We have now done inference on $\beta$ (and $\alpha$), and the prediction of $\gamma$.

What about multiple regression (i.e., multiple $x$'s and $\beta$'s)?

In going from $\hat{y} = \hat{\alpha} + \hat{\beta} x$ (1+1 parm) \# of $\beta$'s.

to $\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$ (k+1 params)

things generalize in a straightforward way.

Basically, all that happens is $df= n-2 \rightarrow df= n- (k+1)$

This happens everywhere, e.g.,

1) The estimate of $\sigma^2_e$ is $\hat{\sigma}_e^2 = \frac{SSE}{n-(k+1)}$

2) The df associated with $t$-test changes: $n-2 \rightarrow n- (k+1)$

Finally, don't forget that the issues of collinearity and interaction all come back again when doing multiple regression, and nonlinearity.

But, the presence of multiple $\beta$'s allows for 2 more tests:

1) $H_0: \beta_i = 0 \quad H_1: \beta_i \neq 0$

Is the $i$th predictor useful?

2) $\begin{cases} H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_a: \text{At least 1 } \beta_i \neq 0 \end{cases}$

Ave any of the predictors useful?

(Teest of "model utility")

In $\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$, if all $\beta_i = 0$, then none of the predictors are useful for predicting $\gamma$.\)
**Good News:** The test for each $\beta_i$ is the same as the t-test for a single $\beta$, except for $df = n-(k+1)$

E.g. suppose we want to test $\beta_3$:

$H_0: \beta_3 = \beta_0$ (e.g. 0)
$H_1: \beta_3 \neq \beta_0$

$t_{obs} = \frac{\hat{\beta}_3 - \beta_0}{se/\sqrt{s_{xx}}}$

C.I. for $\beta_3$: $\hat{\beta}_3 \pm t^{\frac{SS}{n-(k+1)}}$

$p$-value = (1/2) pr ($t \geq t_{obs}$)

$df = n-(k+1)$

Table VI

Technically, in multiple regression $se$ is not $se/\sqrt{s_{xx}}$. The denominator ending up being a more complicated function of $x$'s. But when the predictors are completely uncorrelated, then this formula is ok.

**Note:** even though we are testing ONE $\beta_i$, the $df$ is $n-(k+1)$

**Bad News:** If you test each of the $\beta_i$ separately, you will make many more Type I errors than $\alpha\%$ of the time.

Consider 3 $\beta$'s: $\beta_1$, $\beta_2$, $\beta_3$

Type I errors: $e_1$ (if $\beta_0 \neq 0$) $e_2$ (if $\beta_2 \neq 0$) $e_3$ (if $\beta_3 \neq 0$)

You may commit the errors $e_1$ or $e_2$ or $e_3$

or ($e_1$ and $e_2$) or ($e_1$ and $e_3$) or ($e_2$ and $e_3$) or ($e_1$ and $e_2$ and $e_3$)

It can be shown that the prob of making at least 1 Type I error approaches 1 as the number of tests increases.
**Good News**: Enter the test of model utility!

Tam. \[ F = \frac{R^2/(k)}{(1-R^2)/(n-(k+1))} \sim F\text{-distribution with } \text{df} = (k, n-(k+1)) \]

\[ p\text{-value} = p(F < F_{obs}) \]

Just like in 1-way ANOVA where \( H_0: \text{At least...} \)

Then, if \( p\text{-value} < \alpha \), we can reject \( H_0 (\beta_1 = \beta_2 = \cdots = \beta_k = 0) \) in favor of \( H_1 \) (at least 1 \( \beta_i \) is not zero).

This F-test allows you to do ONE test to find out if any of the predictors are useful for predicting \( y \). This is very useful if \( k \) is large, because it tells you if any of the predictors are useful. I.e. it tells you if there is a "needle in the haystack," to begin with!

**IF** you get a significant result (i.e. \( p\text{-value} < \alpha \)) from the test of model utility, **THEN** there is evidence that at least one of the predictors is useful. **THEN** you can do separate tests on each of the \( \beta \)'s to see which predictors are useful. (see next page).

But **IF** the F-test comes back as non-significant, **THEN** there is no evidence that any of the predictors are useful. **THEN**, you don't have to test each predictor separately. This will not only save time, but more importantly, it will save you from the danger of making multiple Type I errors (i.e. declaring some predictor as useful, when in fact, it is not).
Recall that

- "bad" things happen if you keep adding terms to a regression model. Specifically, overfitting happens.
- Overfitting is not a black and white thing - it happens gradually, and in degrees, as you add more terms.
- Even a complete "garbage" term can lead to overfitting.

What happens to $F$ (and its $p$-value)?

More terms $\rightarrow$ higher $R^2$ $\rightarrow$ higher $F$ $\rightarrow$ lower $p$-value.

I.e. If you keep throwing enough predictors into a model (regression or otherwise), the $F$-test of model utility will find at least 1 useful predictor, regardless of whether or not the predictors are actually useful.

So, you must be thoughtful about adding terms to regression.

The $k$-dependence of the formula for $F$ does complicate things a bit but you can ignore it, because the real problem arises from $R^2$ approaching 1, as the # of predictors increases.

Still, we can pay attention to the $k$-dependence:

$$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))} = \frac{R^2}{1-R^2} \left( \frac{n-(k+1)}{k} \right) = \frac{R^2}{1-R^2} \left( \frac{n-1}{k} - 1 \right)$$

Now, technically $k$ must be less than $(n-1)$, otherwise $F < 0$, which it cannot be. So, $k < n-1$, in which case the largest allowed value of $k$ is $n-2$, and so $\frac{n-1}{n-2}$ is at most $\frac{n-1}{n-2}$, i.e. a constant! Then, we're back to looking at how $R^2$ grows.
The regression equation is
\[ \text{durpr} = -0.912 + 0.161 \text{formconc} + 0.220 \text{catratio} + 0.0112 \text{temp} + 0.102 \text{time} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.9122</td>
<td>0.8755</td>
<td>-1.04</td>
<td>0.307</td>
</tr>
<tr>
<td>formconc</td>
<td>0.16073</td>
<td>0.06617</td>
<td>2.43</td>
<td>0.023</td>
</tr>
<tr>
<td>catratio</td>
<td>0.21978</td>
<td>0.03406</td>
<td>6.45</td>
<td>0.000</td>
</tr>
<tr>
<td>temp</td>
<td>0.011226</td>
<td>0.004973</td>
<td>2.26</td>
<td>0.033</td>
</tr>
<tr>
<td>time</td>
<td>0.10197</td>
<td>0.05874</td>
<td>1.74</td>
<td>0.095</td>
</tr>
<tr>
<td>S = 0.8365 R-Sq = 69.2% R-Sq(adj) = 64.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>39.3769</td>
<td>9.8442</td>
<td>14.07</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>25</td>
<td>17.4951</td>
<td>0.6998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>56.8720</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n - (k+1) = n-1 a) Is the model useful? \( F \)-test:

\[ F_{0.05} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \]

\[ = \frac{0.692}{4} \]

\[ = 14.04 \]

\[ p \text{-value} = \text{prob}(F > F_{0.05}) = \text{prob}(F > 14.04) < 0.001 \]

According to Table VIII, df = (4,25)

\[ \therefore \text{At any reasonable } \alpha, \text{ we can reject } H_0 (\text{that all } \beta_i = 0) \text{ in favor of } H_1 (\text{that at least 1 of the } \beta_i \neq 0). \text{ I.e. The model is useful.} \]
b) Estimate, in a way that conveys information about precision & reliability, the average change in durability press rating associated with a 1-degree increase in curing temperature, when all other predictors remain fixed. (If there is no collinearity)

I.e. what's the CI for $\beta_{\text{temp}}$; $t^{*} = \frac{25}{\sqrt{8.365}}$

95% CI: $\hat{\beta} \pm t_{*} \cdot \text{se} = 0.012 \pm 2.060 \cdot 0.8365$

2-sided $t_{*}$ temp. $\pm \text{se in } \hat{\beta}$ not given!

$0.012 \pm 2.060 \cdot 0.8365 \Rightarrow (-0.001, 0.021)$

This is the interval estimate of $\beta_{\text{temp}}$. It's useful as it is, but we can also see that $\beta_{\text{temp}} = 0$

We can build the CI for all the other $\beta_i$:

C.I. for conc: $0.1607 \pm 2.060 \cdot 0.6617 = (0.02, 0.30)$
C.I. for cat ratio: $0.2178 \pm 0.3406 = (0.15, 0.29)$
C.I. for temp: $0.012 \pm 0.00497 = (0.001, 0.02)$
C.I. for time: $0.10197 \pm 0.0587 = (-0.02, 0.22)$

Note that 3 $\beta_i$s are non-zero.
In part a, we found out that at least one of the $\beta_i \neq 0$. To see which one(s), we test each of them:

$H_0: \beta_i = 0$ vs. $H_i: \beta_i \neq 0$ for each $i$.

c) $E.g.$ $H_0: \beta_{\text{formald.}} = 0$

$H_1: \beta_{\text{formald.}} \neq 0$

$t_{\text{stat}} = \frac{0.6073 - 0}{0.0617} = 2.43$ (check the output!)

de even though testing 1 $\beta$, $df = n - (i + 1) = 25$

$p\text{-value} = 2 \cdot \text{prob}(t > t_{\text{stat}}) = 2(0.012) = 0.024$ (check output!)

So, $p\text{-value} < \alpha \Rightarrow \text{formaldehyde provides useful info}$.

In fact, look at all the $p\text{-values}$:

0.023, 0.000, 0.33, 0.95 (last col. of printout).

At $\alpha = 0.05$, $\beta \neq 0$ formal. etc. tend to tell time

consistent with the conclusions in part b.

Note these $p\text{-values}$ are different from what you would get if you did $y = \alpha + \beta_1 x_1$, $y = \alpha + \beta_2 x_2$, etc., and tested if each of these $\beta_i$ are zero. The multiple regression model is more correct because it does take into account the correlations between predictors. See ch. 3 lects.
We have seen that adding useless predictors to a regression model will increase R². Here, let's examine what our inference methods say if the predictors are in fact useless. Suppose the true/pop fit is \( y = 1 \), (i.e., no \( x \) at all), and so a possible sample from the population could be the following:

\[
\text{set.seed(123)     # Use this line to make sure we all get the same answers.}
\]
\[
n = 20
\]
\[
y = 1 + \text{rnorm(n,0,1)}
\]

a) Write code to make data on 10 useless predictors (and no useful predictors) each from \( \text{unif(-1,+1)} \), fit the model \( y = \alpha + \beta_1 x_1 + \ldots + \beta_{10} x_{10} \), perform the test of model utility, and perform t-tests on each of the 10 coefficients to see if they are zero. Show/turn-in your R code.

b) According to the F-test of model utility, are any of the predictors useful at alpha = 0.1? 

c) According to the t-tests, are any of the predictors useful at alpha = 0.1? See the solns to make sure you understand the moral of this exercise.

The F-ratio appearing in the test of model utility depends on \( R^2 \) of the model. So, if you know the critical value of F (as in part a), then you have a critical value of \( R^2 \).

b) Find the critical value of \( R^2 \) (above which p-value < \( \alpha \)). Moral: The test of model utility can be done by comparing \( R^2_{obs} \) with this critical value, i.e. without p-values.
Optional

Consider a multiple regression problem with \( k \) terms on the right-hand side. Suppose all of the \( k \) predictors are useless. But of course we don't know that, so we test each of the \( \beta \)'s individually. Our hypothesis testing formalism assures that each test has prob. \( \alpha \) of finding the predictor useful (when in fact it's useless).

a) What's the prob. of finding \( j \) useful predictors out of \( k \) predictors?

Hint: You should be able to see a familiar string of words here!

b) What's the prob. that at least 1 of the \( k \) predictors will be found to be useful (when it's not)?

(Make sure you check the soln, later.)