Homework 8

PSTAT 5A: Spring 2023, with Ethan P. Marzban

Instructions

- Please submit your work to Gradescope by no later than **11:59pm on Wednesday, May 31**. As a reminder, late homework will not be accepted.
- Recall that you will be asked to upload a **single** PDF containing your work for *both* the programming and non-programming questions to Gradescope.
  - You can merge PDF files using either Adobe Acrobat, or using adobe's online PDF merger at this link.

Caution

Be aware that some parts may be easier (or, in fact, may *need* to be) computed using Python. If you do use Python for any part, please write down the code you used.

Problem 1: Look at All Those Chickens!

The average weight of an adult male chicken is claimed to be 5.7 lbs. A representative sample of 36 adult male chickens is taken, and it is found that the weights of these sampled chickens have an average of 6.1 lbs and a standard deviation of 0.9 lbs. Suppose that we wish to test the original claims (that the true average weight of an adult male chicken is 5.7 lbs) against a two-sided alternative.

a. Define the parameter of interest.

Solution: The parmaeter of interest is μ, the true average weight of an adult male chicken.

b. State the null and alternative hypotheses in terms of the parameter you defined in part (a).

Solution:

$$\left[\begin{array}{rl} H_0 : & \mu = 5.7 \\ H_A : & \mu \neq 5.7 \end{array} \right.$$

c. What distribution do we use when performing our hypothesis test? Be sure to include any/all relevant parameter(s)!

Solution:

- **Are weights of adult male chickens normally distributed?** No; or, at least, it is not stated that they are.
- **Is our sample large enough?** Yes; $n = 36 \geq 30$.
- **Do we know σ or s?** We know s, not σ.

Based in the answers to these question, we use the t_{35} distribution.

d. Compute the value of the test statistic.

Solution:

$$\text{TS} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{6.1 - 5.7}{0.9/\sqrt{36}} = 2.67$$

e. Compute the p-value of the test statistic.

Solution: Because we are using the $t-$distribution, we need to use Python. Our p-value can be computed using

```
scipy.stats.t.cdf(-2.67, 35)  + (1 - scipy.stats.t.cdf(2.67, 35))
```

or, equivalently,

```
2 * scipy.stats.t.cdf(-2.67, 35)
```

which yields a result of 0.011.

f. Conduct the test using the p-value value, and state the conclusions of the test in the context of the problem using an $\alpha = 0.05$ level of significance.

Solution: We reject whenever the p-value is smaller than the level of significance. Since, $0.011 < 0.05$ we reject the null:

> At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject the claim that the true average weight of an adult male chicken is 5.7 lbs in favor of the alternative that the true average weight is *not* 5.7 lbs.

g. Compute the critical value of the test (again using an $\alpha = 0.05$ level of significance).

Solution: From the table, we see that the critical value is 2.03.

h. Conduct the test using critical value from part (g), and state the conclusions of the test in the context of the problem.

Solution: We reject when the test statistic is larger than the critical value. Since $2.67 > 2.03$, we again reject the null:

> At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject the claim that the true average weight of an adult male chicken is 5.7 lbs in favor of the alternative that the true average weight is *not* 5.7 lbs.

i. Redo the test, now using an $\alpha = 0.01$ level of significance. Do your conclusions change? If so, state the new conclusions in the context of the problem.

Solution: It may be easiest to use the p-value here. Since we are still using a two-sided test, our p-value does not change. Instead, we now compare 0.011 to $\alpha = 0.01$; since $0.011 > 0.01$, we now fail to reject:

> At an $\alpha = 0.01$ level of significance, there was insufficient evidence to reject the claim that the true average weight of an adult male chicken is 5.7 lbs in favor of the alternative that the true average weight is *not* 5.7 lbs.

Intuitively, this makes sense. In reducing α we are reducing our chances of committing a Type I error; i.e. of rejecting the null when the null was in fact true. This means that we reject for fewer values, and hence it is not surprising that we would fail to reject with this new, smaller, level of significance.

Problem 2: Turn On the Light

GauchoBrite-brand lightbulbs are claimed to burn with an average wattage of 60 Watts. In actuality, the distribution of wattages across all *GauchoBrite*-brand lightbulbs is known to be roughly normal with a standard deviation of 27 Watts. A representative sample of 25 lightbulbs was taken; these 25 lightbulbs had a combined average wattage of 57 Watts. **Use a 5% level of significance.**

a. Define the parameter of interest.

Solution: Let μ denote the true average wattage of a *GauchoBrite*-brand lightbulb.

b. State the null and alternative hypotheses in terms of the parameter you defined in part (a).

Solution:

$$\left[\begin{array}{rl} H_0 : & \mu = 60 \\ H_A : & \mu \neq 60 \end{array}\right.$$

c. What distribution do we use when performing our hypothesis test? Be sure to include any/all relevant parameter(s)!

Solution: Since the distribution of wattages is stated to be normal, we can automatically use the $\mathcal{N}(0,\ 1)$ distribution.

d. Compute the value of the test statistic.

Solution:

$$\mathrm{TS} = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{57 - 60}{27/\sqrt{25}} = -0.56$$

e. Compute the p-value of the test statistic.

Solution: Since we are using a two-sided alternative, our p-value can be computed in Python using

```
scipy.stats.norm.cdf(-0.56)  + (1 - scipy.stats.norm.cdf(0.56))
```

or

```
2 * scipy.stats.norm.cdf(-0.56)
```

which gives a result of 0.576.

f. Conduct the test using the p-value value, and state the conclusions of the test in the context of the problem.

Solution: We reject only when the p-value is smaller than the level of significance. In this case, since the p–value is larger than α, we fail to reject:

> At a 5% level of significance, there was insufficient evidence to reject the claims that the Wattage of a *GauchoBrite*-brand lightbulb is 60 Watts in favor of the alternative that the true average Wattage is *not* 60 Watts.

g. Compute the critical value of the test.

Solution: We run

```
-scipy.stats.norm.pdf(0.05)
```

which yields a critical value of 1.96.

h. Conduct the test using critical value, and state the conclusions of the test in the context of the problem.

Solution: We reject only when $|\text{TS}|$ is larger than the critical value; since $|\text{TS}| = |-0.56| = 0.56 \leq 1.96$ we fail to reject:

> At a 5% level of significance, there was insufficient evidence to reject the claims that the Wattage of a *GauchoBrite*-brand lightbulb is 60 Watts in favor of the alternative that the true average Wattage is *not* 60 Watts.

Problem 3: Drinking Water

City officials of *Gauchonia* believe that 15% of households in *Gauchonia* have slightly elevated levels of fluoride in their drinking water. To test this claim, a representative sample of 375 households is taken. It is found that 13.6% of households in this sample have elevated levels of fluoride in their drinking water.

a. Check that the success-failure conditions are met.

Solution:

1) $np_0 = (375)(0.15) = 56.25 \geq 10$

2) $n(1 - p_0) = (375)(1 - 0.15) = 318.75 \geq 10$

b. Assuming the null is correct, what is the distribution of the test statistic?

Solution: Since the success-failure conditions are met, we can invoke the CLT for Proportions to conclude (along with our Standardization result)

$$\text{TS} \overset{H_0}{\sim} \mathcal{N}(0,\ 1)$$

c. Suppose the city officials wish to test their claims against a two-sided alternative at an $\alpha = 0.05$ level of significance. Compute the p-value of the test statistic, and use this to form a conclusion. Be sure to state your conclusion in the context of the problem.

Solution: We first compute the value of the test statistic:

$$\text{TS} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\widehat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{(0.136) - (0.15)}{\sqrt{\frac{(0.15)\cdot(1-0.15)}{375}}} \approx -0.76$$

The $p-$value is therefore computed using

```
2 * scipy.stats.norm.cdf(-0.76)
```

which gives a p–value of approximately 0.447. As this is greater than the 5% level of significance utilized, we fail to reject the null:

> At a 5% level of significance, there was insufficient evidence to reject the null hypothesis that the true proportion of households with elevated levels of Fluoride in their drinking water is 15% in favor of the alternative that the true proportion is *not* 15%.

d. Now, suppose the city officials wish to test their claims against a lower-tailed alternative, still at an $\alpha = 0.05$ level of significance. Compute the p-value of the test statistic, and use this to form a conclusion. Be sure to state your conclusion in the context of the problem.

Solution: The value of the test statistic does not change; what does change is how we compute the p–value:

```
scipy.stats.norm.cdf(-0.76)
```

which yields 0.224. This is still above the 5% level of significance utilized, meaning we still fail to reject the null:

> At a 5% level of significance, there was insufficient evidence to reject the null hypothesis that the true proportion of households with elevated levels of Fluoride in their drinking water is 15% in favor of the alternative that the true proportion is *less than* 15%.

Problem 4: Programming

Part (a): Recap of LaTeX Syntax

❗ Task 1

First, add a second-level header that says `Task 1`. Then, typeset the following set of equations into a Markdown Cell. Pay very close attention to the alignment of equations, and make sure your parentheses display correctly. (Also, you may need to look up how to place a box around text in LaTeX)

$$\begin{aligned}\mathbb{P}(4 \leq X \leq 7) &= \mathbb{P}(X \leq 7) - \mathbb{P}(X \leq 4) \\ &= \mathbb{P}\left(\frac{X - 3}{1.4} \leq \frac{7 - 3}{1.4} \right) - \mathbb{P}\left(\frac{X - 3}{1.4} \leq \frac{4 - 3}{1.4} \right) \\ &= \mathbb{P}\left(\frac{X - 3}{1.4} \leq 2.86 \right) - \mathbb{P}\left(\frac{X - 3}{1.4} \leq 0.71 \right) \\ &= 0.9979 - 0.7611 = \boxed{0.2368}\end{aligned}$$

Solutions

```
\begin{align*}
 \mathbb{P}(4 \leq X \leq 7)  & = \mathbb{P}(X \leq 7) - \mathbb{P}(X \leq 4)   \\
  & = \mathbb{P}\left( \frac{X - 3}{1.4} \leq \frac{7 - 3}{1.4} \right) - \mathbb{P}\left( \f
  & = \mathbb{P}\left( \frac{X - 3}{1.4} \leq 2.86 \right) - \mathbb{P}\left( \frac{X - 3}{1
   & = 0.9979 - 0.7611 = \boxed{0.2368}
\end{align*}
```

Part (b): Numbered Equations

We have not yet talked about how to number equations in LaTeX. The syntax for creating a numbered equation is:

```
\begin{equation}{
  <whatever equation you want}
}\end{equation}
```

For example,

$$f_X(x) = x^2 \tag{1}$$

was created using the syntax

```
\begin{equation}
  f_X(x) = x^2
\end{equation}
```

❗ Task 2

Create a labeled equation (you can use whatever equation you want) in a new Markdown Cell.

Note

Your equation will not appear with a number in your `.ipynb` file; the equation number will only display in your final `.pdf`.

Solutions

Answers will vary. As an example:

```
\begin{equation}{
  \sum_{k=1}^{n} k = \frac{n (n + 1)}{2}
}\end{equation}
```

typesets as

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \tag{2}$$

One of the benefits of labeling your equations is that you can reference them later! To create a labeled equation that is referable, use the syntax

```
\begin{equation}{\label{eq:<name>}
  <your equation>
}\end{equation}
```

Then, to reference the equation later, use `\ref{eq:<name>}` where `<name>` is whatever you called your equation. For example:

$$a^2 + b^2 = c^2 \tag{3}$$

was created using

```
\begin{equation}{\label{eq:pyth}
  a^2 + b^2 = c^2
}\end{equation}
```

meaning I can reference equation (3) using the code `\ref{eq:pyth}`.

Task 3

Create a labeled equation (you can use whatever equation you want) in a new Markdown Cell that is labeled, and then refer to the equation in a markdown cell underneath (just like we did above).

Note

Again, neither the equation number nor the referenced equation number will appear in your `.ipynb` file; they will only appear in your final `.pdf`.

Solutions

Answers will vary. As an example:

```
\begin{equation}{\label{eq:trig1}
  \sin(2\theta) = 2 \sin(\theta) \cos(\theta)
}\end{equation}

We now reference equation (\ref{eq:trig1}).
```

typesets as

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta) \tag{4}$$

We now reference equation (4).