PSY 3307

Probability

Jonathan A. Pedroza, MS, MA

Cal Poly Pomona

2021-09-07

Agenda

1. Understand what probability is
2. Probability Distributions
3. Probability from the Standard Normal Curve
4. Random Sampling & Sampling Error
5. The Decision of Whether a Sample Represents a Population

Finally!!!!!!

Understanding Probability

- Examination of the representative sample of a given population
- **Random Sampling** is the selection of samples so that members of the population have the same chance of being selected
    - Drawing names from an extremely large hat
- **probability** or **p** is the likelihood of an event when a population is randomly sampled
    - equal to the event's' relative frequency in the population

Making Sure it wasn't by chance

- How often an event occurs over the long run
- Knowing the probability of heads when flipping a coin 10 times
- If it happens a lot, it has a high probability
    - if it happens a small amount of times, it has a low probability

Occurs over the long run is how often occurs in the population of all possible events

- When something is happening a certain amount, we are describing the event's' relative frequency
    - proportion of time that the event out of all the events that might occur in the population
- **Probability** = how many times is heads going to occur out of all the total number of events
- **Odds** = how many times is heads going to occur compared to how many times is tails going to occur

Occurs over the long run is how often occurs in the population of all possible events

- probability is the confidence we have that a particular random event will occur
- Example of typing where you have a relative frequency of .85 of your words are spelled correctly
    - we'd expect this to continue
- a relative frequency of 0 is a probability of .0

Important for Inferential Statistics

- If we have a probability of .95, we are confident that in our samples, we would be correct 95% of the time
- Probabilities are addressed as proportions, which are between 0 and 1
- Percentages are between 0% and 100%
- Over the long run, we are comparing our sample of observations to what we believe will occur in the population

Probability Distributions

- indicates the probability of all possible events in a population
- Observing relative frequency of events creates an *empirical probability distribution*
    - empirical just means that its been tested
    - Number of students with pictures vs just their names on zoom
- another way is through the theoretical probability distribution
    - we determine what an expected relative frequency is
    - coin toss where half should theoretically be heads and the other half should be tails
    - sex differences in specific majors

Example of Probabilities

- 52 playing cards
  - 26 black
  - 26 red
  - 13 hearts
  - 13 clubs
  - 13 diamonds
  - 13 spades

Obtaining Probability from Normal Curve

- Probabilities are the same as percentiles but instead of using percentages we are sticking between 0 and 1
- The middle will always be the 50% or .5

Probability of Sample Means

- Can calculate the probability of obtaining particular sample means using the sampling distribution of means
- The sampling distribution shows that there will always be more values in the middle than on the tails
- The distribution shows what occurs by random chance

Review

- We are working with a sampling distribution
- we need to get the standard error

$$ \sigma{\overline{X}} = \frac{\sigma}{\sqrt{N}} $$

- Then we get the z-score from the standard error

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

- Then we can get the proportion/percentile of whatever scores we are looking at

- "The probability of randomly selecting a particular sample mean is also the probability of randomly selecting a sample of participants whose scores produce that mean."

Reminders

- The larger the value of a sample's mean z-score is, the less likely it will occur when samples are drawn from the actual scores being tested in a population
- z-scores are a good way to get the probability

Random Sampling & Sampling Error

- We can never truly be certain that a sample is representative of a population
- **Representative Sample** is a sample where participants and scores reflect the participants and scores of a given population
- The issue is that my dumb ol' luck, you could have a sample that isn't representative of your population
- **Sampling Error** is when random chance produces an unrepresentative sample from a population
    - Makes sample statistics differ from the population parameters
- This can be a problem for researchers and even some fields as a whole due to not being able to replicate

Two possibilities of sampling error

1. We may randomly gotten a sample of high or low scores
2. Our sample just does not represent the population

Decide whether or not our sample represents a population

- If our sample mean is the same as the population mean, that is highly unlikely but possible
- If the value is similar and close to the population mean, then it may be more likely to resemble the population
- If the value is away from the population mean, then you reject that the sample represents the population

Quote from the Book

"THIS IS THE LOGIC USED IN ALL INFERENTIAL PROCEDURES, SO BE SURE THAT YOU UNDERSTAND IT: WE WILL ALWAYS BEGIN WITH A KNOWN, UNDERLYING RAW SCORE POPULATION THAT A SAMPLE MAY OR MAY NOT REPRESENT. FROM THE UNDERLYING RAW SCORE POPULATION, WE ENVISION THE SAMPLING DISTRIBUTION OF MEANS THAT WOULD BE PRODUCED. THEN WE DETERMINE THE LOCATION OF OUR SAMPLE MEAN ON THE SAMPLING DISTRIBUTION. THE FARTHER INTO THE TAIL OF THE SAMPLING DISTRIBUTION THE SAMPLE MEAN IS, THE LESS LIKELY THAT THE SAMPLE COMES FROM AND REPRESENTS THE UNDERLYING RAW SCORE POPULATION THAT WE BEGAN WITH."

Setting up Sampling Distribution

- Two steps to see if our sample represents the population

1. Determine the probability of obtaining our sample from the population
2. decide whether the probability makes the sample likely to actually represent the population

- Using statistics, we decide any sample beyond a specific point is no longer representative
    - region(s) of rejection
- **Region of rejection** is the area of a sampling distribution containing means so unlikely to occur that we reject that they represent the population
    - these regions could be representative of other populations
    - they are on the tails of a sampling distribution contianing sample means less likely to occur
- If the mean looks like it could represent the population, then we retain the sample

Setting up Sampling Distribution

- **Criterion** is the probability that defines whether a sample is reflective of a population or not
    - .05 or .01 is often used in the social sciences
- Sample means within the 5% or 1% of the sampling distribution are unlikely to occur and we reject that our sample is representative of the population
- If we are talking about scores above or below our criterion then we look at both tails
    - if we are only interested in scores above or below then we focus on extreme values in one tail of the sampling distribution

Identifying the Critical Value

- We'll first do this with z-scores
- **Critical value** is the score that marks the inner edge of the region of rejection in a sampling distribution
    - values outside this critical value are rejected as being representative of the population
- We get our critical value by looking at the criterion of .05 (for most cases)
- For our z-scores, we see that .025 of the distribution is beyond the z-score of **1.96**
- We reject that our sample is representative of the population if the sample mean is greater than the absolute value of 1.96

Deciding Whether the Sample is Representative

- We would get the standard error and the z-score to see if it is higher or lower than the critical value of +-1.96

Summary

1. Set up sampling distribution
    - include criterion probability
    - locate the region of rejection and the critical value
2. Compute sample mean and z-score
    - standard error of the mean
    - z-score
3. Compare z-score to critical value
    - if above +-1.96 then reject that your sample is representative of the population
    - if under +-1.96 then retain that your sample is representative of the population

Other Ways to set up the Sampling Distribution

- if only interested in seeing scores lower than the population mean, you would have a criterion of .05 on the lower tail of your distribution
  - this creates a different critical value