Statistics

Texas A&M University Interview

Fatemeh Asgarinejad

Announcements + schedule + links (in an actual class)

- Lectures
  In person, SCHEDULE
- Discussions
  In person, SCHEDULE
- Office Hours
  SCHEDULES
- Course website | Piazza (Q&A) | Canvas (recordings + quizzes) | GradeScope (HW submission) | GitHub (code + notebooks)
- HWs, HW solutions, due dates and exams.

Course Plan: Where We Are Now (in an actual class)

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1 Aug 5 - Aug 11	**Lec 1+ Exit Ticket 1** **HW1 release**	**Lec 2+ Exit Ticket 2** **DI1**	**Lec 3+ Exit Ticket 3**	**Lec 4+ Exit Ticket 4** **DI2**	**Test1**	**HW1due**	**HW1 late due** **HW2 release**
2 Aug 12 - Aug 18	**Lec 5+ Exit Ticket 5**	**Lec 6+ Exit Ticket 6** **DI3** **Review quiz1**	**Lec 7+ Exit Ticket 7**	**Lec 8+ Exit Ticket 8** **DI4**	**Test2**	**HW2due**	**HW2 late due** **HW3 release**
3 Aug 19 - Aug 25	**Lec 9 + Exit Ticket 9**	**Lec 10+ Exit Ticket 10** **Midterm** **DI5**	**Lec 11+ Exit Ticket 11**	**Lec 12+ Exit Ticket 12** **DI6**	**Test3**	**HW3due**	**HW3 late due** **HW4 release**
4 Aug 26 - Sep 1	**Lec 13 + Exit Ticket 13**	**Lec 14+ Exit Ticket 14** **DI7** **Review quiz2**	**Lec 15 + Exit Ticket 15**	**Lec 16 + Exit Ticket 16** **DI8**	**Test4**	**HW4due**	**HW4 late due** **HW5 release**
5 Sep 1 - Sep 8	**Lec 17+ Exit Ticket 17**	**DI9** **Lec 18 + Exit Ticket 18**	**HW5 due** **Lec 19 + Exit Ticket 19**	**Lec 20 + Exit Ticket 20** **DI10**	**Final**		

Today’s Learning Goals

- Review [Statistics]
- Hypothesis Testing (Using statistics to evaluate some hypothesis about a population)
  - What is a Hypothesis?
  - Null and Alternative Hypothesis
  - Test Statistics
  - Type I error
  - Accept or reject Null Hypothesis
    - p-value

Review: What We've Covered So Far

When do we use statistics?

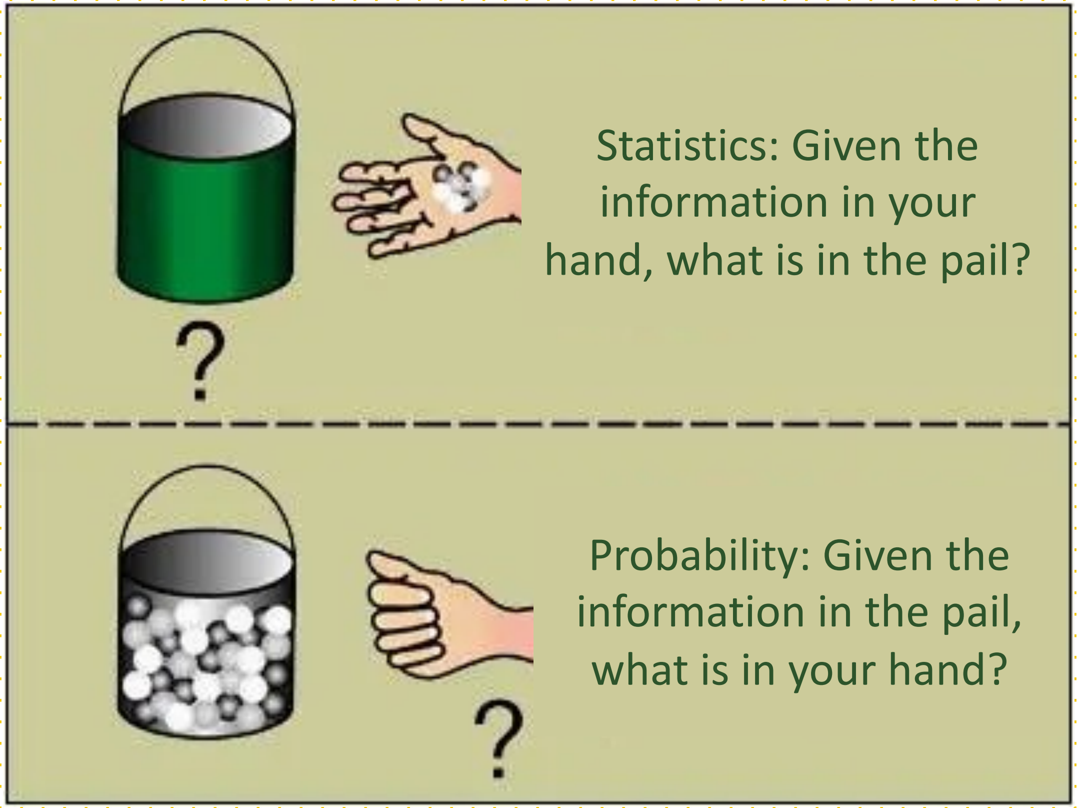
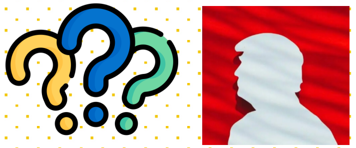


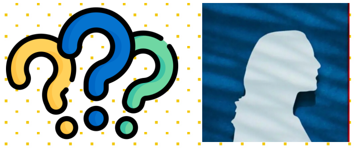
image source: https://teachingstatisticsandprobability.wordpress.com/

Practical Use Cases of Statistics (Election)

Predicting which of the USA presidency candidates will win the election.
We ask a sample of the population and estimate which candidate is likely to win based on their opinions.



President Trump Supporters assume he will earn more electoral votes.



Former Vise-President Harris's supporter assume she will earn more electoral votes.

Sample (here polls) can give us a good **estimate** about who is more probable to win the election.

images sources: first, second

Hypothesis

A **hypothesis** is an assumption (statement) about parameters of a population or distribution that can be True or Not True.

Note: parameter is a numerical or measurable characteristic of a population, e.g., $average\ (\mu), variance\ (\sigma^2), min, max, range, etc.$

Discuss other real-life hypotheses with your friends.

Example: If we have a collection of individuals, hypothesis the max, average, etc. (e.g. average height of Americans is more than 5 feet 8 inches).

Example: We can have an assumption that the average GPA of first year students is > 3.0.

Example: We can hypothesis that men play more video games than women.

Example: We have a coin and our hypothesis is that it is unbiased ($P(Head) = \frac{1}{2}$).

Null and Alternative Hypothesis

We aim to evaluate the effectiveness of a medicine on **a group of patients (population)** using a randomized clinical trial. We give the **actual medicine** to half of the participants, while the other half receive **a placebo**.

6 people
Took medicine

6 people
Took placebo

There will not be any difference between two groups (medicine is not impactful). **Null Hypothesis (H_0)**

There will be a difference between two groups of people (e.g., more people will be cured in the group who took the medicine) **Alternative Hypothesis (H_a or H_1)**

The assumption that is often believed to be true is called **Null Hypothesis (H_0).**

The complementary (often) view is called **Alternative Hypothesis (H_a or H_1).**

Null and Alternative Hypothesis

Null Hypothesis (H_0): The assumption that is often believed to be true.
Alternative Hypothesis (H_a or H_1): The complementary (often) view of Null hypothesis.

Discuss the null and alternative hypotheses in real-life scenarios.

Example: **Null hypothesis (H_0)** : Men play more video games than women.
Alternative hypothesis (H_a) : Women play more video games than men.

Example: **Null Hypothesis (H_0)** : People like Rock and pop music equally.
Alternative hypothesis (H_a) : More than 50% of people prefer Rock music over Pop.

Example: **Null Hypothesis (H_0)** : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).
Alternative hypothesis (H_a) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

Hypothesis Testing

6 people
Took medicine

6 people
Took placebo

There will not be any difference between two groups (medicine is not impactful).
Null hypothesis (H_0)

There will be a difference between two groups of people (e.g., more people will be cured in the group who took the medicine) **Alternative Hypothesis (H_a or H_1)**

Hypothesis Testing: After collecting data, if the data is consistent with the null hypothesis, we **accept the null hypothesis (H_0)**.
Otherwise, if we have an strong evidence for the alternative hypothesis we reject the Null Hypothesis and **accept the Alternative Hypothesis**.

Test Statistics

6 people
Took medicine

4 cured , 2 not cured

after 1 month

6 people
Took placebo

1 cured, 5 not cured

- There will not be any difference between two groups (**Null hypothesis (H_0)**).
- There will be a difference between two groups (**Alternative Hypothesis (H_a or H_1)**).

Which characteristic(s) should we study to measure the medicine's impact?

Test Statistics

6 people
Took medicine

4 cured , 2 not cured

after 1 month

6 people
Took placebo

1 cured, 5 not cured

Which characteristic(s) should we study to measure the medicine's impact?

We want to identify some measurable factors for evaluating the medicine's effectiveness

Test statistics is a function (mean, standard deviation, count, etc.) that helps us determine whether we should reject the Null Hypothesis or not. Here, test statistics can be number of cured patients in a group.

Significance Level

6 people
Took medicine

6 people
Took placebo

4 cured , 2 not cured

after 1 month

1 cured, 5 not cured

Should we allow the pharmaceutical company to release the medicine?

Should we set a threshold for rejecting the Null Hypothesis (H_0) (the medicine is not effective)?

What this threshold should be?

Significance Level

6 people
Took medicine

4 cured , 2 not cured

after 1 month

6 people
Took placebo

1 cured, 5 not cured

We set a threshold for test statistics to reject H_0 with **some confidence** (e.g. 95% or 99% confidence). This confidence is denoted as $1 - \alpha$ where α is significance level.

significance level (α) is the probability of mistakenly rejecting H_0 (when it is actually true, here the medicine is not impactful) **(Often 1% or 5%)**

Significance Level and Type I error

6 people
Took medicine

6 people
Took placebo

4 cured , 2 not cured

← after 1 month →

1 cured, 5 not cured

significance level (α) is the probability of mistakenly rejecting H_0 (when it is actually true, here the medicine is not impactful) **(Often 1% or 5%)**

$\boldsymbol{P(\text{Mistakenly rejecting } H_0 \text{ given the test statistics} | H_0) <= \alpha}$ **Type *I* error**

If : $P(\text{Mistakenly approving medicine}|\text{medicine was not impactful}) <= \ 5\%$
Then: with 95% confidence we can reject Null H$ypothesis$ (H_0) in favor of H_a

Significance Level and Type I error

6 people
Took medicine

6 people
Took placebo

4 cured , 2 not cured

after 1 month

1 cured, 5 not cured

What can we say about a large and small value of:

$\boldsymbol{P}(\textbf{Mistakenly rejecting H}_0\textbf{ based on test statistics}|\boldsymbol{H}_0)$

large : H_0 is consistent with the data so we do not reject it. There's a higher risk of rejecting H_0.

small: H_0 is inconsistent with the data so we reject it in favor of H_a. There's a lower risk of rejecting H_0.

P-Value

6 people
Took medicine

4 cured , 2 not cured

after 1 month

6 people
Took placebo

1 cured, 5 not cured

What can we say about a large and small value of:

$\boldsymbol{P(\text{Mistakenly rejecting } H_0 \text{ based on test statistics}|H_0)}$

$P(X \geq x\ |H_0) \leq \alpha$ $\quad$ $P(X \geq x\ |H_0)$ is called Probability value or p-value

example:

x (#cured)= 5 -> $P(X \geq 5\ |H_0)$ might be very small. So, we can conclude with 1- α certainty that H_0 is accepted.

Cont. Next session: Fisher Exact Test

Testing the Hypothesis - P-value

Null Hypothesis (H_0) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis (H_a) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

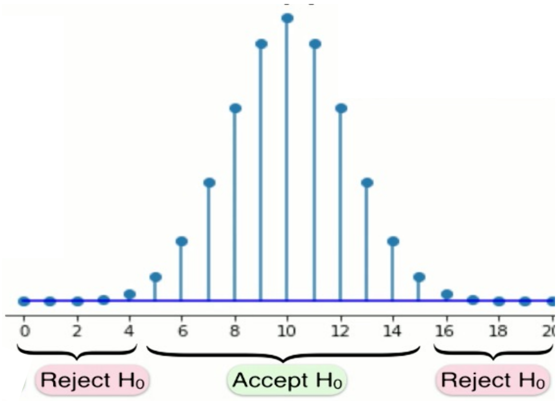
Experiment: we flip the coin **20 times**

Test Statistic: Number of heads (we show it with variable X)

Intuitively we might say if $5 \leq X \leq 15$, we do not reject the **Null hypothesis (H_0)** and we can conclude that the coin is not biased and otherwise we reject the $\boldsymbol{H_0}$ in favor of $\boldsymbol{H_a}$

What range of X confirms the coin is unbiased?

How confident are we in this conclusion?

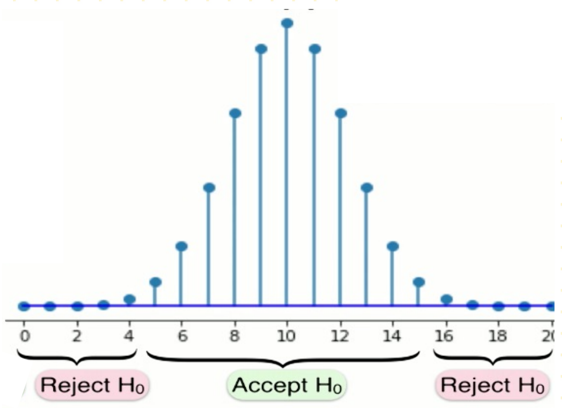


P-value

Null Hypothesis (H_0) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis (H_a) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

#Heads (X)	$P_{H_0}(X = x) = P(X = x \mid H_0)$	$P_{H_0}(X \geq x)$
20		
19		
18		
17		
16		
15		
14		



P-value

Null Hypothesis (H_0) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis (H_a) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

#Heads (X)	$P_{H_0}(X = x) = P(X = x \mid H_0)$	$P_{H_0}(X \geq x)$
20	$\binom{20}{20} * \frac{1}{2^{20}} = 0.0001\%$	0.0001%
19	$\binom{20}{19} * \frac{1}{2^{20}} = 0.0019\%$	0.002%
18	$\binom{20}{18} * \frac{1}{2^{20}} = 0.0181\%$	0.0201%
17	$\binom{20}{17} * \frac{1}{2^{20}} = 0.1087\%$	0.1288%
16	$\binom{20}{16} * \frac{1}{2^{20}} = 0.4621\%$	0.5909%
15	$\binom{20}{15} * \frac{1}{2^{20}} = 1.4786\%$	2.0695%
14	$\binom{20}{14} * \frac{1}{2^{20}} = 3.6964\%$	5.7659%

Significance value = 5%

$$P_{H_0}(X \geq x) \leq 5\%$$

what is the largest x? **15**

what is the smallest x? (if we draw the whole table, hint: use symmetry identity $\binom{a}{b} = \binom{a}{a-b}$) **5**

Conclusion: If $5 \leq X \leq 15$, we accept the null hypothesis. Otherwise, we reject it.

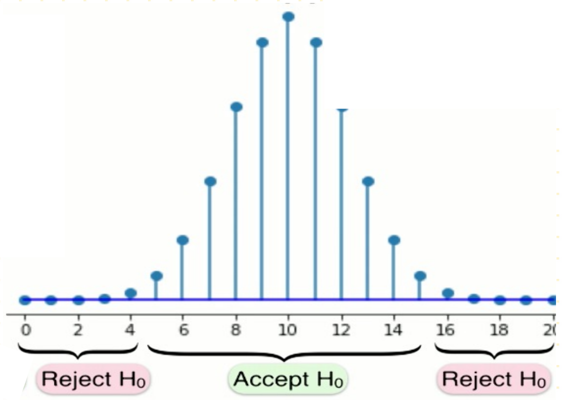
P-value

Null Hypothesis ($\boldsymbol{H_0}$) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis ($\boldsymbol{H_a}$) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

#Heads (X)	$P_{H_0}(X = x) = P(X = x \mid H_0)$	$P_{H_0}(X \geq x)$
20	$\binom{20}{20} * \frac{1}{2^{20}} = 0.0001\%$	0.0001%
19	$\binom{20}{19} * \frac{1}{2^{20}} = 0.0019\%$	0.002%
18	$\binom{20}{18} * \frac{1}{2^{20}} = 0.0181\%$	0.0201%
17	$\binom{20}{17} * \frac{1}{2^{20}} = 0.1087\%$	0.1288%
16	$\binom{20}{16} * \frac{1}{2^{20}} = 0.4621\%$	0.5909%
15	$\binom{20}{15} * \frac{1}{2^{20}} = 1.4786\%$	2.0695%
14	$\binom{20}{14} * \frac{1}{2^{20}} = 3.6964\%$	5.7659%

$\boldsymbol{P_{H_0}(X \geq x)}$ is called P-value of x (probability value) and is the probability of observing value x or more extreme values of x when H_0 holds.



Distribution of #Heads when H_0 holds

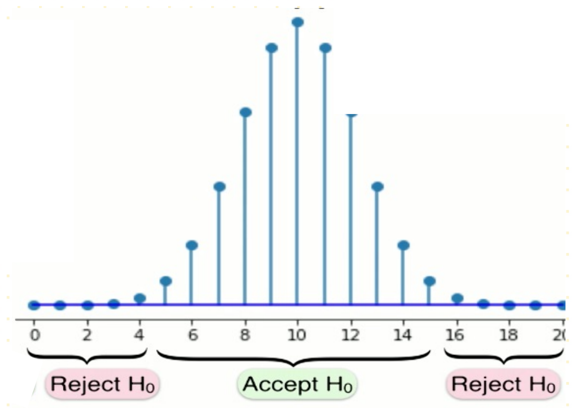
P-value

Null Hypothesis ($\boldsymbol{H_0}$) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis ($\boldsymbol{H_a}$) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

#Heads (X)	$P_{H_0}(X \geq x)$
20	0.0001%
19	0.002%
18	0.0201%
17	0.1288%
16	0.5909%
15	2.0695%
14	5.7659%

$\boldsymbol{P_{H_0}(X \geq x) = P(X \geq x \mid H_0)}$ is called P-value of x (probability value) and is the probability of observing value x or more extreme values of x when H_0 h



Distribution of #Heads when H_0 holds

Example: If p-value is 0.2 it means that you could observe 15 Heads with a 20% probability so we cannot very easily reject the null hypothesis with X=15.

P-value

Null Hypothesis ($\boldsymbol{H_0}$) : The coin used by judge in Soccer games is unbiased ($P(Head) = \frac{1}{2}$).

Alternative hypothesis ($\boldsymbol{H_a}$) : The coin is biased ($P(Head) \neq \frac{1}{2}$).

#Heads (X)	$P_{H_0}(X \geq x)$
20	0.0001%
19	0.002%
18	0.0201%
17	0.1288%
16	0.5909%
15	2.0695%
14	5.7659%

$\boldsymbol{P_{H_0}(X \geq x) = P(X \geq x \mid H_0)}$ is called P-value of x (probability value) and is the probability of observing value x or more extreme values of x when H_0 holds.

Note: If significance level is 5% and P-value is less than that, i.e., $\boldsymbol{P_{H_0}(X \geq x) < 5\%}$, we can claim our results as significant results and can reject H_0.

Exit Ticket – Due Sunday 23:59

Scan me!

Or submit the quiz in Canvas quizzes

fasgarinejad@ucsd.edu

HW bonus question: Testing a hypothesis about a population

Choose a Null Hypothesis and prove/disprove it.

Example: election

Null Hypothesis (H_0) : Harris will win the election (will get more votes).

Alternative Hypothesis (H_a or H_1) : Trump will win the election.

Steps for testing a hypothesis:

1) Design an experiment (Test)
2) Define numerical outcomes (Test statistics to measure election outcomes) (e.g., the number of votes for each candidate, the ratio of votes (e.g. >51%))
3) Gather data (sample of population, please submit you hypothesis by tonight so we can gather data from your classmates)
4) Data consistent with null hypothesis or not?
   A. Calculate P(T=test statistics | H_0) $\leq$ significance value (often 5%, 1%, etc.)

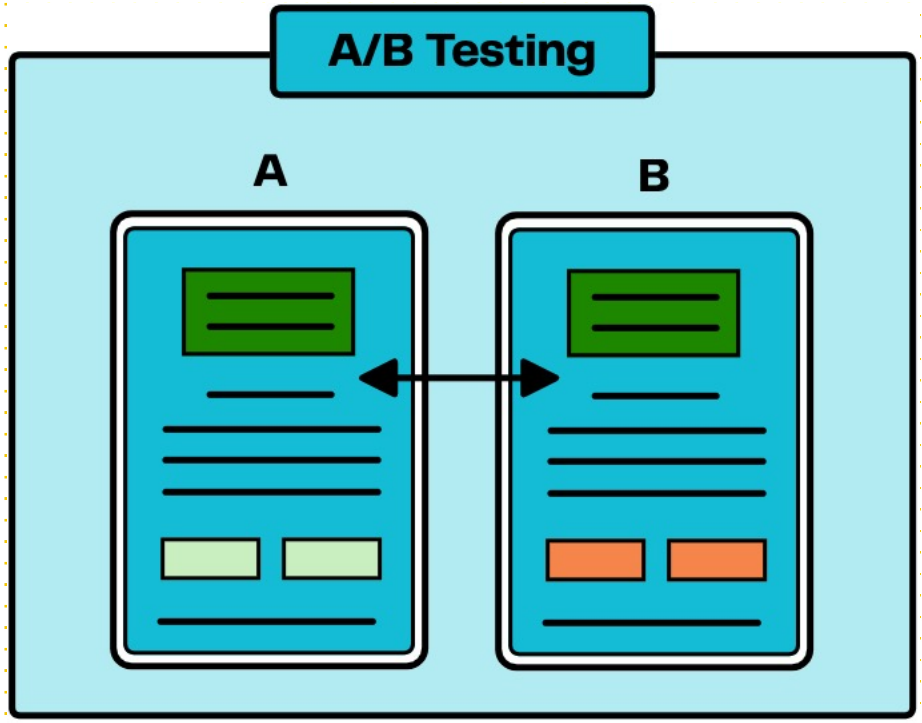
      Yes. H_0 is inconsistent with data (reject Null)

      No. H_0 is consistent with data (accept Null)

HW bonus question example
Practical Use Cases of Hypothesis Testing (A/B Test)

We want to determine which version of our software performs better. We test a sample of users and estimate the preferred version based on their behavior or feedback before the final release.

Null Hypothesis (H_0) : Version A of the software performs better than Version B.
Alternative Hypothesis (H_a or H_1) : Version B of the software performs better than Version A



Next Session

- Today’s follow-up: hypothesis testing using Fisher exact test
- Today’s follow-up: simple vs complex hypothesis
- Z – test
- T – test
- Type II error
- Sensitivity
- Specificity

fasgarinejad@ucsd.edu