Hypothesis testing and p-value

- After learning some many concepts about hypothesis testing, do you know:
- What is the objective of conducting a hypothesis test?
- You want to know some information about the value of the parameters of your distribution.

- What is Parameters of your distribution?
- e.g. mean, variance of your distribution

Hypothesis testing and p-value

main steps:

1. Identify your sampling distribution under the assumption that H_0 is true.

(what is sampling/empirical distribution?

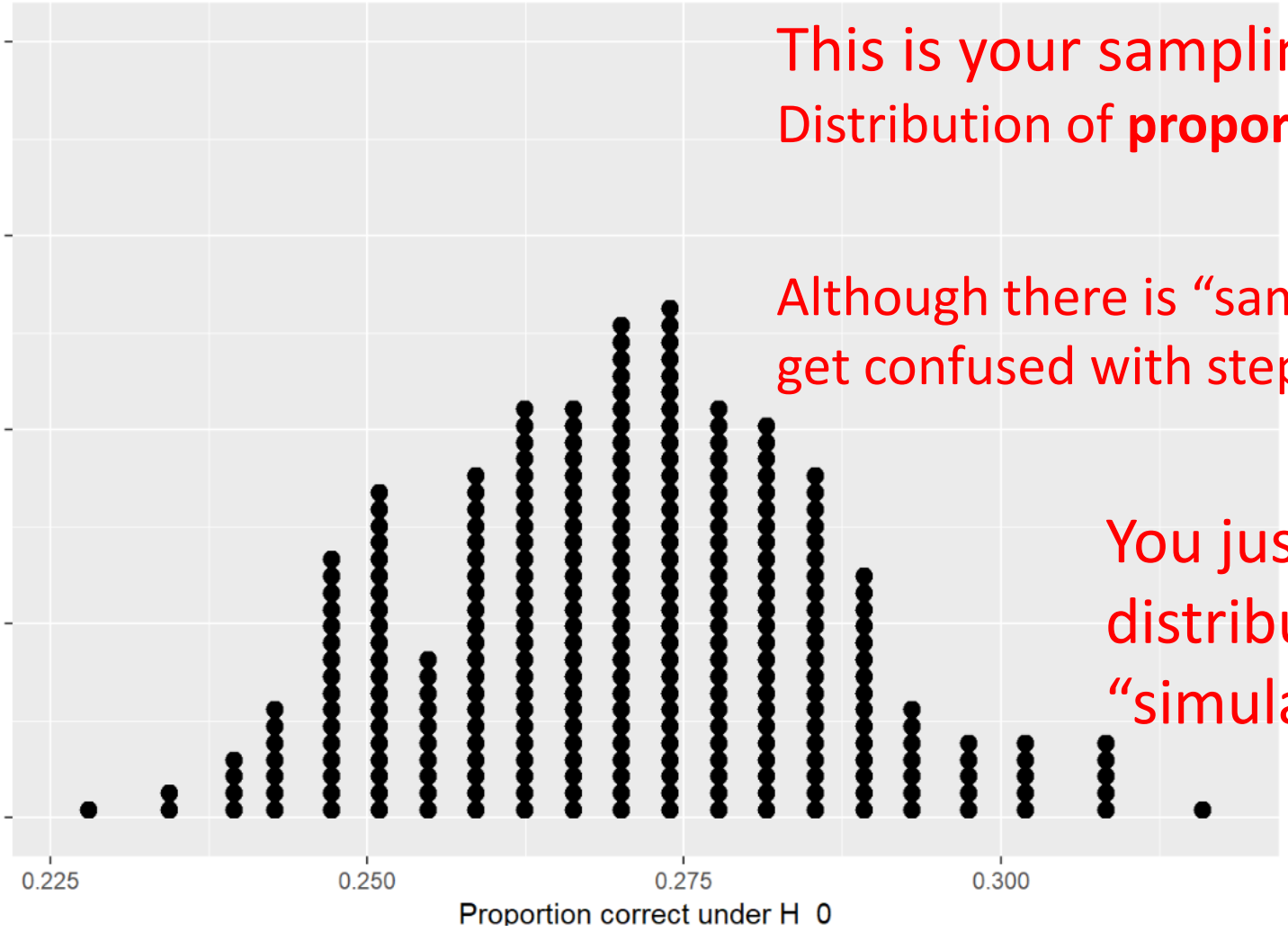
Ans: distribution of some parameter of the distribution you want to study.)

2. Do the sampling. (e.g. One survey showed that among 785 randomly selected subjects who completed four years of college, 144 smoke and 641 do not)

3. calculate the p-value by using R or looking at the sampling distribution (where are the test-stats located in your sampling distribution.)

- For last week's homework, don't get confused on step 1 & 2,

b. Assume you conduct a hypothesis test using simulation and get the following empirical distribution for values of the test statistic $\hat{p}$, assuming the null hypothesis is true. For simplicity, this distribution only shows the results of 300 simulations. There are 300 dots on the plot, one for each simulation (note that in practice, 300 simulations is not sufficient). What does each dot on the plot represent?



This is your sampling/empirical distribution!
Distribution of **proportion of smoker**.

Although there is "sampling" process involved, don't get confused with step 2 of the hypothesis testing!

You just generate the sampling distribution by "sampling" or "simulation"

- There are many ways of getting the sampling distribution.
- Sometimes, the distribution of your population is given.
- For example, we assume, the distribution of student’s height is normal in UofT.

Group Discussion on the article

- Try to answer the following question:

a) Suppose you are interested in testing if 90% of people believe that it is becoming more difficult for young people to buy a home in Toronto.

- (i) Write the null and alternative hypotheses for this test.
- (ii) Based on the poll results summarized in the article, what test statistic would you use for this test?
- (iii) If you wanted to do simulations to test this hypothesis, what sample size (n) would you use?

(b) Suppose you are interested in testing if 70% of people believe that there are not enough homes being built to keep the cost of housing affordable.

- (i) Write the null and alternative hypotheses for this test.
- (ii) Based on the poll results summarized in the article, what test statistic would you use for this test?
- (iii) If you wanted to do simulations to test this hypothesis, what sample size (n) would you use?

STA130 Week 5 Writing exercise:

- **In a short paragraph, summarize your discussion about the article following the steps (you can choose between part A and B, i.e., write about one of them):**
- a. Contextualize the problem (translate it to a statistical problem first).
- b. State the null and alternative hypotheses
- c. Provide the test statistic and the sample size
- d. Assume that p=0.04 for part A, is there evidence to support the null hypothesis? (Assume that p=0.14 for part B, is there evidence to support the null hypothesis)

- **Please avoid using symbols or R codes in your writing.**