Statistical Inference Overview

Young W. Lim

2024-07-24 Wed

Outline

"Understanding Statistics in the Behavioral Sciences" R. R. Pagano

Populations and Samples (1)

- population: everything in the group that we want to learn about.
- sample: a part of the population.

- Examples of populations and a sample from those populations:

Population	Sample
All of the people in Germany	500 Germans
All of the customers of Netflix	300 Netflix customers
Every car manufacturer	Tesla, Toyota, BMW, Ford

https://www.w3schools.com/statistics/statistics_statistical_inference.php

- For good statistical analysis,
  the sample needs to be as *similar*
  as possible to the population.

- If they are *similar enough*, we say
  that the sample is representative of the population.

- The sample is used to make conclusions
  about the whole population.

`https://www.w3schools.com/statistics/statistics_statistical_inference.php`

- If the sample is not *similar enough* to the whole population, the conclusions could be useless.

- Many words have specific meanings in statistics.
- The word population normally refers to a group of people.
- In statistics, it is any specific group that we are interested in learning about.

https://www.w3schools.com/statistics/statistics_statistical_inference.php

- Using <u>data analysis</u> and <u>statistics</u>
  to make <u>conclusions</u> about a <u>population</u> is called
  statistical inference.

- The main <u>types</u> of statistical inference are:
  - Estimation
  - Hypothesis testing

`https://www.w3schools.com/statistics/statistics_statistical_inference.php`

- Statistics from a sample are used
  to estimate population parameters.

- *The most likely* value is called a point estimate.

- There is always uncertainty when estimating.

https://www.w3schools.com/statistics/statistics_statistical_inference.php

- The uncertainty is often expressed as confidence intervals
  defined by a *likely* lowest and highest value
  for the parameter.

- An example could be a confidence interval
  for the number of bicycles a Dutch person owns:
  - The average number of bikes a Dutch person owns
    is between 3.5 and 6.

https://www.w3schools.com/statistics/statistics_statistical_inference.php

- a method to check if a claim about a population is true.
- checks how likely it is that a hypothesis is true is based on the sample data.
- there are different types of hypothesis testing.

https://www.w3schools.com/statistics/statistics_statistical_inference.php

Hypothesis Testing (2)

- the steps of the test depends on:
    - Type of data (categorical or numerical)
    - If you are looking at:
        - a single group
        - comparing one group to another
        - comparing the same group before and after a change

https://www.w3schools.com/statistics/statistics_statistical_inference.php

- a hypothesis is a claim about a population parameter
- a hypothesis test is a formal procedure
  to check if a hypothesis is true or not.
- examples of claims that can be checked:
  - the average height of people in Denmark is more than 170 cm.
  - the share of left handed people in Australia is not 10%.
  - The average income of dentists is less the average income of lawyers.

https://www.w3schools.com/statistics/statistics_hypothesis_testing.php

1. *State* your research hypothesis as a null hypothesis (H_0) and alternate hypothesis and (H_a or H_1).
2. *Collect data* in a way designed to test the hypothesis.
3. *Perform* an appropriate statistical test
4. *Decide* whether to reject or fail to reject your null hypothesis
5. *Present* the findings in your results and discussion section.

https://www.scribbr.com/statistics/hypothesis-testing/

- Statistical tests are used in hypothesis testing.
  They can be used to:
    - determine whether a predictor variable has
      a statistically significant relationship
      with an outcome variable.
    - estimate the difference between two or more groups.

https://www.scribbr.com/statistics/statistical-tests/

- Statistical tests assume a null hypothesis of no relationship
  or no difference between groups.
  Then they determine whether the observed data
  fall outside of the range of values predicted by the null hypothesis.

https://www.scribbr.com/statistics/statistical-tests/

- Statistical tests work by calculating a test statistic
  - a number that describes how much
    the relationship between variables in your test differs
    from the null hypothesis of no relationship
- then calculates a p value (probability value).
  the p-value estimates how likely it is
  that you would see the difference described by the test statistic
  if the null hypothesis of no relationship were true.

https://www.scribbr.com/statistics/statistical-tests/

- If the value of the test statistic is <u>more</u> <u>extreme</u> than the statistic calculated from the null hypothesis, then you can infer a <u>statistically significant</u> relationship between the predictor and outcome <u>variables</u>

- If the value of the test statistic is <u>less extreme</u> than the one calculated from the null hypothesis, then you can infer <u>no</u> <u>statistically significant</u> relationship between the predictor and outcome <u>variables</u>.

https://www.scribbr.com/statistics/statistical-tests/

- You can <u>perform</u> statistical tests on data
  that have been <u>collected</u> in a <u>statistically valid</u> manner
  - either through an experiment,
  - or through observations made using probability sampling methods.

- For a statistical test to be <u>valid</u>,
  your sample size needs to be <u>large</u>
  enough to <u>approximate</u> the true distribution
  of the population being studied.

- To <u>determine</u> which statistical test to use,
  you need to know:
  - whether your <u>data</u> meets certain assumptions.
  - the <u>types of variables</u> that you're dealing with.

`https://www.scribbr.com/statistics/statistical-tests/`

- Statistical tests make some common assumptions about the data they are testing:
  - Independence of observations (a.k.a. no autocorrelation):
  - Homogeneity of variance:
  - Normality of data:

`https://www.scribbr.com/statistics/statistical-tests/`

- Independence of observations (a.k.a. no autocorrelation):
  - the observations / variables you include in your test
    are not related
  - for example,
    multiple measurements of a single test subject are not independent,
    while measurements of multiple different test subjects are independent

https://www.scribbr.com/statistics/statistical-tests/

- Homogeneity of variance:
  - the variance within each group being compared is similar among all groups.
  - If one group has much more variation than others, it will limit the test's effectiveness.

https://www.scribbr.com/statistics/statistical-tests/

- Normality of data:
  - the data follows a normal distribution (a.k.a. a bell curve).
  - This assumption applies only to quantitative data.

https://www.scribbr.com/statistics/statistical-tests/

Types of data (1)

- Data is a specific measurement of a variable
  it is the value you record in your data sheet.
  Data is generally divided into two categories:
    1. Quantitative data represents amounts
    2. Categorical data represents groupings

https://www.scribbr.com/statistics/statistical-tests/

- a variable that contains quantitative data
  is a quantitative variable;
- a variable that contains categorical data
  is a categorical variable.
- Each of these types of variables can be broken down
  into further types.

https://www.scribbr.com/statistics/statistical-tests/

Quantitative variables

- When you collect quantitative data, the numbers you record represent real amounts that can be added, subtracted, divided, etc. There are two types of quantitative variables: discrete and continuous.

https://www.scribbr.com/statistics/statistical-tests/

- Categorical variables represent groupings of some kind. They are sometimes recorded as numbers, but the numbers represent categories rather than actual amounts of things.

- There are three types of categorical variables: binary, nominal, and ordinal variables.

`https://www.scribbr.com/statistics/statistical-tests/`

Independant and dependant variables

- Experiments are usually designed to find out what effect one variable has on another – in our example, the effect of salt addition on plant growth.

- You manipulate the independent variable (the one you think might be the cause) and then measure the dependent variable (the one you think might be the effect) to find out what this effect might be.

- You will probably also have variables that you hold constant (control variables) in order to focus on your experimental treatment.

https://www.scribbr.com/statistics/statistical-tests/

- Hypothesis testing is based on making two different claims about a population parameter.
- The null hypothesis (H_0) and the alternative hypothesis (H_1) are the claims.
- The two claims needs to be mutually exclusive, meaning only one of them can be true.
- The alternative hypothesis is typically what we are trying to prove.
- For example, we want to check the following claim:
  - "The average height of people in Denmark is more than 170 cm."

https://www.w3schools.com/statistics/statistics_hypothesis_testing.php

tests	
• one-sample test	comparing sample mean, population mean
• two-sample test	comparing two independent sample means
• paired test	comparing two related sample means

tests	test conditions
• t-test	1. when the population variance is *known*
	2. when the sample size is *large*
• z-test	1, when the population variance is *unknown*
	2. the sample size is *small*

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

one sample z-test	sample mean, population mean
	known population var / large sample size
one sample t-test	sample mean, population mean
	unknown population var / small sample size
two sample z-test	two *independent* sample means
	known population var / large sample size
two sample t-test	two *independent* sample means
	unknown population var / small sample size
paired t-test	two *related* sample means
	unknown population var / small sample size

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

Summary (3) comparing proportions

one sample propotion test	sample proportion, population proportion when $np \geq 10$ and $n(1-p) \geq 10$
two sample proportion test	two *independent* sample proportions when $np \geq 10$ and $n(1-p) \geq 10$

test conditions
the normal approximation is used
when both $np \geq 10$ and $n(1-p) \geq 10$
(data should have at least 10 "successes" and at least 10 "failures")

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

Summary (4)

compare variances between	
sample variance, known population variance	Chi-square test
two independent sample variances	F-test
observed frequencies, expected frequencies	goodness of fit test
observed frequencies, expected frequencies	contingency tables
means of three or more independent samples	ANOVA (Analysis of Variance

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

Tests for Comparing Means (1)

- One-sample z-test:
  - used to compare the mean of a sample
    to a known population mean
  - used when the population variance is known,
    or the sample size is *large* ($n > 30$).

- Two-sample z-test:
  - used to compare the means of two independent samples.
  - used when the population variances are known,
    or the sample sizes are *large* ($n > 30$).

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

Tests for Comparing Means (2)

- One-sample t-test:
  - used to <u>compare</u> the mean of a <u>sample</u>
    to a known <u>population</u> mean.
  - used when the <u>population</u> variance is <u>unknown</u>,
    and the <u>sample</u> <u>size</u> is *small* ($n < 30$).

- Two-sample t-test:
  - used to <u>compare</u> the means of two independent <u>samples</u>.
  - used when the <u>population</u> variances are <u>unknown</u>,
    and the <u>sample</u> <u>sizes</u> are *small* ($n < 30$).

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

- Paired t-test:
  - used to compare the means of two *related* samples, such as the before and after measurements of the same group of subjects.
  - used when the population variances are unknown, and the sample size is *small* ($n < 30$).

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

Tests for Comparing Proportions (1)

- Let us consider the parameter p
  of the population proportion

    - eg) we might want to know
      the proportion of males
      within a total population of adults
      when we conduct a survey.

- A test of proportion will assess
  whether or not a sample from a population represents
  the true proportion of the entire population

`https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/proportions`

Examples of proportions

- an example
  - newborn babies are more likely to be <u>boys</u> than <u>girls</u>.
  - a random sample found 13,173 <u>boys</u> were born among 25,468 newborn <u>children</u>
  - the sample proportion of <u>boys</u> was 0.5172 ($= \frac{13173}{25468}$)
  - is this sample evidence that the birth of <u>boys</u> is more common than the birth of <u>girls</u> in the entire population? ($0.5172 > 0.4828$)

https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/proportions

- examples involved testing
  whether a single population proportion p equals some value .

- different examples of testing
  whether one population proportion equals
  a second population proportion

https://online.stat.psu.edu/stat415/lesson/9/9.4

- Additionally, most of our examples thus far have involved
  - left-tailed tests in which the alternative hypothesis involved
  - right-tailed tests in which the alternative hypothesis involved

- Here, let's consider an example that tests
  the equality of two proportions
  against the alternative that they are not equal

https://online.stat.psu.edu/stat415/lesson/9/9.4

- Time magazine reported the result of a telephone poll of 800 adult Americans.

- the question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?"

- the results of the survey were:

Non-smokers	Smokers
$n_1 = 605$	$n_2 = 195$
$y_1 = 351$ said yes	$y_2 = 41$ said yes
$\hat{p}_1 = \frac{351}{605} = 0.58$	$\hat{p}_2 = \frac{41}{195} = 0.21$

https://online.stat.psu.edu/stat415/lesson/9/9.4

- If $p_1 =$ the proportion of the <u>non-smoker</u> population who reply "yes"

- and $p_2 =$ the proportion of the <u>smoker</u> population who reply "yes,"

- then we are interested in testing the <u>null hypothesis</u>:
  $H_0 : p_1 = p_2$
  against the <u>alternative hypothesis</u>:
  $H_A : p_1 \neq p_2$

- Before we can actually conduct the hypothesis test, we'll have to derive the appropriate test statistic.

https://online.stat.psu.edu/stat415/lesson/9/9.4

- The overall sample proportion is:
  $\hat{p} = \frac{41+351}{195+605} = \frac{392}{800} = 0.49$

- that implies then that the test statistic for testing:
  $H_0 : p_1 = p_2$ versus $H_A : p_1 \neq p_2$
  is:
  $Z = \frac{(0.58-0.21)-0}{\sqrt{0.49(0.51)(\frac{1}{195}+\frac{1}{605})}} = 8.9$

https://online.stat.psu.edu/stat415/lesson/9/9.4

- One-sample proportion test :
  - used to compare the proportion of a sample
    to a known population proportion.
  - the normal approximation is used
    when both $np \geq 10$ and $n(1 - p) \geq 10$
    (data should have at least 10 "successes" and at least 10 "failures")
    (in some books, it is 5)

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

- Two-sample proportion test :
  - used to <u>compare</u> the proportions of two independent <u>samples</u>.
  - the <u>normal approximation</u> is used
    when both $np \geq 10$ and $n(1-p) \geq 10$
    (data should have at least 10 "successes" and at least 10 "failures")
    (in some books, it is 5)

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

- Time magazine reported the result of a telephone poll of 800 adult Americans.

- The question posed of the Americans who were surveyed was:
  "Should the federal tax on cigarettes be raised to pay for health care reform?"
  The results of the survey were:

`https://online.stat.psu.edu/stat415/lesson/9/9.4`

Non-Smokers	Smokers
$n_1 = 605$	$n_2 = 195$
$y_1 = 351$ said "yes"	$y_2 = 41$ said "yes"
$\hat{p}_1 = \frac{351}{605} = 0.58$	$\hat{p}_2 = \frac{41}{195} = 0.21$

- Is there sufficient evidence at the , say, to conclude that the two populations - smokers and non-smokers - differ significantly with respect to their opinions?

https://online.stat.psu.edu/stat415/lesson/9/9.4

- Errr. . . . that Z-value is off the charts, so to speak. Let's go through the formalities anyway making the decision first using the rejection region approach, and then using the P-value approach. Putting half of the rejection region in each tail, we have:

- That is, we reject the null hypothesis if or if . We clearly reject , since 8.99 falls in the "red zone," that is, 8.99 is (much) greater than 1.96. There is sufficient evidence at the 0.05 level to conclude that the two populations differ with respect to their opinions concerning imposing a federal tax to help pay for health care reform.

`https://online.stat.psu.edu/stat415/lesson/9/9.4`

- That is, the P-value is less than 0.0001. Because, we reject the null hypothesis. Again, there is sufficient evidence at the 0.05 level to conclude that the two populations differ with respect to their opinions concerning imposing a federal tax to help pay for health care reform.

- Thankfully, as should always be the case, the two approaches.... the critical value approach and the P-value approach... lead to the same conclusion

https://online.stat.psu.edu/stat415/lesson/9/9.4

Tests for Comparing Variance

- **Chi-square test** for variance :
  - used to <u>compare</u> the variance of a <u>sample</u>
    to a <u>known</u> <u>population</u> variance

- **F-test** for variance :
  - used to <u>compare</u> the variances of two *independent* samples

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

- Goodness of fit test :
- used to determine whether a <u>sample</u> fits a *specific* distribution.
- used to <u>compare</u> the <u>observed frequencies</u> of a *categorical variable* to the <u>expected frequencies</u> under a *particular* distribution.

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

- Testing for independence of two attributes (Contingency Tables) :
- used to determine whether there is a relationship between two *categorical variables*.

- often used in the form of a chi-square test, which compares the observed frequencies in a contingency table to the expected frequencies under the assumption of independence.

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

- ANOVA (Analysis of Variance) :
- used to compare the means of three or more independent samples.
- used to determine whether there is a significant difference between the means of the groups.

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

- used to test a <u>hypothesis</u> about the *population* mean

- based on the assumption that the <u>sample</u> is drawn from a normally distributed <u>population</u>.

  - the null hypothesis
    the *population* mean is <u>equal</u> to a specific value

  - the alternative hypothesis
    the *population* mean is <u>not</u> <u>equal</u> to that value

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

Two-sample z-test

- based on the assumption that both <u>samples</u> are drawn from normally distributed <u>populations</u> with <u>equal</u> variances.

- the two-sample z-test requires that the <u>population</u> standard deviations be known or that the <u>sample</u> sizes be *large* (30 or more),

  - the null hypothesis
    the means of the <u>two</u> <u>samples</u> are <u>equal</u>

  - the alternative hypothesis
    the means are <u>not</u> <u>equal</u>

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

- used to test a <u>hypothesis</u> about the *population* mean

- based on the assumption that the <u>sample</u> is drawn from a normally distributed <u>population</u>

  - the null hypothesis
    the *population* mean is <u>equal</u> to a specific value

  - the alternative hypothesis
    the *population* mean is <u>not</u> <u>equal</u> to that value

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

- based on the assumption that the samples are drawn from populations with normal distributions.

- the two-sample t-test
  that the population standard deviations need not be known or
  that the sample sizes need not be *large* (30 or more),

  - the null hypothesis
    the means of the two samples are equal

  - the alternative hypothesis
    the means are not equal

`https://www.qualitygurus.com/common-types-of-hypothesis-tests/`

Paired t-test

- used to test a hypothesis about the difference between the means of the two samples
- based on the assumption that the differences between the pairs are normally distributed

- In a dependent two-sample t-test (a paired t-test), the samples in the two groups being compared are *related* in some way.

  - the null hypothesis
    there is no difference between the means of the two samples

  - the alternative hypothesis
    there is a difference between the means

https://www.qualitygurus.com/common-types-of-hypothesis-tests/

- used to test a hypothesis about the <u>difference</u> between the <u>proportions</u> of the two samples and

- based on the assumption that the <u>samples</u> are drawn from populations with a normal distribution

  - the null hypothesis :
    there is <u>no</u> <u>difference</u> between the <u>proportions</u> of the two samples

  - the alternative hypothesis :
    there is a <u>difference</u> between the <u>proportion</u>

https://www.qualitygurus.com/common-types-of-hypothesis-tests/