Analysis of Psychological Data

Lab 3. Hooked on a Distribution: Central Tendency, Variability, Z-Scores, and Percentiles

Ihnwhi Heo (iheo2@ucmerced.edu)
Quantitative Methods, Measurement, and Statistics
Website: https://ihnwhiheo.github.io
Office: https://ucmerced.zoom.us/j/2093557522 (Friday 1:30 - 3:30 pm)

What are we going to do?

Recap to give you a big picture

Central tendency & variability

Z-scores & percentiles

(Primers on the formula included)

Group activity

Reflecting upon three weeks

Fundamentals (design, measurement, analysis)

Some basic stat terms (population, parameter, sample, statistic)

Focus on variables (type, measurement scale)

Fiddling with variables (visualization, distribution)

We are playing the game with the distribution

Distribution

An intuitive way to understand how the values of a variable are spread over

We are interested in describing the distribution

What would be efficient and informative ways?

Distribution

Measures of central tendency

- Reflect where 'values of a variable (i.e., distribution)' are centered
- **Mean**, median, and mode

Measures of variability

- Reflect how much 'values of a variable (i.e., distribution)' are dispersed
- Range, deviation score, sum of squares, **variance**, **standard deviation**

Primers on formula

Notation

- Uppercase Roman letters that are usually near the end of the alphabet denote variables (e.g., X and Y)
  - A numerical subscript represents an individual value of that variable (e.g., X_3 is the third value of the variable X)
- Greek letters are used represent population parameters (e.g., μ, σ^2, σ)
- Roman letters are used to represent sample statistics (e.g., $\bar{X}$, s^2, s)
  - An alphabetical subcript represents the corresponding variables (e.g., σ^2_X, s_X)
- Lowercase n denotes the sample size whereas the uppercase N indicates the population size

Primers on formula

Notation

- Σ (the uppercase Greek letter sigma) indicates the operation of summation (i.e., the addition of scores)
  - $\Sigma_{i=1}^{5} X_i = X_1 + X_2 + X_3 + X_4 + X_5$
- Numerical exponents indicate multiplying that number itself multiple times
  - $X^5 = X \times X \times X \times X \times X$
- (Positive) square root
  - $\sqrt{X^2} = (X^2)^{\frac{1}{2}} = X$
- Summation and exponents combined
  - $\Sigma_{i=1}^{5} (X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + (X_4 - \bar{X})^2 + (X_5 - \bar{X})^2$
    $= (X_1 - \bar{X}) \times (X_1 - \bar{X}) + (X_2 - \bar{X}) \times (X_2 - \bar{X}) + \ldots + (X_5 - \bar{X}) \times (X_5 - \bar{X})$

Central tendency

Summary

	Sample	**Population**
Mean	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$
Median	(When aranged) The $\frac{n+1}{2}$th value	(When aranged) The $\frac{N+1}{2}$th value
Mode	The most frequent value	The most frequent value

Mean vs. median

- When there are outliers (i.e., extreme values), consider using the median as a central tendency measure

Mode

- Useful for variables measured on nominal and ordinal scales

Variability

Range

$max - min$

Deviation score

$X_i - \bar{X}$

Sum of squares

$\Sigma_{i=1}^{n}(deviation\ score)^2 = \Sigma_{i=1}^{n}(X_i - \bar{X})^2$

Variability

Variance

- Sample variance = $s_X^2 = \frac{\Sigma_{i=1}^{n}(deviation\ score)^2}{n-1} = \frac{sum\ of\ squares}{n-1} = \frac{\Sigma_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$
- Population variance = $\sigma_X^2 = \frac{\Sigma_{i=1}^{N}(deviation\ score)^2}{N} = \frac{sum\ of\ squares}{N} = \frac{\Sigma_{i=1}^{N}(X_i-\mu_X)^2}{N}$

Standard deviation

- Sample standard deviation = $\sqrt{sample\ variance} = \sqrt{s_X^2} = s_X = \sqrt{\frac{\Sigma_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$
- Population standard deviation = $\sqrt{population\ variance} = \sqrt{\sigma_X^2} = \sigma_X = \sqrt{\frac{\Sigma_{i=1}^{N}(X_i-\mu_X)^2}{N}}$

Variability

Some tips on interpreting the standard deviation

Average distance from the mean

What if the standard deviation is zero?

Central tendency and variability

Summary

	Sample	**Population**
Mean	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$
Median	(When aranged) The $\frac{n+1}{2}$th value	(When aranged) The $\frac{N+1}{2}$th value
Mode	The most frequent value	The most frequent value
Variance	$s_X^2 = \frac{\Sigma_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$	$\sigma_X^2 = \frac{\Sigma_{i=1}^{N}(X_i - \mu_X)^2}{N}$
Standard deviation	$s_X = \sqrt{\frac{\Sigma_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$	$\sigma_X = \sqrt{\frac{\Sigma_{i=1}^{N}(X_i - \mu_X)^2}{N}}$

Central tendency and variability

Tips

Be sure to see whether you are looking at raw data or frequency data

- Raw data

50, 50, 60, 70, 70, 70, 80, 90, 90, 90, 90, 100

- Frequency data

Scores	**50**	**60**	**70**	**80**	**90**	**100**
Frequency	2	1	3	1	4	1

Shall we calculate the mean from respective data formats?

Z-score

Why do we use the z-score?

So far, we have played with the single distribution. We oftentimes want to compare values from **different distributions** (i.e., measured on different scales)!

How to calculate it?

If the variable is normally distributed, we can convert any value on that variable into a z-score!

$$z_i = \frac{X_i - \mu_X}{\sigma_X}$$

Z-score

Interpretation

How far certain values are from the mean (in standard deviation units)

- Say you have a z-score of 1.5. What does this mean?
- Say you have 4 z-scores (-7.0, 0.5, 2.3, and 6.5). Which score is the farthest from the mean?

How much proportion of the population falls lower than certain values -> Percentile

- Say you have a z-score, which has the percentile of 85th. What does this mean?

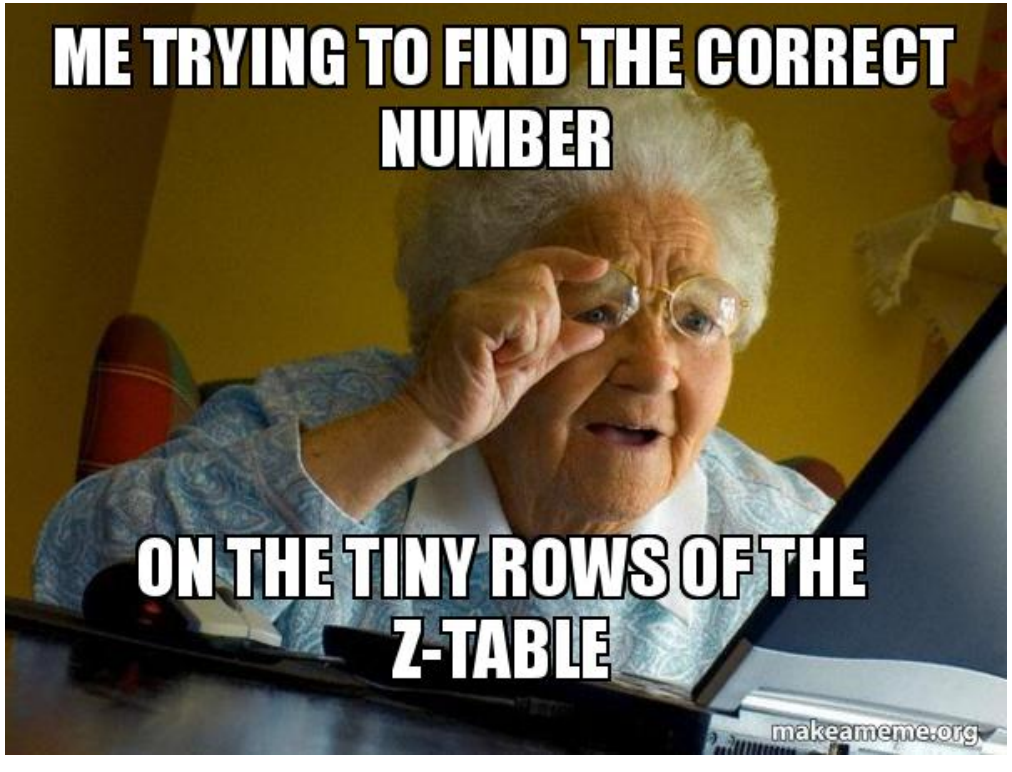
Z-score

We know how to calculate the z-score and interpret it...

What should you do next?

Z-score

OPEN YOUR EYES to see the Z-table...



Percentile

Again, an **easy** definition

The n^{th} percentile means that $n\%$ of the population data lie below that point.

- Say you are a cool person that belongs to the top 5% of the population. What is your percentile?

Z-table is all about finding the percentile of a z-score

1. Find the decimal places of the z-score
2. Find the corresponding percentile

Never forget to print the Z-table for your exam :)

Percentile

Tips

First, be sure if you need to convert a raw score to a z-score or from a z-score to a raw score

Second, be sure to know what you actually need to know

- Are you interested in the percentage of the population below a certain point or above it?
- Are you interested in a certain range, or a point would be enough?

Do not be confused by lots of numbers!

Group activity

Case 1

Stephen, a statistician at the United States Air Force, has collected the data on pilots' satisfaction with aviation training. With the data collected, his research team would like to evaluate the current training system and know how to improve it in the future. See the collected data following:

25, 35, 18, 20, 15, 5, 25, 10, 22, 25

Can you answer below?

- What is $\Sigma_{i=1}^{10} X_i$ (assuming X refers to the score variable)?
- What are the mean, the median, and the mode?
- What are the range, the sum of squares, the variance, and the standard deviation?
- What is the relative frequency of the score 25?
- What is the cumulative frequency of the score equal to 18 or lower?

Group activity

Case 2

Makram is interested in programming, so he took two programming courses in R and Python. His achievement was excellent, but he is curious to compare his performances. In the R programming course, he scored 125 where the mean and the standard deviation are 80 and 18. In the Python programming course, he scored 92 where the mean and the standard deviation are 72 and 10. Assume that all scores are normally distributed.

Can you answer below?

- What are the z-scores in each programming course? Can you interpret them?
- What is the percentage of students who did better than Makram in the Python programming course?
- What course did Makram showed a better achievement relative to his classmates?
- Say Ihnwhi also took the same R programming course, and his z-score is -0.25.
  - What would be Ihnwhi's raw score?
  - What is the percentage of students who did worse than Ihnwhi?
  - What is the percentage of students who scored between Makram and Ihnwhi?

Before you go home...

Any questions or comments?

Thanks! Have a wonderful weekend!