Recitation 2 - Statistics and RCM *

Jon Cohen

14.03 Fall 2022

Contents

1 Statistics

1.1 Expectation and Variance

Sometimes we want to know the "average value" of a random variable X in which case we can compute the expectation:

$$\begin{aligned} E[X] &= \sum_{x_i} P(X = x_i)x_i \\ &= \int_{-\infty}^{\infty} xf(x)dx \end{aligned}$$

We might also not just be interested in the "average value" but how often the distribution departs from this average value, the "variance". We can define this as: $Var(X) = E[(X - E[X])^2]$ and it's handy to define the standard deviation as the square root of this value.

1.2 Conditional Expectation

Another object of interest is the average realization of the random variable x in a subset of events ("conditioning on a subset of events"). This object is called *conditional expectation*, and it is defined

*Thanks to previous TAs for sharing materials from previous years.

in the same way as the mean, but using conditional probabilities.

$$\begin{aligned} E[X|Y=y] &= \sum_{x_i} P(X=x_i|Y=y)x_i \\ &= \int_{-\infty}^{\infty} xf(x|Y=y)dx \end{aligned}$$

1.3 Inference

All the above discussion is nice, but in general we do not observe *populations* (e.g. the universe of people in the US) but just *random samples* taken from these populations. This mean that we will only be able to compute statistics that summarize the data we collected. An example is the sample average. However, often want to do more than just compute averages. One thing we commonly want to know is whether a mean is equal to zero. (for example if we want to know if the treatment effect is zero. We typically do this using a "t-test" on a sample mean. The basic logic is as follows:

- Suppose I assume that true mean of our data is zero.
- I can compute the t-statistic, that has a known distribution conditional on some value of the mean. This means that I can tell what is the probability that I observe certain values of t.[1]
- If I see a value that is unlikely if the data has indeed a zero mean, this suggests the original assumption that the mean is zero is wrong, and I can therefore reject my original hypothesis.

The above procedure is an application of the central limit theorem. The CLT tell us that if we have a sequence of independent and identically distributed or *i.i.d.* random variables, where $E[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$ then

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right) \rightarrow^d N(0, \sigma^2)$$

Where $N(\cdot)$ is the normal distribution. That is that we know the distribution of sample averages converges to the normal distribution. And if that's true, we know that $T = \frac{\bar{x}-\mu}{\sqrt{s^2/n}}$ is distributed approximately as Student-t distribution $t(n-1)$ where s^2 is the sample variance.

So in order to test my hypotheses that the mean is zero, I assume $\mu = 0$. Then compute $t^* = \frac{\bar{x}}{\sqrt{s^2/n}}$. If t^* takes on an unlikely value given the distribution, it weighs against my hypotheses. And in particular to make concrete how "unlikely" I compute a p-value which is $p = P(|T| \leq t^*|\mu = 0)$.[2] Then we have rules of thumb about what "unlikely" is. Say usually $p = 0.05$ which

[1] This distributions of the statistic encodes the sampling uncertainty surrounding it. In particular they tell us the frequency with which we will observe a value of the statistic if we sample infinitely many times from the population and repeat the computation of the statistic.

[2] The p-value therefore tells us that if we extracted infinite samples from a population with mean 0, and computed a t-stat for each of those, we would have that only 5% of the values we computed would fall above the value t^*. This is clearly a frequentist definition.

occurs when $t^* \approx 2$.

A useful rule of thumb to test significance is $|\bar{x} \pm 2 * \sqrt{s^2/n}| > 0$. This means that the (approximate) 5% *confidence interval*:

$$[\bar{x} - 2 * \sqrt{s^2/n}, \bar{x} + 2 * \sqrt{s^2/n}]$$

does not overlap with 0.

2 Review of the Rubin Causal Model

2.1 Objects of Interest

- X_i: treatment actually administered to an individual i. $X_i = 1$ usually refers to treatment being administered, while $X_i = 0$ to no treatment. Therefore all individuals with $X_i = 1$ are part of the *treatment group*, while those with $X_i = 0$ constitute the *control group*;
- Y_{ij} the outcomes of some individual i, after receiving treatment j
- $T_i = (Y_i|X_i = 1) - (Y_i|X_i = 0)$: treatment effect for individual i

Denote the treatment group by **T** and the control by **C**. In the notation above, what we observe in the data is:

- $Y_{j1}|X_j = 1$, i.e. the outcomes of individuals j after receiving treatment, for the group $j \in \mathbf{T}$ that has been treated ($X_j = 1$);
- $Y_{k0}|X_k = 0$, $k \in \mathbf{C}$ i.e. the outcomes of individuals k after NOT receiving treatment, for the group k that has NOT been treated ($X_k = 0$);

This is where selection issues come in. If you worry that receiving the treatment (the value of variable X) is not random, you will also worry that outcomes will be systematically different across treatment and control *groups*, *independently of the nature of the treatment!*

This is the selection bias that we have seen in class:[3]

$$E_{\mathbf{T}}(Y_{j1}|X_j = 1) - E_{\mathbf{C}}(Y_{k,0}|X_k = 0) = \left[E_{\mathbf{T}}(Y_{j1}|X_j = 1) - E_{\mathbf{T}}(Y_{j,0}|X_j = 1)\right] + \left[E_{\mathbf{T}}(Y_{j,0}|X_j = 1) - E_{\mathbf{C}}(Y_{k,0}|X_k = 0)\right] \tag{1}$$

[3]Here I use superscripts $\mathbf{C}, \mathbf{T}$ to stress that averages are taken over control and treatment group. This is exactly the same as the notation in class:

$$E(Y_1|X = 1) - E(Y_0|X = 0) = [E(Y_1|X = 1) - E(Y_0|X = 0)] + [E(Y_1|X = 0) - E(Y_0|X = 0)]$$

The last term in brackets is the selection effect (or bias), the systematic difference in outcomes between the two groups absent treatment. Therefore, if we want to obtain the average treatment-on-the-treated effect ATT:

$$E(Y_1|X = 1) - E(Y_0|X = 1)$$

we have to ensure that the bias term is 0 if we want to use the above difference in means (1). Instead, if we want to use $E_{\mathbf{T}}(Y_{j1}|X_j = 1) - E_{\mathbf{C}}(Y_{k,0}|X_k = 0)$ to compute the average treatment effect ATE:

$$E(Y_1) - E(Y_0)$$

IN addition to a 0 selection effect, we would need to also assume

$$E_{\mathbf{T}}(Y_{j1}|X_j = 1) = E_{\mathbf{C}}(Y_{k1}|X_k = 0)$$

i.e. that the outcomes of the two groups would be the same if administered treatment. Randomization, if successfully carried out, ensures both approximately apply, since the treatment and control groups are very similar to each other.

2.2 Examples

Miguel & Kramer 2004 found deworming medicine increased probability of attending school for a sample of Kenyan school children by 7pp. Let's think through with RCM.

- What are potential outcomes here?
  - Y_{i1} is school attendance for individual i if they take the deworming medicine
  - Y_{i0} is school attendance for individual i if they do not take the deworming medicine
- Should we be worried about the stable unit treatment value assumption (SUTVA)?
  - SUTVA requires the potential outcome of one unit does not depend on treatment assignment for other units.
  - In the deworming context where the mechanism is a communicable disease, we might believe that one person's deworming can affect another person's health (and therefore that other person's schooling).
  - If we were to relax SUTVA, we would need to define potential outcomes in terms of own treatment assignment and others' treatment assignment.
  - Miguel & Kramer 2004 do this, and they implement it by randomizing both whether an individual person is treated *and* the share of people in a community who are treated.
  - For the rest of the discussion, however, let's assume that SUTVA holds.
- What is the ATT?

  - $E[Y_{i1} - Y_{i0}|X_i = 1]$ is the average effect on school attendance for individuals who will choose to take the drug.

- What is the ATE?
  - $E[Y_{i1} - Y_{i0}]$ is the average effect on school attendance for all individuals

- Which one is likely bigger and why?
  - The ATT is likely bigger as individuals who have worms are more likely to take the medicine. While individuals without worms likely wouldn't benefit from the medicine and are unlikely to take it.