

# STA 6360, Report 7.1

**Carson Slater** *Baylor University*

## **Problem**

A study was performed for the Educational Testing Service to analyze the effects of special coaching programs on test scores. Separate randomized experiments were performed to estimate the effects of coaching programs for the SAT-V (Scholastic Aptitude Test-Verbal) in each of eight high schools.

The outcome variable in each study was the score on a special administration of the SAT-V, a standardized multiple choice test administered by the Educational Testing Service and used to help colleges make admissions decisions; the scores can vary between 200 and 800, with mean about 500 and standard deviation about 100. The SAT examinations are designed to be resistant to short-term efforts directed specifically toward improving performance on the test; instead they are designed to reflect knowledge acquired and abilities developed over many years of education.

Nevertheless, each of the eight schools in this study considered its short-term coaching program to be successful at increasing SAT scores. Also, there was no prior reason to believe that any of the eight programs was more effective than any other or that some were more similar in effect to each other than to any other.

The results of the experiments are summarized in Table 1. All students in the experiments had already taken the PSAT (Preliminary SAT), and allowance was made for differences in the PSAT-M (Mathematics) and PSAT-V test scores between coached and uncoached students. In particular, in each school the estimated coaching effect and its standard error were obtained by an analysis of covariance adjustment (that is, a linear regression was performed of SAT-V on treatment group, using PSAT-M and PSAT-V as control variables) appropriate for a completely randomized experiment. A separate regression was estimated for each school.

Although not simple sample means (because of the covariance adjustments), the estimated coaching effects, which we label  $y_j$ , and their sampling variances,  $\sigma_j^2$ , play the same role in our model as  $y_j$  and  $\sigma_j^2$  in the previous section.

The estimates  $y_j$  are obtained by independent experiments and have approximately normal sampling distributions with sampling variances that are known, for all practical purposes, because the sample sizes in all of the eight experiments were relatively large, over thirty students in each school (recall the discussion of data reduction in Section 4.1). Incidentally, an increase of eight points on the SAT-V corresponds to about one more test item correct.

Hierarchical models and multiple comparisons:

Table 1: Observed effects of special preparation on SAT-V scores in eight randomized experiments. Estimates are based on separate analyses for the eight experiments.

School	Treatment Estimate $y_j$	$\hat{\sigma}_j$
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18

**Part a (i)**

Reproduce the computations in Section 5.5 for the educational testing example. Use the posterior simulations to estimate (i) for each school  $j$ , the probability that its coaching program is the best of the eight; and (ii) for each pair of schools,  $j$  and  $k$ , the probability that the coaching program in school  $j$  is better than that in school  $k$ .

We use Stan to calculate the posterior. Note that Stan will assume a uniform prior (on the domain of the parameter) unless otherwise specified.

S4 class stanmodel 'stan\_report7.1' coded as follows:

```
data \{
  int<lower = 0> J; // number of schools
  vector[J] y; // estimated treatment effects
  vector<lower = 0>[J] sigma; // standard errors
\}

parameters \{
  real mu; // pop mean
  real<lower = 0> tau; // pop std deviation
  vector[J] eta; // school-level errors
\}

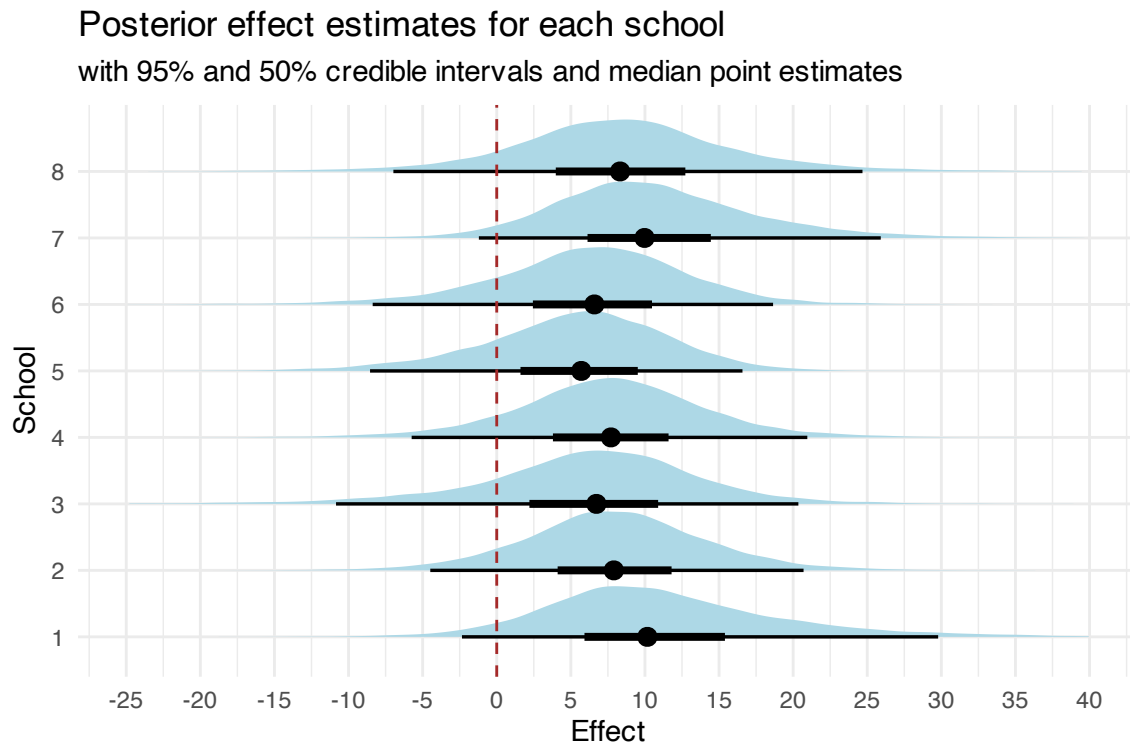
transformed parameters \{
  vector[J] theta = mu + tau * eta; // school effects
\}

model \{
  tau ~ unif(0, 20);
  eta ~ normal(0, 1);
  y ~ normal(theta, sigma);
\}
```

Table 2: School Distribution

Var1	Freq
1	0.247
2	0.104
3	0.085
4	0.095
5	0.051
6	0.076
7	0.201
8	0.142

```
## Warning: Removed 155 rows containing missing values or values outside the s
## ('stat_slabinterval()').
```

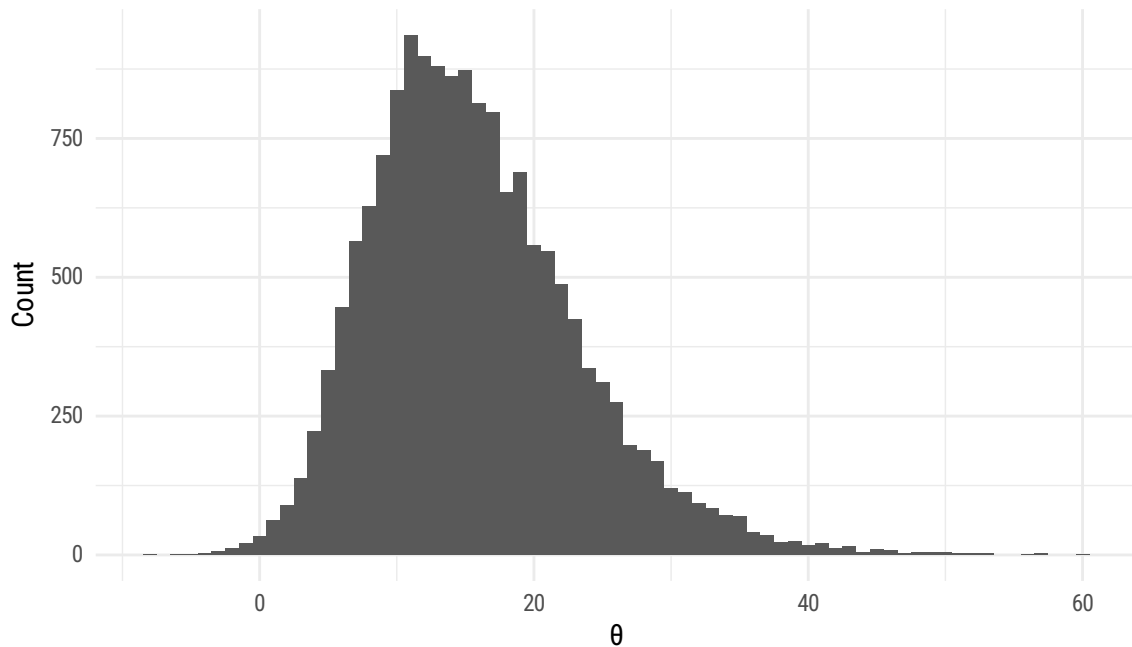


### *Determining the Best Coaching Program*

We compute the relative frequency of each school having the best coaching program from each set of draws. The results are in the table below.

## Posterior Distribution of $\theta$ Values for Best Coaching Programs

\*Note this is per draw\*



### Part a (ii)

j	1	2	3	4	5	6	7	8
2	0.5357807	NA	0.4711350	0.5988464	0.5544739	0.4553709	0.5118505	0.4894822
3	0.5541383	0.5288650	NA	0.6033621	0.5693868	0.4938329	0.5377175	0.5136376
4	0.4549957	0.4011536	0.3966379	NA	0.4474073	0.3645279	0.4121180	0.4198908
5	0.4948981	0.4455261	0.4306132	0.5525927	NA	0.4024873	0.4578930	0.4528170
6	0.5701023	0.5446291	0.5061671	0.6354721	0.5975127	NA	0.5558143	0.5212390
7	0.5265117	0.4881495	0.4622825	0.5878820	0.5421070	0.4441857	NA	0.4813476
8	0.5367590	0.5105178	0.4863624	0.5801092	0.5471830	0.4787610	0.5186524	NA
1	NA	0.4642193	0.4458617	0.5450043	0.5051019	0.4298977	0.4734883	0.4632410

### Part b (i)

We use Stan to calculate the posterior. Note that Stan will assume a uniform prior (on the domain of the parameter) unless otherwise specified. We diffuse the prior of  $\mu$  to be  $\text{Unif}(-\infty, \infty)$ , the default for Stan when a parameter is not specified.

S4 class stanmodel 'stan\_report7.1.2' coded as follows:

```
data \{
  int<lower = 0> J; // number of schools
```

```

vector[J] y; // estimated treatment effects
vector<lower = 0>[J] sigma; // standard errors
\}

parameters \{
  real mu; // pop mean
  real<lower = 0> tau; // pop std deviation
  vector[J] eta; // school-level errors
\}

transformed parameters \{
  vector[J] theta = mu + tau * eta; // school effects
\}

model \{
  // notice no tau here now
  eta ~ normal(0, 1);
  y ~ normal(theta, sigma);
\}

## Warning: Removed 185 rows containing missing values or values outside the s
## ('stat_slabininterval()').

```

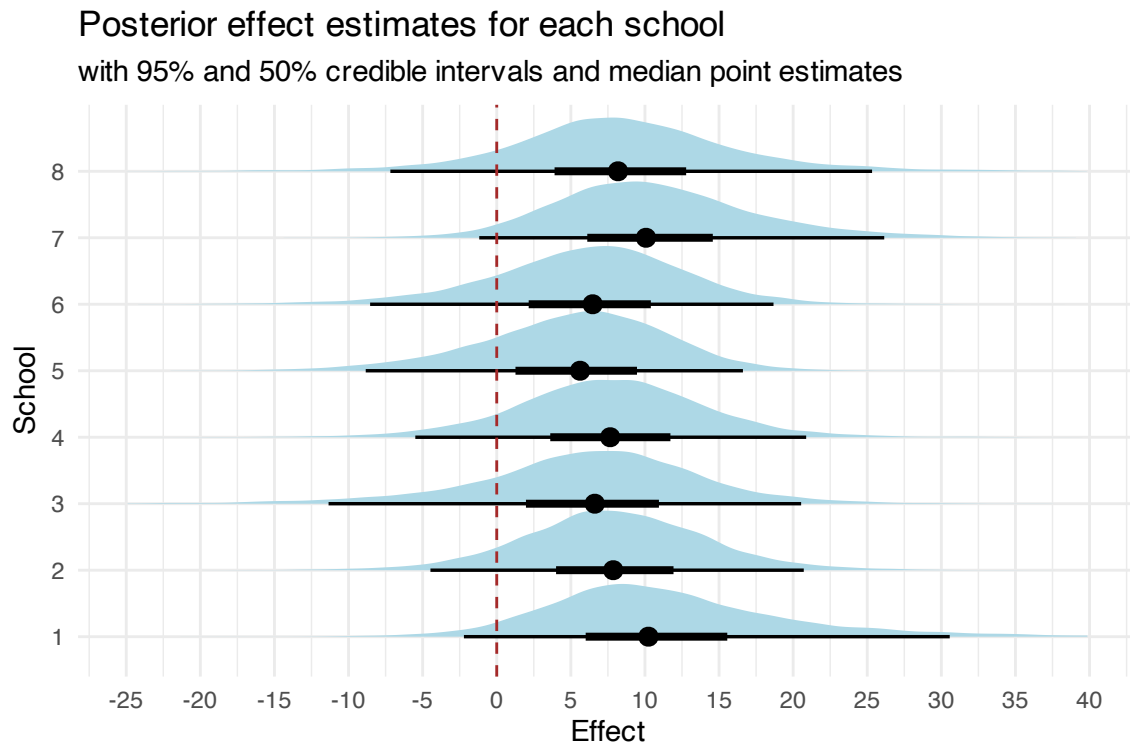
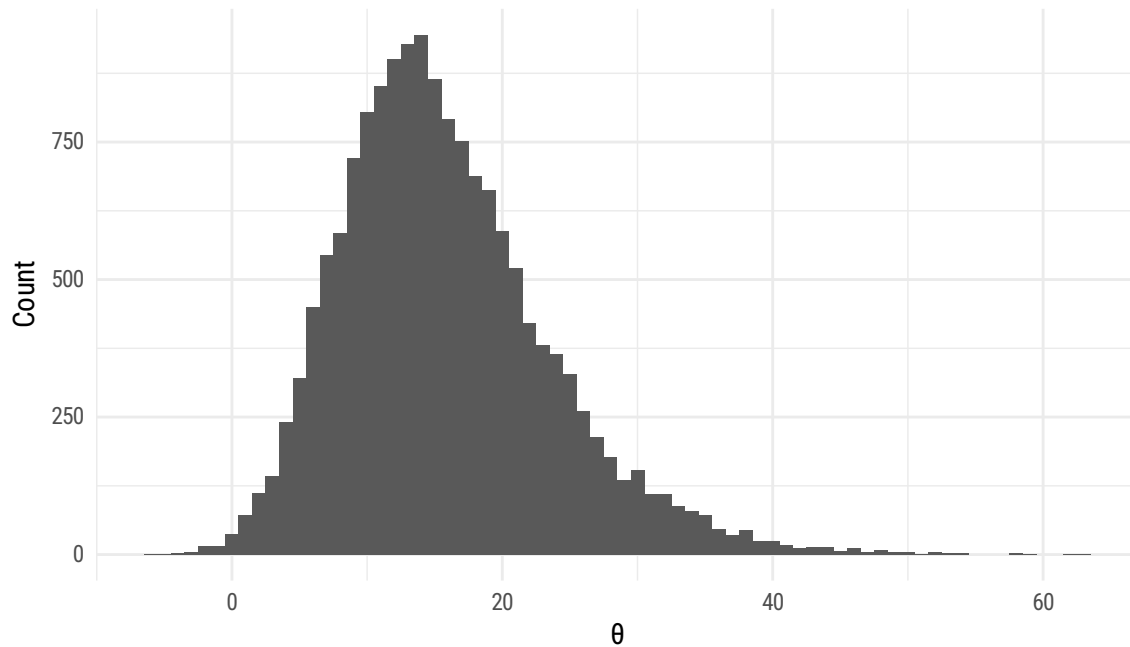


Table 4: School Distribution

Var1	Freq
1	0.250
2	0.106
3	0.085
4	0.097
5	0.053
6	0.071
7	0.201
8	0.137

Posterior Distribution of  $\theta$  Values for Best Coaching Programs

\*Note this is per draw\*



Part b (ii)

j	1	2	3	4	5	6	7	8
2	0.5551261	NA	0.5780833	0.5420792	0.4056479	0.6057658	0.4870795	0.4910683
3	0.4778597	0.4219167	NA	0.4559987	0.3529216	0.5055828	0.4124404	0.4311381
4	0.5198974	0.4579208	0.5440013	NA	0.3682338	0.5617512	0.4457176	0.4616310
5	0.6279719	0.5943521	0.6470784	0.6317662	NA	0.6937957	0.5810984	0.5543667
6	0.4680884	0.3942342	0.4944172	0.4382488	0.3062043	NA	0.3824260	0.4164702
7	0.5651479	0.5129205	0.5875596	0.5542824	0.4189016	0.6175740	NA	0.4999430
8	0.5502664	0.5089317	0.5688619	0.5383690	0.4456333	0.5835298	0.5000570	NA
1	NA	0.4448739	0.5221403	0.4801026	0.3720281	0.5319116	0.4348521	0.4497336