HW2, Theory of Inference 2015/6

Jonathan Rougier
School of Mathematics
University of Bristol UK

This homework is about some of the inferential issues which arise when doing experiments to assess causal effects. In this case, the question is whether drinking a glass of milk in the morning causes children to grow faster.

Note that where there is no mark given, this type of question is unlikely to come up in an exam.

1. The experimental protocol is as follows. The experiment will run for one school year, for all of the children in one class. The heights of the children will be measured at the beginning and end of the year; the outcome will be the change in height for each of the children. The 'treatment' group will be given a 100 ml glass of milk each morning, at the start of class. The 'control' group will be given a 100 ml glass of water. Half of the boys will be allocated to the treatment group, and half of the girls will be allocated to the treatment group. All of the other children will be in the control group. Selection of the subset of boys and girls in the treatment group will be random.

   (a) Why does the control group get a glass of water rather than nothing at all?

   (b) Why is allocation to treatment or control 'blocked' by boy or girl?

   (c) Why is allocation to treatment or control within blocks done at random?

(d) Provide an R snippet that identifies the children to be in the treatment group.

2. The standard notation for a statistical model is

$$\boldsymbol{Y} \sim f(\theta) \quad \theta \in \Omega$$

where f is the name of the model, θ are the parameters, and Ω is the parameter space, which is usually a convex subset of Euclidean space. This is synomymous with the statement

$$\Pr(\boldsymbol{Y} \doteq \boldsymbol{y}; \theta) = f(\boldsymbol{y}; \theta)$$

(a) The *Maximum Likelihood Estimator (MLE)* of θ is defined as

$$\hat{\theta}(\boldsymbol{y}) := \operatorname*{argmax}_{\theta \in \Omega} f(\boldsymbol{y}; \theta).$$

Show that if $g : \theta \mapsto \psi$ is bijective, then the MLE of ψ satisfies $\hat{\psi}(\boldsymbol{y}) = g(\hat{\theta}(\boldsymbol{y}))$. That is, the MLE is invariant to bijective transformations. This is one of its most attractive features. [10 marks]

(b) If $\theta = (\theta_1, \theta_2)$, then the MLE of θ_1 is defined to be the first component of the MLE of θ, and similarly for generalisations. Show that if $g : \theta \mapsto \psi$ is *not* bijective, but can be embedded within a bijective mapping by adding additional elements, then the MLE of ψ is $g(\hat{\theta}(\boldsymbol{y}))$. [5 marks]

3. Let $Y_1, \ldots, Y_m$ be the change in heights of the control group, and $Z_1, \ldots, Z_n$ be the change heights of the treatment group. A standard model would be

$$f(\boldsymbol{y}, \boldsymbol{z}; \mu_c, \mu_t, \sigma^2) = \prod_{i=1}^{m} \phi(y_i; \mu_c, \sigma^2) \cdot \prod_{j=1}^{n} \phi(z_j; \mu_t, \sigma^2)$$

where ϕ is the Normal density function with specified expectation and variance. The treatment effect is defined as $\psi := \mu_t - \mu_c$. Show that the MLE for the treatment effect is

$$\hat{\psi}(\boldsymbol{y}, \boldsymbol{z}) = \bar{y} - \bar{z}$$

where $\bar{y}$ and $\bar{z}$ are the sample means of $\boldsymbol{Y}$ and $\boldsymbol{Z}$. [10 marks]

Obviously, this is a very attractive model, since it has such an intuitive measure of the treatment effect. Note that the sample sizes have no effect in the estimator of the treatment effect, although, as we will see later, they have an important effect in the uncertainty about the treatment effect.

4. The growth rate is not constant with age, and so it makes sense to incorporate the age of each child in the model for the change in height. How might you modify the above statistical model to allow for this? [5 marks]

5. Ideally, the estimated treatment effect could be used to decide whether or not to make milk available in the school. Briefly consider the difficulties in generalising from the experiment to the school.

   This question does not have a simple answer. Its purpose is to illustrate that even in this rather simple situation, generalising from an experiment on a sample to an inference about a population is hard.