Analyze the Data

Based on your analysis of the original, unmodified training data:

Answer the following questions based on summary statistics you have calculated from the training dataset. **Where applicable, be sure to round your answers to** *four decimal places* (e.g., 5.3002).

Minimum prevalence of undernourishment

1.0/1.0 point (graded)

2.4934 ✔

2.4934

Submit You have used 2 of 2 attempts

Maximum prevalence of undernourishment

1.0/1.0 point (graded)

59.0898 ✔

59.0898

Submit You have used 2 of 2 attempts

Mean prevalence of undernourishment

1.0/1.0 point (graded)

15.5107 ✔

15.5107

Submit You have used 2 of 2 attempts

Median prevalence of undernourishment

1.0/1.0 point (graded)

12.1187 ✔

12.1187

Submit You have used 2 of 2 attempts

Standard deviation of undernourishment prevalence

1.0/1.0 point (graded)

11.6104 ✔

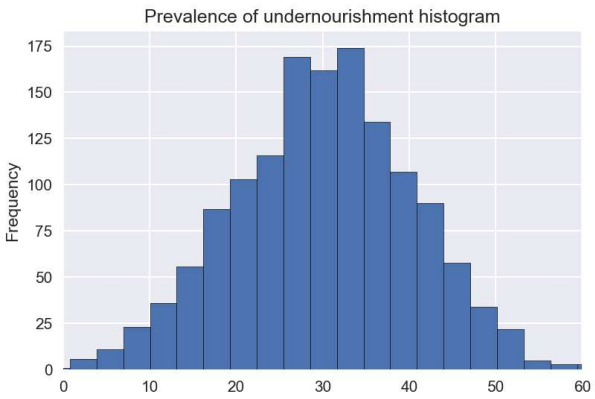
11.6104

Submit You have used 2 of 2 attempts

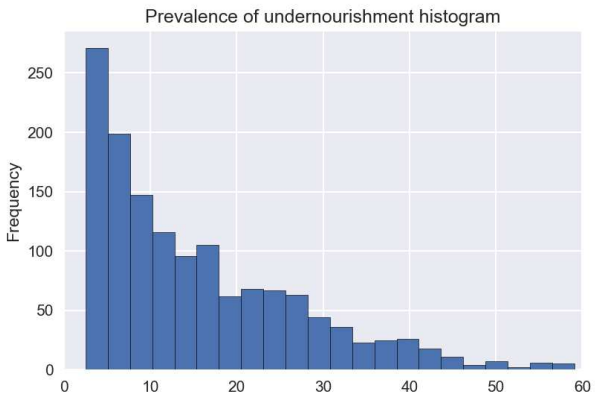
Distribution of undernourishment

1/1 point (graded)

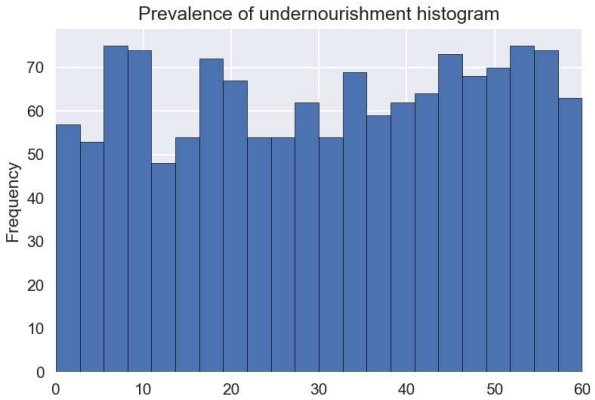
Which of these histograms most closely resembles the distribution of undernourishment prevalence?



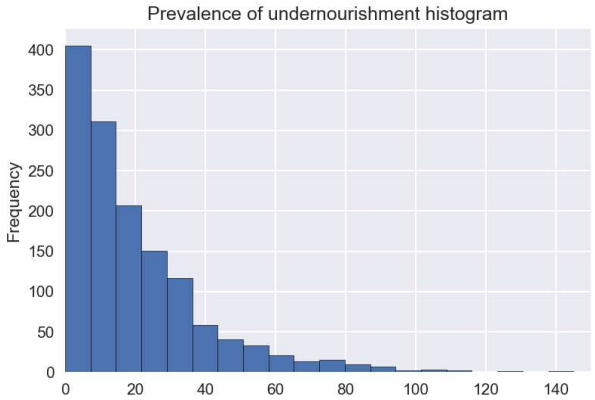
Histogram A



Histogram B



Histogram C



Histogram D

- ○ Histogram A
- ◉ Histogram B ✔
- ○ Histogram C
- ○ Histogram D

Submit

You have used 2 of 2 attempts

Undernourishment in 2000 vs. 2010

2.0/2.0 points (graded)
Which **two** of the following statements are true?

- [ ] The median prevalence of undernourishment is higher in 2010 than in 2000.
- [x] The median prevalence of undernourishment is lower in 2010 than in 2000.
- [x] The mean prevalence of undernourishment in 2010 is 13.9.
- [ ] The mean prevalence of undernourishment in 2010 is 16.2.

✔

Submit You have used 2 of 2 attempts

Access to improved water sources vs. prevalence of undernourishment

2.0/2.0 points (graded)
Which of the following best describes the relationship between access to improved water sources and prevalence of undernourishment?

- ○ A — Higher access to improved water sources is associated with a higher prevalence of undernourishment, on average.
- ◉ B — Higher access to improved water sources is associated with a lower prevalence of undernourishment, on average. ✔
- ○ C — There is not a strong and obvious correlation between access to improved water sources and prevalence of undernourishment, on average.

Submit You have used 2 of 2 attempts

Urban population (total and share) vs. undernourishment

0.0/1.0 point (graded)

Total urban population refers to the number of people living in urban areas. Urban population **share** refers to the fraction of the total population living in urban areas (i.e. total urban population / total population). Which of the following statements is true?

- A — Total urban population (but not urban population share) is negatively correlated with prevalence of undernourishment.
- B — Urban population share (but not total urban population) is negatively correlated with prevalence of undernourishment.
- C — Both urban population share and total urban population are negatively correlated with prevalence of undernourishment.
- D — Neither urban population share nor total urban population are negatively correlated with prevalence of undernourishment.

Submit You have used 2 of 2 attempts

Prevalence of undernourishment in a given year vs. 10 years prior

0.0/2.0 points (graded)

For each country, we have the prevalence of undernourishment for a range of years (between 5 and 16 years of data depending on the country). Within each country, we want to explore the correlation between the prevalance of undernourishment in a given year with the prevalence of undernourishment 10 years prior. This means that

for countries wtih 16 years of data, there will be 6 years where we can calculate this difference.

Which of the following statements is true about the apparent relationship between the prevalence of undernourishment in a given year and the prevalence of undernourishment 10 years prior, on average?

- A — There is a strong positive correlation between prevalence of undernourishment in a given year and its value 10 years prior. Pearson's r (correlation coefficient) is 0.92.
- B — There is a strong positive correlation between prevalence of undernourishment in a given year and its value 10 years prior. Pearson's r (correlation coefficient) is 0.86.
- C — There is a strong negative correlation between prevalence of undernourishment in a given year and its value 10 years prior. Pearson's r (correlation coefficient) is -0.86.
- D — There is a strong negative correlation between prevalence of undernourishment in a given year and its value 10 years prior. Pearson's r (correlation coefficient) is -0.92.

Submit You have used 0 of 2 attempts

Ethics question: 1

2/2 points (graded)

Data ethics: With great power comes great responsibility. Part of the responsibilities of a data scientist includes thinking about the ethical implications of your work, such as was discussed during the MPP course. It is therefore important to be able to spot when different ethical issues may arise.

For the following scenario, familiarize yourself with, then use this Data Science Ethics Checklist, (part of the open source deon command line tool) to find the most relevant

ethical concerns.

The United Nations wants to use a model that estimates national levels of chronic hunger to direct food aid appropriately. This model is trained on publicly accessible, annual, country-level data. Which set of checklist items best captures the most salient ethical concerns?

- (•) D4: Explainability, C3: Honest representation, C1: Missing perspectives ✔
- ( ) B2: Right to be forgotten, E2: Roll back, A3: Limit PII exposure
- ( ) B1: Data security, A3: Limit PII exposure, C4: Privacy in analysis

Submit You have used 2 of 2 attempts

✔ Correct (2/2 points)

Ethics question: 2

2/2 points (graded)

For the following scenario, familiarize yourself with, then use this Data Science Ethics Checklist, (part of the open source deon command line tool) to choose the best answer.

A certain country's Ministry of Health requires more fine-grained maps of hunger than the UN's national estimates provide. They therefore decide to collect data on malnourishment at a subnational level. The data collection involves weighing and measuring the height of participants to calculate their body mass index. Which option best captures the most salient ethical concerns?

- ( ) D4: Explainability, E2: Roll back, B2: Right to be forgotten
- (•) A1: Informed consent, A3: Limit PII exposure, C4: Privacy in analysis ✔

Submit You have used 1 of 1 attempt

✔ Correct (2/2 points)

Ethics question: 3

2/2 points (graded)

For the following scenario, familiarize yourself with, then use this Data Science Ethics Checklist, (part of the open source deon command line tool) to choose the best answer.

A certain country builds a model to estimate food stamp usage to inform the design of welfare policies. A politician running for office uses these estimates to dole out free food in areas of high food stamp usage to try to sway voters to vote for him. What checklist item best captures what is happening here?

- ○ C2: Dataset bias
- ○ A2: Collection bias
- ○ D3: Metric selection
- ◉ E4: Unintended use ✔

Submit You have used 2 of 2 attempts

✔ Correct (2/2 points)

Learn About Verified Certificates