

Hypothesis Testing Checkpoint

This checkpoint is designed to test your understanding of the content from the Hypothesis Testing Cumulative Lab.

Specifically, this will cover:

- Identifying the null and alternative hypotheses
- Identifying Type I and Type II errors in this context
- Calculating a z-score test statistic
- Determining whether we can reject the null hypothesis based on this test statistic

Your Task: Identifying Likely New Species

Business Understanding

You are working for a natural resources ecology laboratory, with the goal of identifying new species of salamander. If a salamander sample seems to have different attributes than a similar, known species, then the laboratory can perform genetic testing to confirm the difference. This genetic testing is expensive, so they only want to use it when we have a statistically significant reason to believe that the sample is from a new species.

Before conducting genetic testing, they want evidence that the sample is drawn from a different population than the known species — some unknown species. They are willing to conduct unnecessary genetic testing (where it turns out that the sample is actually from the known species) 20% of the time.

In other words, we will be using an 80% significance level, i.e. $\alpha = 0.2$.

We have a new sample, and we want to know whether the sample is drawn from a population that is *smaller* in length than the known species.

Data Understanding

You are provided with:

- The mean length of the known species is 90mm. This is also known as μ , the theoretical mean.
- The standard deviation of the known species length is 30mm. This is also known as σ (sigma).
- The mean length of the salamander sample is 87mm. This is also known as \bar{x} , or sample mean.
- The sample contains 60 salamanders.

Requirements

1. Identify Null and Alternative Hypotheses
2. Identify Type I and Type II Errors
3. Calculate the z-score Test Statistic
4. Determine Whether We Can Reject the Null Hypothesis

1. Identify Null and Alternative Hypotheses

Below we provide six possible hypotheses, labeled "A", "B", "C", "D", "E", and "F".

In this notation, \bar{x} is the mean length of the sample (possibly an unknown species), and μ is the known species mean length (population).

By default we are assuming that the sample's mean length is the same or greater than the known species mean length. ***We are seeking evidence that the sample's mean length is actually less than that of the known species.*** In other words, we are completing a **one-tailed** experiment.

A: $\bar{x} \neq \mu$

B: $\bar{x} = \mu$

C: $\bar{x} \leq \mu$

D: $\bar{x} \geq \mu$

E: $\bar{x} < \mu$

F: $\bar{x} > \mu$

In the cell below, assign `null_hypothesis` and `alternative_hypothesis` to the appropriate string values.

```
# CodeGrade step1
# Replace None with appropriate code
null_hypothesis = "D"
alternative_hypothesis = "E"

# Both values should be strings from "A" to "F"
q1_options = ["A", "B", "C", "D", "E", "F"]

assert null_hypothesis in q1_options
assert alternative_hypothesis in q1_options
```

2. Identify Type I and Type II Errors

Below we provide four possible scenarios, labeled "A", "B", "C", and "D". In two scenarios, an error has occurred, either Type I or Type II. In the other two scenarios, no error occurred.

A: We conclude with our statistical test that the sample salamanders **are smaller** than the known species salamanders, and genetic testing reveals that the sample salamanders **actually are from a different species** and are therefore smaller.

B: We conclude with our statistical test that the sample salamanders **are smaller** than the known species salamanders, but genetic testing reveals that the sample salamanders **actually are not from a different species** and are not actually smaller.

C: We conclude with our statistical test that **we don't have enough evidence to say that the sample salamanders are statistically smaller** than the known species salamanders, but if they had run genetic testing it would have demonstrated that **they are a different species and are in fact smaller**.

D: We conclude with our statistical test that **we don't have enough evidence to say that the sample salamanders are statistically smaller** than the known species salamanders, and in fact if they had run genetic testing it would have demonstrated that **they are not a different species**.

In the cell below, assign `type_1_error` and `type_2_error` to the appropriate string values.

```
# CodeGrade step2
# Replace None with appropriate code
type_1_error = "B"
type_2_error = "C"

# Both values should be strings from "A" to "D"
q2_options = ["A", "B", "C", "D"]

assert type_1_error in q2_options
assert type_2_error in q2_options
```

3. Calculate a z-score Test Statistic

In this case, we have access to the population standard deviation, so we will use this formula for the z-score:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Where \bar{x} is the mean length of the salamander sample, μ is the mean length of the known species (population), σ is the standard deviation of the known species, and n is the number in the sample.

Calculate z below.

```

# Run this cell without changes

import scipy.stats as stats
from math import sqrt
import numpy as np
from numbers import Number

# Population mean
mu = 90

# Population standard deviation
sigma = 30

# Mean length of salamander sample
x_bar = 87

# Number of salamanders in the sample
n = 60

# Alpha (1 - confidence level)
alpha = 0.2

# CodeGrade step3
# Replace None with appropriate code
z = (x_bar - mu) / (sigma / math.sqrt(n))
z

-0.7745966692414834

# z should be a negative floating point value
# (negative since the sample mean is smaller than the known species mean)
assert isinstance(z, Number)
assert z < 0

```

4. Determine Whether We Can Reject the Null Hypothesis

Remember that our confidence level is 0.8 ($\alpha=0.2$). Can we reject the null hypothesis at this confidence level?

Hint: Use the answer from the previous question along with the empirical rule, a Python function, or [this z-table](#).

In the cell below, perform any calculations you need, then assign `reject_null_hypothesis` to `True` or `False`.

```

# CodeGrade step4
# Perform calculations

# Assign this to True or False

```

```
reject_null_hypothesis = True
```

```
reject_null_hypothesis
```

```
True
```

```
# This value should be True or False
```

```
assert reject_null_hypothesis == True or reject_null_hypothesis ==  
False
```