

~~Wen~~ No office hours today.

New homework in Gradescope.

Remarks on Grading:

We will look at "Heard on the Street"
by T. Crack.

One piece of advice: ATQ!

Answer the Question!

If the question has a concrete answer, give
that answer at the outset.

Another remark on Grading:

There will be more questions on interpretation
of regression coefficients.

We'll look at some wrong answers to

5(b) on the midterm.

5 (b) :

(A) "If the car's weight increases by 1 pound, then the MPG will decrease by 6".

(B) "When the MPG goes up by 1, the weight must go down by 6 thousand pounds."

(C) "If the car weighs 1 thousand pounds, subtract 6 thousand pounds to produce the expected weight of the car."

(D) "The mpg goes up as the weight goes down."

(A) is wrong because the units are thousands of pounds and because the sentence implies a causal relationship.

③ Numbers are wildly wrong:

it is not possible to reduce the weight of a 3500 pound car by 6000 pounds.

Explanatory variable and response exchanged.

④ Makes no sense and does not mention the response MPG.

⑤ No numerical information from regression output.

None of these answers say "holding horsepower constant" or "among the cars with similar horsepower ratings".

Variations on this type of question:

Interpret regression output when the model is $\log(Y) \sim X_1 + X_2$.

or when one or more of $X_1, X_2, \dots$ is a categorical variable.

a categorical variable is one that takes on non-numeric values e.g

"Ford" "Toyota" "Rivian"

R will still do the regression.

What do the coefficients mean now?

Today: Ch. 6 - so you can do the homework.

The 4 standard diagnostic plots are obtained with `plot(lmod)`. (and `hit` return 4x).

#1: fitted vs residuals (most important)

any pattern you see here suggests an improvement to your model.

#2: Q-Q plot

Tests for normality of errors.

#3: Scale-location: $\sqrt{\text{Std resid}}$ vs. fitted.
Pattern suggests improvement.

#4: Std resid vs leverage or "Cook's Dist."
checks for ~~infl~~ influential points.

Regression ~~at~~ atpt may be wholly dependent
on only a few points (even if you have
10,000 rows of data!)