Handout 6b

STAT3530 Applied Linear Models

How should we assess and test model quality? Two approaches:

1. P_____
2. E_____

Prediction error

For a single new observation Y_0 at x_0, the prediction is

$$\hat{Y}_0 =$$

One measurement of accuracy is mean square prediction error, which decomposes as

$$\mathbb{E}[(Y_0 - \hat{Y}_0)^2] =$$

3. Interpret each term:
   a. $\text{var}(Y_0)$
   b. $\text{var}(\hat{Y}_0)$
   c. $\text{cov}(Y_0, \hat{Y}_0)$
   d. $\mathbb{E}[Y_0] - \mathbb{E}[\hat{Y}_0]$
4. What controls the last two terms?
   a. The bias term is zero when _____.
   b. The covariance term is zero when _____.
5. Suppose the model is correctly specified. Fill in both cases.

$$\mathbb{E}[(Y_0 - \hat{Y}_0)^2] = \begin{cases} \sigma^2 + \sigma^2 h_{00} - _____ = \sigma^2(_____) & \text{in sample} \\ \sigma^2 + \sigma^2 h_{00} \qquad\qquad = \sigma^2(_____) & \text{out of sample} \end{cases}$$

6. So in-sample prediction error is _____ than out-of-sample prediction error. Explain why.

Measuring prediction error in practice

```
# in sample: every cheese helped fit the model that predicts it
mean(resid(full)^2)
```

```
[1] 88.94704
```

```
# out of sample: predict each cheese from a model fit without it
loo <- sapply(1:30, function(i)
  predict(lm(taste ~ ., data = cheddar[-i, ]), cheddar[i, ]))
mean((cheddar$taste - loo)^2)
```

```
[1] 113.408
```

7. Using the output above:

   a. The ratio of out-of-sample average to in-sample average is about _____.

   b. For this model $p = 4$ and $n = 30$, so the average leverage is $\bar{h} = p/n =$ _____. Notice that

   $$\frac{1+\bar{h}}{1-\bar{h}} =$$

   and compare this to the ratio you just computed.

8. Leaving one observation out does not require refitting the model:

   $$Y_i - \hat{Y}_{i(i)} =$$

```
mspe <- mean((resid(full) / (1 - hatvalues(full)))^2)

c(mspe = mspe, rmspe = sqrt(mspe))
```

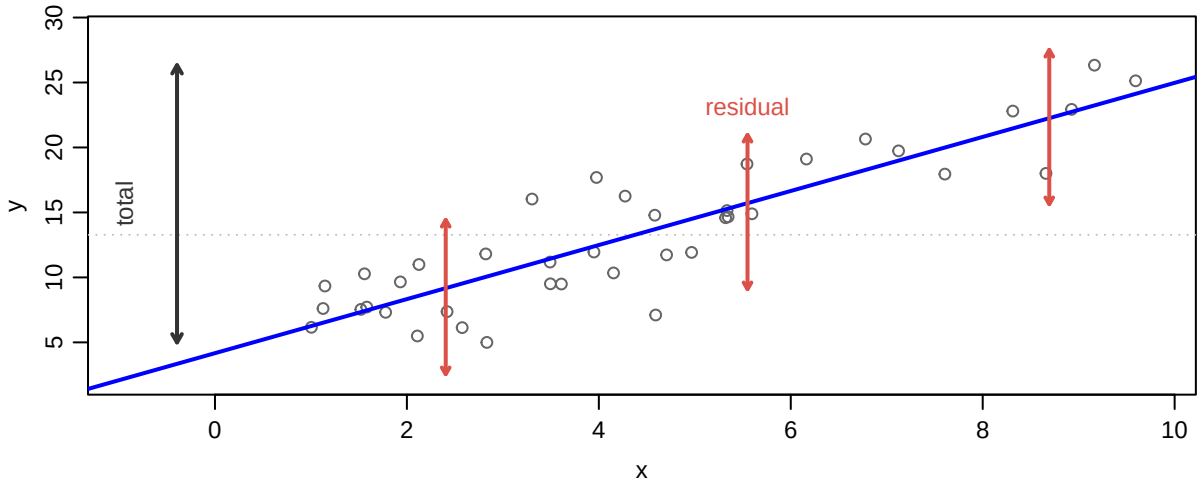
```
     mspe     rmspe
113.40802  10.64932
```

9. Using the output above:

   a. Why take a square root before interpreting?

   b. Interpret the result in plain language.

Measuring fit



10. How would you measure the labeled parts of the picture?

    a. Total variation:

    b. Residual variation:

    c. Interpret the difference:

    d. Decompose the total variation into the residual and the difference:

$$\sum_i (Y_i - \bar{Y})^2 =$$

Label each term in the decomposition.

    e. The share of total variation explained by the model is:

$$R^2 =$$

    f. This is monotonic in the number of predictors. An adjusted version is given by:

$$R^2_{\text{adj}} =$$

Compare R^2 and R^2_{adj} if a noise variable is added to the cheddar model:

```
set.seed(92926)
extra <- update(full, ~ . + rnorm(n = 30, mean = 0, sd = 4))

fit.stats <- function(m) c(R2 = summary(m)$r.squared,
                           adjusted = summary(m)$adj.r.squared)

round(cbind(original = fit.stats(full), with.noise = fit.stats(extra)), 4)
```

```
         original with.noise
R2         0.6518     0.6540
adjusted   0.6116     0.5987
```

The omnibus F test

The idea: test whether the model explains a *significant* amount of variation.

11. Framing. Suppose the model explains no variation: $\mathbb{E}[SSR/(p-1)] = \sigma^2$. We can understand this in terms of the collection of all coefficients.

    a. If no predictor is associated with the response, the coefficients are all zero. Express this as a testable hypothesis and specify the alternative.

    H_0 : vs H_A :

    b. Under H_0, the model collapses to

    $\mathbb{E}[Y] =$

    and every fitted value becomes _____.

    c. If H_0 were true, what would you expect to see when you compare SSR to SSE?

    d. What would you expect if H_0 were false?

12. The test.

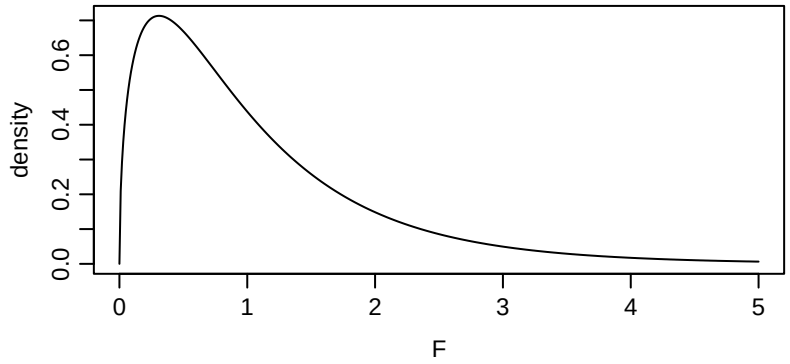
    a. Each sum of squares carries its own *degrees of freedom.*

    SST : _____ SSR : _____ SSE : _____

    b. Compare the variation explained by the model to the error variation as a ratio adjusted for degrees of freedom and identify the sampling distribution under H_0 and error normality:

    $F =$ $\sim F_{__,__}$

c. Below is the F distribution for the cheddar model. Shade the region corresponding to a p-value for $F = 2.5$. Which values of F count as evidence against H_0 and why?



13. Perspectives. The same test can be described three equivalent ways. Fill in what H_0 claims under each.

	in terms of…	H_0 says
1.	coefficient values	
2.	explanatory significance	
3.	model comparison	

14. Why is this *not* the same as running a t test on each coefficient and seeing whether any is significant?

Here is the cheddar model again.

```
summary(full)
```

```
Call:
lm(formula = taste ~ Acetic + H2S + Lactic, data = cheddar)

Residuals:
    Min      1Q  Median      3Q     Max
-17.390  -6.612  -1.009   4.908  25.449

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
Acetic        0.3277     4.4598   0.073  0.94198
H2S           3.9118     1.2484   3.133  0.00425 **
Lactic       19.6705     8.6291   2.280  0.03108 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared:  0.6518,    Adjusted R-squared:  0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

15. Using the output above:

    a. Locate the statistic, both degrees of freedom, and the p-value. Check the degrees of freedom against your answer to 12a.

    b. The statistic can also be written in terms of R^2. Verify the reported value using $R^2 = 0.6518$:

$$F = \frac{n-p}{p-1} \cdot \frac{R^2}{1-R^2} =$$

    c. State the conclusion in context.

A discrepancy

```
summary(seat)
```

```
Call:
lm(formula = hipcenter ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.827 -22.833  -3.678  25.017  62.337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared:  0.6866,    Adjusted R-squared:  0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05
```

16. This is the car seat data: 38 drivers, eight body measurements.

    a. What does the omnibus test say? What do the individual t tests say?

    b. Explain how both can be true at once.

    c. What is this a symptom of, and what would you do about it?