MSAA 5300 Group 6 Final Project

Using Corn Syrup in Coca-Cola: A New Opportunity in the U.K. Market

Jianyu Yue, Kate Wang, Chenning Zhang, Siyu Zhou, Yuanchi Li, Yuting Zhang

Part 1: Research Proposal

1.1 Executive Summary / Abstract

To: Coca-Cola UK Management Team
From: Research & Development Team
Date: 12/12/2020
Subject: Cost Saving Opportunity

Coca-Cola just announced the report of the third quarter of Fiscal 2020 and reported a decline in revenue. The business has been affected by the ongoing pandemic of Covid-19. With the lockdown in some markets, restaurants, and cinemas, the performance of the third quarter will be most devastated. This situation highlights the gap between the demanding intense amounts of fundings and the cold reality of market arithmetic. We would like to help **Coca-Cola UK in addressing how to cut costs and maintain a strong presence in the soft drink industry during COVID-19.** Since the cost of corn syrup is cheaper than that of table sugar, we can help Coca-Cola to increase profitability by replacing the original table sugar with the corn syrup, hence reducing the raw materials cost. In order to help our company achieve ongoing profit, our team will conduct an experimental research in the U.K. market and understand whether our consumers can taste the difference between original table sugar coke and corn syrup coke.

Participants

Two hundred United Kingdom coke drinkers aged from 12 to 30 without diabetes and corn allergies history will be randomly selected as our population of interest. Subjects were evenly distributed across for gender proportions.

Material

The experiment will involve two kinds of coke which are original coke and corn syrup coke, in which the corn syrup coke is going to be diluted into three different levels of sweetness 42%, 48%, and 55%.

Design

A proportional test will be used for the study by examining the proportion of correct judgement between taste original coke and the coke with different concentration levels of corn syrup. We would randomly assign 100 participants to the control group in each experiment, and they are going to taste the original coke. The remaining 100 participants are assigned to taste the coke with 42% concentration level of corn syrup for the first experiment. We are going to replace the concentration level of corn syrup in the coke to 48% in experiment two. Furthermore, we are using a 55% concentration level of corn syrup coke in experiment three. In the end, 200 participants will need to identify what kind of coke that they drink on a survey.

We suggest addressing this problem by collecting data from our experiment and analyzing whether participants can taste the differences between the original coke and different concentration levels of corn syrup coke. We will then make recommendations accordingly to your team. More specifically, if the participants cannot taste the difference between the original coke and the specific sweetness level of corn syrup coke. We can recommend the company replace the sugar with corn syrup to reduce the raw material cost and thus increase the net profits per beverage. Otherwise, if the UK customers could taste the difference between corn syrup and original sugar added in the drinks, Coca-Cola might need to look for another alternative sweetener if they want to lower their cost in the future. We hope this research will provide insights to Coca-Cola U.K. for our business problem, and look forward to discussing more details with you soon.

Yours sincerely,
R&D Team

1.2 Statement of the Problem

Almost 30 years ago, Coca-Cola switched over from sugar to high-fructose corn syrup to sweeten America's favorite carbonated soft drink in the US market (Thomson, 2012). With corn subsidized by the US government, the sugary syrup has become a more affordable alternative for the beverage company (Thomson, 2012). However, the international markets, especially in the UK, have not made this change in their ingredients (Coca-Cola coke ingredient section on the UK website). But, we are concerned that the alternative corn syrup will affect the sales since we are not sure about how different it would be taste-wise.

Therefore, in order to help Coca-Cola to solve its business problem, we would like to perform experimental research on the taste of corn syrup and sugar (added in the soft drink) and

investigate whether the UK customers can detect the difference between the original taste and the corn syrup taste. This study will be experimental research where 200 customers will be included in our sample since based on a general research rule, sample sizes of 200 to 300 participants will provide an acceptable margin of error and fall before the point of diminishing returns (Lyons, 2015). First, we will randomly assign 100 customers to taste the corn syrup coke first and then the original sugar coke. Then, the rest 100 controlled customers are assigned to taste the original sugar coke, and then the corn syrup coke. The participants will not be notified beforehand which coke they are drinking. Finally, all the 200 customers are required to tell the observer which is the corn syrup coke and which is the table sugar coke after the experiment. After data collection, we can analyze the result from the experiment to understand whether the customers can feel the difference between corn syrup coke and the original sugar coke.

We hope that this research will provide insights to Coca-Cola in their business problem of whether corn syrup will impact their sales and if customers can taste the difference. Otherwise, if the UK customers could taste the difference between corn syrup and original sugar and that they would prefer the original one more, Coca-Cola would need to find another alternative sweetener if they want to lower their cost in the future.

1.3 Research Questions and Hypotheses

Research Question 1:

Does 42% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 42% concentration of corn syrup and group of sugar have similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 42% concentration of corn syrup and group of sugar have NO similarity. Proportion(A)≠Proportion(B)

Research Question 2:

Does 48% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 48% concentration of corn syrup and group of sugar have similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 48% concentration of corn syrup and group of sugar have NO similarity. Proportion(A)≠Proportion(B)

Research Question 3:

Does 55% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 55% concentration of corn syrup and group of sugar have similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 55% concentration of corn syrup and group of sugar have NO similarity. Proportion(A)≠Proportion(B)

1.4 Importance of the Study

Based on the goal of organization, the ultimate aim is to make more money. Also, saving money on the raw material means making more money for one company. Since the corn farms are subsidized by the US government, the price of corn syrup is much lower than the table sugar. Moreover, this study will figure out what percentage of corn syrup will be the best alternative towards the original table sugar for the UK customers. Therefore, this study significantly will impact Coca-Cola's future marketing strategy and sales in the UK market, if this study could prove that the UK customers would not taste the difference between the corn syrup and the table sugar. Coca-Cola will expand the change which is replacing the table sugar to corn syrup into the UK market. Afterwards, after implementing this change in the UK market, Coca-Cola will save more money on raw materials and use the rest of the money to expand the business in other essential areas like logistics, packaging, supply chain, flavors of coke and so on.

All in all, cost saving will be used to broaden the organization that would generate more jobs and overall industrial prosperity. Also, cost saving also provides the basis for further returns to owners, and wealth management is a strong expectation. Besides, cost saving tends to enhance process standards when improvisation methods influence the essence of the operation of existing processes that improve product formation standards. What's more, cost saving improves efficiency at certain stages, when workers who have been in the comfort zone for a long time get a signal that the organization is in micro-management and that it will be processed right away, but that may also be micro-management for employees.

1.5 Literature Review

The price of sugar increased dramatically after World War II due to the shortage of it (Bode et al., 2014). Soft drink brands like Coca-Cola and Pepsi started to replace sugar with the HFCS-55 in their products in 1978 (Bode et al., p137-148). HFCS stands for High Fructose Corn Syrup, and 55 stands for the level of sweetness (Bode et al., p137-148). By 2010, producers of HFCS applied to the FDA to rename HFCS "corn sugar" (Barclay et al., p53). Nowadays, corn sugar or syrup (the liquid form of sugar) are widely used in the soft drink industry, not just because they are inexpensive, but also because they have very practical attributes - softening textures, adding volume, preventing sugar from crystallizing, and enhancing flavour (Barclay et al., p54).

There are some controversies going on regarding the taste of corn syrup/HFCS compared to that of table sugar. According to a research on the Journal of Chiropractic Medicine, researchers who measured the relative sweetness of natural and artificial sweeteners

determined that HFCS is 1.5 times as sweet as table sugar (Melina, 2010, p2), however, the Corn Refiners Association argued that HFCS is not sweetener than sugar, and that it was formulated to provide equal sweetness to table sugar so that consumers could not tell the difference in product sweetness and taste (Melina, p2).

A causal link between increased consumption of HFCS and obesity has been hypothesized, and although animal research and a small number of human studies support this hypothesis, extensive human research on this topic is limited (Duffey et al., 2009, p2). Therefore researchers could not conclude that there is a direct relationship between consumption of sugar and health effects, and ongoing research shows that there were too many other different aspects that lead to obesity. In addition, in the study of Added Sugars and Health: Evidence from Prospective Cohort Studies and Controlled Dietary Trials (Sievenpiper et al., 2014), researchers ran a prospective cohort study on HFCS. After observation of the data, researchers suggested that there is no harm in applying HFCS in soft drinks even if it contributed minimum nutrition value when people were on diet (Sievenpiper et al., p113-123). This finding of the previous study is important to support our own research in terms of health concerns.

In addition, the reason why we are manipulating the concentration levels of corn syrup is that people have different preferences and sensations towards levels of sweetness. We deliberately choose concentration levels of corn syrup in this study (42%, 48%, 55%), because soft drink manufacturers have already adapted 55% corn syrup in the U.S (Bode et al., p137-148). Further, 42% is the standard accepted by food processors for food products (Bode et al., p137-148). We hope that in using these levels of sweetness in our independent variable, we can figure out the most appropriate sweetness concentration levels of corn syrup in the United Kingdom. However, according to research, the environment might be one of the factors affecting people's sweetness preference (Venditti et.al ., p1-28), so location could be one of the limitations and uncertainties in our study since we are performing the study in the U.K. while using the standard of the U.S.'s.

1.6 Research Plan

This section will include description of our proposed research investigation in technical terms.

- **Population of Interest**: The United Kingdom Coca-Cola target consumers (age of 12-30)

- **Sample Selection**: Since we are partnering with Coca-Cola UK in this study, we will be able to access their consumer information. Based on the historical customers data and the literature reviews, the 200 United Kingdom coke drinkers with age of 12-30 without diabetes and corn allergies history will be randomly selected as our sample

population. From the sample population, we will randomly select 100 participants into control groups A, C, E and 100 participants into treatment groups B, D, F. Each selection will be independent. The participants in each set of the control and treatment groups are different and randomized in order to minimize the effect of confounding variables. Participants will not be informed about the type of coke they are drinking prior to the experiment.

- **Sample Size**: Based on a general research rule, sample sizes of 200 to 300 subjects will have a reasonable margin of error and will decline before the point of decreasing return. The specific sample size in each treatment and control group will be 100 participants.

- **Operational Procedures**: The Participants will be awarded with a $50 gift card for Coca-Cola as incentives. The company will distribute their coupons to the participants' after the experiment.

  In the first part of the experiment, 200 participants will be randomly assigned into 1 control group (100 participants) and 1 treatment group (100 participants). We called them Group A and B. Participants will be called by the researchers into the room. Each participant will have 5 minutes in the room to taste the drink, and after each participant's taste test, researchers will clean the table and chair before the next participant comes in. The process will remain the same for experiment 2 and 3.

  The first control group will be tasting the original Coca-Cola with table sugar as sweetener, and the treatment group will be tasting a coke with 42% corn syrup as sweetener (all the other ingredients remain the same). After each group is done with the taste test, we will then have the participants write down choose which coke they taste on a survey (option A: original Coca-Cola with table sugar as sweetener, option B: coke with corn syrup as sweetener). Researchers will collect the survey and let the participants rest for an hour while they rinse their months with a bottle of plain water.

  After an hour, researchers will again randomly assign the 200 participants into 1 control group 100 participants) and 1 treatment group (100 participants) in the second part of the experiment. We called them Group C and D. The control group will be tasting the original Coca-Cola with table sugar as sweetener, and the treatment group will be tasting a coke with 48% corn syrup as sweetener (all the other ingredients remain the same). After each group is done with the taste test, we will then have the participants judge which coke they taste on a survey (option A: original Coca-Cola with table sugar as sweetener, option B: coke with corn syrup as sweetener). Researchers will again collect the survey and let the participants rest for an hour while they rinse their months with a bottle of plain water.）

  The third part of the experiment will repeat the same process as the first and second part, but with control group E and treatment group F. The control group will be

drinking the original Coca-Cola, whereas the treatment group will be tasting a coke with 55% corn syrup as sweetener.

Lastly, the researchers will translate the participants' answers from the 3 experiments into dummy variables called "Decision", in which "1" represents a subject correctly labelling the drinks' type, whereas "0" means a subject fails to do so. The data stored in "Decision" would become our dependent and target variable for our study.

Experiments will take 13 days to complete (7 days for pre-test logistics, 3 days for the experiments itself and 3 days for collecting data and transforming them into "1" and "0" by the researchers). Brief schedule will be outlined below.

- **Brief Schedule**: At the onset of the study, we will randomly assign the 200 participants into 3 sets of control and treatment groups for our experiments. The participants in each set of the control and treatment groups are different and randomized in order to minimize the effect of confounding variables. Participants will not be informed about the type of coke they are drinking prior to the experiment.

Phase	Date	Activites
Pre-testing	11/1-11/7	Confirm participants, location and time of experiments; Train researchers and acquire consent forms
Experiment 1	11/8	Conduct Experiement
Experiment 2	11/9	Conduct Experiement
Experiment 3	11/10	Conduct Experiement
Post-testing	11/11-11/13	Collect Data and Transform Answers for Analysis

- **Data Collection**: As stated above, in each experiment, we will make the concentration levels of corn syrup in the drinks independent variables in our study. Each participant will write down and choose which coke they taste on a survey based on their own judgement (option A: original Coca-Cola with table sugar as sweetener, option B: coke with corn syrup as sweetener). For further investigation (not our independent variables), we are going to record the gender and age from each subject. Then, researchers are going to collect data from the survey of three experiments. We will translate the participants' answers from three experiments into dummy variables called "Decision", in which "1" represents a subject correctly labelling the drinks' type, whereas "0" means a subject fails to do so.

- **Data Security**: We will make sure that the researchers who monitor the experiment received standardized and appropriate training (e.g. not to show preference and facial expressions when distributing the drinks, and keep the data confidential) by our research team. Participants and researchers must sign the confidential form before

they participate in the experiment process. We will make sure each researcher has their own identity card which they will need to scan before entering the experiment room. Further, the researchers who are in charge of the surveys will announce to each participant that they will not need to leave their name on the surveys, and that their information (judgement on drinks, gender, age) will remain confidential. Researchers will also make sure to remain silent and have no hint on their facial expression when passing drinks to the participants. Then, the collected data will be sealed into 3 separate envelopes and the envelops will remain sealed until the researchers start to analyze the result. Lastly, researchers who did not have the authorization will not be able to open the survey data. The data will be stored in two different locations and only permitted researchers will have access to it.

- **Variables**
    - **Outcomes (Dependent Variables)**: In our study, the dependent variable is the judgement of participants on the type of coke. In our three experiments, participants will be randomly assigned into three control and treatment groups (group A&B, C&D, E&F) to taste coke with different concentrations of corn syrup and original coke. Participants are going to put a “check mark” under the type of coke they think they are drinking. Researchers will then repeat the same process for all three experiments, since the levels of concentration corn syrup are different for each experiment. Lastly, researchers are going to calculate the proportion of correct judgement for each experiment.
    - **Treatments (Independent Variables)**: The type of coke participants are having is the independent variable of our study (corn syrup coke [42%, 48%, 55%], and the original coke). In the first experiment of our study, the treatment is 42% corn syrup as sweetener (all other ingredients remain the same for the coke) .
    - **Other Variables:**
        - Gender (self-identified): Female, Male, Transgender,Gender Neutral, Do not wish to identify)
        - Age (self-identified): age groups 12-30

1.7 Statistical Analysis Plan

As the point of view that whether the probability that the event will occur in the group is different from the event will occur in another group is the common question that we would like to learn in this corn syrup research study, it means that the probabilities of the correct judgement of 3-pair groups will be different or not. On the other hand, based on this corn syrup research study, since we would like to figure out which kind of concentration of corn syrup is most close to the original flavor of the table sugar that Coca-Cola added into coke

during the previous time and this corn syrup research is trying to prove that customer will not be able to distinguish the difference between the original sugar coke and the corn syrup coke of a certain level of concentration, it is appropriate that we will use two-sample proportion z tests, which uses binary attribute data obtained from simple random sampling in the form of "0 or 1" that represents the rate of success.

From the perspective of this corn syrup research study, there are 200 participants from the United Kingdom who will be treated as the research population of interest. After all, participants have tasted different kinds of flavors of coke three times, observers will record the results and transfer the results into dummy variables for each time, which "1" represents that this participant makes a correct judgement of the kind of the coke, and otherwise "0" represents that this participant makes a wrong judgement of the kind of the coke.

1.8 Limitations and Uncertainties

There are a few limitations in our study. First, since we cannot manipulate gender and age, in the data analysis process, we leave out these two other variables in this study. However, these two variables can be confounding ones because gender differences might have a significant impact towards the judgement of taste. In addition, age could be a confounding variable because people who are older might have less interest in soft drinks. And, people who are older might have health conditions that do not allow them to consume soft drinks.

Secondly, our study is exclusive to people who have diabetes, allergy, or any other relative health concern issues, under the assumption that these people are most likely not going to purchase soft drinks due to their body condition. But there is uncertainty in this assumption because some of these people might still buy Coca-Cola.

In addition, as stated above in literature review, the use of U.S. standards in corn syrup sweetener might induce some confounding variables to this study, as we are performing the study in the U.K. while using the standard of the U.S.'s. According to other researches, environment might be one of the factors affecting people's sweetness preference (Venditti et.al ., p1-28), but we are not sure of the level of impact location will have on people's taste. These are some of the limitations and uncertainties in our study that are worth further investigation and discussion on.

Part 2: Simulation of Effects

Simulation: As mentioned in the first part of the report, our research simulations aimed to solve three primary research questions, including:

Research question 1:

Does 42% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 42% concentration of corn syrup and the group of sugar have a similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 42% concentration of corn syrup and the group of sugar has NO similarity. Proportion(A)≠Proportion(B)

Research question 2:

Does 48% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 48% concentration of corn syrup and the group of sugar have a similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 48% concentration of corn syrup and the group of sugar has NO similarity. Proportion(A)≠Proportion(B)

Research question 3:

Does 55% concentration of corn syrup taste different from sugar?

H0: Proportion of correct judgement between the group of 55% concentration of corn syrup and the group of sugar have a similarity. Proportion(A)=Proportion(B)

H1: Proportion of correct judgement between the group of 55% concentration of corn syrup and the group of sugar has NO similarity. Proportion(A)≠Proportion(B)

For each research question, we have two scenarios to reach, whether there is an expected effect or no effect. We want to use corn syrup to replicate the taste of regular sugar coke as close as we can, and our treatment groups use different levels of concentration corn syrup.

In scenario 1, there is no effect. Our proposed treatment has no impact on the population's outcomes. In this scenario, the taste of corn syrup is identified as different from regular sugar.

If the outcome did not reach our standard, which tastes different from regular sugar, it would be included in scenario 1.

In scenario 2, there is an expected effect that our proposed treatment did impact the population's outcomes. In this scenario, the taste of corn syrup is identified as the same as regular sugar. If the outcome reaches our standard, which tastes the same as regular sugar, it will be included in scenario 2.

Simulation Assumptions:

Our data are simulated under following assumptions:

1. If corn syrup tastes the same with sugar, the customers would make their judgements by guessing. We assume random guessing has 50% probability of success.

2. American customers do not share the similar taste of sweetness with the UK's customers.

In our experiment, we will apply the 42%, 48%, and 55% American corn syrup concentration standard, since a research article from the literature review session mentions that the US coke treats these concentrations of corn syrup as the optimal sugar substitutes. However, we do not know whether these American concentration standards would also be as effective in the UK's market. Thus, during the experiment, we will adjust the customers' acceptance of these three corn syrup levels.

3. The probability of the UK customers tasting the difference between and 42%, 48%, and 55% diluted corn syrup and sugar are 70%, 55%, and 80%.

4. Our expected effect of the study is that the UK customers are not able to taste the difference between sugar and corn syrup.

Data simulation:

The experiment is evaluated under the proportion test (prop.test()) in R. To increase the overall processing efficiency, we first created a function called "analyze.experiment" as shown in the figure below:

```r
analyze.experiment <- function(the.dat) {
  require(data.table)
  setDT(the.dat)

  tab <- the.dat[, .(count = sum(BP), n = .N, proportion = mean(BP)), by = Group]
  the.test <- prop.test(x = tab[, count], n = tab[, n])

  the.effect <- the.test$estimate[1] - the.test$estimate[2]
  upper.bound <- the.test$conf.int[2]
  lower.bound <- the.test$conf.int[1]
  p <- the.test$p.value

  result <- data.table(effect = the.effect, upper_ci = upper.bound, lower_ci = lower.bound,
                       p = p)

  return(result)
}
```

The function takes a data frame as an input and returns a data table with the test results including:

1. True effect of the scenario

2. 95% confidence interval

3. P-value in each simulated study

After setting up the function, we started constructing the experiments' parameters:

```r
##constructing parameters
B <- 1000
n <- 200

fourtytwo = 0.7
fourtyeight = 0.55
fiftyfive = 0.8
```

As the codes shown above, variables "B" and "N" represent number of simulation trail and sample size. In our study, we simulated the study 1000 times on the experiment with 200 samples. Variables "forty-two", "forty-eight", and "fifty-five" indicate different corn syrup concentration levels. These parameters will be used to generate the experimental data.

Research question 1 (scenario 1):

The following codes were used for simulating the data:

```
RNGversion(vstr = 3.6)
set.seed(seed = 42)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = fourtytwo,size = 1))]
```

We first created a data table called “sim.dat” which contained sample data in all simulation trials. Each row would be randomly assigned to either treatment or control group. Each group accounted for half of the sample size in each trail. Then, in each group of data, we use the “rbinom()” function to randomly generate “1” or “0” based on the probability of success. “1” represents the customer label the type of sweetener in the drink, whereas, “0” represents the customer failing to identify the type of sweetener in the drink.

In the first scenario of question 1, every data in the control group would have 50% of probability to get a “1” in the “Decision” column, and each treatment group’s data had 70% probability of success. The sample data are shown below:

Experiment	Group	Dicision
1	Control	0
1	Control	0
1	Control	0
1	Control	1
1	Control	0
1	Control	0
1	Control	0
1	Control	0
1	Control	0
1	Control	0
1	Control	0
1	Control	1

Then, we applied the “analyze.experiment” function to generate the results of all studies:

```
##Apply the function
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]

DT::datatable(data = round(x = exp.results, digits = 3),
              rownames = F)
```

The sample results:

Show 10 entries Search:

Experiment	effect	upper_ci	lower_ci	p
1	-0.084	0.063	-0.231	0.292
2	-0.032	0.116	-0.181	0.755
3	-0.036	0.113	-0.186	0.711
4	0.116	0.263	-0.03	0.131
5	0.052	0.2	-0.096	0.552
6	0.021	0.169	-0.127	0.877
7	0.051	0.199	-0.098	0.566
8	-0.088	0.06	-0.237	0.271
9	0.03	0.179	-0.118	0.773
10	-0.046	0.101	-0.193	0.609

Next, we utilized the results data to calculate:

1. **The true effect of the scenario that you selected.**

   Calculated by averaging all experiments' effect

2. **A 95% confidence interval for the estimated effect from the simulated results.**

   Represented by the interval between the average upper lower bound.

3. **The percentage of true negatives.**

   Measured by the proportion of all data with a greater than or equal to 0.05 P-value

4. **The percentage of false positives**

   One minus the true negative rate

5. **The percentage of true positives.**

   Calculated by the proportion of all data with a less than 0.05 P-value

6. **The percentage of false negative rate**

   One minus the true positives rate

The codes for calculating the above indexes are shown below:

```
#result
exp.results[, mean(p < 0.05)]##true positive rate
1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)] ##The true effect of the scenario that you selected.
exp.results[, mean(upper_ci)] ##average upper bound
exp.results[, mean(lower_ci)] ##average lower bound
```

Research question 1 (scenario 2):

In scenario 2, we assume that the 42% corn syrup concentration would taste the same as the sugar. Under such an assumption, the subjects would make their decision by guessing. Therefore, when generating random "0" and "1", we assigned 50% probability of success in both groups:

```
###42%S2
RNGversion(vstr = 3.6)
set.seed(seed = 42)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
dim(sim.dat)

exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)
```

The rest data analysis codes remain the same with the codes in scenario 1:

```
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)

#result
exp.results[, mean(p < 0.05)]##true positive rate
mean(exp.results$p)
1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)]
exp.results[, mean(upper_ci)]
exp.results[, mean(lower_ci)]
```

Research question 2 (scenario 1):

In scenario 1 question 2, we assume that the samples in the treatment group would have 55% probability of distinguishing 48% corn syrup and sugar. The simulation codes are shown below:

```
###48%S1
RNGversion(vstr = 3.6)
set.seed(seed = 48)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = fourtyeight,size = 1))]
```

We apply the same data analytic function on this question:

```
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)

#result
exp.results[, mean(p < 0.05)]##true positive rate
1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)]
exp.results[, mean(upper_ci)]
exp.results[, mean(lower_ci)]
```

Research question 2 (scenario 2):

In scenario 2, since our expected effect is that the customer can not taste the difference between 48% corn syrup and sugar, we modify the codes by setting the probability of success in each group equal to 50%.

The following codes used to simulate and analyze the study:

```
###48%S2
RNGversion(vstr = 3.6)
set.seed(seed = 48)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
```

```
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)

#result
exp.results[, mean(p < 0.05)]##true positive rate
1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)]
exp.results[, mean(upper_ci)]
exp.results[, mean(lower_ci)]
```

Research question 3 (scenario 1)

In research question 3, we want to find out whether 55% corn syrup would taste the same as sugar. Thus, during the simulation phase, when generating binary “1” and “0”, we change the probability of success in the treatment group to 80%. The simulation codes are shown below:

```
###55%S1
RNGversion(vstr = 3.6)
set.seed(seed = 55)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = fiftyfive,size = 1))]
```

Then, we apply the following data analytic codes to the scenario:

```
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)

#result
exp.results[, mean(p < 0.05)]##true positive rate
1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)]
exp.results[, mean(upper_ci)]
exp.results[, mean(lower_ci)]
```

Research Question 3 (scenario 2)

In scenario 2, we expect the samples would think 55% corn syrup tastes the same as sugar. Therefore, the probability of success in both sample groups are set to be 50%. Below are the simulation code:

```
###55%S2
RNGversion(vstr = 3.6)
set.seed(seed = 55)
Experiment <- rep.int(x = 1:B, times = n)
Group = sample(x = c("Treatment", "Control"), size = n * B, replace = T)
sim.dat <- data.table(Experiment, Group)
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Dicision := round(x = rbinom(n = .N, prob = 0.5,size = 1))]
sim.dat[Group == "Treatment", Dicision := round(x = rbinom(n = .N, prob = 0.52,size = 1))]
```

The followings are the analysis codes:

```
exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]
DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
              rownames = F)

#result
exp.results[, mean(p < 0.05)]##true positive rate

1-exp.results[, mean(p < 0.05)]##false negative rate
exp.results[, mean(p >= 0.05)]##true negative rate
1-exp.results[, mean(p >= 0.05)]##false positive rate

exp.results[, mean(effect)]
exp.results[, mean(upper_ci)]
exp.results[, mean(lower_ci)]
```

Analytical results:

Based on the data that we collected and the analysis that we tested, the following table presents every research question and their two different scenarios. The analysis parameters

include mean effect in simulated data, 95% confidence interval of mean effect, percentage of false positives, percentages of true negatives, percentage of false negatives, and percentages of true positives

Research Question	Scenario	Mean Effect in Simulated Data	95% Confidence Interval of Mean Effect	Percentage of False Positives	Percentage of True Negatives	Percentage of False Negatives	Percentage of True Positives
Q1	No Effect	-0.2002	-0.3427906 To -0.0575323	80.8%	19.2%	19.2%	80.8%
Q1	Effect	-0.0026	-0.1504176 To 0.1451469	3.3%	96.7%	96.7%	3.3%
Q2	No Effect	-0.0494	-0.1969586 To 0.09808464	8.5%	91.5%	91.5%	8.5%
Q2	Effect	-0.0002	-0.1479422 To 0.1475977	3.6%	96.4%	96.4%	3.6%
Q3	No Effect	-0.3027	-0.4376882 To -0.1677029	99.4%	0.6%	0.6%	99.4%
Q3	Effect	-0.0240	-0.1717564 To 0.1237882	4.6%	95.4%	95.4%	4.6%

The ultimate goal is to find out a correct concentration of corn syrup to replicate the taste of the original sugar, and the proportion of the correct judgement between the treatment group and control group have a similarity. Thus, an ideal result is statistically rejected H1 and accepts H0. The probability to have an ideal result is a True Negative rate. Based on the previous table, the Percentage of True Negative is the parameter we need concern about.

In the results table, we would like to analyze the percentage of true negatives. Our expected effect is that the subjects cannot distinguish between a certain level of diluted corn syrup and sugar. The higher true negative rate represents more experiments revealing that a certain level of diluted corn syrup has a similar taste as sugar.

In research question 1, the percentage of the true negative of scenario 1 is 19.2%, and the percentage of the true negative of scenario 2 is 96.7%. Compared with the result of scenario 2 96.7%, the probability to have an ideal result in scenario 1 only has 19.2% which is quite small, and the gap between the two scenarios is huge.

In research question 2, the percentage of the true negative of scenario 1 is 91.5%, and the percentage of the true negative of scenario 2 is 96.4%. Compared with the result of scenario 2 96.7%, the probability to have an ideal result in scenario 1 have 91.5% which is quite large, and the gap between the two scenarios is small.

In research question 3, the percentage of the true negative of scenario 1 is 0.6%, and the percentage of the true negative of scenario 2 is 95.4%. Compared with the result of scenario 2

95.4%, the probability to have an ideal result in scenario 1 only has 0.6% which is extremely small, and the gap between the two scenarios is extremely huge.

Key Findings & Recommendations

According to our results table, In the effect scenario, most of the simulated study shows that 42% diluted corn syrup tastes the same as sugar. However, in both scenarios analysis, 48% diluted corn syrup has above 90% of true negatives rate. It also has the smallest true negatives rate difference between both scenarios.

Due to the slight true negatives' variance, 48% may most likely share a similar taste with sugar. Therefore, under current simulation results, we recommend the UK coke company to start switching the primary sweetener ingredients in the original coke from sugar to 48% diluted corn syrup. In such a way, the company could save a substantial amount of money from manufacturing and operating. Moreover, the tasting similarity would minimize the company's ingredient switching costs.

References

Alan Barclay, Philippa Sandall, Claudia Shwide-Slavin (2014). The Ultimate Guide to Sugars and Sweetener.

Bode J.W., Empie M.W., Brenner K.D (2014). Evolution of High Fructose Corn Syrup Within the Sweeteners Industry. In: Rippe J. (eds) Fructose, High Fructose Corn Syrup, Sucrose and Health. Nutrition and Health. Humana Press, New York, NY. https://doi-org.ezproxy.cul.columbia.edu/10.1007/978-1-4899-8077-9_9.

Coca-Cola Inc. Ingredients https://www.coca-cola.co.uk/brands/coca-cola-original-taste

Kevin Lyons (2015) Does Your Sample Size Matter? https://www.lipmanhearne.com/does-your-sample-size-matter/.

Kiyah J. Duffey and Barry M. Popkin (2009). High-Fructose Corn Syrup: Is this what's for dinner? doi: 10.3945/ajcn.2008.25825C.

Remy Melina (2010). What's the Difference Between Corn Syrup and Sugar? Live Science. Life's Little Mysteries.

Sievenpiper J.L (2014). Added Sugars and Health: Evidence from Prospective Cohort Studies and Controlled Dietary Trials. In: Rippe J. (eds) Fructose, High Fructose Corn Syrup, Sucrose and Health. Nutrition and Health. Humana Press, New York, NY. https://doi-org.ezproxy.cul.columbia.edu/10.1007/978-1-4899-8077-9_7.

Thomson (2012). Coca-Cola Taste Test: High Fructose Corn Syrup vs. Sugar https://www.huffpost.com/entry/coca-cola-taste-test_n_1324282.

Venditti, C., Musa-Veloso, K., Lee, H. Y., Poon, T., Mak, A., Darch, M., . . . Maia, J. (2020). Determinants of sweetness preference: A scoping review of human studies. Nutrients, 12(3), 718. doi:http://dx.doi.org.ezproxy.cul.columbia.edu/10.3390/nu12030718.

Vendredi (2012). Coca Cola targeting and positioning. Retrieved from: http://anandahussein.lecture.ub.ac.id/files/2015/09/Article-3.pdf.