Advanced Statistics
DS2003 (BDS-4A)
Lecture 06

Instructor: Dr. Syed Mohammad Irteza

Assistant Professor, Department of Computer Science, FAST

03 March, 2022

Previous Lecture

- Central Limit Theorem
  - CLT conditions
- Confidence Intervals for a proportion (p)
  - 95% confidence intervals → z=1.96
- Meaning of “95% confident”?
- Changing the confidence level
- Interpreting confidence intervals

Difference between SD and SE

- Standard Deviation vs. Standard Error: What's the Difference? (statology.org)
- This is a good resource to understand the difference between the standard deviation and the standard error

- The **standard deviation** measures how spread out values are in a dataset.
- The **standard error** is the standard deviation of the mean in repeated samples from a population.

Moving On…

- Hypothesis Testing
- Quick Review of Conditional Probability
- Testing hypotheses using confidence intervals
- Decision errors

Hypothesis Testing for a Proportion

- Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

$\hat{p}_{\text{males}}$ = 21 / 24 = 0.88

$\hat{p}_{\text{females}}$ = 14 / 24 = 0.58

Hypothesis Testing for a Proportion

- Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

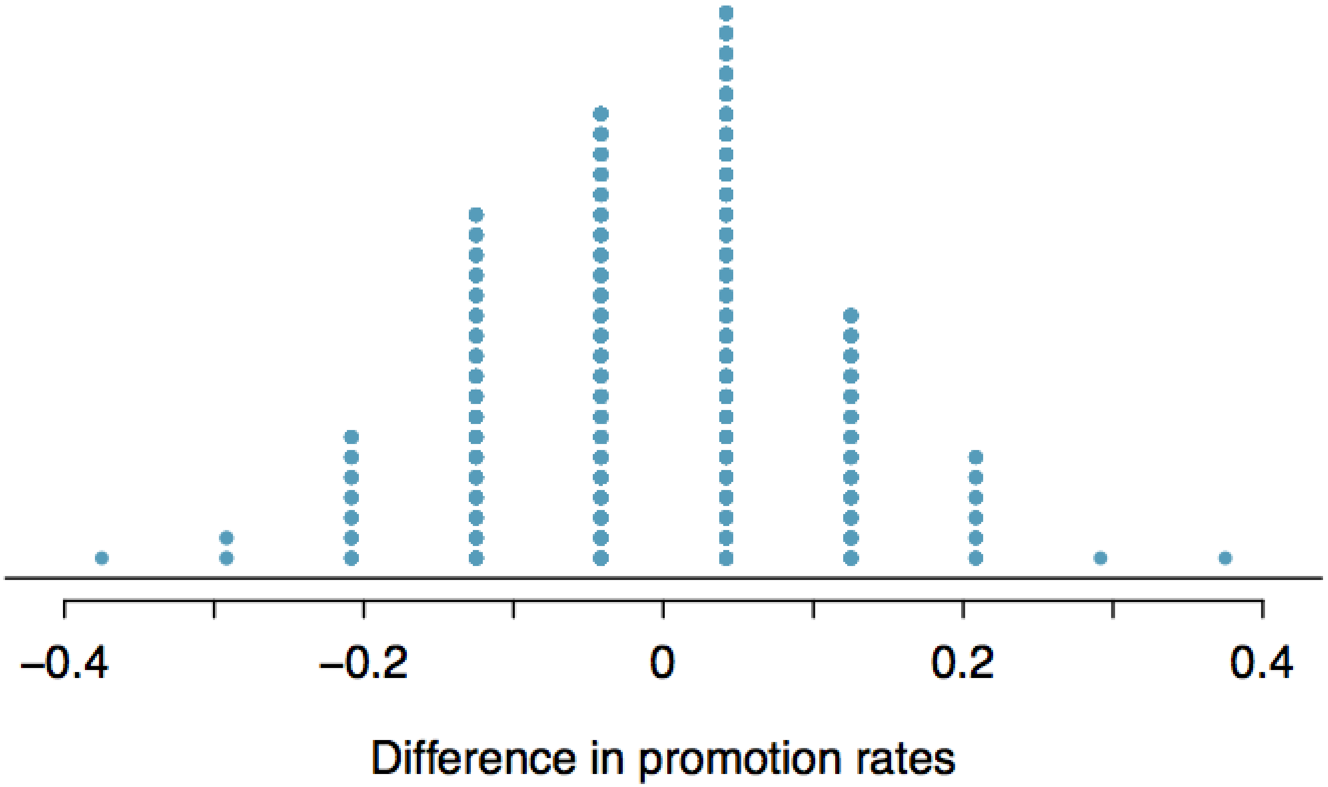
$\hat{p}_{males}$ = 21 / 24 = 0.88

$\hat{p}_{females}$ = 14 / 24 = 0.58

Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.
  → null (nothing is going on)
- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.
  → alternative (something is going on)

Result



Since it was quite unlikely to obtain results like the actual data or something more extreme in the simulations (*male promotions being 30% or more higher than female promotions*):
→ we decided to *reject the null hypothesis* in favor of the *alternative*.

Recap: Hypothesis Testing Framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the *null hypothesis is true*, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the *alternative hypothesis*, we stick with the *null hypothesis*. If they do, then we reject the *null hypothesis* in favor of the *alternative*.

- We'll formally introduce the hypothesis testing framework using an example on testing a claim about a population mean.

Quick Review

Quick Review of Conditional Probability

- Roll two 6-sided dice, yielding values $D1$ and $D2$.

Let E be event: $D1 + D2 = 4$.

What is $P(E)$?

$|S| = 36$

$E = \{(1,3), (2,2), (3,1)\}$

$P(E) = 3/36 = 1/12$

- Let F be event: $D1 = 2$.

What is $P(E)$, given F already observed?

Conditional Probability

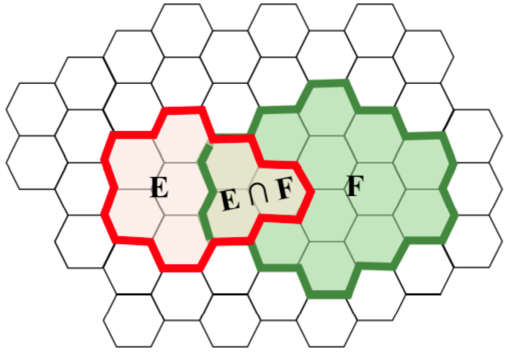
- The conditional probability of E given F is the probability that E occurs *given that F has already occurred*. This is known as *conditioning on F*.
- Written as: $P(E|F)$
- What does this mean? → "$P(E)$, given F already observed "
- Sample space? → all possible outcomes consistent with F (i.e. $S \cap F$)
- Event? → all outcomes in E consistent with F (i.e. $E \cap F$)

Conditional Probability, equally likely outcomes

- The conditional probability of E given F is the probability that E occurs given *that F has already occurred*. This is known as *conditioning on F*.

- With *equally likely outcomes*:

$$P(E|F) = \frac{\#\ of\ outcomes\ in\ E\ consistent\ with\ F}{\#\ of\ outcomes\ in\ S\ consistent\ with\ F} = \frac{|E\ \cap F|}{|S \cap F|}$$



$$P(E|F) = \frac{|EF|}{|F|}$$

$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$

Slicing up the spam

$P(E|F) = \frac{|EF|}{|F|}$ for equally likely outcomes

- 24 emails are sent, 6 each to 4 users
  - 10 of the 24 emails are spam.
  - All possible outcomes are equally likely.

Let E = user 1 receives 3 spam emails. → What is $P(E)$?

Let F = user 2 receives 6 spam emails. → What is $P(E|F)$?

Let G = user 3 receives 5 spam emails. → What is $P(G|F)$?

Slicing up the spam

$P(E|F) = \frac{|EF|}{|F|}$ for equally likely outcomes

- 24 emails are sent, 6 each to 4 users
  - 10 of the 24 emails are spam.
  - All possible outcomes are equally likely.

Let E = user 1 receives 3 spam emails. → What is $P(E)$?

$$P(E) = \frac{\left(\binom{10}{3}\binom{14}{3}\right)}{\binom{24}{6}} = 0.3245\ (approx.)$$

Let F = user 2 receives 6 spam emails. → What is $P(E|F)$?

Let G = user 3 receives 5 spam emails. → What is $P(G|F)$?

Slicing up the spam

$P(E|F) = \frac{|EF|}{|F|}$ for equally likely outcomes

- 24 emails are sent, 6 each to 4 users
  - 10 of the 24 emails are spam.
  - All possible outcomes are equally likely.

Let E = user 1 receives 3 spam emails. → What is $P(E)$? [Answer = 0.3245]

Let F = user 2 receives 6 spam emails. → What is $P(E|F)$?

$$P(E|F) = \frac{\left(\binom{4}{3}\binom{14}{3}\right)}{\binom{18}{6}} = 0.0784\ (approx.)$$

Let G = user 3 receives 5 spam emails. → What is $P(G|F)$?

Slicing up the spam

$P(E|F) = \frac{|EF|}{|F|}$ for equally likely outcomes

- 24 emails are sent, 6 each to 4 users
  - 10 of the 24 emails are spam.
  - All possible outcomes are equally likely.

Let E = user 1 receives 3 spam emails. → What is $P(E)$? [Answer = 0.3245]

Let F = user 2 receives 6 spam emails. → What is $P(E|F)$? [Answer = 0.0784]

Let G = user 3 receives 5 spam emails. → What is $P(G|F)$?

$$P(G|F) = \frac{\left(\binom{4}{5}\binom{14}{1}\right)}{\binom{18}{6}} = 0$$

Conditional probability in general

- General definition of conditional probability:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

- The Chain Rule (aka Product rule):

$$P(EF) = P(F)P(E|F)$$

These properties hold even when outcomes are not equally likely.

Law of Total Probability

- (Theorem) Let F be an event where $P(F) > 0$. For any event E,

$$P(E) = P(E|F)\, P(F) + P(E|F^C)\, P(F^C)$$

Bayes' Theorem

- Theorem: For any events E and F where $P(E) > 0$ and $P(F) > 0$,

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

Back to Hypothesis Testing

Testing *hypotheses* using confidence intervals

- Earlier we calculated a 95% confidence interval for the proportion of American Facebook users who think Facebook categorizes their interests accurately as *64% to 70%*. Based on this confidence interval, do the data support the hypothesis that majority of American Facebook users think Facebook categorizes their interests accurately.

The associated hypotheses are:

H_0: $p = 0.50$: 50% of US Facebook users think FB categorizes their interests accurately

H_A: $p > 0.50$: More than 50% of US Facebook users think FB categorizes their interests accurately

If the null value is not included in the interval → reject the null hypothesis.

This is a quick-and-dirty approach for *hypothesis testing*, but it doesn't tell us the likelihood of certain outcomes under the null hypothesis (p-value)

Decision Errors

- Hypothesis tests are not flawless.
- In the court system innocent people are sometimes wrongly convicted, and the guilty sometimes walk free.
- Similarly, we can make a wrong decision in statistical hypothesis tests as well.
- The difference is that we have the tools necessary to quantify how often we make errors in statistics.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		✓

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true		✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	Type 1 Error
	H_A true	Type 2 Error	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

  Type 1 error

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

  Type 1 error

Which error do you think is the worse error to make?

"better that ten guilty persons escape than that one innocent suffer"

- William Blackstone

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

- This is why we prefer small values of α -- increasing α increases the Type 1 error rate.

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

Setting the hypotheses

The *parameter of interest* is the proportion of all American Facebook users who are comfortable with Facebook creating categories of interests for them.

There may be two explanations why our sample proportion is lower than 0.50 (minority).

- The true population proportion is different than 0.50.
- The true population mean is 0.50, and the difference between the true population proportion and the sample proportion is simply due to natural sampling variability.

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

Setting the hypotheses

We start with the assumption that 50% of American Facebook users are comfortable with Facebook creating categories of interests for them

H_0: $p = 0.50$

We test the claim that the proportion of American Facebook users who are comfortable with Facebook creating categories of interests for them is different than 50%.

H_A: $p \neq 0.50$

Sources

- openintro.org/os (Chapter 5)
- https://web.stanford.edu/class/archive/cs/cs109/cs109.1208/lectures/04_cond_bayes_blank.pdf