ANOVA table and geometry

- The simplest ANOVA table partitions the total sum of squares in y into a component explained by the variation in $\mathbf{x}$ and a residual
- We are considering the situation with n observations and k predictor variables.

Source	Sum of Squares	Degrees of Freedom
Regression	$SSR = \sum(\hat{y}_i - \bar{y})^2$	k
Residual	$SSE = \sum(y_i - \hat{y}_i)^2$	$n - k - 1$
Total	$SST = \sum(y_i - \bar{y})^2$	$n - 1$

- The ANOVA table shows how much of the variation in $\mathbf{y}$ is explained by the predictors, and is often measured as $R^2 = SSR/SST$.
- The total and residual sums of squares give measures of fit respectively for a model containing a constant mean

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon} = \beta_0 \mathbf{1} + \boldsymbol{\epsilon}$$

and a model with a mean depending on all the predictors

- Let $\mathbf{1}$ be the vector of 1's of length n, and let $\mathbf{J}$ be the matrix of 1's.
- Then the hat matrix for the simple model with intercept only is
$$\mathbf{J}/n = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T.$$
- Recall that the hat matrix gives fitted values, $\hat{y} = \mathbf{H}\mathbf{y}$.
- The residuals are $\mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$.
- The matrices $\mathbf{H}, \mathbf{I} - \mathbf{H}, \mathbf{J}/n, \mathbf{I} - \mathbf{J}/n$, and $\mathbf{H} - \mathbf{J}/n$ are all *projection* matrices, which map vectors into subspaces of sample space.
- Projection matrices are symmetric and idempotent, which means that when you multiply the matrix by itself, it doesn't change.
  - For example $\mathbf{H}\mathbf{H} = \mathbf{H}$.

- There are several different matrix expressions for the sums of squares

$$SST = \mathbf{y}^T(\mathbf{I} - \mathbf{J}/n)\mathbf{y} = \mathbf{y}^T\mathbf{y} - n\bar{y}^2$$

$$\begin{aligned} SSR &= \mathbf{y}^T(\mathbf{H} - \mathbf{J}/n)\mathbf{y} = \hat{\mathbf{y}}^T\hat{\mathbf{y}} - n\bar{y}^2 \\ &= \hat{\beta}^T\mathbf{X}^T\mathbf{y} - n\bar{y}^2 \end{aligned}$$

and

$$\begin{aligned} SSE &= \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y} \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y} \end{aligned}$$

Cochran's Theorem

Theorem: Suppose that the variables y_i are independent $N(\boldsymbol{\mu}_i, 1)$, $i = 1, \ldots, n$, and let $\mathbf{y}^T Q_1 \mathbf{y}, \mathbf{y}^T Q_2 \mathbf{y}, \ldots, \mathbf{y}^T Q_s \mathbf{y}$ be quadratic forms in the y's such that

$$\sum_{i=1}^{n} y_i^2 = \mathbf{y}^T Q_1 \mathbf{y} + \mathbf{y}^T Q_2 \mathbf{y} + \ldots + \mathbf{y}^T Q_s \mathbf{y}$$

Let $n_j = \text{rank } Q_j$. Then the quadratic forms will have independent noncentral chi-squared distributions with degrees of freedom $n_1, n_2, \ldots, n_S$ respectively, if and only if, $\sum_{j=1}^{s} n_j = n$. (the noncentrality parameter of $\mathbf{y}^T Q_1 \mathbf{y}$ is given by $\boldsymbol{\mu}^T Q_1 \boldsymbol{\mu}$)

Application of Cochran's theorem

- By construction $SST = SSR + SSE$.
- As $SST = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2$, it follows that

$$\sum_{i=1}^{n} y_i^2 = n\bar{y}^2 + SSR + SSE = \mathbf{y}^T(\mathbf{J}/n)\mathbf{y} + SSR + SSE \quad (1)$$

- The matrices of the quadratic forms in SSR and SSE are projection matrices The rank of a projection matrix is the trace of the matrix.

- $\text{rank}(\mathbf{J}/n) = tr(\mathbf{J}/n) = n/n = 1$
-
$$\begin{aligned} \text{rank}(\mathbf{I} - \mathbf{H}) &= tr(\mathbf{I} - \mathbf{H}) = tr(I_{n\times n}) - tr(\mathbf{H}) = n - tr(\mathbf{H}) = \\ &= n - tr(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) \\ &= n - tr(\mathbf{I}_{(k+1)\times(k+1)}) = n - (k+1) \end{aligned}$$

  using the facts that $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$ and $tr(\mathbf{AB}) = tr(\mathbf{BA})$.
-
$$\text{rank }(\mathbf{H} - \mathbf{J}/n) = tr(\mathbf{H} - \mathbf{J}/n) = tr(\mathbf{H}) - tr(\mathbf{J}/n) = (k+1) - 1 = k$$

- as $n = 1 + [(k+1) - 1] + [n - (k+1)]$, the conditions of Cochran's theorem hold, and it follows that $\frac{n\bar{y}^2}{\sigma^2}$, $\frac{SSR}{\sigma^2}$ and $\frac{SSE}{\sigma^2}$ have independent χ^2 distributions with 1, k and $n-1-k$ degrees of freedom
- and additionally, under $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$, the distributions are "central" chi-squared distributions

test of overall significance: $H_0 : \beta_1 = \ldots = \beta_k = 0$

- Under the null hypothesis, SSE/σ^2 and SSR/σ^2 have independent χ^2 distributions with $n-1-k$ and k degrees of freedom, respectively.
- It follows by definition of the F distribution that under H_0, $F = \frac{SSR/k}{SSE/(n-1-k)}$ has an F distribution with k numerator and $n-1-k$ denominator degrees of freedom
- and the p-value for the test is $P(F_{k,n-1-k} > F_{obs})$.

Geometry

- Geometrically, the deviation of the response vector from its mean is decomposed into a component in the expectation surface orthogonal to the vector of 1's and a component perpendicular to the expectation surface

$$(\mathbf{I} - \mathbf{J}/n)\mathbf{y} = (\mathbf{H} - \mathbf{J}/n)\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y}$$

- Multiplying each side by its transpose, and using the facts that projection matrices are symmetric and idempotent, gives the ANOVA identity

$$\begin{aligned}\mathbf{y}^T(\mathbf{I} - \mathbf{J}/n)\mathbf{y} &= \mathbf{y}^T(\mathbf{H} - \mathbf{J}/n)\mathbf{y} \\ &\quad + \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y}\end{aligned}$$

- The cross product term vanishes because $(\mathbf{I} - \mathbf{H})(\mathbf{H} - \mathbf{J}/n) = \mathbf{0}$. (Note that $\mathbf{HJ/n} = \mathbf{J}/n$, because the vector of 1's is one of the columns of $\mathbf{X}$.)

Interpretation

- The ANOVA decomposition is just Pythagoras' theorem for the right angle triangle with hypotenuse $(\mathbf{I} - \mathbf{J}/n)\mathbf{y}$, and perpendicular sides $(\mathbf{H} - \mathbf{J}/n)\mathbf{y}$ and $(\mathbf{I} - \mathbf{H})\mathbf{y}$.
- (Note: the dimension of the subspace onto which a projection matrix projects is given by the trace (sum of diagonal entries) of the matrix.)
- From Cochran's theorem the degrees of freedom are equal to the dimension of the subspaces into which $\mathbf{y}$ is projected.
- For SST, we project perpendicular to the vector $\mathbf{1}$, so the degrees of freedom is/are $n - 1$.
- For SSE, the projection is orthogonal to the expectation surface, and so the degrees of freedom is/are $n - k - 1$
- For SSR, the projection is into the component of the expectation surface perpendicular to the vector $\mathbf{1}$, so the degrees of freedom is/are k.