Field Experiments: Design, Analysis and Interpretation Solutions for Chapter 11 Exercises

Alan S. Gerber and Donald P. Green*

Question 1

Important concepts:

a) Explain the distinction between a sample average treatment effect and a population average treatment effect. Why might a researcher be primarily interested in one rather than the other?
Answer:
Define a population as a set of subjects from which an experimental sample is drawn. Depending on how a sample is drawn, the ATE for the sample may be similar or different from the ATE for the broader population; large, random samples tend to have similar ATEs to their parent populations. Researchers may be interested in the ATE for the sample because their primary goal is to figure out how the subjects in the experiment respond to the treatment. Or researchers may be interested in the ATE for the population because they seek to draw generalizations about how the intervention would work were it applied to others in the population.

b) What is a meta-analysis? Why is meta-analysis a better way to summarize research findings than comparing the number of studies that show significant estimated treatment effects to the number of studies that show insignificant estimated treatment effects?
Answer:
Meta-analysis refers to statistical procedures designed to summarize the results of research literatures. Meta-analysis is sometimes described as a "systematic" method for constructing a literature review because it summarizes research findings based on a replicable formula. Specifically, when meta-analysis is used to pool several studies, each study's experimental result is weighted according to a formula that follows from an underlying statistical model. In this chapter, the model involves random sampling from a large population, and the formula (fixed effects meta analysis) weights each study to the inverse of its precision, or squared standard error. This procedure is superior to a count of studies that show significant or insignificant results because the latter potentially accords too much weight to small studies that produce statistically insignificant results and too little weight to large studies that convincingly demonstrate an effect when other, smaller studies fail to do so. Another advantage of meta-analysis over this head-count method is that meta-analysis generates a point estimate and confidence interval, which is more informative than a summary statement about statistical significance.

c) Using equations (11.2), (11.3), and (11.4), provide a hypothetical example to illustrate how uncertainty about the possibility of bias affects the way in which prior beliefs are updated in light of new evidence.
Answer:

---

*Solutions prepared by Peter M. Aronow and revised by Alexander Coppock

Suppose that a researcher were to conduct a study on the effects of SAT prep classes on SAT scores using an observational design that compares a national random sample of high school seniors who take the class to those who do not. The researcher's normal prior about the ATE is centered at 30 points with a standard deviation of 15 points. The researcher's normal prior about the bias of the design is 15 points with a standard deviation of 10 points. The study's results suggest that the course increases performance by 65 points with a standard deviation of 5 points. In other words, $g = 30$, $\sigma_g^2 = 225$, $\beta = 15$, $\sigma_\beta^2 = 100$, $x_e = 65$, and $\sigma_{x_e}^2 = 25$. Plugging these numbers into equation (11.3) gives:

$$\sigma_{\bar{\tau}|x_e}^2 = \frac{1}{\frac{1}{\sigma_g^2} + \frac{1}{\sigma_\beta^2 + \sigma_{x_e}^2}} = \frac{1}{\frac{1}{225} + \frac{1}{100+25}} = 80.36$$

Plugging these numbers into equation (11.4) gives:

$$p_1 = \frac{\sigma_{\bar{\tau}|x_e}^2}{\sigma_g^2} = \frac{\sigma_\beta^2 + \sigma_{x_e}^2}{\sigma_g^2 + \sigma_\beta^2 + \sigma_{x_e}^2} = \frac{100 + 25}{225 + 100 + 25} = 0.357 \qquad p_2 = \frac{\sigma_{\bar{\tau}|x_e}^2}{\sigma_\beta^2 + \sigma_{x_e}^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\beta^2 + \sigma_{x_e}^2} = 1 - p_1 = 0.643$$

Finally, plugging these numbers into equation (11.2) gives the posterior estimate:

$$E[\bar{\tau}|X_e = x_e] = p_1 * g + p_2(x_e - \beta)$$
$$= 0.357 * 30 + 0.643 * (65 - 15) = 42.87$$

In the absence of uncertainty about bias (i.e., if $\sigma_\beta^2 = 0$), the weight given to the new evidence (p_2) would have been much greater: $\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\beta^2 + \sigma_{x_e}^2} = \frac{225}{225+0+25} = 0.9$. The posterior would have more strongly shaped by the observational results:

$$E[\bar{\tau}|X_e = x_e] = p_1 * g + p_2(x_e - \beta)$$
$$= 0.1 * 30 + 0.9 * (65 - 15) = 48$$

d) What does it mean to conduct a hypothesis test that compares two "nested" models?
Answer:
Models are said to be "nested" when one model can be written as a special case of another model. For example, if one conducts an experiment with three groups, a control group and two treatments, one could estimate the ATE of each treatment, or one could estimate a nested model in which both treatments are assumed to have the same ATE. When expressed in regression form (with indicator variables for each treatment), the first model is

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \epsilon_i$$

and the second model is:

$$Y_i = \beta_0 + \beta_1(D_{1i} + D_{2i}) + \epsilon_i$$

Question 2

Question 3

Suppose one were to sample N subjects at random from a population of $N*$ people. An experiment is performed whereby m of the N subjects are assigned to receive a treatment, and the remaining $N-m$ are assigned to the control group. Suppose that sometime after the treatment is administered, outcomes are measured for all $N*$ people.

a) Suppose one estimates the population ATE by comparing the mean outcome among the m subjects in the treatment group to the mean outcome among the $N^* - m$ subjects who were not assigned to the treatment. Is this estimator unbiased?
Answer:
Yes. The subjects assigned to the treatment and control groups are each random samples from the pool of N^* subjects in the population. Therefore, they have the same expected potential outcomes.

b) Would the appropriate standard error of this difference-in-means estimator be equation (11.1), equation (3.4), or neither?
Answer:
The correct formula is a modified version of equation (3.4) in which N^* replaces N.

Question 4

Question 5

Using the Bayesian updating equations, show algebraically how the priors represented in Figure 11.1 combine with the experimental results depicted in order to form a posterior distribution with a mean of 8 and a standard deviation of 0.89.
Answer:
In this example, $g = 0$, $\sigma_g^2 = 4$, $\beta = 0$, $\sigma_\beta^2 = 0$, $x_e = 10$, and $\sigma_{x_e}^2 = 1$
Plugging these numbers into equation (11.3) gives:

$$\sigma_{\bar{\tau}|x_e}^2 = \frac{1}{\frac{1}{\sigma_g^2} + \frac{1}{\sigma_\beta^2 + \sigma_{x_e}^2}} = \frac{1}{\frac{1}{4} + \frac{1}{0+1}} = 0.8$$

Plugging these numbers into equation (11.4) gives:

$$p_1 = \frac{\sigma_{\bar{\tau}|x_e}^2}{\sigma_g^2} = \frac{\sigma_\beta^2 + \sigma_{x_e}^2}{\sigma_g^2 + \sigma_\beta^2 + \sigma_{x_e}^2} = \frac{0+1}{4+0+1} = 0.2$$

$$p_2 = \frac{\sigma_{\bar{\tau}|x_e}^2}{\sigma_\beta^2 + \sigma_{x_e}^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\beta^2 + \sigma_{x_e}^2} = 1 - p_1 = 0.8$$
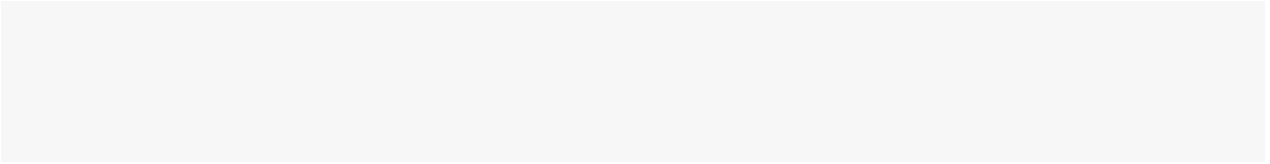
Finally, plugging these numbers into equation (11.2) gives the posterior estimate:

$$E[\bar{\tau}|X_e = x_e] = p_1 * g + p_2(x_e - \beta)$$
$$= 0.2 * 0 + 0.8 * (10) = 8$$

Thus, Figure 11.1 depicts the posterior as centered at 8 with a standard error of $\sqrt{0.8} = 0.89$.

Question 6

Question 7

According to the logistic regression coefficients reported in Table 11.2, the intercept in Region 1 is 8.531 and the slope is -1.978. Based on these numbers, what proportion of those offered a price of 100 shillings is expected to buy a bed net? How does this compare to the actual rate of purchases at this price?

Answer:

The logistic model for Region 1 is

$$Pr[Y_i = 1] = \frac{1}{1 + e^{-(\beta_0 + \beta_1 ln[D-i])}}$$
$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 ln[100])}}$$
$$= 0.359$$

The corresponding empirical value from this region is 0.340.

4

Question 8

Question 9

Because the log transformation of price is undefined when price is zero, we excluded the zero price condition from the analysis of bed net purchases in Tables 11.2 and 11.3.

a) If we exclude zero prices from our experimental analysis, will our estimate of the causal effect of price be biased?
Answer:
No. The exclusion is based on the treatment not on the results. Because those receiving the zero price are a random subset of all subjects, excluding these observations does not lead to biased estimates of the ATE. Thinking back to Chapter 7, missingness here is unrelated to potential outcomes.

b) Suppose we reasoned that a nominal price of zero nevertheless involves some transaction cost, as villagers have to make the effort to redeem their vouchers. For a given subject, we may model the probability of making a purchase as:

$$Pr[Y = 1] = \frac{1}{1 + e^{-(\beta_0 + \beta_1 ln[V_i + \gamma])}}$$

where γ represents the transaction cost of redeeming the voucher. In order to estimate γ, insert a positive value of γ, and use logistic regression to estimate the revised model; note the value of the log-likelihood for this model. Repeat this exercise for different values of γ. Obtain the "maximum likelihood estimate" of γ by finding the value of γ that maximizes the log-likelihood.
Answer:
We tried different values of γ until we came upon the value 19, which maximized the log-likelihood:

```
In [1]: import delim ./data/chapter11/Chapter_11_Dupas_2010_Dataset,clear

In [2]: gen purchased =.
        replace purchased = 1 if purchasednet=="yes"
        replace purchased = 0 if purchasednet=="no"
        rename cfw_id region

        matrix t=J(100, 2, .)
        matrix colnames t=gammas lls

        forvalues i = 1/100 {
```

```
                              gen log_price_star_`i' = log(price + `i')
                              qui glm purchased log_price_star_`i' i.region,
                                       family(binomial) link(logit)

                              matrix t[`i', 1] = `i'
                              matrix t[`i', 2] = `e(ll)'

         }

In [3]: svmat double t, names(col)
        qui sum lls
        list gammas if lls==r(max)

        scatter lls gammas, xline(19)
        graph export ../results/chapter11/exercise_11_9_graph.pdf

      +--------+
      | gammas |
      |--------|
  19. |    19  |
      +--------+
```
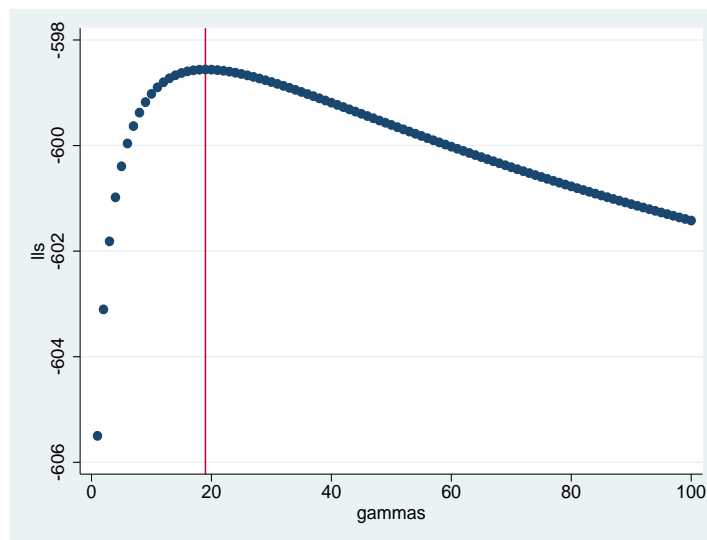

c) What is the substantive interpretation of the maximum likelihood estimates of γ and β_1? (Note that the standard errors using this method understate the true sampling variability because they are conditional on a particular choice of γ. Ignore the reported standard errors, and just interpret the estimates.)

Answer:

The maximum likelihood estimate of γ is 19 shillings, which suggests that redeeming the voucher involves some transaction cost even when the nominal price is very small. The coefficient on the treatment variable, $ln(price + 19)$, is -1.98, which suggests that for every one unit change in this rescaled version of the treatment, the log-odds of purchase declines by -1.98. For example,

suppose that the offer price rises from 0 to 100 shillings. The $ln(price + 19)$ would change from 2.94 to 4.78, and this change would reduce the log-odds of purchase by 3.63. To illustrate what this means in terms of percentage points, suppose that a person has a 50% chance of making a purchase at a price of zero (50% implies a log-odds of 0). If that person were to be offered a price of 100, the predicted probability of a purchase is $1/(1 + e^(- 3.63)) = 0.026$, or 2.6%.

Question 10