Hypothesis Testing

Young W. Lim

2019-05-25 Tue

Outline

1. Based on
2. Hypothesis Testing
3. Power
4. Definitions

Based on

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

Statistical Hypothesis

- a statistical hypothesis is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables.
- A statistical hypothesis test is a method of statistical inference

Statistical Inference

- Statistical inference is the process of using data analysis to deduce properties of an underlying probability distribution.
- Inferential statistical analysis infers properties of a population, for example by testing hypotheses and deriving estimates
- It is assumed that the observed data set is sampled from a larger population

Statistical Relationship

- Commonly, two statistical data sets are compared, or
  a data set obtained by sampling is compared
  against a synthetic data set from an idealized model.

- a hypothesis is proposed for the statistical relationship
  between the two data sets, and
- this is compared as an alternative
  to an *idealized* null hypothesis
  that proposes no relationship between two data sets

Statistical Significance

- the comparison is deemed statistically significant if the relationship between the data sets would be an unlikely realization of the null hypothesis according to a threshold probability - the significance level

Hypothesis tests

- Hypothesis tests are used when determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance

Alternative Hypothesis (H_1)

- alternative hypothesis is the one that claims the difference in results between conditions is due to the independent variable
  - directional
    does specify the direction of the effect
  - non-directional
    does not specify the direction of the effect

Null Hypothesis (H_0)

- null hypothesis
  set up to be the logical counterpart of the alternative hypothesis
- if the null hypothesis is false, the alternative hypothesis is true
- the alternative and null hypotheses must be mutually exclusive and exhaustive

Null Hypothesis (H_0)

- the null hypothesis is a general statement or default position that there is no relationship between two measured phenomena, or no association among groups
- Testing (accepting, approving, rejecting, or disproving) the null hypothesis—and thus concluding that there are or are not grounds for believing that there is a relationship between two phenomena (e.g. that a potential treatment has a measurable effect) -

Non-directional Alternative and Null Hypotheses

- for the *non-directional* alternative hypothesis,
  - the independent variable has an effect on the dependent variable
  - the null hypothes :
    the independent variable has no effect on the dependent variable

Directional Alternative and Null Hypotheses

- for the *directional* alternative hypothesis,
  - the independent variable has an effect on the dependent variable in the direction specified by the alternative hypothesis
  - the null hypothes :
    the independent variable has no effect on the dependent variable
    in the direction specified by alternative hypothesis
    no effect or effect in the opposite direction

Decision Rule (α level) (1)

- always evaluate the results of an experiment by assessing the null hypothesis
- because we can calculate the probability of chance events for the null hypothesis, but there is no way to calculate the probability of the alternative hypothesis
- evaluate the null hypothesis by assuming it is true and testing the reasonableness of this assumption by calculating the probability of getting the results if chance alone is operating

Decision Rule (α level) (2)

- if the obtained probability turns out to be equal or less than a critical probability level (α level) we reject the null hypothesis

- if the obtained probability $\leq \alpha$, reject H_0
- if the obtained probability $> \alpha$, retain H_0

Type I and Type II Errors

- A Type I Error
  a decision to reject the null hypothesis
  when the null hypothesis is true
- A Type II Error
  a decision to retain the null hypothesis
  when the null hypothesis is false

Type I Error

- the incorrect rejection of a true null hypothesis.
- Usually a type I error leads one to conclude that a supposed effect or relationship exists when in fact it doesn't.

Type I Error - false positive

- when the null hypothesis (H_0) is true, but is rejected
- false hit :
  asserting something (*reject* H_0) that is absent (*true* H_0)
- false positive : a result that indicates that a given condition is present (*reject* H_0) when actually not present (*true* H_0)

Type I Error - example

- an investigator may see the wolf when there is none
  raising a false alarm where H_0 comprises "There is no wolf".
- test that shows a patient to have a disease
  when in fact the patient does not have the disease,
- a fire alarm going on indicating a fire
  when in fact there is no fire, or
- an experiment indicating that a medical treatment
  should cure a disease when in fact it does not.

Type I Error Rate - significance level

- The type I error rate or
  significance level or
  α level is
  the probability of rejecting the null hypothesis
  given that it is true
- Often, the significance level is set to 0.05 (5%), implying that it is acceptable to have a 5% probability of incorrectly rejecting the null hypothesis.

α level

- the alpha level the scientist sets at the beginning of the experiment is the level to which he / she wishes to limit the probability of making type I error

Type II Error

- the failure to reject a false null hypothesis.
- a blood test failing to detect the disease it was designed to detect, in a patient who really has the disease;
- a fire breaking out and the fire alarm does not ring; or
- a clinical trial of a medical treatment failing to show that the treatment works when really it does.

Type II Error - false negative

- when the null hypothesis (H_0) is false,
  but *erroneously* fails to be rejected (*accepted*)
- false miss
  failing to assert (*accept* H_0) what is present (*false* $H_0)
- false negative : an actual 'hit' (*false* H_0) was disregarded
  by the test and seen as a miss (*accept* H_0) in a test

Type II Error - example

- when we fail to believe a true alternative hypothesis
- an investigator may fail to detect the metaphoric "wolf"
  when in fact a wolf is present
  (and therefore fail to raise an alarm)
- the wolf either exists or does not exist within a given context
  the only question is, do we correctly detect the wolf or do we fail,
  either failing to detect him when he is present, or
  detecting him when he is not present

Type II Error Rate - ($1 - \textit{power}$)

- The rate of the type II error is denoted by the Greek letter β and related to the power of a test (which equals $1 - \beta$).

probability evaluations

- the evaluation should always be two-tailed unless the experimenter will retain H_0 when results are extreme in the direction opposite to the predicted direction

Tables of error types

	True H_0	False H_0
fail to reject H_0	Correct Inference (True Negative) $(1-\alpha)$	Type II error (False Negative) (β)
reject H_0	Type I error (False Positive) (α)	Correct Inference (True Positive) $(1-\beta)$

Power Definition

- the power of a binary hypothesis test is the probability that the test rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_1) is true.
- The statistical power ranges from 0 to 1

Power and Type II error

- as statistical power increases,
  the probability of making a type II error decreases
    - *wrongly* failing to reject the null hypothesis decreases
- for a type II error probability of β
  the corresponding statistical power is $1 - \beta$

Power examples

- if experiment 1 has a statistical power of 0.7 ($\beta_1 = 0.3$), and experiment 2 has a statistical power of 0.95 ($\beta_2 = 0.05$),
- then there is a stronger probability ($\beta_1 > \beta_2$) that experiment 1 had a type II error than experiment 2, and
- experiment 2 is more reliable than experiment 1 due to the reduction in probability of a type II error.

Power detects a specific effect

the statistical power can be equivalently thought of as
the probability of accepting the alternative hypothesis (H_1)
when the alternative hypothesis is true

- the ability of a test to detect a specific effect,
  if that specific effect actually *exists*

power of an experiment

- the probability that the results of an experiment will allow rejection of the null hypothesis if the independent variable has a real effect

P_{null} and P_{real}

- P_{null} is the probability of getting a plus with any subject in the sample of the experiment when the independent variable has no effect
- P_{real} is the probability of getting a plus with any subject in the sample of the experiment when the independent variable has real effect

P_{null}

- P_{null} is the probability of getting a plus
  with any subject in the sample of the experiment
  when the independent variable has no effect
- it is also the proportion of pluses in the population
  if the experiment were done on the entire population and
  independent variable as a real effect

Power and β

- $p(rejecting\ \ H_0\ \ if\ \ it\ \ is\ \ false) + p(retaining\ \ H_0\ \ if\ \ it\ \ is\ \ false) = 1$
- $Power = p(rejecting\ \ H_0\ \ if\ \ it\ \ is\ \ false)$
  $Beta = p(retaining\ \ H_0\ \ if\ \ it\ \ is\ \ false)$
- $Power + Beta = 1$
  $Beta = 1 - Power$

α - β and reality

- p(correctly concluding) = p(retaining H_0) = 1 - α
- p(correctly concluding) = p(rejecting H_0) = power = 1 - β

OOO

definitions (1)

- Null hypothesis (H0) A hypothesis associated with a contradiction to a theory one would like to prove.
- Alternative hypothesis (H1) A hypothesis (often composite) associated with a theory one would like to prove.
- Critical value The threshold value delimiting the regions of acceptance and rejection for the test statistic.

definitions (2)

- Power of a test $(1 - \beta)$ The test's probability of correctly rejecting the null hypothesis.
  - the complement of the false negative rate, β.
  - power is termed sensitivity in biostatistics.
  - *This is a sensitive test. Because the result is negative,*
    we can confidently say that the patient does not have the condition
  - See sensitivity and specificity and Type I and type II errors
    for exhaustive definitions.

defintiions (3)

- p-value The probability, assuming the null hypothesis is true, of observing a result at least as extreme as the test statistic. In case of a composite null hypothesis, the worst case probability.

defintiions (4)

- statistical significance test

  A predecessor to the statistical hypothesis test
  An experimental result was said to be statistically significant
  if a sample was sufficiently inconsistent with the null hypothesis
  this provides mathematical rigor and philosophical consistency
  to the concept by making the alternative hypothesis explicit.
  The term is loosely used to describe the modern version
  which is now part of statistical hypothesis testing.

defintiions (5)

- Exact test A test in which the significance level or critical value can be computed exactly, i.e., without any approximation. In some contexts this term is restricted to tests applied to categorical data and to permutation tests, in which computations are carried out by complete enumeration of all possible outcomes and their probabilities.

defintiions (6)

- Uniformly most powerful test (UMP) A test with the greatest power for all values of the parameter(s) being tested, contained in the alternative hypothesis.