Correlations

1.1 Setting

Consider the linear model:

$$Y = X\beta + \epsilon, \tag{1.1}$$

where Y is an $n \times 1$ outcome, X is an $n \times k$ design matrix, assumed to include an intercept, and $\epsilon \sim N(0, \sigma^2 I)$ is an $n \times 1$ residual vector. Model (1.1) is described as the *full-model*, in contrast to the *reduced-model*, which includes an intercept only:

$$Y = 1\beta_0 + \varepsilon \tag{1.2}$$

1.2 Sum of Squares Decomposition

The projection matrix for the full model is $P_X = X(X'X)^{-1}X'$, and that for the reduced model is $P_0 = 1(1'1)^{-1}1'$. Note that the full matrix X is assumed to contain an intercept. The projection of Y onto X is $\hat{Y}_X = P_X Y$, and that onto 1 is $\hat{Y}_0 = P_0 Y$. The **total sum of squares** is defined as:

$$||Y - \hat{Y}_0||^2 = ||(I - P_0)Y||^2 = Y'(I - P_0)Y.$$

Since $Y - \hat{Y}_X \in \text{im}(X)^\perp$ and $\hat{Y}_X - \hat{Y}_0 \in \text{im}(X)$, the total sum of squares decomposes as:

$$\begin{aligned} ||Y - \hat{Y}_0||^2 &= ||(I - P_0)Y||^2 \\ &= ||(I - P_X + P_X - P_0)Y||^2 \\ &= ||(I - P_X)Y||^2 + ||(P_X - P_0)Y||^2 \\ &= ||Y - \hat{Y}_X||^2 + ||\hat{Y}_X - \hat{Y}_0||^2. \end{aligned}$$

Here $||Y - \hat{Y}_X||^2 = Y'(I - P_X)Y$ is the **residual sum of squares** while $||\hat{Y}_X - \hat{Y}_0||^2 = Y'(P_X - P_0)Y$ is the **model sum of squares**.

1.3 Coefficient of Determination

The **coefficient of determination** for the full model (1.1) is defined as:

$$R^2 = \frac{||\hat{Y}_X - \hat{Y}_0||^2}{||Y - \hat{Y}_0||^2}.$$

This is the proportion of total variation explained by the columns of X other than the intercept. Note that:

$$R^2 = 1 - \frac{||Y - \hat{Y}_X||^2}{||Y - \hat{Y}_0||^2}.$$

1.4 Snedecor's Statistic

The ***F*-statistic** comparing the full (1.1) and reduced (1.2) models is:

$$T_F = \frac{||\hat{Y}_X - \hat{Y}_0||^2/(k-1)}{||Y - \hat{Y}_X||/(n-k)} \overset{H_0}{\sim} F_{k-1,n-k}(0).$$

Under the null hypothesis $\mathbb{E}(Y) \in \text{im}(1)$, T_F follows a central F distribution with numerator and denominator degrees of freedom $k-1$ and $n-k$.

1.5 Distribution of R^2

The F-statistic may be expressed in terms of the coefficient of determination:

$$T_F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}.$$

Likewise, R^2 may be expressed using the F-statistic:

$$R^2 = \frac{(k-1)T_F}{(k-1)T_F + (n-k)}.$$

For $T_F \sim F_{\nu_1,n_2}(0)$, $\nu_1 = k-1$, $\nu_2 = n-k$, the random variable $\nu_1 T_F/(\nu_1 T_F + \nu_2)$ follows a beta distribution with parameters $\alpha = \nu_1/2$ and $\beta = \nu_2/2$.

1.6 Adjusted R^2

Now, under H_0, $R^2 \sim B(\nu_1/2, \nu_2/2)$, and has expectation:

$$\mathbb{E}(R^2) = \frac{\nu_1}{\nu_1 + \nu_2} = \frac{k-1}{n-1}.$$

However, the expected value of R^2 is non-zero. Thus, R^2 is upward biased in general. To correct for this, consider the **adjusted $\boldsymbol{R^2}$**, defined as:

$$R_a^2 = R^2 + (1-R^2)\frac{(k-1)}{(n-k)}.$$

Observe that, in contrast to R^2, R_a^2 has expectation zero under H_0:

$$\begin{aligned}\mathbb{E}(R_a^2) &= \frac{k-1}{n-1} + \left(1 - \frac{k-1}{n-1}\right)\frac{(k-1)}{(n-k)} \\ &= \frac{k-1}{n-1} + \left\{\frac{(n-1)-(k-1)}{n-1}\right\}\frac{(k-1)}{(n-k)} = 0.\end{aligned}$$

1.7 (Semi) Partial R^2

1.7.1 Projection Decomposition

Let X_k denote the kth column of X, and let $X_{(-k)}$ denote the design matrix excluding column k. The projection onto X can be decomposed as:

$$\hat{Y}_X = P_X Y = (P_{X_{(-k)}} + P_{Q_{(-k)}X_k})Y = P_{X_{(-k)}}Y + P_{X_k^{\perp}}Y = \hat{Y}_{(-k)} + \hat{Y}_{X_k^{\perp}}.$$

Here $\hat{Y}_{(-k)} = P_{X_{(-k)}}Y$ denotes projection of Y onto all columns of X except k,

$$Q_{(-k)} = I - X_{(-k)}(X'_{(-k)}X_{(-k)})^{-1}X'_{(-k)}$$

is projection onto the orthogonal complement of $\mathrm{Im}(X_{(-k)})$, $X_k^{\perp} = Q_{(-k)}X_k$ is the portion of X_k orthogonal to the span of $X_{(-k)}$. To obtain the projection onto $X_k^{\perp}$ write:

$$Y = X_k\beta_k + X_{(-k)}\beta_{(-k)} + \epsilon.$$

Projecting first by $Q_{(-k)}$ to remove $X_{(-k)}$:

$$Q_{(-k)}Y = Q_{(-k)}X_k\beta_k + \tilde{\epsilon}.$$

The least squares estimator of β_k is:

$$\hat{\beta}_k = (X_k'Q_{(-k)}X_k)^{-1}X_k'Q_{(-k)}Y,$$

and the projection of Y onto $X_k^{\perp}$ is expressible is:

$$\hat{Y}_{X_k^{\perp}} = X_k^{\perp}\hat{\beta}_k = \frac{\langle X_k^{\perp}, Y\rangle}{\langle X_k^{\perp}, X_k^{\perp}\rangle}X_k^{\perp}.$$

Here $\langle X_k^{\perp}, Y\rangle = (X_k^{\perp})'Y = X_k'Q_{(-k)}Y$ and $\langle X_k^{\perp}, X_k^{\perp}\rangle = (X_k^{\perp})'X_k^{\perp} = X_k'Q_{(-k)}X_k$.

1.7.2 Semi Partial R^2

Define the **semi-partial** R^2 for X_k as:

$$\delta R_k^2 = R_X^2 - R_{(-k)}^2,$$

where R_X^2 is the coefficient of determination for the full model, and $R_{(-k)}^2$ is that for $X_{(-k)}$, i.e. the model excluding X_k.

Since $\hat{Y}_X$ and $Y - \hat{Y}_X$ are orthogonal:

$$||Y||^2 = ||Y - \hat{Y}_X + \hat{Y}_X||^2 = ||Y - \hat{Y}_X||^2 + ||\hat{Y}_X||^2.$$

Similarly, since $\hat{Y}_{(-k)}$ and $\hat{Y}_{X_k^\perp}$ are orthogonal:

$$||\hat{Y}_X||^2 = ||\hat{Y}_{(-k)}||^2 + ||\hat{Y}_{X_k^\perp}||^2.$$

Decomposing the full model R^2:

$$\begin{aligned}
R_X^2 &= 1 - \frac{||Y - \hat{Y}_X||^2}{||Y - \hat{Y}_0||^2} \\
&= 1 - \frac{||Y||^2 - ||\hat{Y}_X||^2}{||Y - \hat{Y}_0||^2} \\
&= 1 - \frac{||Y||^2 - ||\hat{Y}_{(-k)}||^2 - ||\hat{Y}_{X_k^\perp}||^2}{||Y - \hat{Y}_0||^2} \\
&= 1 - \frac{||Y||^2 - ||\hat{Y}_{(-k)}||^2}{||Y - \hat{Y}_0||^2} + \frac{||\hat{Y}_{X_k^\perp}||^2}{||Y - \hat{Y}_0||^2} \\
&= 1 - \frac{||Y - \hat{Y}_{(-k)}||^2}{||Y - \hat{Y}_0||^2} + \frac{||\hat{Y}_{X_k^\perp}||^2}{||Y - \hat{Y}_0||^2} \\
&= R_{(-k)}^2 + \frac{||\hat{Y}_{X_k^\perp}||^2}{||Y - \hat{Y}_0||^2}.
\end{aligned}$$

Thus, the semi-partial R^2 for X_k is:

$$\delta R_k^2 = R_X^2 - R_{(-k)}^2 = \frac{||\hat{Y}_{X_k^\perp}||^2}{||Y - \hat{Y}_0||^2}$$

To simplify the right-hand side, note that:

$$||\hat{Y}_{X_k^\perp}||^2 = \frac{\langle X_k^\perp, Y\rangle^2}{||X_k^\perp||^4}||X_k^\perp||^2 = \frac{\langle X_k^\perp, Y\rangle^2}{||X_k^\perp||^2} = \langle X_k^\perp/||X_k^\perp||, Y\rangle^2.$$

Therefore:

$$\delta R_k^2 = \frac{\langle X_k^\perp/||X_k^\perp||, Y\rangle^2}{||Y - \hat{Y}_0||^2} = \widehat{\mathrm{Cor}}^2\big(Y, X_k^\perp\big).$$

1.7.3 Partial R^2

The **partial** R^2 is the improvement in R^2 due to X_k relative to the maximum possible improvement:

$$R_k^2 = \frac{\delta R_k^2}{1 - R_{(-k)}^2} = \frac{R_X^2 - R_{(-k)}^2}{1 - R_{(-k)}^2}.$$

Expressing the numerator and denominator in terms of sums of squares:

$$R_X^2 - R_{(-k)}^2 = \frac{\langle X_k^\perp, Y\rangle^2}{||X_k^\perp||^2||Y - \hat{Y}_0||^2}, \qquad 1 - R_{(-k)}^2 = \frac{||Y - \hat{Y}_{(-k)}||^2}{||Y - \hat{Y}_0||^2}.$$

Thus:

$$R_k^2 = \frac{R_X^2 - R_{(-k)}^2}{1 - R_{(-k)}^2} = \frac{\langle X_k^{\perp}, Y \rangle^2}{||X_k^{\perp}||^2 ||Y - \hat{Y}_{(-k)}||^2}.$$

The inner product is expressible as:

$$\langle X_k^{\perp}, Y \rangle = X_k' Q_{(-k)} Y = X_k' Q_{(-k)} (Y - \hat{Y}_{(-k)}) = \langle X_k^{\perp}, Y - \hat{Y}_k \rangle,$$

Now:

$$R_k^2 = \frac{\langle X_k^{\perp}, Y - \hat{Y}_{(-k)} \rangle^2}{||X_k^{\perp}||^2 ||Y - \hat{Y}_{(-k)}||^2} = \left\langle \frac{X_k^{\perp}}{||X_k^{\perp}||}, \frac{Y - \hat{Y}_{(-k)}}{||Y - \hat{Y}_{(-k)}||} \right\rangle^2 = \widehat{\mathrm{Cor}}^2\big(Q_{(-k)} X_k, Q_{(-k)} Y\big).$$

Therefore, the partial R_k^2 is the R^2 for regression of $Q_{(-k)} Y$ onto $Q_{(-k)} X_k$.