GMMA 863 – Team Assignment

I recommend that each student think about the structure for each question before speaking with colleagues. The work on problems like this can be divided into conceptual and technical. The conceptual work is where the real value is – but if you divide and conquer the assignment before thinking through the conceptual problems, you will miss some key learning opportunities. (Arguing about it with your colleagues may also be a good learning experience.)

The results should be provided in a single ‘main’ document (probably a PDF) that is **print-ready** with a single supporting Excel file if required. You should assume that any supporting documentation may not be reviewed, so your entire answer should be in the main document.

The main document may contain scanned versions of hand-drawn diagrams / formulas if you like.

1. Replacing lightbulbs in a shopping mall can be quite a bit more expensive than you would expect. It requires a unionized crew to come in with specialized equipment. Believe it or not, a bulb replacement costs on average $20. While consulting to a property management firm, I was asked to provide input into the problem of lightbulbs.

The company was purchasing lightbulbs for $1.00 each. From experience they know the bulbs lasted, on average, 3 years before burning out. The sales representative of The Better Bulb company proposed that they could buy Better Bulbs for $2.00 which he claimed would last three times as long – though he might not be the most reliable person, given that he was trying to sell them lightbulbs. That said, The Better Bulb Company provided test data (see the Bulbs tab) on how long a random selection of lights lasted.

* 1. Construct the test appropriate to the client’s question using the process outlined in class.

First is to establish a hypothesis based on the client’s question, work out the cost of failure and opportunity costs.

The decision is whether to switch to Better Bulbs. In English, it means, <the client will switch to Better Bulbs> if <their bulbs cost less over the lifecycle compares to the current bulbs>.

Life cycle cost of current bulb over its life cycle = (cost of bulb + replacement cost) / (years of service) = (1 + 20 ) / 3 = $7.0

The test would then be H0: life cycle cost of the Better Bulb is equal or more than $7.0.

If we can reject the H0 with a reasonable alpha, the client should switch to Better Bulb.

**Cost of failure**. If we made a type I error and advised the client to made the switch when they shouldn’t. Assume the Better Bulb were in fact Worse Bulb, and they only last 3 years. Then the cost would be (2 + 20) / 3 = $7.3. For 2,000 light bulbs of the shopping mall, that could cost 0.3\*2000 = $600 per annum more over the life of the bulbs. The amount is not significant in the overall picture. 0.05 alpha level seems appropriate.

**Opportunity Cost**. If we made type II error and asked the client to not to switch when they should be switching. If the Better Bulb is indeed better, then its lifecycle cost could be as advertised (2 + 20) / 9 = $2.4, that means (7.0 – 2.4) = $4.6 saving per bulb p.a., thus 4.6\*2000 = $92,000 saving p.a. this saving amount would warrant further measurement and testing if the current data proves inconclusive.

* 1. Conduct the test and report the results with an alpha of 5%. Be sure to explain the relevant assumptions as required.

After transforming the burn time into an annual life cycle cost and test its average against the current of 7.0, using one-tail t-test, excel gave the below table as a result. With a p-value of 7.6E-124, the H0 of the Better Bulb cost more or equal to $7.0 p.a. were rejected.

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Equal Variances | |  |
|  |  |  |
|  | New annual cost | current |
| Mean | 4.476293157 | 7 |
| Variance | 0.202199025 | 0 |
| Observations | 100 | 100 |
| Pooled Variance | 0.101099512 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 198 |  |
| t Stat | -56.12409833 |  |
| P(T<=t) one-tail | 7.6265E-124 |  |
| t Critical one-tail | 1.652585784 |  |
| P(T<=t) two-tail | 1.5253E-123 |  |
| t Critical two-tail | 1.972017478 |  |

The assumptions here multiple:

* + 1. Random sampling of bulbs. It is given in the description, but it is provided by the vendor, a better understanding of the sampling methodology would be required
    2. Sample mean should follow normal distribution. this is satisfied with large random sample and CLT
    3. Scale of measurements. All are continuous data.
    4. The test used assumed equal variance of data, the second sample is actually just a single number here, 7, so this is satisfied.
  1. Should they switch to The Better Bulbs?

The client should switch. Cost of failure is small, and test is favourable.

1. Your manager doesn’t know much about analytics. That is too bad, but don’t worry, you’ll have his job soon enough. In the meantime he has made the following comments. If his suggestions / comments are good ones, explain why; if they are bad ones, explain why and what you should do to fix them or do better; if they do not make sense, make sense of them. Make sure your answer demonstrates that you have a sophisticated understanding of the issues involved.
   1. “The type 1 error is the probability that you will reject null and shouldn’t have.”
   2. “If you have a common source of variation in two data sets, you must use matched pairs or your results are going to be wrong.”
   3. “When picking a sample size, if you need to be within 2.5% of the true proportion, a sample of just over 1500 is a good choice. Well, at least that’s what someone told me.”
2. I offer a home inspection service for prospective homebuyers. I do not actually perform the home inspections myself, but rather pay one of several retired inspectors I know $500 per inspection to do them – they seem pretty happy with the relationship. I am happy with it too because I really like making a profit!

Clients typically shop around for inspection services and there are several local inspection firms. Historically I have quoted $700 for an inspection and have had clients accept that proposal 30% of the time. Recently I have suspected that I should raise my prices by $100, so for the past 30 jobs, I have proposed $800 and have sold the jobs 8 jobs.

* 1. Using this information, what is the 95% confidence interval for the proportion of jobs I will get at the new price?
  2. Suppose I really want to consider the option of increasing the price. Using the method I described in class, develop the test appropriate to this problem. You do not need to actually conduct the test, just explain to me what I should do, and being sure to cover all the important decisions involved in constructing the test.
  3. What would power mean in the context of this test?
  4. Is it possible for you to test this with the given information?

1. I was walking a dog (Banksy) who liked to stop to sniff things occasionally. On average, he did this about two times per 100 m we walked. I walk about 1km per hour – I know, that is slow, but with all the sniffing a walk takes time…
2. If I take him on a 40 min walk, what is the probability that he stops to sniff thinks more than 8 times?
3. If I break a walk into 15 blocks of 40m each, what is the probability that he will stop to smell something in more than 10 of those blocks?
4. Bonus Question (1 Mark): Suppose that in a job interview you were provided the above information from part a above and told to model the number of times Banksy stopped to sniff something as a normal distribution.
   1. What normal distribution would you use? (Hint: you should attempt to match the parameters of the normal to the distribution you want to approximate.)
   2. What probability would you estimate the probability of exactly 8 sniffs to be? (HINT: when using a continuous distribution to model a discrete distribution, it is common to use a ‘continuity correction factor’ to account for the discrete distribution’s probability mass on specific values. So P(X=4) in a discrete model would be calculated as P(3.5 < x < 4.5) in a continuous distribution.)
5. In my experience, developing a question and its answer is a great way to ensure you really know your material. To this end, develop one or more questions similar in difficulty to the ones I have provided.
   1. Provide a background story that can be resolved to a model that is in-scope for our class. The background story should contain sufficient information to answer the question, but also information that is not required. This forces the students to think through what is an is not important.
   2. Provide any data you require (I recommend choosing a question that does not require data if possible since data generation is an art in itself and is not in scope for this course.)
   3. Provide a solution for your problem.
   4. Write an explanation of why solution fits the problem. This part will require a bit more effort than you expect.
   5. If you like, please indicate that I may use this problem in future classes. You are absolutely free to say ‘no’ and there will be no consequences for those who do.