13. Covariance of regression parameters:
   a. Calculate $\text{Cov} [\hat{\beta}]$
   b. Use above fact with $\hat{\beta} = CY$
   c. Components of $Y$ are independent
   d. $\text{Cov} [Y] = \sigma^2 I$
      i. Hence covariance matrix is $C(\sigma^2 I)C^\top = \sigma^2 (X\top X)^{-1}(X\top X)(X\top X)^{-1} = \sigma^2 (X\top X)^{-1}$

14. (Multivariate) Normality
   a. If the response variables are exactly normal, $\hat{\beta}$ is multivariate normal
   b. If the response variables are close to normal, and $n$ large, a central limit theorem proves $\hat{\beta}$ is approximately multivariate normal

15. Regression of residuals on existing explanatory variable gives all zeros
   a. Residuals are $\hat{\epsilon} = (I - X(X\top X)^{-1}X\top)Y$
   b. Parameter estimates for residuals based on existing explanatory variables are $(X\top X)^{-1}$ times $X\top \hat{\epsilon} = X\top (I - X(X\top X)^{-1}X\top)Y$
   c. Use distributive property to give second factor as $X\top Y - X\top X(X\top X)^{-1}X\top Y$
   d. Use fact $(X\top X)(X\top X)^{-1} = I$
   e. Hence $X\top \hat{\epsilon} = 0$
   f. Hence parameter estimate is zero
   g. Hence fitted values are all zero.

16. Orthogonal Regressors
   a. Recall that with one explanatory variable, subtracting off mean gives $X\top X$ matrix diagonal
   b. With two covariates, make new $n \times 3$ matrix $Z$
      i. First column is all 1, as before
      ii. Second column is $X_{1i} - \bar{X}_1$, first covariate with mean removed
   c. Can make same fitted values by letting new constant term coefficient be $\beta_0 + \beta_1 \bar{X}_1$
   d. Replace third column by residuals after regressing on constant and first regressor
   e. Again, get same fitted value by adjusting new constant and first regression coefficients
   f. Preceeding fact shows that sums of products of columns of $Z$ are 0.
   g. Hence $Z\top Z$ is diagonal with $\sum_j z^2_{ji}$ as entry in row and column $i$
   h. Regression of $Y$ on $Z$ is easy
   i. Hence fitted values are all zero.
   j. Can then reconstruct regression of $Y$ on $X$

17. Orthonormal Regressors
   a. Can rescale each of the new regressors to make $\sum_{j} z^2_{ji} = 1$ for all $i$
   b. This makes $Z\top Z = I$
   c. Columns of $Z$ are called orthonormal.
   d. $\text{MPV}: 3.10$

18. Can happen that residuals of explanatory variable on earlier explanatory variable all become zero.

b. When regressors are not orthogonal, this is manifest by $X\top X$ not having an inverse.

19. Colinearity sometimes holds approximately.

   a. In such cases, $(X\top X)^{-1}$ has large values
   b. Hence coefficient standard errors become large
      i. “inflated”.
      ii. Measured by constructing matrix of standardized regressors $X_s$
      iii. Calculate $(X_s\top X_s)^{-1}$
      iv. Marginally, for each variable, standard error is $(n - 1)\sigma$
      v. Ratio of diagonal entry of $(X_s\top X_s)^{-1}$ to $n - 1$ is called variance inflation factor.
   c. Reflects difficulty in separating causation between two sources with almost the same information.

C. Estimation of the Dispersion Parameter

1. Mean Square Error in terms of the Hat Matrix
   a. $SS_{\text{Res}} = \sum_j (Y_j - \hat{Y}_j)^2 = \sum_j \hat{\epsilon}_j^2$
   b. Write in matrix form as $\hat{\epsilon}\top \hat{\epsilon}$
   c. As before, $\hat{\epsilon}$ is a column vector, treated as a vector with 1 column
   d. New notation: $\hat{\epsilon}\top$ is called a row vector, treated as a matrix with 1 row.
   e. Then the product is the sum of the values in column $j$ of $\hat{\epsilon}\top$ and row $j$ of $\hat{\epsilon}\top$
   f. This product is $\sum_j \hat{\epsilon}^2_j$
   g. As previously, transpose of product is product of transpose in reverse order.
   h. Then $SS_{\text{Res}} = Y\top(I-H)(I-H)Y = Y\top(I-H)Y$

2. Filling Out the Regression Matrix
   a. Let $k$ be the number of original regressors, including the constant.
   b. Make $n - k$ fake regressors $Z_f$ by
      i. Convert regressors to an orthonormal set as above.
      ii. Convert the set of vectors $e_j$ with 1 for component $j$ and 0 everywhere else sequentially into more orthogonal covariates
      iii. You can prove that you can find $n - k$ of these that are not zero
      iv. Make these orthonormal by dividing by the sum of squares.
   c. The combined matrix $U = (Z, Z_f)$ is invertible.
      i. Hence $Y$ is the vector of fitted values for regression model $Y = U\gamma + \epsilon$

3. Distribution of MSE
   a. Then $Y\top Y = Y\top U(U\top U)^{-1}U\top Y$
   b. Note
      i. $U\top U = I$
      ii. $UU\top = ZZ\top + Z_fZ_f\top$
      iii. $H = I - ZZ\top$
   c. Then $Y\top Y = Y\top UU\top Y$
D. Hypothesis Testing and Confidence Intervals

1. Tests of a Single Slope Parameter
   a. $\hat{\beta}_j \sim N(\beta_j, d_j\sigma^2)$ for $d_j$ the element in row $j$ and column $j$ of $(X^\top X)^{-1}$.
   b. Hence $(\hat{\beta}_j - \beta_j)/SE(\hat{\beta}_j) \sim t_{n-k}$ for $SE(\hat{\beta}_j) = \hat{\sigma}\sqrt{a_j}$.
   c. Test of level $\alpha$ rejects $H_0: \beta_j = \beta^0_j$ if $\hat{\beta}_j > \beta^0_j + t_{n-k,\alpha/2}SE(\hat{\beta}_j)$ or if $\hat{\beta}_j < \beta^0_j - t_{n-k,\alpha/2}SE(\hat{\beta}_j)$.
   d. Confidence interval with confidence $1 - \alpha$ is $\hat{\beta}_j \pm t_{n-k,\alpha/2}SE(\hat{\beta}_j)$.
   e. This is called univariate Wald test.

2. Tests involving multiple Parameters
   a. Again, number parameters from 0 to $k-1$.
   b. Without loss of generality, to test $\beta_j = \beta^0_j$ for $j = k-m+1, \ldots, k$
   i. Let $\gamma$ represent the vector of these values.
   ii. Let $\gamma^0$ represent the corresponding null values.
   c. Let $W$ be the part of matrix $(X^\top X)^{-1}$.
   d. Then $\gamma \sim N(\gamma^0, W\sigma^2)$.
   e. Then $(\gamma - \gamma^0)^\top W^{-1}(\gamma - \gamma^0)/\sigma^2 \sim \chi^2_m$.
   i. Recall that a $\chi^2_m$ distribution is the distribution of the sum of $m$ independent squares of standard normal variables.
   ii. This is called multivariate Wald test.
   f. Then $((\gamma - \gamma^0)^\top W^{-1}(\gamma - \gamma^0)/m)/\sigma^2 \sim F_{m,n-k}$.

3. Case with $H_0: \beta_j = 0 \forall j > 0, \beta_0$ an intercept parameter
   a. Subtract off mean from each of the regressors: “centered”.
   b. Let
   i. $X_c$ represent centered covariates without the constant term.

Lecture 5

ii. $X = (1, X_c)$.
   - A partitioned matrix, not a vector
   - Multiplies in the same way that vectors do.
iii. $\beta^* = (\beta_1, \ldots, \beta_p)$
   c. Then model is $E[Y \mid X] = \beta_0 + X_c\beta^*$

d. $X^\top X = \begin{pmatrix} n & \mathbf{0} \\ \mathbf{0} & X_c^\top X_c \end{pmatrix}$

e. $(X^\top X)^{-1} = \begin{pmatrix} \mathbf{1}/n & \mathbf{0}^\top \\ \mathbf{0} & (X_c^\top X_c)^{-1} \end{pmatrix}$
   f. $(X^\top X)^{-1}X^\top = \begin{pmatrix} \mathbf{1}/n & \mathbf{0}^\top \\ X_c^\top X_c^{-1}X_c^\top \end{pmatrix}$

4. Wald Test in this Restrictive Case
   a. $W^{-1} = X_c^\top X_c/\sigma^2$
   b. Statistic is $F = \beta^*\top X_c^\top X_c\beta^*/\sigma^2 = (\mathbf{Y} - \bar{Y})^\top (\mathbf{Y} - \bar{Y})/(k-1))/\sigma^2 = MS_R/MS_{Res}$
   c. $P$-value is the probability that an $F_{k-1,n-k}$ variable exceeds $F$.
   d. Summarized by Analysis of Variance (ANOVA) table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$SS_R$</td>
<td>$m$</td>
<td>$MS_R$</td>
<td>$F$</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>$SS_{Res}$</td>
<td>$n - 1 - m$</td>
<td>$MS_{Res}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_t$</td>
<td>$n - 1$</td>
<td>$MS_t$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>