A/B Testing

CMSC320

In this exercise you will experiment with the application of statistical inference in A/B testing. You are a Data Scientist at jsFrameworksRUs and you are tasked with conducting an experiment to measure the effect of a webpage redesign on click rate for a link of interest. You decide to use hypothesis testing to analyze the data you gather from the experiment.

Part 1: Compare to known click rate ($p_A = 0.5$)

In the first case, you assume the click rate for the original version of the page (version A) is $p_A = .5$. The experiment you carry out is pretty simple: show the webpage to $n = 50$ subjects and record whether they click on the link of interest or not. You will use this experiment to estimate your parameter of interest: p_B, the click rate for the new page design (version B).

When you carry out your experiment, you record that $s = 30$ subjects clicked on the link of interest.

Based on our discussion in class, you treat this as $n = 50$ draws from a $\mathrm{Bernoulli}(.5)$ random variable, and use the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{30}{50} = 0.6$ as your estimate $\hat{p}_B$.

You remember that the hypothesis testing framework is setup in a way where you use your experiment to *reject* the hypothesis that the new design *does not* increase click rate. Therefore, you want to test the (null) hypothesis $p_B \leq p_A = 0.5$ and *reject* it if $P(\overline{X} > \hat{p}_B) \leq \alpha$ under this hypothesis. Remember, α is the rejection level, and we will use $\alpha = 0.05$ here.

To compute $P(\overline{X} > \hat{p}_B)$ under the null hypothesis you will use the normal approximation given by the Central Limit Theorem (CLT).

(a) Derive expressions for $\mathbb{E}\overline{X}$ and $\mathrm{var}[\overline{X}]$ under the null hypothesis in terms of p_A. You will need to use the properties of expectations and variances described below. Here, I give you the derivation for $\mathbb{E}[\overline{X}]$, you need to do the same for $\mathrm{var}[\overline{X}]$.

$$
\begin{aligned}
\mathbb{E}[\overline{X}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] && (1)\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] && (2)\\
&= \frac{1}{n}(np_A) && (3)\\
&= p_A && (4)
\end{aligned}
$$

(b) Based on your derivation, compute values for $\mathbb{E}[\overline{X}]$ and $\mathrm{var}[\overline{X}]$ based on $p_A = 0.5$ and $n = 50$. Use R or python to do this.

(c) Using the result above, you can now use the CLT by approximating the distribution of $\overline{X}$ as $N(\mathbb{E}[\overline{X}], \sqrt{\mathrm{var}(\overline{X})})$. Based on this approximation, compute $P(\overline{X} > \hat{p}_B)$. Use the R function `pnorm`, or `norm.cdf` in `scipy.stats` to compute this.

(d) Should you reject the null hypothesis $p_B \leq p_A$? Why?

(e) What if you had observed the same $\hat{p}_B = 0.6$ but with $n = 100$ samples. Should you reject the null hypothesis in this case? Why?

(f) What is the *smallest* value $\hat{p}_B$ you would reject the null hypothesis with $n = 100$. Use the `qnorm` function in R or `norm.ppf` in `scipy.stats` for this. Denote this *smallest* value as q_B.

(g) Based on (f), the smallest detectable improvement for $p_A = 0.5$ with $n = 100$ is then $q_B - p_A$. What is the smallest detectable improvement in your experiment (that is, with $n = 50$)?

Part 2: Compare to known click rate ($p_A = 0.75$)

In this second case, you also assume the click rate for the original version is known, but is $p_A = 0.75$. The data recorded for the experiment is the same. You showed the new design to $n = 50$ subjects and recorded that $s = 30$ clicked on the link of interest.

You want to test the hypothesis $p_B \leq 0.75$ and reject it if $P(\overline{X} > \hat{p}_B) < 0.05$ under this hypothesis. Note the probability in this case is different since $p_A = 0.75$.

(a) What are the values of $\mathbb{E}[\overline{X}]$ and $\text{var}(\overline{X})$ under the null hypothesis in this case.

(b) Based on the CLT approximation, compute $P(\overline{X} > \hat{p}_B)$ under the null hypothesis.

(c) Should you reject the null hypothesis $p_B \leq 0.75$? Why?

(d) What if you had observed the same $\hat{p}_B = 0.6$ but with $n = 100$ samples. Should you reject the null hypothesis in this case? Why?

(e) What is the *smallest* value $\hat{p}_B$ you should reject the null hypothesis with $n = 100$. Denote this *smallest* value as q_B.

(f) Based on (e), the smallest detectable improvement for $p_A = 0.75$ with $n = 100$ is then $q_B - p_A$. What is the smallest detectable improvement in your experiment ($n = 50$)?

Part 3

Consider your answers for parts (1g) and (2f). Is the smallest *detectable* improvement in Question (1g) larger or smaller than in Question (2f)? Explain why this makes sense mathematically.

Part 4: Comparing to estimated click rate p_A.

In this more realistic case you estimate click rates for both page designs in your experiment. The experiment you carry out is as follows: when a customer visits the site, they are randomly (and independently from other customers) shown design A or B, and you record if they click on the link of interest or not. You did this for $n = 100$ customers and recorded the following data:

design	number shown	number clicked
A	$n_A = 55$	$s_A = 35$
B	$n_B = 45$	$s_B = 35$

The null hypothesis we want to test in this case is that $p_B - p_A \leq 0$. That is, that the new design *does not* improve the click rate. How can we use what we know about the CLT in this case?

What we will do is treat estimates using sample means $\hat{p}_A = \overline{X}_A$ and $\hat{p}_B = \overline{X}_B$ as random variables and define a new random variable $Y = \overline{X}_B - \overline{X}_A$ corresponding to the *difference in click rates* $p_B - p_A$. With that, we derive $\mathbb{E}[Y]$ and $\mathrm{var}(Y)$ under the null hypothesis that $p_B - p_A = 0$.

(a) Derive expressions for $\mathbb{E}[Y]$ and $\mathrm{var}(Y)$ under the null hypothesis in terms of $p_A = p_B = p$. You will need to use the properties of expectations and variances described below. Here, I give you the derivation for $\mathbb{E}[Y]$, you need to do the same for $\mathrm{var}(Y)$.

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\left[\overline{X}_B - \overline{X}_A\right] && (5)\\
&= \mathbb{E}[\overline{X}_B] - \mathbb{E}[\overline{X}_A] && (6)\\
&= p_B - p_A && (7)\\
&= 0 && (8)
\end{aligned}$$

(b) It looks like we will need an estimate of $p_A = p_B = p$ for our CLT approximation. Luckily, under the null hypothesis all $n = 100$ observations from this experiment can be treated as independent identically distributed (iid) draws from a $\mathrm{Bernoulli}(p)$ distribution. Based on this observation, what would be your estimate of $p_A = p_B = p$?

(c) Now that you have an estimate of p, compute a value for $\mathrm{var}(Y)$.

(d) What is your estimate $\hat{y}$ of $p_B - p_A$ based on the data your recorded for this experiment?

Now, we can reject the null hypothesis of no improvement if $p(Y > \hat{y}) \leq \alpha$ under the null hypothesis.

(e) Using the CLT approximation, what is $P(Y > \hat{y})$

(f) Can you reject the null hypothesis of no improvement in this case? Why? Remember, we are using $\alpha = 0.05$.

Bonus: Smallest detectable improvement for estimated click rates

We could compute smallest detectable improvements in parts 1 and 2 above because we assumed p_A was known. For part 4, we don't know p_A and instead estimate it, so we cannot compute a smallest detectable improvement before the experiment is run because we don't know $p_B = p_A = p$. We can however, compute what the smallest detectable difference *would be* for different values of p.

(a) Make a line plot, with p in the x-axis and the smallest detectable difference as a function of p in the y-axis. You should assume $n_A = 55$ and $n_B = 45$ as above.

Expectation and variance properties

Properties of expectation

(i) $\mathbb{E}[aX] = a\mathbb{E}[X]$ for constant a and random variable X
(ii) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for random variables X and Y

Properties of variance

(i) $\mathrm{var}[aX] = a^2\mathrm{var}[X]$ for constant a and random variable X
(ii) $\mathrm{var}[X + Y] = \mathrm{var}[X] + \mathrm{var}[Y]$ for *independent* random variables X and Y

Submission

Prepare an Rmarkdown file or Jupyter notebook with your derivations and answers, including code you used to get your answers. Knit to PDF (or save HTML to PDF) and submit to ELMS.