Lecture 15: Hypothesis Testing Part II

Chapter 4.3

Previously... Statistical Hypothesis Testing

A hypothesis test is a method for using sample data to decide between two competing hypotheses about the population parameter:

- A null hypothesis H_0.
  i.e. the status quo that is initially assumed to be true, but will be tested.
- An alternative hypothesis H_A. i.e. the challenger.

There are two potential outcomes of a hypothesis test. Either we

- reject H_0
- fail to reject H_0

Previously... Decision Errors

Hypothesis tests will get things right sometimes and wrong sometimes:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	OK	Type I Error
	H_A true	Type II Error	OK

Two kinds of errors:

- Type I Error: a false positive (test result)
- Type II Error: a false negative (test result)

Type I Errors: US Criminal Justice System

Defendants must be proven "guilty beyond a reasonable doubt": in theory they would rather 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 is putting an innocent person in jail (considered worse)
- Type II error is letting a guilty person go free.

Type II Errors: Airport Screening

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

H_0 : passenger X does not have a weapon

H_A : passenger X has a weapon

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.

Goals for Today

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

Significance Level

Thought experiment: p-Values

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

Intuitively, a p-value quantifies how extreme an observation is given the null hypothesis.

The smaller the p-value, the more extreme the observation, where the meaning of extreme depends on the context.

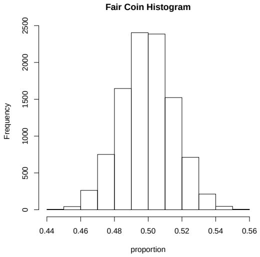
Note the p-value is different than the population proportion p (bad historical choice).

p-Values

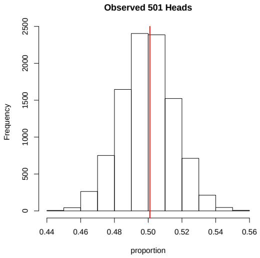
Recall our Coin Example

Sampling Distribution of $\widehat{p}$

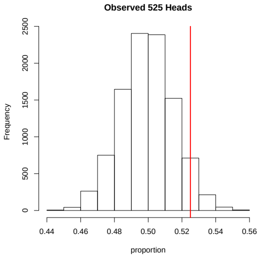
Under H_0 that the coin is fair i.e. $p = p_0 = 0.5$, the sampling distribution of $\widehat{p}$ when $n = 1000$ is:



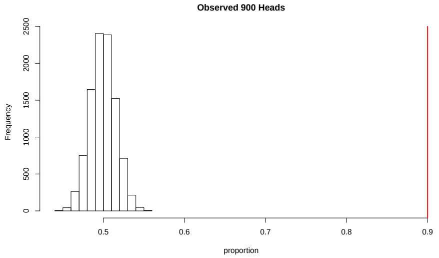
Say we observe...



Say we observe...

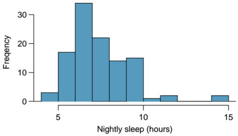


Say we observe...



Example about Sleep Habits

A poll found that college students sleep about 7 hours a night. Researchers suspect that Reedies sleep more. They want to investigate this claim at a pre-specified $\alpha = 0.05$ level. They sample $n = 110$ Reedies and find that $\overline{x} = 7.42$ and $s = 1.75$ and the histogram looks like:

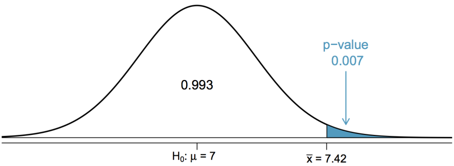


Example about Sleep Habits

Example about Sleep Habits

In our case, since $H_A : \mu > 7$, more extreme means to the right of $z = 2.47$.

Hence, the p-value is 0.007:



Example about Sleep Habits

Since the p-value $0.007 < 0.05 = \alpha$, the pre-specified significance level, it has a high degree of extremeness, and thus we reject H_0.

Interpretation: we reject (at the $\alpha = 0.05$ significance level) the hypothesis that the average # of hours of Reedies sleep is 7, in favor of the hypothesis that sleep more.

Example about Sleep Habits

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.

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

Two-Sided Alternative Hypothesis

Say instead we had a two-sided alternative hypothesis:

- $H_0 : \mu = 7$
- $H_A : \mu \neq 7$

The the p-value would be double: $2 \times 0.007 = 0.014$. Picture:

Next Time

- How big a sample size to I need? i.e. power calculations
- Statistical vs practical significance