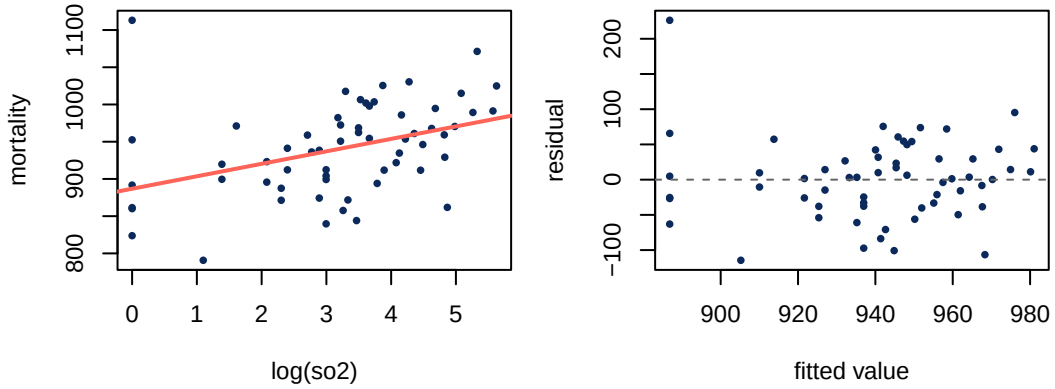
Handout 10

STAT3530 Applied Linear Models

This week we will cover variable selection: choosing which predictors belong in a model when there are too many candidates to consider them all. The procedures are easy. The difficulty is what they do to the inference afterwards.

Does sulfur dioxide pollution increase mortality?

Age-adjusted mortality and air quality in 60 US metropolitan areas, 1959–61. The response `mortality` is deaths per 100,000 residents and `so2` is a relative pollution potential for sulfur dioxide. Fourteen other variables record climate, population composition, and two further pollutants. All three pollutants are strongly right-skewed, so we work with them on the log scale.




```
round(summary(fit.slr)$coefficients, 4)
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 886.8237    17.5944 50.4037   0.0000
lso2         16.7436     4.9909  3.3548   0.0014
```

1. If `lnox` is included in the model, what is the chance of finding a significant association, assuming all six-predictor models are equally likely to be selected?

2. There are fourteen additional predictors available. How many models are possible with…
   a. Adjusting for exactly six of the fourteen?
   b. Adjusting for *any* subset of the fourteen?
   c. If two-way interactions are included in each of (a) and (b)?

Model selection

Parametric inference relies on the sampling distribution

$$\hat{\beta} \sim N(_____,\ _____)$$

If instead we *select* a model M from a set of candidates $\mathcal{M}$, then X is no longer fixed and the distribution is conditional on the selected model:

$$\hat{\beta} \mid M \ \sim \ N(_____,\ _____)$$

3. Two analysts select different models from $\mathcal{M}$ and each reports an estimate of "the association with sulfur dioxide." These are shown below. Are the estimands the same for each model? Why or why not?

$$M_1: \quad \widehat{\texttt{mort}} = 745.80 + 19.60\,\texttt{lso2} + 3.53\,\texttt{precip}$$
$$M_2: \quad \widehat{\texttt{mort}} = 707.30 + \ \ 7.88\,\texttt{lso2} + 4.24\,\texttt{precip} + 21.21\,\texttt{lnox}$$

4. A model is **confirmatory** if

5. A model is **exploratory** if

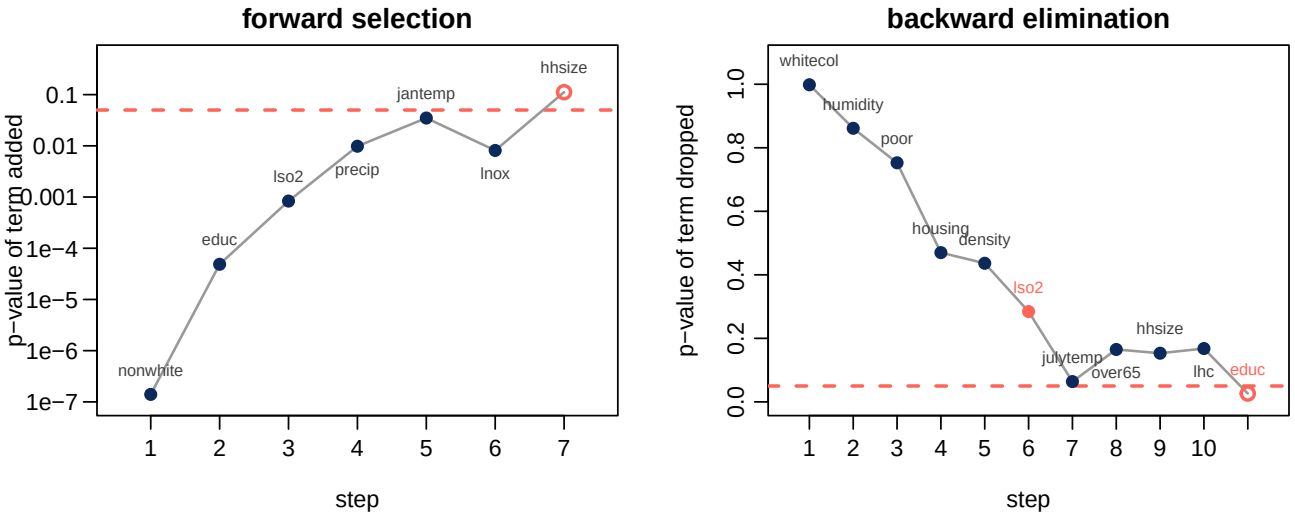
6. Confirmatory or exploratory? Classify each example below.
   a. Test for an interaction and include it only if the test is significant.
   b. Drop a predictor whose p-value is large.
   c. Log-transform a predictor after checking residual diagnostics.
   d. Adjust for income after noticing it changes the estimates.

Stepwise procedures are "classical" model selection algorithms that add or remove variables based on sequential F tests using a significance threshold as a stopping rule.

7. **Forward selection** begins with an intercept-only model and adds one predictor at a time. Write the algorithm.

8. **Backward elimination** begins with the largest candidate model and removes one predictor at a time. Write the algorithm.

The output below shows the algorithms run on the mortality data. Each point is the p-value of the predictor entering (left) or leaving (right) at that step. The dashed line is the threshold $\alpha = 0.05$.



9. Write the model selected by each algorithm, and describe how they differ.

forward : $\texttt{mort}_i =$

backward : $\texttt{mort}_i =$

What criteria should we use to decide between them?

Model criteria

Two standard criteria for comparing models are Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both transform the model errors and penalize for model complexity.

10. Write the formula for **AIC** in terms of n, RSS and the number of estimated coefficients p.

AIC $=$

11. Write the formula for **BIC**.

BIC $=$

`summary()` reports $\hat{\sigma}$, the residual standard error, rather than RSS; we can compute AIC and BIC from this estimate.

12. Write RSS/n in terms of $\hat{\sigma}^2$, n and p.

$$\frac{\text{RSS}}{n} =$$

13. Substitute to rewrite both criteria in terms of $\hat{\sigma}$.

$$\text{AIC} =$$

$$\text{BIC} =$$

14. The forward model has $p = 7$ coefficients and the output below. Compute its AIC.

```
sigma(fit.sel.fwd)
```

```
[1] 33.13098
```

$$\text{AIC} =$$

15. AIC and BIC prioritize different objectives.

    a. AIC optimizes for…

    b. BIC optimizes for…