INFERENCE FOR MULTIPLE LINEAR REGRESSION

**Terminology:** Similar to terminology for simple linear regression

- \( \hat{y}_i = \hat{\eta}_T u_i \) (\( i^{th} \) fitted value or \( i^{th} \) fit)
- \( \hat{e}_i = y_i - \hat{y}_i \) (\( i^{th} \) residual)
- \( \text{RSS} = \text{RSS}(\hat{\eta}_T) = \sum (y_i - \hat{y}_i)^2 = \sum \hat{e}_i^2 \) (residual sum of squares)

**Results similar to those in simple linear regression:**

- \( \hat{\eta}_j \) is an unbiased estimator of \( \eta_j \).
- \( \hat{\sigma}^2 = \frac{1}{n-k} \text{RSS} \) is an unbiased estimator of \( \sigma^2 \).
- \( \hat{\sigma}^2 \) is a multiple of a \( \chi^2 \) distribution with \( n-k \) degrees of freedom -- so we say \( \hat{\sigma}^2 \) and \( \text{RSS} \) have \( \text{df} = n-k \).

*Note:* In simple regression, \( k = 2 \).

Example: Haystacks

**Additional Assumptions Needed for Inference:**

1. \( Y \mid x \) is normally distributed
   (Recall that this will be the case if \( X,Y \) are multivariate normal.)
2. The \( y_i \)'s are independent observations from the \( Y \mid x_i \)'s.

**Consequences of Assumptions (1) - (4) for Inference for Coefficients:**

- \( Y \mid x \sim N(\eta^T u, \sigma^2) \)
- There is a formula for s.e. (\( \hat{\eta}_j \)). (We'll use software to calculate it.)
- \( \frac{\hat{\eta}_j - \eta_j}{\text{s.e.}(\hat{\eta}_j)} \sim t(n-k) \) for each \( j \).

Example: Haystacks

**Inference for Means:**

In simple regression, we saw

\[
\text{Var} \left( \hat{\mu}(Y \mid x) \right) = \text{Var}(\hat{\mu}(Y \mid x) \mid x_i, \ldots, x_n) = \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right).
\]

So

\[
\text{s.e.} \left( \hat{\mu}(Y \mid x) \right) = \hat{\sigma} \sqrt{ \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} } = \hat{\sigma} \text{ times a function of } x \text{ and the } x_i \text{'s (but not the } y_i \text{'s)}
\]
An analogous computation (best done by matrices -- see Section 7.9) in the multiple regression model gives

$$\text{Var}(\hat{E}(Y|x)) = \text{Var}(\hat{E}(Y|x)|x_1, \ldots, x_n) = h \sigma^2,$$

where $h = h(u)$ (= $h(x)$ by abuse of notation) is a function of $u_1, u_2, \ldots, u_n$, called the leverage. (The name will be explained later.)

In simple regression,

$$h(x) = \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX}.$$  

Note that $(x - \bar{x})^2$ (hence $h(x)$) is a (non-linear) measure of the distance from $x$ to $\bar{x}$. Similarly, in multiple regression, $h(x)$ is a type of measure of the distance from $u$ to the centroid

$$\bar{u} = \begin{bmatrix} 1 \\ u_1 \\ \vdots \\ u_{k-1} \end{bmatrix},$$

(that is, it is a monotone function of $\sum (u_j - \bar{u}_j)^2$.) In particular:

The further $u$ is from $\bar{u}$, the larger $\text{Var}(\hat{E}(Y|x))$ is, so the less precisely we can estimate $E(Y|x)$ or $y$. (Thus an outlier could give a large $h$, and hence make inference less precise.)

Example: 1 predictor

Define:

$$\text{s.e.}(\hat{E}(Y|x)) = \hat{\sigma} \sqrt{h(u)}$$

Summarizing:

- The larger the leverage, the larger s.e. $(\hat{E}(Y|x))$ is, so the less precisely we can estimate $E(Y|x)$.
- The leverage depends just on the $x_i$'s, not on the $y_i$'s.

Similarly to simple regression:

$$\frac{\hat{E}(Y|x) - E(Y|x)}{\text{s.e.}(\hat{E}(Y|x))} \sim t(n-k).$$

Thus we can do hypothesis tests and find confidence intervals for the conditional mean response $E(Y|x)$. 
**Prediction**: Results are similar to simple regression:

- Prediction error = $Y|\mathbf{x} - \hat{E}(Y|\mathbf{x})$
- $\text{Var}(Y|\mathbf{x} - \hat{E}(Y|\mathbf{x})) = \sigma^2(1 + h(\mathbf{u})) = \sigma^2 + \text{Var}(E(Y|\mathbf{x}))$
- Define s.e. $(Y_{\text{pred}}|\mathbf{x}) = \hat{\sigma}\sqrt{1 + h}$
- $\frac{Y|\mathbf{x} - \hat{E}(Y|\mathbf{x})}{se(y_{\text{pred}} | \mathbf{x})} \sim t(n-k)$, so we can form prediction intervals.

Example: Haystacks