COMP 680
Statistics for Computing and Data Science
Week 6: Hypothesis Testing I

Su Chen, Assistant Teaching Professor,
Rice D2K Lab

Outline

Review of Statistical Inference

- Goal: to estimate a population parameter
  - mean, median, standard deviation...

Review of Statistical Inference

- Goal: to estimate a population parameter
  - mean, median, standard deviation...
- Inference: point estimate and a confidence interval

Review of Statistical Inference

- Goal: to estimate a population parameter
  - mean, median, standard deviation...
- Inference: point estimate and a confidence interval
  - point estimate is a statistic
  - need to quantify its sampling distribution

Review of Statistical Inference

- Goal: to estimate a population parameter
  - mean, median, standard deviation...
- Inference: point estimate and a confidence interval
  - point estimate is a statistic
  - need to quantify its sampling distribution
- Method: parametric vs. nonparametric
  - MLE + asymptotic normality
  - bootstrap for standard error and CI

Hypothesis Testing

- Apply statistical inference to decision making
  - make conclusion about population
  - or the data generating mechanism

Hypothesis Testing

- Apply statistical inference to decision making
  - make conclusion about population
  - or the data generating mechanism
- Two hypotheses: H_0 vs. H_1
  - which one is better supported by data?
  - binary decision: reject H_0 or fail to reject

Hypothesis Testing

- Apply statistical inference to decision making
  - make conclusion about population
  - or the data generating mechanism
- Two hypotheses: H_0 vs. H_1
  - which one is better supported by data?
  - binary decision: reject H_0 or fail to reject
- Rational: if H_0 is true, data should look like...
  - data could have come out differently
  - inference quantifies the uncertainty

General Framework

1. Frame your hypothesis:
   - problem-dependent with some convention
   - H_0: a predefined conjecture
   - H_1: negate H_0, a new discovery

General Framework

1. Frame your hypothesis:
   - problem-dependent with some convention
   - H_0: a predefined conjecture
   - H_1: negate H_0, a new discovery
2. Pick a test statistic T:
   - one number summary of your data

General Framework

1. Frame your hypothesis:
   - problem-dependent with some convention
   - H_0: a predefined conjecture
   - H_1: negate H_0, a new discovery
2. Pick a test statistic T:
   - one number summary of your data
3. What is the value of T based on observed data?
   - calculate the observed test statistic

General Framework

1. Frame your hypothesis:
   - problem-dependent with some convention
   - H_0: a predefined conjecture
   - H_1: negate H_0, a new discovery
2. Pick a test statistic T:
   - one number summary of your data
3. What is the value of T based on observed data?
   - calculate the observed test statistic
4. What would be the value of T if H_0 is true?
   - quantify the sampling distribution of T under H_0
   - often by simulation

General Framework

1. Frame your hypothesis:
   - problem-dependent with some convention
   - H_0: a predefined conjecture
   - H_1: negate H_0, a new discovery
2. Pick a test statistic T:
   - one number summary of your data
3. What is the value of T based on observed data?
   - calculate the observed test statistic
4. What would be the value of T if H_0 is true?
   - quantify the sampling distribution of T under H_0
   - often by simulation
5. Compare 3 (one number) to 4 (a distribution):
   - does 3 looks like a "usual value" from 4

Rationale

- Null hypothesis H_0 is our assumption of population distribution:
  - data is from $N(\mu = 100, \sigma^2 = 15^2)$
  - two sets of data are from the same distribution
  - data of two variables are independent

Rationale

- Null hypothesis H_0 is our assumption of population distribution:
  - data is from $N(\mu = 100, \sigma^2 = 15^2)$
  - two sets of data are from the same distribution
  - data of two variables are independent
- Null hypothesis is usually specific about how data was generated
  - if null is true and we repeatedly collect data under H_0
  - how would those hypothetical data look like?
  - how does hypothetical data compare to the observed data at hand?

Rationale

- Null hypothesis H_0 is our assumption of population distribution:
  - data is from $N(\mu = 100, \sigma^2 = 15^2)$
  - two sets of data are from the same distribution
  - data of two variables are independent
- Null hypothesis is usually specific about how data was generated
  - if null is true and we repeatedly collect data under H_0
  - how would those hypothetical data look like?
  - how does hypothetical data compare to the observed data at hand?
- Do the math:
  - calculate sampling distribution of the test statistic under H_0

Rationale

- Null hypothesis H_0 is our assumption of population distribution:
    - data is from $N(\mu = 100, \sigma^2 = 15^2)$
    - two sets of data are from the same distribution
    - data of two variables are independent
- Null hypothesis is usually specific about how data was generated
    - if null is true and we repeatedly collect data under H_0
    - how would those hypothetical data look like?
    - how does hypothetical data compare to the observed data at hand?
- Do the math:
    - calculate sampling distribution of the test statistic under H_0
- Do the simulation:
    - simulate the empirical distribution of the the test statistic under H_0

Outline

Example: Jury Selection

- Swain v.s. Alabama 1965
  - Talladega County, Alabama
  - Robert Swain, a black man convicted of crime
  - appeal: one factor was all white jury
  - 26% of population in the county were black
  - Swains jury panel consisted of 100 men
  - 8 men on the panel were black

Example: Jury Selection

- Swain v.s. Alabama 1965
  - Talladega County, Alabama
  - Robert Swain, a black man convicted of crime
  - appeal: one factor was all white jury
  - 26% of population in the county were black
  - Swains jury panel consisted of 100 men
  - 8 men on the panel were black
- Supreme Court wrote:
  - "... the overall percentage disparity has been small and reflects no studied attempt to include or exclude a specified number of ..."
  - appeal denied

Examples: Jury Selection

- Null hypothesis H_0:

Examples: Jury Selection

- Null hypothesis H_0:
  - the jury panel of 100 men is a random sample from the population.

Examples: Jury Selection

- Null hypothesis H_0:
  - the jury panel of 100 men is a random sample from the population.
  - H_0 needs to specify the population distribution.
  - the population distribution has 26% blacks.

Examples: Jury Selection

- Null hypothesis H_0:
  - the jury panel of 100 men is a random sample from the population.
  - H_0 needs to specify the population distribution.
  - the population distribution has 26% blacks.
- Alternative hypothesis H_1: not H_0

Examples: Jury Selection

- Null hypothesis H_0:
  - the jury panel of 100 men is a random sample from the population.
  - H_0 needs to specify the population distribution.
  - the population distribution has 26% blacks.
- Alternative hypothesis H_1: not H_0
  - The jury panel contains too few blacks.

Examples: Jury Selection

- Null hypothesis H_0:
  - the jury panel of 100 men is a random sample from the population.
  - H_0 needs to specify the population distribution.
  - the population distribution has 26% blacks.
- Alternative hypothesis H_1: not H_0
  - The jury panel contains too few blacks.
  - H_1 does not specify the population distribution.
  - H_1 can specify a direction of how the sample is different from H_0.
  - this is called "one-sided" test

Example: A Genetic Model

- Gregor Mendel (1822- 1884)
- Pea plants with purple or white flowers

Example: A Genetic Model

- Gregor Mendel (1822- 1884)
- Pea plants with purple or white flowers
- Mendels model
  - each plant is purple-flowering with P=75%
  - regardless of the colors of others

Example: A Genetic Model

- Gregor Mendel (1822- 1884)
- Pea plants with purple or white flowers
- Mendels model
  - each plant is purple-flowering with P=75%
  - regardless of the colors of others
- Mendels data
  - 705 out of 929 pea plants with purple flowers

Examples: A Genetic Model

- Null hypothesis H_0:

Examples: A Genetic Model

- Null hypothesis H_0:
  - the population distribution has 75% purple flowers

Examples: A Genetic Model

- Null hypothesis H_0:
  - the population distribution has 75% purple flowers
  - 929 pea plants is a random sample from that population distribution

Examples: A Genetic Model

- Null hypothesis H_0:
  - the population distribution has 75% purple flowers
  - 929 pea plants is a random sample from that population distribution
- Alternative hypothesis H_1: not H_0

Examples: A Genetic Model

- Null hypothesis H_0:
  - the population distribution has 75% purple flowers
  - 929 pea plants is a random sample from that population distribution
- Alternative hypothesis H_1: not H_0
  - the population distribution is not 75% purple flowers

Examples: A Genetic Model

- Null hypothesis H_0:
  - the population distribution has 75% purple flowers
  - 929 pea plants is a random sample from that population distribution
- Alternative hypothesis H_1: not H_0
  - the population distribution is not 75% purple flowers
  - notice in this case, does not specify $> 75\%$ or $< 75\%$
  - this is called "two-sided" test

Examples: Stats 101

- Stats101 at UT-Austin is divided into 12 sections
- Same course material / assignments/ exams, taught by different GAI
- After midterm exam, section 3 has the lowest average grade

Examples: Stats 101

- Stats101 at UT-Austin is divided into 12 sections
- Same course material / assignments/ exams, taught by different GAI
- After midterm exam, section 3 has the lowest average grade
- Question: are section 3 grades really lower?

Examples: Stats 101

- Stats101 at UT-Austin is divided into 12 sections
- Same course material / assignments/ exams, taught by different GAI
- After midterm exam, section 3 has the lowest average grade
- Question: are section 3 grades really lower?
  - are section 3 grades like a random sample from the entire population distribution (what is the population distribution) ?

Examples: Stats 101

- Stats101 at UT-Austin is divided into 12 sections
- Same course material / assignments/ exams, taught by different GAI
- After midterm exam, section 3 has the lowest average grade
- Question: are section 3 grades really lower?
  - are section 3 grades like a random sample from the entire population distribution (what is the population distribution) ?
- H_0: section 3 grade is the same as others
- H_1: section 3 grade is significantly lower than others

General One Sample Test - Assessing a Model

- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$ a distribution (PDF) with unknown parameter θ

General One Sample Test - Assessing a Model

- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$ a distribution (PDF) with unknown parameter θ
- H_0: $\theta = \theta_0$

General One Sample Test - Assessing a Model

- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$ a distribution (PDF) with unknown parameter θ
- H_0: $\theta = \theta_0$
- θ_0 specifies the population distribution
  - $H_1 : \theta \neq \theta_0$ two-sided test
  - $H_1 : \theta > \theta_0$ or $\theta < \theta_0$ one-sided test

General One Sample Test - Assessing a Model

- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$ a distribution (PDF) with unknown parameter θ
- H_0: $\theta = \theta_0$
- θ_0 specifies the population distribution
  - $H_1 : \theta \neq \theta_0$ two-sided test
  - $H_1 : \theta > \theta_0$ or $\theta < \theta_0$ one-sided test
- Often times, θ is about the population mean.

Examples: Baby Weights

- Data from Kaiser Study in the 1960s
- Birth weights and mother's smoking status

Examples: Baby Weights

- Data from Kaiser Study in the 1960s
- Birth weights and mother's smoking status
- Question: are smoker mothers' babies have lower birth weights on average?

Examples: Baby Weights

- Data from Kaiser Study in the 1960s
- Birth weights and mother's smoking status
- Question: are smoker mothers' babies have lower birth weights on average?
  - are babies from smoker and non-smoker mothers have same average birth weights?

Examples: Baby Weights

- Data from Kaiser Study in the 1960s
- Birth weights and mother's smoking status
- Question: are smoker mothers' babies have lower birth weights on average?
    - are babies from smoker and non-smoker mothers have same average birth weights?
- H_0: yes they are from "the same" population distribution
- H_1: birth weights of babies from smoker mothers have lower average.

General Two Sample Test - Comparing Distributions

- Two samples from unknown distribution(s)
- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(\cdot)$ and $Y_1, Y_2, \cdots Y_m \overset{\text{i.i.d.}}{\sim} f_Y(\cdot)$

General Two Sample Test - Comparing Distributions

- Two samples from unknown distribution(s)
- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(\cdot)$ and $Y_1, Y_2, \cdots Y_m \overset{\text{i.i.d.}}{\sim} f_Y(\cdot)$
- H_0: they have the same population mean $\mu_X = \mu_Y$

General Two Sample Test - Comparing Distributions

- Two samples from unknown distribution(s)
- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(\cdot)$ and $Y_1, Y_2, \cdots Y_m \overset{\text{i.i.d.}}{\sim} f_Y(\cdot)$
- H_0: they have the same population mean $\mu_X = \mu_Y$
    - $H_1 : \mu_X \neq \mu_Y$ two-sided test
    - $H_1 : \mu_X > \mu_Y$ or $\mu_X < \mu_Y$ one-sided test

General Two Sample Test - Comparing Distributions

- Two samples from unknown distribution(s)
- $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} f_X(\cdot)$ and $Y_1, Y_2, \cdots Y_m \overset{\text{i.i.d.}}{\sim} f_Y(\cdot)$
- H_0: they have the same population mean $\mu_X = \mu_Y$
  - $H_1 : \mu_X \neq \mu_Y$ two-sided test
  - $H_1 : \mu_X > \mu_Y$ or $\mu_X < \mu_Y$ one-sided test
- Can also test if they have the same median.
- In general, can test whether two samples are from the same distribution.

Outline

Test Statistic

- Jury Selection: $X_1, X_2, \cdots X_{100} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- p is the probability of being black and $X_i = 1$ is black

Test Statistic

- Jury Selection: $X_1, X_2, \cdots X_{100} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- p is the probability of being black and $X_i = 1$ is black
  - $H_0 : p = 26\%$
  - $H_1 : p < 26\%$

Test Statistic

- Jury Selection: $X_1, X_2, \cdots X_{100} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- p is the probability of being black and $X_i = 1$ is black
    - $H_0 : p = 26\%$
    - $H_1 : p < 26\%$
- test statistic $T = \sum_{i=1}^{100} X_i$: number of blacks in a sample of 100
    - small number of T is evidence against H_0, or support H_1
    - observed test statistic $T_{obs} = 9$

Test Statistic

- Jury Selection: $X_1, X_2, \cdots X_{100} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- p is the probability of being black and $X_i = 1$ is black
  - $H_0 : p = 26\%$
  - $H_1 : p < 26\%$
- test statistic $T = \sum_{i=1}^{100} X_i$: number of blacks in a sample of 100
  - small number of T is evidence against H_0, or support H_1
  - observed test statistic $T_{obs} = 9$
- Simulate data under the null:
  - each time simulate a sample of 100 and calculate T_{H_0}

Test Statistic

- Jury Selection: $X_1, X_2, \cdots X_{100} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- p is the probability of being black and $X_i = 1$ is black
  - $H_0 : p = 26\%$
  - $H_1 : p < 26\%$
- test statistic $T = \sum_{i=1}^{100} X_i$: number of blacks in a sample of 100
  - small number of T is evidence against H_0, or support H_1
  - observed test statistic $T_{obs} = 9$
- Simulate data under the null:
  - each time simulate a sample of 100 and calculate T_{H_0}
  - sampling distribution of T_{H_0} under the null

The Significance Level

- Compare T_{obs} with the sampling distribution of T_{H_0}
  - if H_0 is true, T_{obs} is "one of " the T_{H_0}
  - reject H_0 if T_{obs} is very "unlikely"

The Significance Level

- Compare T_{obs} with the sampling distribution of T_{H_0}
  - if H_0 is true, T_{obs} is "one of" the T_{H_0}
  - reject H_0 if T_{obs} is very "unlikely"
- How "unlikely"?
  - significance level α defines the cut-off threshold
  - convention: 5% or 1%

The P-Value

- If H_0 is true, where is T_{obs} in the sampling distribution of T_{H_0}?
  - p-value is a probability
  - to observe the data or something even more extreme if H_0 is true
  - "more extreme" in the sense of against H_0 or supporting H_1

The P-Value

- If H_0 is true, where is T_{obs} in the sampling distribution of T_{H_0}?
  - p-value is a probability
  - to observe the data or something even more extreme if H_0 is true
  - "more extreme" in the sense of against H_0 or supporting H_1
- Example of Jury Selection:
  - if 100 panel is indeed a random sample from the county population
  - what is the probability of getting 9 blacks or even fewer?

The P-Value

- If H_0 is true, where is T_{obs} in the sampling distribution of T_{H_0}?
  - p-value is a probability
  - to observe the data or something even more extreme if H_0 is true
  - "more extreme" in the sense of against H_0 or supporting H_1
- Example of Jury Selection:
  - if 100 panel is indeed a random sample from the county population
  - what is the probability of getting 9 blacks or even fewer?
- Compare p-value to the significance level
  - p-value $< \alpha \Rightarrow$ reject H_0
  - p-value $\geq \alpha \Rightarrow$ fail to reject H_0

Error Probability

- Decision making:

Error Probability

- Decision making:

Reject H_0?

	No	Yes
H_0 is true	☺	Type I error
H_1 is true	Type II error	☺

Table 11.1: Type I and II errors.

P-value as an Error Probability

- P-value is NOT the probability of H_0 being true!!!

P-value as an Error Probability

- P-value is NOT the probability of H_0 being true!!!
- What is p-value really???

P-value as an Error Probability

- P-value is NOT the probability of H_0 being true!!!
- What is p-value really???
  - the probability of making a Type I error when reject H_0
  - the small chance of observing such data when H_0 is true
- What is the significance level α?

P-value as an Error Probability

- P-value is NOT the probability of H_0 being true!!!
- What is p-value really???
  - the probability of making a Type I error when reject H_0
  - the small chance of observing such data when H_0 is true
- What is the significance level α?
  - the upper bound of probability of making a Type I error
  - guaranteed control of Type I error

Power

- What about Type II error?
  - fail to reject when you should have

Power

- What about Type II error?
  - fail to reject when you should have
- Statistical power $\beta = 1$ - Type II error
  - how likely you will correctly reject H_0

Power

- What about Type II error?
  - fail to reject when you should have
- Statistical power $\beta = 1$ - Type II error
  - how likely you will correctly reject H_0
- Intuition: more data, more power
  - power analysis and sample size calculation
- A test with more power is better
  - given all else equal

Outline

Next week:

- More hypothesis testing
  - one sample vs. two-sample test
  - parametric vs. non-parametric test
  - implementation in Python
- Which test should I use?
- Multiple testing and FDR