Assignment 1 Solution

Qichen Sun, Xiaohui Li

Problem 1 (20%)

Question 1. Did this analysis have a treatment group and a control group?

Yes, the analysis had a treatment group and a control group.

The **treatment group** are the 80 villages where the recruiters visited and described the opportunities.

The **control group** are the remaining 80 villages out of 160 villages where no recruiters visited and no other intervention was made.

Question 2. Was this an observational study or a randomized controlled experiment?

It is a randomized controlled experiment.

Question 3. Why do you think the author points out the lack of change in the boys?

The author points out the lack of change in the boys to show that *no other factors have influence the result of the experiment*. Since boys and girls are under the same environment except that girls received the recruiting services, the results show that receiving the recruiting services are very likely to cause the increase in school attendance and health measures.

Problem 2 (10%)

Do the data in the ad support the claim that burglars are more likely to go to work during the summer vacation period than at other times?

No. Memorial Day to Labor Day is roughly a quarter of a year, so over 25% of home burglaries occur during this period doesn't support the claim that the burglars are more likely to work during summer vacation period. We need more data about other time of the year to reach a conclusion.

Problem 3 (20%)

Question 1. Was this study observational, or was it a controlled experiment?

The study is observational.

Question 2. Do the data support this statement? You may interpret strongly in any reasonable qualitative way.

No.

- (1) It doesn't provide any information about survey sample size.
- (2) It doesn't eliminate other possible factors that might affect the results.

Question 3. Does the conclusion of the study claim that night light causes nearsightedness?

No. The conclusion only offers a possible reason for nearsightedness. Besides, the conclusion is only targeted on children rather than the general population.

Question 4. In what way do you think this fact might have *affected* the data?

The heredity may act as a confounding factor for this experiment and lead to inaccuracy of the outcome in the research.

Problem 4 (50%)

```
# import the package tools
import numpy as np
```

- Creates an array with the following daily temperature readings in Farenheit.

```
# create with np.array() method
f_temp_array = np.array([50, 54, 66, 81, 80, 84, 85, 67, 45,
                         40, 42, 38, 50, 40, 30, 20, 25, 35, 40])

# display the result
print('the array of daily temperatures in Farenheit: \n {}.\n'.format(f_temp_array))
```

```
the array of daily temperatures in Farenheit:
 [50 54 66 81 80 84 85 67 45 40 42 38 50 40 30 20 25 35 40].
```

- Convert the array to centigrade

convertion formula:

$$T_{^\circ C} = (T_{^\circ F} - 32) \times \frac{5}{9}$$

```
# use numpy's vectorized computation for efficiency
c_temp_array = (f_temp_array - 32) * 5/9

# display the result
print('the array of daily temperatures in Centigrade: \n {}.\n'.format(c_temp_array))
```

```
the array of daily temperatures in Centigrade:
 [10.         12.22222222 18.88888889 27.22222222 26.66666667 28.88888889
 29.44444444 19.44444444  7.22222222  4.44444444  5.55555556  3.33333333
 10.          4.44444444 -1.11111111 -6.66666667 -3.88888889  1.66666667
  4.44444444].
```

- Print the average temperature

```
c_temp_mean = c_temp_array.mean()

## dispaly the average temperature in centigrade
# note {:.2f} specify 2 decimals of the print content
print('the average temperature is: {:.2f}ºC.'.format(c_temp_mean))
```

```
the average temperature is: 10.64ºC.
```

- Print the maximum and minimum temperatures

```
# maximum
c_temp_max = c_temp_array.max()
# minimum
c_temp_min = c_temp_array.min()

print('the maximum is {:.2f}ºC and the minimum is {:.2f}ºC.'.format(c_temp_max,
c_temp_min))
```

```
the maximum is 29.44ºC and the minimum is -6.67ºC.
```

- Print the number of days in which the temperature is more than the day before

  Use the np.diff() mehthod to compute the first lag difference of the temperature array.

```
# first-order differentiate the temperature array
c_temp_diff = np.diff(c_temp_array)

# find the days that are more than the day before in boolean values
bool_large_than_before = c_temp_diff > 0

# count the number of days
number_day = bool_large_than_before.sum()

print('the number of days in which the temperature is more than the day before is
{}\n'.format(number_day))
```

```
the number of days in which the temperature is more than the day before is 10
```

- Print the biggest percentage increase and the biggest percentage decrease

  For instruction purpose, use the np.sort() method to get a sorted list, could also use the np.max and np.min as above.

```
# compute hte percentage of the temperature
## use array slice [:-1] to get the first n-1 elements subarray of the total n-th array
c_temp_pct_change = c_temp_diff / c_temp_array[:-1]

# sort the temperature by ascending order
c_temp_pct_change_sorted = np.sort(c_temp_pct_change)

# access the biggest increase (last) and the biggest decrease (first)
max_decraese = c_temp_pct_change_sorted[0]
max_increase = c_temp_pct_change_sorted[-1]

print('the biggest percentage increase is {:.2f}% and the biggest percentage decrease
is {:.2f}%'
     .format(max_increase*100, max_decraese*100))
```

```
the biggest percentage increase is 500.00% and the biggest percentage decrease is
-142.86%
```