Scheffe's method.

Let $A'_{p\times d} = (a_1, a_2, \ldots, a_d)$ where $a_1, a_2, \ldots, a_d$ are linearly independent and $a_{d+1}, \ldots, a_k$ are linearly dependent on them. Then $d \le \min\{k, r\}$. Let $\phi = A\beta$ and $\hat{\phi} = A\hat{\beta}$. Then

$$F(\beta) = \frac{(\hat{\phi}-\phi)'\,(A(X'X)^{-}A')^{-1}\,(\hat{\phi}-\phi)/d}{\mathrm{RSS}/(n-r)} \sim F_{d,n-r}.$$

Therefore,

$$\begin{aligned} 1-\alpha &= P[F(\beta) \le F_{d,n-r}(1-\alpha)] \\ &= P\left\{(\hat{\phi}-\phi)'\left(A(X'X)^{-}A'\right)^{-1}(\hat{\phi}-\phi) \le d\frac{\mathrm{RSS}}{n-r}\,F_{d,n-r}(1-\alpha)\right\}. \end{aligned}$$

This gives an ellipsoid as before, but consider the following result.

Result. If L is positive definite,

$$b'L^{-1}b = \sup_{h\neq 0}\frac{(h'b)^2}{h'Lh}.$$

Proof. Note that

$$\frac{(h'b)^2}{h'Lh} = \frac{(h'L^{1/2}L^{-1/2}b)^2}{h'Lh} \le \frac{h'Lhb'L^{-1}b}{h'Lh} = b'L^{-1}b.$$

Therefore,

$$1-\alpha = P\left\{\sup_{h\neq 0}\frac{\left\{h'(\phi-\hat{\phi})\right\}^2}{h'\left(A(X'X)^{-}A'\right)h} \le \frac{d}{n-r}\mathrm{RSS}\,F_{d,n-r}(1-\alpha)\right\}$$

$$\begin{aligned} &= P\left\{\frac{\left\{h'(\phi-\hat{\phi})\right\}^2}{h'\left(A(X'X)^{-}A'\right)h} \le \frac{d}{n-r}\mathrm{RSS}\,F_{d,n-r}(1-\alpha) \text{ for all } h\neq 0.\right\} \\ &= P\left\{\frac{|h'(\phi-\hat{\phi})|}{\sqrt{\frac{\mathrm{RSS}}{n-r}}\sqrt{h'\left(A(X'X)^{-}A'\right)h}} \le \{dF_{d,n-r}(1-\alpha)\}^{1/2} \text{ for all } h\neq 0.\right\} \\ &= P\left\{|h'(\phi-\hat{\phi})| \le \{dF_{d,n-r}(1-\alpha)\}^{1/2}\ \text{s.e.}(h'\hat{\phi}) \text{ for all } h\neq 0.\right\}, \end{aligned}$$

where $\text{s.e.}(h'\hat{\phi}) = \sqrt{\frac{\mathrm{RSS}}{n-r}}\sqrt{h'\left(A(X'X)^{-}A'\right)h}$. Therefore,

$$a_i'\hat{\beta} \pm \{dF_{d,n-r}(1-\alpha)\}^{1/2}\sqrt{\frac{\mathrm{RSS}}{n-r}}\sqrt{a_i'(X'X)^{-}a_i}, i = 1, 2, \ldots, k$$

is a simultaneous $100(1-\alpha)\%$ confidence set for $a_1'\beta, a_2'\beta, \ldots, a_k'\beta$, by noting that

$$P\left(a_i'\beta \in a_i'\hat{\beta} \pm \{dF_{d,n-r}(1-\alpha)\}^{1/2}\,\text{s.e.}(a_i'\hat{\beta}, i = 1, 2, \ldots, k\right) \geq$$

$$P\left\{|h'(\phi - \hat{\phi})| \leq \{dF_{d,n-r}(1-\alpha)\}^{1/2}\ \text{s.e.}(h'\hat{\phi}) \text{ for all } h \neq 0.\right\} = 1 - \alpha.$$

Many other methods are also available.

Regression diagnostics

Lack of fit. Suppose the true model is $Y = f(X) + \epsilon$, $\epsilon \sim N_n(0, \sigma^2 I_n)$, whereas we fit $Y = X\beta + \epsilon$. We do get $\hat{\beta} = (X'X)^{-1}X'Y$ and $\hat{\sigma}^2 = \text{RSS}/(n-r)$. σ^2 is supposed to account for only the statistical errors (ϵ_i), and not model misspecification. Therefore, if $f(X) \neq X\beta$, we have statistical errors, ϵ_i, as well as the bias, $f(X) - X\beta$. Then, $\hat{\sigma}^2 = \text{RSS}/(n-r)$ will estimate a quantity which includes σ^2 as well as $(\text{bias})^2$. If σ^2 is known, then comparing $\hat{\sigma}^2$ with σ^2 can act as a check for lack of fit. In other words,
$\text{RSS}/\sigma^2 \sim \chi^2_{n-r}$ if the model, $Y = X\beta + \epsilon$, $\epsilon \sim N_n(0, \sigma^2 I_n)$ is true. Therefore to test
$H_0 : Y = X\beta + \epsilon, \epsilon \sim N_n(0, \sigma^2 I_n)$ versus $H_1 : Y$ has some other model, use RSS/σ^2 as the test statistic. If the observed value is too large compared to χ^2_{n-r}, there is evidence against H_0.

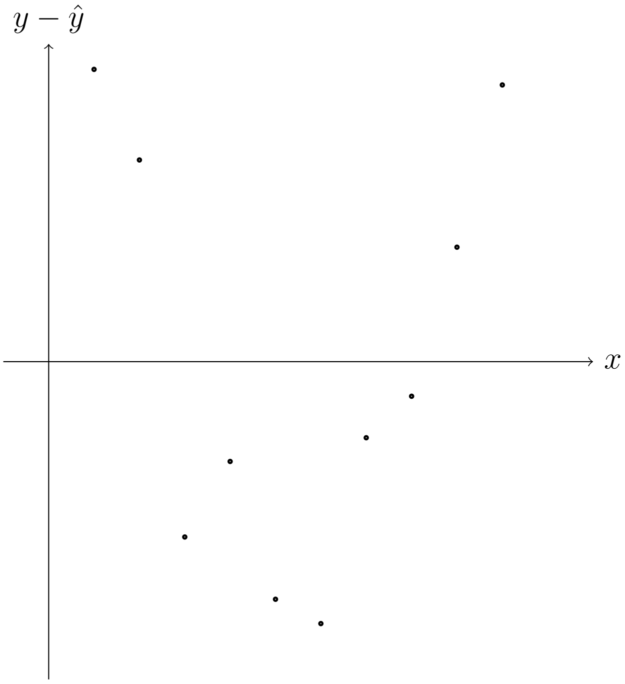
Consider a simulation study where data are generated from $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$, with $\beta_0 = 5$, $\beta_1 = \beta_2 = 2$ and $\sigma^2 = 2^2$:

x	.5	1	1.5	2	2.5	3	3.5	4	4.5	5
y	8.68	12.85	10.71	18.54	21.67	27.3	37.56	44.64	54.09	63.83

Regress Y on X. i.e., fit $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Then we get $\hat{\beta}_0 = -3.925$, $\hat{\beta}_1 = 12.33$ and the ANOVA table:

source	d.f	SS	MS	F	R^2
Regression	1	3134.2	3134.2	130.76	94.2%
Error	8	191.7	24.0		
Total	9	3325.9			

These are very good results, but $\text{RSS}/\sigma^2 = 191.7/4 = 47.925 >> \chi^2_8(.99) = 20.08$. $R^2 = 94.2\%$ is high, and F-ratio of 130.76 at (1, 8) d.f. is very high, indicating that X is a very useful predictor of Y. However this does not mean that the fitted model is the correct one. Check the residual plot:



Now regress Y on X and X^2.

source	d.f	SS	MS	F	R^2
Regression	2	3305.7	1652.8	572.28	99.4%
Error	7	20.2	2.9		
Total	9	3325.9			

RSS$/\sigma^2 = 20.2/4 = 5.5 << \chi^2_7(.90) = 12.02$.

σ^2 is usually unknown, so this test is difficult, but what this indicates is that residual plots are useful for checking lack of fit (see plot above). Another possibility is to check for any pattern between fitted values and residuals. Yet another reason to explore this is the following.
$\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta} = (I-P)Y$ and $\hat{Y} = X\hat{\beta} = PY$ are uncorrelated (since $(I-P)P = 0$) if $Cov(Y) = \sigma^2 I_n$. If one sees significant correlation and some trend, then the model is suspect. What if $Var(y_i) = \sigma_i^2$, not a constant? This is called heteroscedasticity (as against homoscedastcity), a problem discussed in Sanford Weisberg: *Applied Linear Regression* in the context of regression diagnostics.