Challenge Question: Hypothesis Testing & Confidence Interval

During the European football championships in 2008, and the football World Cup in 2010, **an octopus called Paul** living at an aquarium in Oberhausen, Germany, was used to predict the outcome of football matches, mostly involving the German national football team. To obtain Pauls' predictions, his keepers at the aquarium would present him with two boxes of food before each match. Each box was covered in the flag of the two nations that were participating, and the box that Paul chose to feed from first determined which nation he predicted would win.

Paul was asked to predict the outcome of $\mathbf{14}$ matches, $\mathbf{12}$ of which involved Germany. He correctly predicted the outcomes of $\mathbf{12}$ matches, only incorrectly guessing that Germany would beat Croatia in the Euro 2008 group stage, and that Germany would beat Spain in the Euro 2008 final. Some people claimed he was an **"animal oracle"**:

Answer the questions below **regarding the claim that Paul the Octopus is an oracle**. Provide working, reasoning or explanations that you have used, as appropriate.

1. Calculate an estimate of Paul's success rate at predicting football matches. Calculate a $95\ \%$ confidence interval for this estimate, and summarise/describe your results appropriately. Show all workings as required.
2. Using hypothesis testing, test the hypothesis that Paul just `got lucky` and was randomly guessing (probability for success 0.5) the outcomes of the football matches. Write down explicitly the hypothesis that you are testing, and then calculate a **p-value** using the approximate approach for testing a Bernoulli population. What does this **p-value** suggest?
3. **Paul has a sister named Emily**. The aquarium also wants to know whether Emily has the same clairvoyant ability similar to that of Paul's. Thus, they also conduct another similar experiment on Emily. Out of $\mathbf{14}$ matches, Emily correctly guessed $\mathbf{8}$ of them. Test the hypothesis that Emily has the same abilities as Paul's (i.e. difference between the two Bernoulli population), calculate the appropriate **p-value** for the null hypothesis, and give a conclusion on the hypothesis as well.
4. Given your analysis, do you think that Paul is an oracle (can see the future!), or can you identify any potential weaknesses in the experimental setup or way in which the data was sampled, feel free to perform a search for Paul and his predictions, you can start using this Wikipedia article (https://en.wikipedia.org/wiki/Paul_the_Octopus)?

From the question it is given that:

1. Total number of Matches Paul was asked to predict the outcome of is 14.
2. Paul correctly predicted the outcome of 12 matches.
3. Out of 14 matches, 12 involved Germany.
4. Incorrect predictions by Paul were 2 (Germany beats Croatia, Germany beats Spain)

Solution 1 :

We proceed as follows:

Let us take S_{Paul} such that it represents the success rate of Paul's prediction for a match.

We know Paul predicted 12 macthes correctly out of 14 matches. Hence,

$$S_{Paul} = \frac{12}{14}$$

$$S_{Paul} = 0.85714$$

Also for non-success rate $S_{\bar{Paul}}$,

$$S_{\bar{Paul}} = 1 - S_{Paul}$$

$$S_{\bar{Paul}} = 1 - 0.85714$$

$$S_{\bar{Paul}} = 0.14286$$

Now for Standard Error (SE), we proceed as:

$$SE = \sqrt{\frac{S_{Paul} \times S_{\bar{Paul}}}{TotalMatches}}$$

$$SE = \sqrt{\frac{0.85714 \times 0.14286}{14}}$$

$$SE = \sqrt{\frac{0.12245}{14}}$$

$$SE = 0.09352$$

Now for a Confidence Interval of 95%, we proceed as:

For $\alpha = 100 - 0.95 = 0.05$,

$z_{\frac{\alpha}{2}} = Q(1 - \frac{\alpha}{2})$, Where Q is the $QuantileFunction$ and $z_{\frac{\alpha}{2}}$ is the $100(1 - \frac{\alpha}{2})$ percentile of the unit normal

That is,

$z_{0.025} = Q(0.975)$ which is ≈ 1.96

Therefore,

For 95% Confidence Interval (CI):

$$95\%\ CI = S_{Paul} \pm 1.96(SE)$$

$$95\%\ CI = 0.85714 \pm 1.96 \times 0.09352$$

$$95\%\ CI = (1.0404, 0.6738)$$

Hence, From the Confidence Interval, it is observed that there is only 5% chance for the success rate of Paul ($S_{Paul} = 0.85714$) to fall out of the range 0.6738 to 1.0404.

Solution 2 :

From the question, it is given that probability of success where Paul was randomly guessing and got lucky is 0.5.

For the Hypothesis, we proceed as follows:

From the given information, we compute for Null Hypothesis (H_0) and Alternative Hypothesis (H_a)

$For H_0$

$\mu = 0.5$

$For H_a$

$\mu \neq 0.5$

For Standard Error (SE):

$SE = \frac{\sqrt{0.5 \times 0.5}}{14}$

$SE = 0.13363$

Now, we compute the z_{test}:

From Part 1, we have $S_{Paul} = 0.85714$,

$z_{test} = \frac{(S_{Paul} - 0.5)}{SE}$

$z_{test} = \frac{(0.85714 - 0.5)}{0.13363}$

$z_{test} = 2.6726 \approx 2.67$

Now, we proceed as:

$p_{value} = 2 \times Pr(Z \geq 2.67)$

From the Z-score Table, the z_{score} value is computed to be 0.9962, Hence for the p-value, it is : 1- 0.9962 = 0.0038.

Hence,

$p_{value} = 2 \times 0.0038$

$p_{value} = 0.0076$

From the above value, our informal grading of p_{value} resulted in telling that $p_{value} < 0.05$, defining that it is the strong evidence against the Null.

Henceforth, we can reject the Null Hypothesis and can conclude that there is more than enough evidence to say that Paul is clairvoyant/prophetic, that is, Paul has the ability to predict the future happenings.

Solution 3 :

It is given that:

1. Paul has a sister named Emily.
2. Out of 14 matches Emily correctly guessed 8 of them.

According to question, for the Hypothesis, we proceed as follows:

Let us take H_0 as Null Hypothesis and H_a as Alternate Hypothesis, such that:

For H_0:

$\mu_{Paul} = \mu_{Emily}$

For H_a:

$\mu_{Paul} \neq \mu_{Emily}$

From Solution 2.1, we know that:

$$S_{Paul} = 0.85714$$

Simillarly,

$$S_{Emily} = \frac{8}{14} = 0.57142$$

Hence,

$$\mu_{combined} = \frac{12+8}{14+14}$$

$$\mu_{combined} = \frac{20}{28}$$

$$\mu_{combined} = 0.71428$$

It was understood that the number of observation n was the same for both the cases in Paul and Emily as they both were given a total of 14 matches to predict. Therefore,

$$n_{Paul} = n_{Emily} = 14$$

Now, for Standard Error (SE), we proceed as:

$$SE = \sqrt{\mu_{combined}(1 - \mu_{combined}) \times (\frac{1}{n_{Paul}} + \frac{1}{n_{Emily}})}$$

$$SE = \sqrt{0.71428 \times (0.28572) \times (2 \times \frac{1}{14})}$$

$$SE = \sqrt{(0.20408) \times (0.14285)}$$

$$SE = \sqrt{(0.029153)}$$

$$SE = 0.17074$$

Then,

$$z_{test} = \frac{S_{Paul} - S_{Emily}}{SE}$$

$$z_{test} = \frac{0.85714 - 0.57142}{0.17074}$$

$$z_{test} = 1.67342 \approx 1.67$$

Now, we proceed as:

$$p_{value} = 2 \times Pr(Z \geq 1.67)$$

From the Z-score Table, the z_{score} value is computed to be 0.9525, Hence for the p-value, it is : 1- 0.9525 = 0.0475.

$$p_{value} = 2 \times 0.0475$$

$$p_{value} = 0.095$$

Since $p_{value} > 0.05$, we conclude to have weak/no evidence against the Null, therefore, we reject the Null Hypothesis and conclude that there is not enough proof to state that, Emily has different abilities than her brother Paul.

Solution 4 :

Refering to Paul's Wikipedia article (https://en.wikipedia.org/wiki/Paul_the_Octopus) and FIFA World Ranking (2008) Article (http://www.football-rankings.info/2008/10/fifa-ranking-november-2008-preview-ii.html) , the following points were observed:

1. Paul's keepers and home was in Germany and the predictions made involved all the matches with Germany, there can be a substantial motive that Paul got used to seeing German Flag everytime he was given the food that he choose German Flag.
2. It is also known that the German Football Team is considered one of the best teams and have a higher statistical win advantage over others. As ranked by FIFA world rankings, German Football Team was ranked at 2nd place in the world in the year 2008.

In conclusion, Paul cannot be considered as an oracle as there are various other factors that are rather not justified. If Paul was to predict all the mathces that took place in the World Cup and still showed a higher successful prediction rate, then it could have been stated otherwise.

References

1. LATEX Math for Undergrads. Retrieved on 10/10/2020 from: http://tug.ctan.org/info/undergradmath/undergradmath.pdf (http://tug.ctan.org/info/undergradmath/undergradmath.pdf)
2. Z-Table. Retrieved on 10/10/2020 from: http://www.z-table.com/ (http://www.z-table.com/)
3. Football rankings. Christian. Updated on 20/10/2008. Retrieved on 10/10/2020 from: http://www.football-rankings.info/2008/10/fifa-ranking-november-2008-preview-ii.html (http://www.football-rankings.info/2008/10/fifa-ranking-november-2008-preview-ii.html)

In [ ]: