Estimation of σ^2, the variance of ϵ

- The variance of the errors σ^2 indicates how much observations deviate from the fitted surface.

- If σ^2 is small, parameters $\beta_0, \beta_1, ..., \beta_k$ will be reliably estimated and predictions $\hat{y}$ will also be reliable.

- Since σ^2 is typically unknown, we estimate it from the sample as:

$$\hat{\sigma}^2 = S^2 = \frac{SSE}{n - \text{ number of parameters in model}} = \frac{SSE}{n - (k+1)}.$$

Estimation of σ^2 (cont'd)

- As in the case of simple linear regression:

$$SSE = \sum_i (y_i - \hat{y}_i)^2$$

- Also as before, the predicted value of y for a given set of x's is:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + ... + b_k x_{ki}.$$

Estimation of σ^2 (cont'd)

- As in simple regression, SAS and JMP call S^2 the **MSE** and S the **RMSE**.

- See example on page 169 of text. We have:
  - Three predictors of home sale price, so $k = 3$.
  - Sample size $n = 20$.
  - From SAS output, $MSE = 62,718,204$ and $RMSE = 7,919$.

- If we had wished to compute MSE by hand, we would have done so as:

$$MSE = \frac{SSE}{n-(k+1)} = \frac{1003491259}{20-(3+1)}.$$

Inferences about β parameters

- A $(1-\alpha)\%$ confidence interval for β_j is given by:

$$b_j \pm t_{\frac{\alpha}{2},n-(k+1)} \text{ standard error of } b_j$$

- We use $\hat{\sigma}_{b_j}$ or S_{b_j} to denote the standard error of b_j, and obtain its value from SAS or JMP output.

- The standard errors of the regression estimates are given in the column labeled Standard Error, both in SAS and in JMP.

- We can obtain confidence intervals for any of the regression coefficients in the model (and also for the intercept).

Inferences about β parameters

- Example: see example on page 169. We wish to obtain a 90% CI for β_2:

$$\begin{aligned} b_2 &\pm t_{0.05,16}\hat{\sigma}_{b_2} \\ 0.82 &\pm 1.746(0.21), \end{aligned}$$

or $(0.45,\ 1.19)$.

- As before, we say that the interval has a 90% probability of covering the true value of β_2.

Inferences about β parameters (cont'd)

- We can also test hypotheses about the β's following the usual steps:
  1. Set up the hypotheses to be tested, either one or two-tailed.
  2. Choose level α, determine critical value and set up rejection region.
  3. Compute test statistic.
  4. Compare test statistic to critical value and reach conclusion.
  5. Interpret results.

- Hypotheses: The null is always $H_0 : \beta_j = 0$.
  - Alternative for a two-tailed test: $H_a : \beta_j \neq 0$.
  - Alternative for a one-tailed test: $H_a : \beta_j < 0$ or $H_a : \beta_j > 0$.

Inferences about β parameters (cont'd)

- Critical value: For a two-tailed test, the critical values are $\pm t_{\alpha/2,n-k-1}$. For a one-tailed test, the critical value is $-t_{\alpha,n-k-1}$ (if $H_a : \beta_j < 0$) or $+t_{\alpha,n-k-1}$ (if $H_a : \beta_j > 0$).

- Critical or rejection region:
  - For a two-tailed test: Reject H_0 if test statistic $t < -t_{\alpha/2,n-k-1}$ or $t > t_{\alpha/2,n-k-1}$.
  - For a one-tailed test: Reject H_0 if $t < -t_{\alpha,n-k-1}$ [or $t > +t_{\alpha,n-k-1}$ for a "'bigger-than"' alternative.]

- Test statistic: As in simple linear regression:

$$t = \frac{b_j}{\hat{\sigma}_{b_j}}$$

Inferences about β parameters (cont'd)

- How do we interpret results of hypotheses tests?

- Suppose we reject $H_0 : \beta_j = 0$ while conducting a two-tailed test. Conclusion: Data suggest that the response y and the jth predictor x_j are linearly associated when other predictors are held constant.

- If we fail to reject H_0, then reasons might be:

  1. There is no association between y and x_j.
  2. A linear association exists (when other x's are held constant) but a Type II error occurred (defined as concluding H_0 when H_a is true).
  3. The association between y and x_j exists, but it is more complex than linear.

Multiple coefficient of determination R^2

- The multiple coefficient of determination R^2 is a measure of how well the linear model fits the data.

- As in simple linear regression, R^2 is defined as:

$$R^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}},$$

  and $0 \leq R^2 \leq 1$.

- The closer R^2 is to one, the better the model fits the data.

- If R^2 is equal to 0.65 (for example) we say that about 65% of the sample variability observed in the response can be attributed to (or explained by) the predictors in the model.

The *adjusted* R^2

- As it happens, we can artifially increase R^2 simply by adding predictors to the model.

- For example, if we have $n = 2$ observations, a simple linear regression of y on x will result in $R^2 = 1$ even if x and y are not associated.

- To get R^2 to be equal to 1 all we need to do is fit a model with n parameters to a dataset of size n.

- Then, R^2 makes sense as a measure of goodness of fit only if n is a lot larger than k.

- We can "'penalize"' the R^2 every time we add a new predictor to the model. The penalized R^2 is called the *adjusted* R^2 and it is sometimes more useful than the plain R^2.

The *adjusted* R^2 (cont'd)

- The adjusted R^2 is denoted R_a^2 and is computed as:

$$
\begin{aligned}
R_a^2 &= 1 - \left[\frac{(n-1)}{n-(k+1)}\right]\left(\frac{SSE}{SS_{yy}}\right) \\
&= 1 - \left[\frac{(n-1)}{n-(k+1)}\right](1-R^2).
\end{aligned}
$$

The *adjusted* R^2 (cont'd)

- Note that
  - As k increases, $n-(k+1)$ increases and SSE decreases.
  - If new predictor contributes information about y, SSE will decrease faster than the increase in $n-(k+1)$, so R_a^2 will increase. Else, R_a^2 will *decrease* when we add a new predictor to the model. [The ordinary R^2 always increases with added predictors even if new predictors contribute no information about y.]

- R_a^2 "'adjusts"' for sample size and number of predictors k and cannot be forced to be equal to 1.

- $R_a^2 < R^2$ always.