Lecture 15: Hypothesis Testing Part II

Chapter 4.3

Goals for Today

- Define significance level
- Tie-in p-Values with sampling distributions
- Example

Type I Errors: US Criminal Justice System

Defendants must be "guilty beyond a reasonable doubt": better to let a guilty person go free, than put an innocent person in jail.

Type I Errors: US Criminal Justice System

Defendants must be "guilty beyond a reasonable doubt": better to let a guilty person go free, than put an innocent person in jail.

- H_0: the defendant is innocent
- H_A: the defendant is guilty

Type I Errors: US Criminal Justice System

Defendants must be "guilty beyond a reasonable doubt": better to let a guilty person go free, than put an innocent person in jail.

- H_0: the defendant is innocent
- H_A: the defendant is guilty

thus "rejecting H_0" is a guilty verdict $\Rightarrow$ putting them in jail

Type I Errors: US Criminal Justice System

Defendants must be "guilty beyond a reasonable doubt": better to let a guilty person go free, than put an innocent person in jail.

- H_0: the defendant is innocent
- H_A: the defendant is guilty

thus "rejecting H_0" is a guilty verdict $\Rightarrow$ putting them in jail

In this case:

- Type I error $=$ jailing an innocent person (worse)
- Type II error $=$ letting a guilty person go free.

Type II Errors: Airport Screening

An example of where Type II errors are more serious: airport screening.

Type II Errors: Airport Screening

An example of where Type II errors are more serious: airport screening.

$$\begin{aligned} H_0: &\quad \text{passenger X does not have a weapon} \\ H_A: &\quad \text{passenger X has a weapon} \end{aligned}$$

Type II Errors: Airport Screening

An example of where Type II errors are more serious: airport screening.

$$
\begin{aligned}
H_0 &: \quad \text{passenger X does not have a weapon} \\
H_A &: \quad \text{passenger X has a weapon}
\end{aligned}
$$

Failing to reject H_0 when H_A is true is not "patting down" passenger X when they have a weapon.

Type II Errors: Airport Screening

An example of where Type II errors are more serious: airport screening.

$$\begin{aligned} H_0 &: \quad \text{passenger X does not have a weapon} \\ H_A &: \quad \text{passenger X has a weapon} \end{aligned}$$

Failing to reject H_0 when H_A is true is not "patting down" passenger X when they have a weapon.

Hence the long lines at airport security.

Significance Level

Thought experiment: Coin Flips

Say you flip a coin you think is fair 1000 times. Say you observe

Thought experiment: Coin Flips

Say you flip a coin you think is fair 1000 times. Say you observe

- 501 heads? Do you think the coin is biased?

Thought experiment: Coin Flips

Say you flip a coin you think is fair 1000 times. Say you observe

- 501 heads? Do you think the coin is biased?
- 525 heads? Do you think the coin is biased?

Thought experiment: Coin Flips

Say you flip a coin you think is fair 1000 times. Say you observe

- 501 heads? Do you think the coin is biased?
- 525 heads? Do you think the coin is biased?
- 900 heads? Do you think the coin is biased?

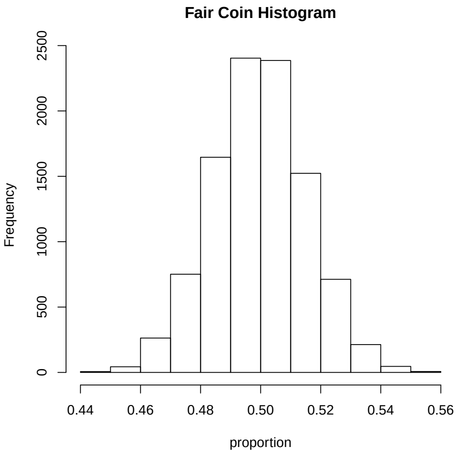
Thought experiment: p-Values

p-Values

Thought experiment: Coin Flips

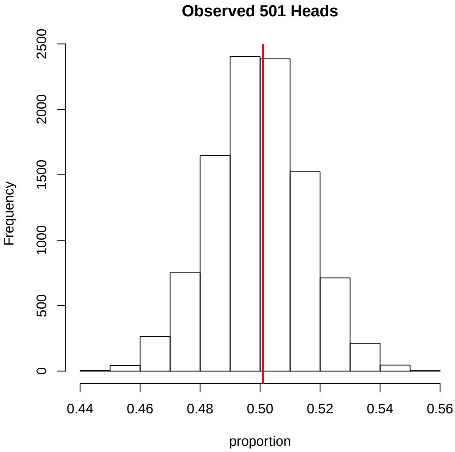
Sampling Distribution of $\widehat{p}$

Under H_0 the sampling distribution of $\widehat{p}$ when $n = 1000$ is:



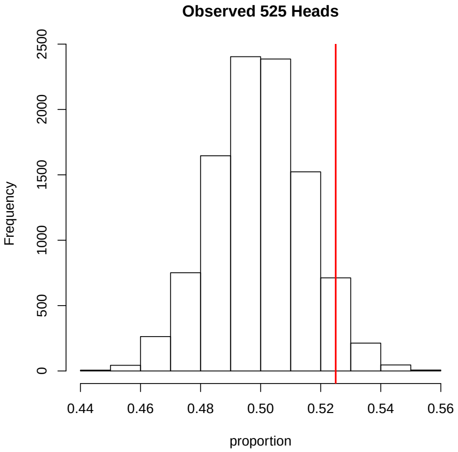
Say we observe...

$$\widehat{p} = \frac{501}{1000}$$



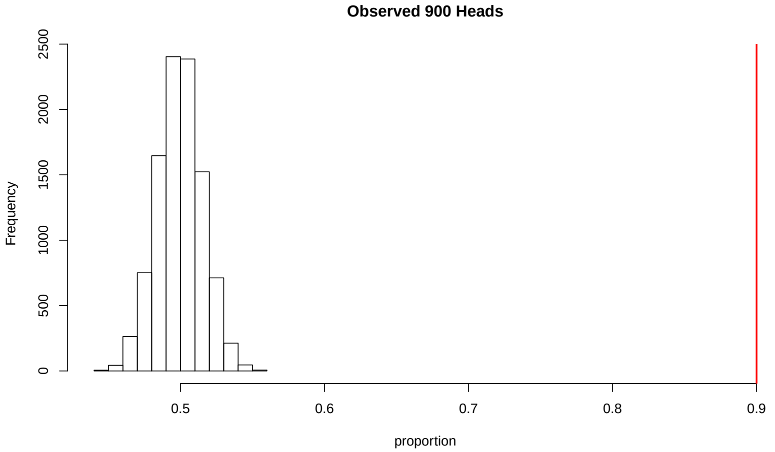
Say we observe...

$$\widehat{p} = \frac{525}{1000}$$



Say we observe...

$$\widehat{p} = \frac{900}{1000}$$



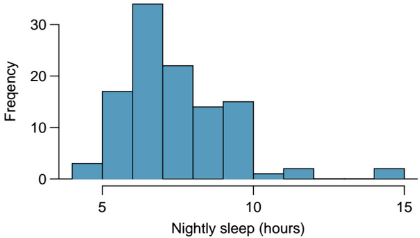
Example about Sleep Habits

A US-wide poll found that college students sleep about 7 hours a night. You suspect that Midd Kids sleep more and investigate this claim at a pre-specified $\alpha = 0.01$ level.

Example about Sleep Habits

A US-wide poll found that college students sleep about 7 hours a night. You suspect that Midd Kids sleep more and investigate this claim at a pre-specified $\alpha = 0.01$ level.

You sample $n = 110$ Midd Kids and find that $\overline{x} = 7.42$ and $s = 1.75$ with a histogram that looks like:



Example about Sleep Habits

Example about Sleep Habits

Conclusion: we reject at the $\alpha = 0.01$ significance level the hypothesis that the average # of hours Midd Kids sleep is 7, in favor of the hypothesis that they sleep more.

Example about Sleep Habits

Conclusion: we reject at the $\alpha = 0.01$ significance level the hypothesis that the average # of hours Midd Kids sleep is 7, in favor of the hypothesis that they sleep more.

Correct interpretation of the p-value: If the null hypothesis is true ($\mu = 7$), the probability of observing a sample mean $\overline{x} = 7.42$ or greater is 0.007 (small).

Example about Sleep Habits

Conclusion: we reject at the $\alpha = 0.01$ significance level the hypothesis that the average # of hours Midd Kids sleep is 7, in favor of the hypothesis that they sleep more.

Correct interpretation of the p-value: If the null hypothesis is true ($\mu = 7$), the probability of observing a sample mean $\overline{x} = 7.42$ or greater is 0.007 (small).

Incorrect interpretation of the p-value: The probability that the null hypothesis ($\mu = 7$) is true is 0.007.