STATISTICAL INFERENCE

ETHAN LEVIEN

Contents

1. Statistical inference for Bernoulli random variables

Let's imagine we do in fact conduct a survey of $n = 20$ students and find $k = 3$ students identify as republicans.

Exercise 1. *Based only on this information (and not your previous experience) what is your best guess of q?*

Can we make this a little more precise? Recall that if each individual data pointy_i has a Bernoulli distribution, then the a number of people, k, who responded saying they are repbulican's follows a Binomial distribution:

$$p(k|q) = \sum \binom{n}{k} q^k (1-q)^{N-k} \tag{1}$$

This equation tells us how likely it is to observe k yeses among n people surveyed. Then, it seems reasonable that this number should not be very small, since that would mean our survay results are an anomoly. In fact, the larger this number is, the more "typical" our results are. This provides us with a way to determine q: We can take as our estimate $\hat{q}$ the value which makes $p(k|q)$ largest. You can do this using calculate

$$\hat{q} = \frac{k}{n} \tag{2}$$

This is not the only way to estimate q, but it is the most natural and widely used. We call estimatoes like this one, which are obtained by maximizng the probability distribtion evaluated at the data, or **likelihood**, maximum likelhood estimates, or MLEs. They are useful, but as we learn later on, they are not the entire story.

A question we often ask about any sort of estimate is: How accurate is this? For example, if we survey $n = 10000$ people and fine 5000 respond YES, our estimate of $\hat{q} = 1/2$ is clearly more reiable than if we had surveyed $n = 4$ and received 2 YES. In classical statistics, we measure accuracy using the standard error, denoted $\mathrm{se}(\hat{q})$. Roughly speaking, if we performed many experienbts and measured $\hat{q}$, the measurments will typically differ by $\mathrm{se}(\hat{q})$. I

$$\mathrm{se}(\hat{q}) = \sqrt{\frac{\hat{q}(1-\hat{q})}{n}} \tag{3}$$

2. Statistical inference for Normal random variables

Now let's think about statistical inference for a Normal random variable.

Date: April 2022.