cdot \|Y - \bar{Y}\|} \\
&= \frac{(\hat{Y} - \bar{Y})'(Y - \hat{Y} + \hat{Y} - \bar{Y})}{\|\hat{Y} - \bar{Y}\| \cdot \|Y - \bar{Y}\|} \\
&= \frac{(\hat{Y} - \bar{Y})'(Y - \hat{Y}) + (\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})}{\|\hat{Y} - \bar{Y}\| \cdot \|Y - \bar{Y}\|} \\
&= \frac{\|\hat{Y} - \bar{Y}\|}{\|Y - \bar{Y}\|}.
\end{aligned}$$

Simple linear regression

In this special case,

$$\|\hat{Y} - \bar{Y}\|^2 = \|\hat{\beta}(X - \bar{X})\|^2 = (n-1)\widehat{\text{cov}}(Y,X)^2/\hat{\sigma}_X^2$$

and

$$\|Y - \bar{Y}\|^2 = (n-1)\hat{\sigma}_Y^2,$$

so the coefficient of determination is

$$\frac{\widehat{\text{cov}}(Y,X)^2}{\hat{\sigma}_X^2\hat{\sigma}_Y^2} = \hat{\rho}_{XY}^2.$$

For simple linear regression, $\widehat{\text{cor}}(Y,X)^2 = \widehat{\text{cor}}(Y,\hat{Y})^2$.

Relationship to the F statistic

The F-statistic for the null hypothesis

$$\beta_1 = \ldots = \beta_p = 0$$

is

$$\frac{\|\widehat{Y} - \bar{Y}\|^2}{\|Y - \widehat{Y}\|^2} \cdot \frac{n-p-1}{p} = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p},$$

which is an increasing function of R^2.

Adjusted R^2

The sample R^2 is an estimate of the population R^2:

$$1 - \frac{\text{var}(Y|X)}{\text{var}(Y)}.$$

Since it is a ratio, the plug-in estimate R^2 is biased, although the bias is not large unless the sample size is small. The adjusted R^2 is an approximately unbiased estimate of the population R^2:

$$1 - (1 - R^2)\frac{n-1}{n-p-1}.$$

The adjusted R^2 is always less than the unadjusted R^2. The adjusted R^2 is always less than or equal to one, but can be negative.

Improvement in R^2 due to one covariate

Let $X_k^\perp$ be the residual of X_k after regressing it against all other covariates (including the intercept).

In other words, let P_{-k} be the projection onto span($\{X_j, j \neq k\}$), so

$$X_k^\perp = (I - P_{-k})X_k.$$

Let

$$\hat{Y}_{-k} = P_{-k}Y$$

be the fitted values in the model omitting covariate k.

Let R^2 denote the multiple R^2 for the full model, and let R^2_{-k} be the multiple R^2 for the regression omitting covariate X_k. The value of

$$R^2 - R^2_{-k}$$

tells us how much "unique information" is in X_k that is not captured by the other covariates.

We would like to have a simple expression for $R^2 - R^2_{-k}$.

Since the fitted values and residuals are independent,

$$\begin{aligned} \|Y\|^2 &= \|Y - \widehat{Y}_{-k} + \widehat{Y}_{-k}\|^2 \\ &= \|Y - \widehat{Y}_{-k}\|^2 + \|\widehat{Y}_{-k}\|^2 \end{aligned}$$

so

$$\|Y - \widehat{Y}_{-k}\|^2 = \|Y\|^2 - \|\widehat{Y}_{-k}\|^2.$$

Since $X_k^\perp$ is orthogonal to the other covariates,

$$\widehat{Y} = \widehat{Y}_{-k} + \frac{\langle Y, X_k^\perp \rangle}{\langle X_k^\perp, X_k^\perp \rangle} X_k^\perp,$$

and

$$\|\widehat{Y}\|^2 = \|\widehat{Y}_{-k}\|^2 + \langle Y, X_k^\perp \rangle^2 / \|X_k^\perp\|^2.$$

Now if we look at the overall R^2, we get

$$\begin{aligned}
R^2 &= 1 - \frac{\|Y - \widehat{Y}\|^2}{\|Y - \bar{Y}\|^2} \\
&= 1 - \frac{\|Y\|^2 - \|\widehat{Y}\|^2}{\|Y - \bar{Y}\|^2} \\
&= 1 - \frac{\|Y\|^2 - \|\widehat{Y}_{-k}\|^2 - \langle Y, X_k^{\perp}\rangle^2 / \|X_k^{\perp}\|^2}{\|Y - \bar{Y}\|^2} \\
&= 1 - \frac{\|Y - \widehat{Y}_{-k}\|^2}{\|Y - \bar{Y}\|^2} + \frac{\langle Y, X_k^{\perp}\rangle^2 / \|X_k^{\perp}\|^2}{\|Y - \bar{Y}\|^2} \\
&= R^2_{-k} + \frac{\langle Y, X_k^{\perp}\rangle^2 / \|X_k^{\perp}\|^2}{\|Y - \bar{Y}\|^2}.
\end{aligned}$$

Partial and semi-partial R^2

The "semi-partial R^2" is

$$R^2 - R^2_{-k} = \frac{\langle Y, X_k^\perp \rangle^2 / \|X_k^\perp\|^2}{\|Y - \bar{Y}\|^2} = \widehat{\text{cor}}(Y, X_k^\perp)^2 = \widehat{\text{cor}}(Y, \hat{Y}_k)^2,$$

where $\widehat{Y}_k$ is the fitted value for regressing Y on $X_k^\perp$.

The "partial R^2" is

$$\frac{R^2 - R^2_{-k}}{1 - R^2_{-k}} = \frac{\langle Y, X_k^\perp \rangle^2 / \|X_k^\perp\|^2}{\|Y - \widehat{Y}_{-k}\|^2}.$$

The semi-partial R^2 for covariate k is the improvement in R^2 resulting from including covariate k.

The partial R^2 for covariate k is the fraction of the maximum possible improvement in R^2 that is contributed by covariate k.

Partial R^2 as a squared correlation coefficient

Let $\widehat{Y}_{-k}$ be the fitted values for regressing Y on all covariates except X_k. Since $\widehat{Y}'_{-k}X_k^{\perp} = 0$,

$$\frac{\langle Y - \widehat{Y}_{-k}, X_k^{\perp}\rangle^2}{\|Y - \widehat{Y}_{-k}\|^2 \cdot \|X_k^{\perp}\|^2} = \frac{\langle Y, X_k^{\perp}\rangle^2}{\|Y - \widehat{Y}_{-k}\|^2 \cdot \|X_k^{\perp}\|^2}.$$

The expression on the left is the usual R^2 that would be obtained when regressing $Y - \widehat{Y}_{-k}$ on $X_k^{\perp}$.

The semi-partial R^2 is the same as the usual R^2 for Y regressed on $(I - P_{-k})X_k$.

The partial R^2 is the same as the usual R^2 for $(I - P_{-k})Y$ regressed on $(I - P_{-k})X_k$.

Decomposition of R^2

In a multiple regression, how much of the Y variance is explained by a particular covariate?

Orthogonal case: If the design matrix is orthogonal, the projection P onto col(X) can be decomposed as

$$P = \sum_{j=0}^{p} P_j = \frac{11'}{n} + \sum_{j=1}^{p} X_j X_j',$$

where X_j is the j^{th} column of the design matrix.

$$P_j = X_j X_j'$$

is the projection onto span(X_j) (and P_0 is the projection onto the span of the vector of 1's). Furthermore, by orthogonality, $P_j P_k = 0$ unless $j = k$. Since

$$\widehat{Y} - \bar{Y} = \sum_{j=1}^{p} P_j Y,$$

by orthogonality

$$\|\widehat{Y} - \bar{Y}\|^2 = \sum_{j=1}^{p} \|P_j Y\|^2.$$

The R^2 for simple linear regression of Y on X_j is

$$R_j^2 \equiv \|\widehat{Y} - \bar{Y}\|^2 / \|Y - \bar{Y}\|^2 = \|P_j Y\|^2 / \|Y - \bar{Y}\|^2,$$

so we see that for orthogonal design matrices,

$$R^2 = \sum_{j=1}^{p} R_j^2.$$

That is, the overall coefficient of determination is the sum of univariate coefficients of determination for all the explanatory variables.

Non-orthogonal case: If X is not orthogonal, the overall R^2 will not generally be the sum of single covariate R^2's.

It's not surprising that $\sum_j R_j^2$ can be bigger than R^2. For example, suppose that

$$Y = X_1 + \epsilon$$

is the data generating model, and X_2 is highly correlated with X_1.

For large sample sizes, the multiple R^2 of Y regressed on X_1 and X_2 will be approximately $1 - \sigma^2/\text{var}(Y)$.

The R^2's of Y regressed on either X_1 or X_2 separately will also be approximately $1 - \sigma^2/\text{var}(Y)$.

Thus $R_1^2 + R_2^2 \approx 2R^2$.

Enhancement

Surprisingly, $\sum_j R_j^2$ can be less than R^2, a situation called *enhancement*. As an example of this phenomenon, suppose the data generating model is

$$Y = Z + \epsilon,$$

but we don't observe Z (for simplicity assume $EZ = 0$). Instead, we observe a value X_2 with mean zero that is independent of Z and ϵ, and a value X_1 which satisfies

$$X_1 = Z + X_2.$$

Since X_2 is independent of Z and ϵ, it is also independent of Y. Therefore $R_2^2 \approx 0$ for large n.

The multiple R^2 of Y on X_1 and X_2 is approximately $\sigma_Z^2/(\sigma_Z^2 + \sigma^2)$ for large n, since the fitted values will converge to $\hat{Y} = X_1 - X_2 = Z$.

Enhancement (continued)

To calculate R_1^2, first note that for the regression of Y on X_1,

$$\widehat{\beta} \to \frac{\mathrm{cov}(Y, X_1)}{\mathrm{var}(X_1)} = \frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_{X_2}^2}$$

and

$$\widehat{\alpha} \to 0.$$

Enhancement (continued)

Therefore for large n,

$$
\begin{aligned}
n^{-1}\|Y-\widehat{Y}\|^2 &\approx n^{-1}\|Z+\epsilon-\sigma_Z^2X_1/(\sigma_Z^2+\sigma_{X_2}^2)\|^2 \\
&= n^{-1}\|\sigma_{X_2}^2Z/(\sigma_Z^2+\sigma_{X_2}^2)+\epsilon-\sigma_Z^2X_2/(\sigma_Z^2+\sigma_{X_2}^2)\|^2 \\
&= \sigma_{X_2}^2\sigma_Z^2/(\sigma_Z^2+\sigma_{X_2}^2)+\sigma^2.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
R_1^2 &\approx 1-\frac{n^{-1}\|Y-\widehat{Y}\|^2}{n^{-1}\|Y-\bar{Y}\|^2} \\
&= \frac{\sigma_Z^2}{(\sigma_Z^2+\sigma^2)(1+\sigma_{X_2}^2/\sigma_Z^2)}
\end{aligned}
$$

Enhancement (continued)

so

$$R_1^2/R^2 \approx 1/(1+\sigma_{X_2}^2/\sigma_Z^2),$$

which is strictly less than one if $\sigma_{X_2}^2 > 0$.

In the example, $R^2 > R_1^2 + R_2^2$.

The reason is that while X_2 contains no directly useful information about Y (hence $R_2^2 = 0$), it can remove the "measurement error" in X_1, making it a better predictor of Z.

Partial R^2 example

Suppose the design matrix satisfies

$$X'X/n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & r \\ 0 & r & 1 \end{pmatrix}$$

and the data generating model is

$$Y = X_1 + X_2 + \epsilon$$

with $\text{var}\, \epsilon = 1$.

Partial R^2 example (continued)

We will calculate the partial R^2 for X_1, using the fact that the partial R^2 is the regular R^2 for regressing

$$(I - P_{-1})Y$$

on

$$(I - P_{-1})X_1$$

where P_{-1} is the projection onto span$(\{1, X_2\})$.

Since this is a simple linear regression, the partial R^2 can be expressed

$$\text{cor}((I - P_{-1})Y, (I - P_{-1})X_1)^2.$$

Partial R^2 example (continued)

The numerator of the partial R^2 is the square of

$$\begin{aligned}\text{cov}((I - P_{-1})Y, (I - P_{-1})X_1) &= Y'(I - P_{-1})X_1/n \\ &= (X_1 + X_2 + \epsilon)'(X_1 - rX_2)/n \\ &= 1 - r^2 + (X_1 - rX_2)'\epsilon/n.\end{aligned}$$

The denominator contains two factors. The first is

$$\begin{aligned}\|(I - P_{-1})X_1\|^2/n &= X_1'(I - P_{-1})X_1/n \\ &= X_1'(X_1 - rX_2)/n \\ &= 1 - r^2.\end{aligned}$$

Partial R^2 example (continued)

The other factor in the denominator is $Y'(I - P_{-1})Y$. This term is complex, but if we eliminate terms with expected value zero we get

$$\begin{aligned} EY'(I - P_{-1})Y/n &= E\epsilon'(I - P_{-1})\epsilon/n + (X_1 + X_2)'(I - P_{-1})(X_1 + X_2)/n \\ &= (n-2)/n + X_1'(I - P_{-1})X_1 \\ &= (n-2)/n + 1 - r^2. \end{aligned}$$

Thus we get that the partial R^2 is approximately equal to

$$\frac{1 - r^2}{(n-2)/n + 1 - r^2}.$$

If $r = 1$ then the result is zero (X_1 has no unique explanatory power), and if $r = 0$, the result is approximately 1/2, suggesting that after controlling for X_2, around half of the outcome variance is uniquely explained by X_1 (the other half is due to ϵ).

Another partial R^2 example

Suppose

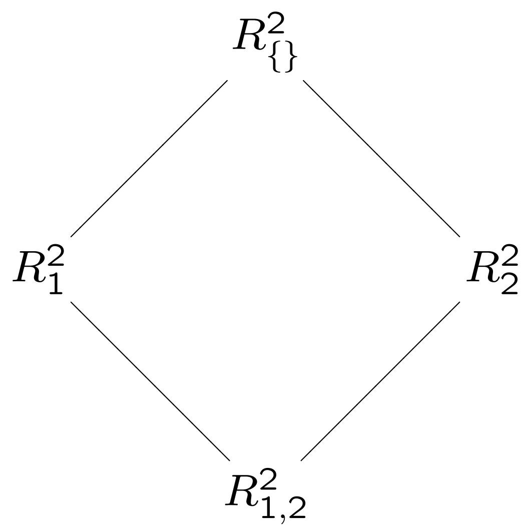
$$Y = bX_1 + \epsilon_1 \qquad\qquad X_2 = X_1 + \epsilon_2$$

where $E\epsilon_1 = E\epsilon_2 = 0$, var($\epsilon_k$) $= \sigma_k^2$, $EX_1 = 0$, var(X_1) $= 1$, and X_1 is independent of ϵ_1 and ϵ_2.

X_1 is "causal" and X_2 is a "surrogate."

Another partial R^2 example (continued)

The four R^2's for this model are related as follows, where $R^2_{\{\}}$ is the R^2 based only on the intercept.



Another partial R^2 example (continued)

We can calculate the limiting values for each R^2:

$$R^2_{\{\}} = 0$$

$$R^2_1 = R^2_{1,2} = \frac{b^2}{b^2 + \sigma_1^2}$$

Another partial R^2 example (continued)

For the regression on X_2, the limiting value of the slope is

$$\begin{aligned}\frac{\text{cov}(Y, X_2)}{\text{var}(X_2)} &= \frac{b \cdot \text{cov}(X_1, X_2) + \text{cov}(\epsilon_1, X_2)}{1 + \sigma_2^2} \\ &= \frac{b}{1 + \sigma_2^2}.\end{aligned}$$

Therefore the residual mean square is approximately

$$\begin{aligned}n^{-1}\|Y - \hat{Y}_2\|^2 &= n^{-1}\|bX_1 + \epsilon_1 - b(X_1 + \epsilon_2)/(1 + \sigma_2^2)\|^2 \\ &= n^{-1}\|\frac{b\sigma_2^2}{1 + \sigma_2^2}X_1 + \epsilon_1 - \frac{b}{1 + \sigma_2^2}\epsilon_2\|^2 \\ &\to \frac{b^2\sigma_2^2}{1 + \sigma_2^2} + \sigma_1^2.\end{aligned}$$

Another partial R^2 example (continued)

So,

$$\begin{aligned} R_2^2 &\rightarrow 1 - \frac{b^2\sigma_2^2/(1+\sigma_2^2) + \sigma_1^2}{b^2+\sigma_1^2} \\ &= \frac{b^2 - b^2\sigma_2^2/(1+\sigma_2^2)}{b^2+\sigma_1^2} \\ &= \frac{b^2/(1+\sigma_2^2)}{b^2+\sigma_1^2}. \end{aligned}$$

If $\sigma_2^2 = 0$ then $X_1 = X_2$, and we recover the usual R^2 for simple linear regression of Y on X_1.

Another partial R^2 example (continued)

With some algebra, we get an expression for the partial R^2 for adding X_1 to a model already containing X_2:

$$\frac{R^2_{1,2} - R^2_2}{1 - R^2_2} = \frac{b^2\sigma^2_2}{b^2\sigma^2_2 + \sigma^2_1 + \sigma^2_1\sigma^2_2}.$$

If $\sigma^2_2 = 0$, the partial R^2 is 0.

If $b \neq 0$, $\sigma^2_2 > 0$ and $\sigma^2_1 = 0$, the partial R^2 is 1.

Summary

Each of the three "R^2" values can be expressed either in terms of variance ratios, or as a squared correlation coefficient.

	Multiple R^2	Semi-partial R^2	Partial R^2
VR	$\|\hat{Y} - \bar{Y}\|^2 / \|Y - \bar{Y}\|^2$	$R^2 - R^2_{-k}$	$(R^2 - R^2_{-k})/(1 - R^2_{-k})$
Correlation	$\widehat{\text{cor}}(\hat{Y}, Y)^2$	$\widehat{\text{cor}}(Y, X_k^\perp)^2$	$\widehat{\text{cor}}((I - P_{-k})Y, X_k^\perp)^2$