Lecture 26 (Ch. 11)

We have been talking about inference on β (and α), the conditional mean of y, given x, and a single y, at x.

What about multiple regression (i.e. multiple x's)?

In going from $y = \alpha + \beta x$ (1+1 parm)

to $Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ (k+1 params) # of β's.

things generalize in a straight forward way.

Basically, all That happens is $(n-2) \longrightarrow n-(k+1)$.

This happens every where, e.g.

1) The estimate of σ_ϵ^2 is $s_e^2 = \dfrac{SSE}{n-(k+1)}$

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

3) Even, CI & PI formulas are The same, except for

df $= n-(k+1)$

$$\hat{y} \pm t^* s_{\hat{y}} \; , \quad \hat{y} \pm t^* \sqrt{s_{\hat{y}}^2 + s_e^2} \qquad s_e^2 = \frac{SSE}{n-(k+1)}$$

In This class we did not develop a formula for $s_{\hat{y}}$ in multiple regression, but it does exist. As far as you are concerned, you can always get $s_{\hat{y}}$ from R.

Finally, don't forget That The issues of collinearity and interaction come back again when doing multiple regression.

The presense of multiple β's allows for one more (last!) test:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
H_1: At least 1 β_i is $\neq 0$ $\Big\}$ Test of "model utility".

In $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k$, if all $\beta_i = 0$, then none of the predictors are useful for prediction y.

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!

So, if you have a lot of possible/potential predictors, but you are not sure if any of them are useful, what you can do is to include all of them in the regression model, and do the F-test of model utility. $\boxed{\text{IF}}$ you get a significant result (ie. p-value $< \alpha$), then there is evidence that at least one of the predictors is useful. $\boxed{\text{THEN}}$ you can do separate tests on each of the β's to see which predictors are the useful ones. But $\boxed{\text{IF}}$ the F-test comes back as non-significant, then there is no evidence that any of the predictors are useful. $\boxed{\text{THEN}}$, you don't have to test each predictor separately.

The main Theorem for this F-test is:

Thm: $F = \dfrac{R^2/[k]}{(1-R^2)/[n-(k+1)]}$ (= "F-ratio")

has an F-distr. with $df = (k, n-(k+1))$

$\therefore$ p-value = Right area under F

Just like in Chi-squared, when H_1 is "at least ...", then the p-value is just the right tail.

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

$\boxed{\text{IF}}$ the p-value $< \alpha$, we can start testing each of the β_i separately. Fortunately, the test for each β_i is the same as the test for a single β, except for $df = n-(k+1)$

E.g. suppose we want to test β_3:

$H_0: \beta_3 \square \beta_0$ (e.g. 0)

$H_1: \beta_3 \square \beta_0$

$t_{obs} = \dfrac{\hat{\beta}_3 - \beta_0}{s_e/\sqrt{S_{x_3x_3}}}$

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

= ---

$df = n-(k+1)$ in Table VI (or IV)

C.I. for β_3:

$\hat{\beta}_3 \pm t^* \dfrac{s_e}{\sqrt{S_{x_3x_3}}}$

$s_e = \sqrt{\dfrac{SSE}{n-(k+1)}}$

$df = n-(k+1)$

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

$\boxed{\text{Q1:}}$

The F-ratio that shows-up in the test of model utility, i.e.

$F = \dfrac{R^2/k}{(1-R^2)/[n-(k+1)]}$, can be written in a way that resembles the ANOVA F-ratio of Ch. 9, i.e. $F = \dfrac{SS_{between}/\cdots}{SS_{within}/\cdots}$

In fact, the former is equal to

a) $\dfrac{SS_{between}/(k-1)}{SS_{within}/(n-k)}$ $\quad$ $\boxed{\text{b) } \dfrac{SS_{between}/k}{SS_{within}/[n-(k+1)]}}$ $\quad$ c) $\dfrac{R^2/(k-1)}{(1-R^2)/(n-k)}$

where $SS_{between} = SS_{explained}$, $SS_{within} = SS_{unexplained}$

$$F = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} = \frac{\left(\frac{SS_{expl.}}{SST}\right)/k}{\left(1-\frac{SS_{expl.}}{SST}\right)/[n-(k+1)]}$$

$$= \frac{SS_{expl.}/k}{SS_{unexp.}/[n-(k+1)]} = \frac{SS_{between}/k}{SS_{within}/[n-(k+1)]}$$

instead of $(k-1)$

instead of $(n-k)$

SS = Sum-Squared

$$= \frac{MS_{between}}{MS_{within}}$$

MS = Mean-Squared

11.66

```
The regression equation is
durpr = -0.912 + 0.161 formconc + 0.220 catratio + 0.0112 temp + 0.102 time

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

Analysis of Variance
Source        DF       SS        MS       F       P
Regression     4  39.3769    9.8442   14.07   0.000
Error         25  17.4951    0.6998
Total         29  56.8720
```

In This example you are going to learn what all These numbers mean!

Also $F = \dfrac{MS_{expl}}{MS_{Error}} = \dfrac{9.8442}{0.6998}$

a) Is The fitted model useful?

F-test: $F_{obs} = \dfrac{R^2/k}{(1-R^2)/(n-(k+1))} = \dfrac{\left(\frac{.692}{4}\right)}{\left(\frac{1-.692}{30-(4+1)}\right)} = 14.04$

From print out (Tricky!) $\Rightarrow$ $\boxed{29 = n-1}$

p-value = prob$(F > F_{obs})$ = prob$(F > 14.04) < .001$

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

So, at any reasonable α, we can reject The null hyp. (That all β_i are zero) in favor of The alt. (That at least one of The β_i is not zero)

I.e. The model is useful.

b) Estimate, in a way that conveys info 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 C.I for β_{temp} ; t^* at $df = n-(k+1) = 25$

95% C.I: $\hat{\beta} \pm t^* \underbrace{\dfrac{s_e}{\sqrt{S_{xx}}}}_{\text{std. err in } \hat{\beta}} = .0112 \pm 2.060 \dfrac{0.8365}{\sqrt{?}}$

2-sided

S_{xx} → temp.

? ← not given!

from printout.

$\therefore\ .0112 \pm 2.060\ (.004973) \Rightarrow \underline{(.001, .021)}$

This is The interval estimate for β. It's useful as it is, but we can also see That $\beta_{temp} \neq 0$. We can build The C.I. for The other β's:

C.I. form conc	.1607	± 2.060	(.06617)	=	(0.02, 0.30)
cat ratio	.2198	± "	.03406	=	(0.15, 0.29)
temp	.0112	± "	.00497	=	(0.001, 0.02)
time	.10197	± "	.0587	=	(−0.02, 0.22)

Note That 3 β's are non-zero.
At least 1 ✓

In part a, we found out that atleast one of the $\beta_i \neq 0$.
To see which one(s), we test each of them!

$H_0: \beta_i = 0$ vs. $H_1: \beta_i \neq 0$ for each i.

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

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

$t_{obs} = \dfrac{.16073 - 0}{\boxed{.06617}} = 2.43$ Same as output in book, even though testing 1 β.

$\dfrac{s_e}{\sqrt{S_{xx}}}$

formald. (from Table)

p-value $= 2\,\text{prob}(t > t_{obs}) = 2(.012)$ $\boxed{df = n-(k+1) = 25}$

$= .024 =$ in printout

So, p-value $< \alpha \implies$ formaldehyde provides useful info.

In fact, look at all the p-values:

.023 , .000 , .033 , .095 ← look at last Col. of printout.

At $\alpha = .05$: $\beta \neq 0$ (formald.), $\beta \neq 0$ (cat.), $\beta \neq 0$ (temp), β= cannot Tell (time)

$\underbrace{\qquad\qquad\qquad\qquad}_{}$ consistent with The conclusions in part b.

Note These p-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 β_i are zero.

hw_lect 26-1

Recall the hw problem (hw_lect14_1) where you showed that adding useless predictors to a regression model can increase R2. This time, 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:

```
set.seed(123)     # Use this line to make sure we all get the same answes.
n = 20
y = 1 + rnorm(n,0,1)
```

a) In the old hw, there was one useful predictor (x1), and 4 useless predictors (x2,x3,x4,x5). Here, revise that code to have data on 10 useless predictors(and no useful predictors), fit the model y = alpha + beta1 x1 + ... + beta10 x10, performs the test of model utility, and perform t-tests on each of the 10 coefficients to see if they are zero. Just show your code. By R

b) According to the F-test of model utilitt, 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.