1 Review

500 days of Summer (2009 Movie, from Wikipedia)

On January 8, Tom meets Summer, the new assistant to his boss. Tom trained as an architect but works as a writer at a greeting card company, living in Los Angeles. Following a karaoke night, Tom’s co-worker, McKenzie, lets slip that Tom is attracted to Summer...

Q) What could Tom have done, if he had known Statistics? What ending could he have?

Hint: Let $X_i =$ Tom’s Day k’s i-th conversation with Summer in a particular day. The parameter of interest is $\mu =$ Summer’s true personality. Use CLT.

ANS: By the way, why don’t you just use CI 99 %, instead of CI 60 % or 80 %? CI 99 % sounds better, because it is more certain than CI 60 % or 80 %.

ANS: The more 100(1 - $\alpha$) % is bigger, the wider the CI is. There is a trade off. Consider an extreme case below.

eg) Mike took a mid exam 1. You think your score will be 50/100. But, of course, it could be 60, or 45. Mom “How’s your mid term 1? Can you guess?”

Case1: Mike: ”My score is 50 $\pm 50 = [0,100]$ with 100 % certainty.”

Case2: Mike: ”My score is 50 $\pm 2 = (48,52)$, by not sure ( = 10 % certainty).” (I made up the number 10 %)

If you are mom, what do you think about the Case1 answer? Mike’s answer is perfect in the sense that his statement is definitely true. Because everyone’s score is in the range [0,100]. Hence the interval (CI) is no use for mom.

Look at Case2. the interval (48,52) is very short compared to Case1. But of course, Mike himself does not believe that his score will be in that range. Neither does mom. Too short interval is also useless because it is too uncertain .

Hence, in reality, we commonly use 95 % C.I. or gives and compares 90, 95, 98 % C.I. at the same time. We do not use 50, 60, 70 % C.I. because they are too uncertain. (as far as I know)

2 Ch8.4 Hypotheses test

• Introduction

Someone sued Aki that he stole a pizza. In the court, there is one of two conclusions.

Conclusion 1: Aki is guilty ( $\iff$ Reject $H_0$ )

Conclusion 2: Aki is not guilty ( $\iff$ Do not reject $H_0$ )

Judges never say "Aki is innocent". Remember, Aki is innocent $\neq$ Aki is not guilty

"Aki is not guilty " just means that there is not enough evidence to conclude "Aki is guilty"

In reality, there is always possibility of mistakes in our decision.

Case1:
Truth: Aki did not steal a pizza.
Conclusion: Aki is guilty
($\equiv$ Type 1 error $\equiv \alpha \in [0,1]$)

Case2:
Truth: Aki stole a pizza.
Conclusion: Aki is not guilty
($\equiv$ Type 2 error $\equiv \beta \in [0,1]$)

Hypothesis test is everywhere in our life. We need to know related terminologies:

• Alternative Hypothesis $H_1$ (p340): the claim we wish to establish.

(eg. "Aki is guilty")

Note: Aki does not wish to establish this conclusion but the society does wish to establish this conclusion.

• Null Hypothesis $H_0$ (p340): negation of the claim.

(eg. "Aki is NOT guilty")
• **Type 1 Error** (p345): Rejection of $H_0$ when $H_0$ is true.
  ≡ Level of significance
  ≡ $\alpha$
  eg) An innocent person gets guilty.

• **Type 2 Error** (p345): Rejection of $H_1$ when $H_1$ is true.
  ≡ $\beta$
  eg) A criminal gets no charge.

Ideally, we want $\alpha = 0$, $\beta = 0$ (innocent people get no charge, and criminal get guilty). But you will see it is impossible to achieve this. In other words, we can not control (decrease) $\alpha$ and $\beta$ at the same time. If we decrease $\alpha$, then $\beta$ will increase. Vice versa.

In planet A, suppose judges always say "Guilty!" to everyone. In this case, $\beta = 0$ because all criminals are guilty. But all innocent people are guilty, hence $\alpha = 1$.

In planet B, suppose judge always says "Not guilty!" to everyone. In this case, $\beta = 1$ because all criminals are free. All innocent people are free, hence $\alpha = 0$.

Then automatically, we have

$H_0 : \mu = 270$

I mean, we will never have $H_0 : \mu = 120$, or $H_0 < 30$, etc, in our class.

Step 3) Decide level of significance $\alpha \in (0, 1)$ (= Type 1 error).
We usually use $\alpha = 0.05$

Step 4) Get $Z_{obs}$ (TEST STATISTIC for $\mu$).

$$Z_{obs} = \frac{\sqrt{n}(\bar{X} - 270)}{S}$$  (1)

where $S$= sample sd. i.e.,

$$S = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n - 1}.$$  (2)

Step 5-a) Find critical point = $c$
$c$= location of $N(0,1)$ which has right area = $\alpha$. Use Z-table.

Picture:

Then step 5-b) Find rejection region
Look at the above picture.

Step 6) Conclude your hypothesis test (Only two options !)
Option 1: If $Z_{obs} \in $ Rejection Region, then
Conclusion: "$\text{Reject } H_0.$"  

Option 2: Otherwise,
Conclusion: "$\text{NOT reject } H_0.$"

Let’s consider 3 kinds of hypothesis test for population mean. (p353) Use Bee example,

<table>
<thead>
<tr>
<th>Case</th>
<th>nick name</th>
<th>$H_0$</th>
<th>$H_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>One sided hypothesis</td>
<td>$H_0 : \mu = 270$</td>
<td>$H_1 : \mu &gt; 270$</td>
</tr>
<tr>
<td>C2</td>
<td>One sided hypothesis</td>
<td>$H_0 : \mu = 270$</td>
<td>$H_1 : \mu &lt; 270$</td>
</tr>
<tr>
<td>C3</td>
<td>Two sided hypothesis</td>
<td>$H_0 : \mu = 270$</td>
<td>$H_1 : \mu \neq 270$</td>
</tr>
</tbody>
</table>
The above example was C1. For C2, C3, all we need is modify Step1 and Step5.

- C2
  Step 1) $H_1: \mu < 270$
  Step 5-a) Find critical point = $c$
  $c$ = location of $N(0,1)$ which has left area = $\alpha$. Use Z-table.
  Picture:

  Step 5-b) Find rejection region
  Look at the above picture.

- C3
  Step 1) $H_1: \mu \neq 270$
  Step 5-a) Find critical point = $c$
  $c$ = location of $N(0,1)$ which has right area = $\alpha/2$.
  Use Z-table.
  Picture:

  Step 5-b) Find rejection region
  Look at the above picture.

Eg1) For Bee example, do C1. with $H_0: \mu = 270$, and $\alpha = 0.05$
Eg2) For Bee example, do C1. with $H_0: \mu = 270$, and $\alpha = 0.10$
Eg3) For Bee example, do C2. with $H_0: \mu = 270$, and $\alpha = 0.05$
Eg4) For Bee example, do C3. with $H_0: \mu = 270$, and $\alpha = 0.05$
Eg5) For Bee example, do C1. with $H_0: \mu = 230$, and $\alpha = 0.05$