Chapter 5: Stat Inference with Simple Models

DATA = MODEL + ERROR

In this chapter, models are simple, mean estimates the model. So when we are asking questions about a simple model, we are asking about the value of a mean. one-sample t-test.

$$\text{MODEL C: } Y_i = B_0 + \varepsilon_i$$

$$\text{MODEL A: } Y_i = \beta_0 + \varepsilon_i$$

We know how to calculate estimate b0, calculate SSE's, calculate PRE. Remaining question:

When is PRE big enough to warrant rejecting C in favor of A?

In this chapt we will answer that question definitively.

Example:

teacher of new math curric
15 students
principle wants evidence new scores beat old norm of 65

$$\text{MODEL C: } Y_i = 65 + \varepsilon_i$$

$$\text{MODEL A: } Y_i = \beta_0 + \varepsilon_i$$

Note: Can ALWAYS write down models BEFORE looking at DATA. DO SO!

Now, look at descriptive stats of data
Ex. 5.1, p. 73

MODEL C: $\hat{Y}_i = 65$

MODEL A: $\hat{Y}_i = 78.1$

$$SSE(C) = \sum_{i=1}^{n} (Y_i - 65)^2$$

$$SSE(A) = \sum_{i=1}^{n} (Y_i - 78.1)^2$$

See calculations in Ex. 5.2, p. 74
Note a couple of BIG errors—watch out!
Can get from USS and CSS, but won't notice BIG errors

$$PRE = \frac{SSE(C) - SSE(A)}{SSE(C)} = \frac{10,403 - 7815.7}{10,403} = .25$$

That is, MODEL A using the mean to estimate the unknown parameter has 25% less error than MODEL C, using the old parameter value. Is this big enough?

Digression: ANOVA decomposition

Source	SS	PRE
Reduce, Mod A	SSR	SSR/SSE(C)
Error for Mod A	SSE(A)	
Total	SSE(C)	

Source	SS	PRE
Reduce, Beta0	2,587	.25
Error for Mod A	7,816	
Total	10,403	

$$SSR = SSE(C) - SSE(A)$$

$$SSR = \sum_{i-1}^{n} (\hat{Y}_{iC} - \hat{Y}_{iA})^2$$

Explain conceptual importance of this formula
The more different the predictions the greater SSR and the greater PRE. MODEL A must make different predictions from MODEL C to be interesting!

$$SSR = \sum_{i-1}^{n} (65 - 78.133)^2 = \sum_{i=1}^{15} (-13.133)^2 = 2{,}587.1$$

Sampling Distribution of PRE
now we will do it! is PRE=.25 "big enough?"

Start with MODEL C

$$\text{MODEL C: } Y_i = 65 + \varepsilon_i$$

Assume it is true. Is it possible for the errors to be such that we COULD get a mean of 78.1? Is that likely? Is it likely that with that MODEL C we could get a PRE = .25? That is our question!

Just as won't get Mean = 65 every time, won't get PRE = 0 every time. In fact, will on average get a PRE bigger than 0 because MODEL A will be a little bit better. If MODEL C is really correct, then the TRUE proportional reduction in error is

$$\eta^2 = 0$$

Note: this means PRE is biased. We will deal with that later.

Let's assume PRE=0, i.e., MODEL C is correct and play our Nature simulation game again.

What should go in bag of error tickets?
We now know typical errors.

$s^2 = 7815.7/14 = 558.3$

Use this info to put tickets in bag. assume normal dist, independent, identically distributed.

We already did this in Chapter 4!
tickets were

-9 -37 -17 22 45
....

and resulting values were

56 28 48 87 110
....

mean = 63
SSE(C) = 8048
SSE(A) = 7988

PRE = (8048-7988)/8048 = .0075

That is Round 1
Now let's do this over and over to get sampling dist.

See Ex. 5.5 and 5.6, pp. 80-81

Can now answer question! Is PRE big enough to reject Model C. Rule: If PRE would is surprising, i.e., if PRE would occur less than 5% of time by chance, then reject Model C.

In this case, reject MODEL C because p(PRE>=.25) = .95

Critical Values
- **really interested in those cutpoints at .05.**
- **it will be different for other values of n, PC, and PA**
- **don't want big table like Ex 5.5 every time.**

  - **Appendix tables for .05 and .01 surprise**

F*

PRE tables are rare (at least for now!)

derive

PRE/additional parameter = PRE/(PA-PC)

1-PRE/parameters that *could* be added = (1-PRE)/(n-PA)

$$F^* = \frac{\frac{PRE}{PA-PC}}{\frac{1-PRE}{n-PA}} = \frac{\frac{.25}{1}}{\frac{.75}{14}} = 4.66$$

explain in terms of junk params

Sampling distribution for F* has same info as sampling distribution for PRE. If one rejects MODEL C, so will the other!

critical values of F*, Ex 5.8

note for .95 about 4 times better than junk

F* vs F distribution

If assumptions met, then F* has F distribution

look at more general tables in Appendix

Alternative F* formula
ubiquitous so we have to learn it

$$F^* = \frac{SSR/(PA-PC)}{SSE(A)/(n-PA)} = \frac{MSR}{MSE}$$

Source	SS	df	MS	F*	PRE	p
Reduce	SSR	PA-PC	MSR	MSR/MSE	SSR	
Error	SSE(A)	n-PA	MSE		SSE(C)	
Total	SSE(C)	n-PC				

Source	SS	df	MS	F*	PRE	p
Reduce	2587	1	2587	4.66	.25	.05
Error	7816	14	558			
Total	10403	15				

Statistical Decisions

all stat decisions are fallible!

		True State Model C Correct	Model C wrong
Statistical Decision	"Reject C"	Type I Error	Correct Decision
	"Don't reject C"	Correct Decision	Type II Error

dollar bill changer
reject good bill---Type I error alpha
accept bad bill----Type II error beta
reject bad bill---OK
accept good bill---OK

p(don't make Type II error) = power = 1-beta

should pay attention to relative costs
can slide criteria to change probabilities of two types of error
but usually select alpha and let beta be whatever it turns out to be

Estimating Statistical Power
don't want to do studies where we can't see anything! must have adequate power

got alpha assuming MODEL C is Correct
now want to assume MODEL C is INCorrect
how do we do that?
assume MODEL A is correct, but what does that mean?
need to specify that MODEL A is correct so that there is a true PRE of a certain amount

So do "What IF" analysis. What if true PRE is xxxx? What would our power be?

play sampling game again with different values of True PRE! distributions of PRE as function of TruePRE
Ex. 5.12, p. 90

Coordinate this info with critical values for specific alphas,
Ex. 5.13
then Appendix C

So how do we get values of True PRE to play the what-if game?

1. **prior studies in same domain**
   correction for bias

$$\hat{\eta}^2 = 1 - (1 - PRE)\left[\frac{n - PC}{n - PA}\right]$$

$$\hat{\eta}^2 = 1 - (1 - .25)\left[\frac{15 - 0}{14 - 1}\right] = .20$$

2. **Cohen**

"small"	**.02**	**[.03]**
"medium"	**.13**	**[.1]**
"large"	**.26**	**[.3]**

do for math curriculum problem

small power = .09
medium power = .21
large power = .63

3. **Parameter guesses**
   guess beta0, SSE(C), and calculate PRE
   see example in book, not common enough

Improving Power

turn up power on microscope

1. Reducing ERROR

DATA = MODEL + ERROR

a. better DATA (measured with less ERROR)

b. better MODEL (more variables in MODEL)

e.g., teacher student had last year

2. Increasing alpha.

more type I errors but fewer type II errors

tradeoff

policy importance

costs of Type I error low

Ex 5.14, p. 98, for math curric

3. Incresing *n*

See Ex. 5.15, p. 99

but

a. may be infeasible due to cost

b. find trivial effects

Confidence Intervals

CI = set of possible values for B_0 for which MODEL C would NOT be rejected

e.g., mean = 78.1, if used that for B_0 would not reject C! so 78.1 IS in CI

but rejected MODEL C for B_0=65 so 65 is NOT in CI

how about B_0=70?

SSE(C) = 8808

SSE(A) = 7815.7 [as before!]

PRE = .11 < crit value of .247

SO 70 IS in CI

could continue this game but there is a formula

$$b_0 \pm \sqrt{\frac{F_{1,n-1;\alpha} MSE}{n}}$$

$$78.1 \pm \sqrt{\frac{(.4.\,6)558.\,3}{15}} \quad \text{or} \quad 78.1 \pm 13.08$$

This is 95% CI of [65.02,91.18]
Any B_0 not in this interval would reject MODEL C
Any B_0 IN this interval would NOT reject MODEL C

Formula for CI shows the three components of statistical power. Whatever makes the CI narrower, increases power.

Equvalence to the t-test

$$F^*_{1,n-1} = \frac{SSR}{MSE} = \frac{n(B_0 - \bar{Y})^2}{s^2}$$

$$t^*_{n-1} = \sqrt{F^*_{1,n-1}} = \frac{\sqrt{n}(\bar{Y} - B_0)}{s} = \frac{\sqrt{15}(78.1 - 65)}{23.6} = 2.15$$

2.15^2 = 4.62, which is F we got before (within rounding)

EXAMPLE

from Appendix A
SATV = self-reported SAT verbal score
ASATV = actual SAT math score
SATVDIFF = SATV - ASATV
n = 170

MODEL C: $SATVDIFF_i = 0 + \varepsilon_i$

MODEL A: $SATVDIFF_i = \beta + \varepsilon_i$

Power analysis:
PA-PC = 1, n-PA = 169 (n-PA=150 is close)
small .57
medium .98
large > .995
might miss small effect but great power for everthing else

MODEL C: $SATVDIFF_i = 0$	**SSE(C)=1861.25**
MODEL A: $SATVDIFF_i = 1.03$	**SSE(A)=1682.13**

PRE = .096 $F^{*}(1,169) = 18.0$ $p < .0001$

Statistical: Reject MODEL C
Substantive: On average, freshmen overstated their SAT-Verbal scores by approximately 10.3 points [on the 200-800 scale].

Do the same thing for SATMDIFF
mean = -.27
PRE = .010 $F(1,169) = 1.73$, $p = .19$
do not reject MODEL C
tendency to understate SATM is not reliable

Punchline

This same machinery for statistical inference will work for all the rest of the more complicated models we will consider! PRE's, F*'s, crit value, power tables ALL THE SAME!