Popular choices of scores:

- Mallow's C_p: RSS $+ 2\hat{\sigma}^2_{\text{full}} \times p$ [a]
- AIC: $-2\text{loglik} + 2p$ [b]
- BIC: $-2\text{loglik} + (\log n)p$

Note that when n is large, adding an additional predictor costs a lot more in BIC than AIC. So AIC tends to pick a bigger model than BIC. C_p performs similar to AIC.

---

[a] $\hat{\sigma}^2$ is estimated from the full model (i.e., the model with all the predictors).

[b] In the context of linear regression with normal errors, we can replace -2loglik by $\log \text{RSS}$.