Properties of normal variables:

a) See Section 5.7 in text

b) Linear combinations of normals are themselves jointly normal

c) Being uncorrelated (covariance 0) is equivalent to being independent (Theorem 5.34)

d) Suppose $Z_1, \ldots, Z_n$ are $\mathcal{N}(0, I)$, that is, $Z = [Z_1, \ldots, Z_n]^T$ and $Y = UZ$ with $U$ an $m \times n$ matrix satisfying $UU^T = I_m$. Then $Y_1, \ldots, Y_m$ are $\mathcal{N}(0, I)$. 
Chi-squared.

**Definition.** Suppose $Z_1, \ldots, Z_n$ are $N(0,1)$, then

$$\chi^2 = Z_1^2 + \ldots + Z_n^2$$

where $n$ is the degrees of freedom.

**Theorem.** Let $H$ be idempotent, $H^2 = H$, with rank $m$. Let $\tilde{Z} = [\tilde{Z}_1^T, \ldots, \tilde{Z}_m^T]^T$. Then

$$\tilde{Z}^T H \tilde{Z} = \chi^2_m$$

**Proof.** One can write $H = O^T D O$ with $O$ orthogonal and $D = \text{diag} \{1, \ldots, 1, 0, \ldots, 0\}$, $m$ ones appearing.
Suppose $Z_1, \ldots, Z_{m+n}$ are IN$(0, 1)$, then

$$F_{m,n} = \frac{(Z_1^2 + \ldots + Z_m^2)/m}{(Z_{m+1}^2 + \ldots + Z_{m+n}^2)/n} = \frac{\chi_m^2 / m}{\chi_n^2 / n}$$

with $\chi_m^2 \perp \chi_n^2$.

Note:

$$\chi_{m+n}^2 = Z_1^2 + \ldots + Z_m^2 + Z_{m+1}^2 + \ldots + Z_{m+n}^2$$

$$= \chi_m^2 + \chi_n^2$$

c.f. ANOVA tables to be discussed later.
Most of the ANOVA and F-test results in Chapters 7-11 come from the following development based on partitioned matrices and vectors and properties of linear transforms of normal variates.

**Normal regression model**

\[
\begin{align*}
Y & = X \beta + \epsilon \\
\epsilon & \sim N(0, \sigma^2)
\end{align*}
\]

with \(X, \beta\) constant and the entries of \(\epsilon \sim N(0, \sigma^2)\).

Suppose the \(p\) \(X\)-variables partition orthogonally

\[
X = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \quad \text{with} \quad X_1^T X_2 = 0
\]

The normal equations \(X^T X \hat{\beta} = X^T Y\) become

\[
\begin{bmatrix} X_1^T X_1 & 0 \\ 0 & X_2^T X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1^T Y \\ X_2^T Y \end{bmatrix}
\]

i.e.,

\[
\begin{align*}
\hat{\beta}_1 & = (X_1^T X_1)^{-1} X_1^T Y \\
\hat{\beta}_2 & = (X_2^T X_2)^{-1} X_2^T Y \\
& = \hat{\beta}_1 + (X_1^T X_1)^{-1} X_1^T \epsilon \\
& = \hat{\beta}_2 + (X_2^T X_2)^{-1} X_2^T \epsilon
\end{align*}
\]

If the inverses exist.
The residuals are
\[ \hat{\mathbf{w}} = \hat{\mathbf{y}} - \mathbf{x} \hat{\mathbf{\beta}} = \mathbf{y} - \mathbf{x}, \hat{\beta}_1, - \mathbf{x}, \hat{\beta}_2. \]

The variates \( \hat{\beta}_1, \hat{\beta}_2, \hat{\mathbf{w}} \) are all normal (being linear combinations of normals) and independent (having covariances 0).

Define the hat matrices
\[ H = \mathbf{x} (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T, \quad H_1 = \mathbf{x}, (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x},, \quad H_2 = \]

Then
\[ \hat{\mathbf{y}} = \mathbf{x}, \hat{\mathbf{\beta}} = H \hat{\mathbf{y}} \]
\[ = \mathbf{x}, \hat{\beta}_1 + \mathbf{x}, \hat{\beta}_2 = H_1 \hat{\mathbf{y}} + H_2 \hat{\mathbf{y}} \]
\[ \hat{\mathbf{w}} = (I - H_1 - H_2) \hat{\mathbf{y}} = (I - H_1 - H_2) \hat{\mathbf{w}} \]
Because of the orthogonality

\[(X_2 \hat{\beta}_2)^T X_2 \hat{\beta}_2 = \hat{\beta}_2^T X_2^T X_2 \hat{\beta}_2 = 0\]

\[(X_2 \hat{\beta}_2)^T (Y - X_2 \hat{\beta}_2) = \hat{\beta}_2^T (X_2^T Y - X_2^T X_2 \hat{\beta}_2) = 0\]

\[\begin{pmatrix} X_2 \hat{\beta}_2 \end{pmatrix}^T \begin{pmatrix} \end{pmatrix} = 0\]

one has the sum of squares identity

\[\|
\begin{align*} Y \end{align*} \|
ew{2} = \| X_1 \hat{\beta}_1 \| \new{2} + \| X_2 \hat{\beta}_2 \| \new{2} + \| Y - X \hat{\beta} \| \new{2} \]

Next,

\[\| Y - X \hat{\beta} \| \new{2} = \| (I - H) W \| \new{2} = \sigma^2 \| Z^T (I - H) Z \| \new{2}\]

as \( I - H \) is idempotent, the distribution is \( \sigma^2 \chi^2_\nu \) where \( \nu = \text{rank} \left( Z^T - H \right) = n - r \left( H \right) = n - r \left( \frac{X}{X^T (X^T X)^{-1} X^T} \right) = n - p \)

Likewise

\[\| X_2 \hat{\beta}_2 \| \new{2} = \| H_2 W \| \new{2}\]

\[= \| H_2 X_2 \hat{\beta}_2 + H_2 W \| \new{2}\]

\[= \sigma^2 Z^T H_2 Z \quad \text{if} \quad \hat{\beta}_2 = 0\]
Its distribution is then $\sigma^2 \chi^2_{r(H_0)} = \sigma^2 \chi^2_{p_2}$.

One has the ANOVA table:

<table>
<thead>
<tr>
<th>Source</th>
<th>$SS$</th>
<th>$DF$</th>
<th>$MS=SS/DF$</th>
<th>$F$</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$\sum_1 \hat{X}_1^2$</td>
<td>$p_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>$\sum_2 \hat{X}_2^2$</td>
<td>$p_2$</td>
<td>$\sigma^2$</td>
<td>$2$</td>
<td>$2/3$</td>
</tr>
<tr>
<td>Residuals</td>
<td>$\sum (Y - \hat{X})^2$</td>
<td>$n-p$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y$</td>
<td>$\sum Y^2$</td>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One sees that under $H_0: \beta_i = 0$, the $F$-statistic is distributed as $F_{p_2, n-p}$.

Notes:

1. $X$ might be partitioned into $\sum X_1, X_2, X_3, \ldots$ and similar results hold. The ANOVA Table has more rows.
2. The text take the last row of the table, ("Total") as corresponding to
\[ \| \mathbf{y} - \bar{y} \|_2^2 = \sum (y_i - \bar{y})^2 \]
This corresponds to taking \( x_1 = \frac{1}{s} \) above and
the columns of \( X_2 \) having mean 0. (This
makes \( x_1 \perp x_2 \).) As
\[ \| \mathbf{y} \|_2^2 = \| \mathbf{y} - \bar{y} \|_2^2 + \bar{y}^2 \| \mathbf{I} \|_2^2 \]
there is no conflict.

3. When \( \beta_2 \neq 0 \)
\[ \| \hat{x}_2 \beta_2 \|_2^2 = \| x_2 \beta_2 + \hat{w}_2 \|_2^2 \]
has a non-central chi-squared distribution, \( F \)
has a non-central \( F \) distribution and is
stochastically larger.
4. Sometimes there are duplicate rows in the \( X \) matrix. Supposing there are \( d \) distinct rows, write

\[
Y_{ki} = \mu_k + W_{ki}, \quad i = 1, \ldots, n_k \text{ and } k = 1, \ldots, d
\]

One can break the residual sum of squares, \( \| Y - \bar{Y} \| ^2 \), into a within sum of squares, \( \sum (Y_{ki} - \bar{Y}_k)^2 \), and a lack of fit sum of squares, \( \sum (\bar{Y}_{ki} - \bar{Y}_k)^2 \), and have a goodness-of-fit statistic

\[
F = \frac{\text{LSS} / (d-p)}{\text{WSS} / (n-d)}
\]

for the model \( (\mu(x) \in \mathcal{A}) \). Its null distribution is \( F_{d-p, n-d} \).
Sum of squares,
WSS: within

BSS: between

RSS: residual

LSS: lack-of-fit

FSS: fitted

Identities:
TSS = BSS + WSS = FSS + LSS + WSS
RSS = LSS + WSS
BSS = FSS + LSS
identifiable: $X$ is of full rank

saturated: $p = d$