DSE 210: Worksheet #10 - Hypothesis Testing

Professor: A. Enis Çetin

Teaching Assistant: Shivani Agrawal

Joshua Wilson **A53228518**

Problem 1

No, the increase does not necessarily show that the public's health got worse.

Problem 6

Given : 5,400 heads in 10,000 tosses

(a) The null hypothesis is H_0: The coin is not biased (i.e. $p(H) = 0.5$).
The alternative hypothesis is H_a: The coin is biased (i.e. $p(H) \neq 0.5$).

(b) The observed number of heads can be modeled as $N(\hat{\mu} = np, \hat{\sigma}^2 = np(1-p))$, where p equals the probability of a single head outcome and n is the number of tosses.

If $p(H) = 0.5$ and $n = 1,000$, then the distribution of the observed number of heads in 1,000 tosses is $N(10,000 \times 0.5, 10,000 \times 0.5 \times 0.5) = N(5,000, 2,500)$, so the estimated standard deviation $\hat{\sigma} = \sqrt{2,500} = 50$.

We can calculate a z statistic as $\dfrac{\text{(observed outcome} - \text{expected outcome)}}{\text{standard deviation}}$.

Since we observed 5,400 head outcomes, the z statistic $= \dfrac{5,400 - 5,000}{50} = \dfrac{400}{50} = 8$. The associated p-value is much lower than 0.001.

(c) The probability of getting 5,400 heads using an unbiased coin tossed 10,000 times if H_0 is true is equal to the p-value calculated in part b, so we reject H_0.

Problem 8

Since we reject the null hypothesis with a sufficiently low p-value, and fail to reject the null hypothesis if the p-value is too high, a higher p-value is better for the null hypothesis.

Problem 10

Let X_1 be the distribution of hours worked for freshmen at public universities, and let X_2 be the distribution for private universities.

H_0: The average hours worked is the same.

$X_1 \sim N(\mu, \sigma_1 = \frac{10.5}{\sqrt{10,000}})$,

$X_2 \sim N(\mu, \sigma_2 = \frac{9.9}{\sqrt{10,000}})$,

$(X_1 - X_2) \sim N(0, \sqrt{\sigma_1^2 + \sigma_2^2} \approx \sqrt{0.332^2 + 0.313^2} \approx 0.4564$

The observed change is $12.2 - 9 = 3$, and the expected change under H_0 is 0, so the z-stat $= \frac{3}{0.4564} \approx 6.57$, and the associated p-value is less than 0.001.

It is therefore very unlikely that the null hypothesis is true and the difference between the hours worked is due to chance.