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.

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

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.

Goals for Today

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

Significance Level

Hypothesis testing is built around rejecting or failing to reject the null hypothesis.
i.e. we do not reject H_0 unless we have strong evidence.

As a rule of thumb, when H_0 is true, we do not want to incorrectly reject H_0 more than 5% of the time.
i.e. $\alpha = 0.05 = 5\%$ is the significance level.

With 95% confidence intervals from earlier, we expect it to miss the true population parameter 5% of the time. This corresponds to $\alpha = 0.05$.

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

Definition: The p-value or observed significance level is the probability of observing a test statistic as extreme or more extreme (in favor of the alternative) as the one observed, assuming H_0 is true.

It is NOT the probability of H_0 being true. This is the most common misinterpretation of the p-value.

Recall our Coin Example

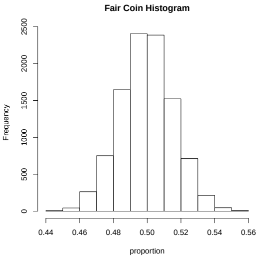
You have a coin that test for fairness with $n = 1000$ flips. Set $p_0 = 0.5$ and define a "success" as getting heads. i.e.

$$\begin{aligned} & H_0 : p = p_0 \text{ i.e. coin is fair} \\ vs \quad & H_A : p \neq p_0 \end{aligned}$$

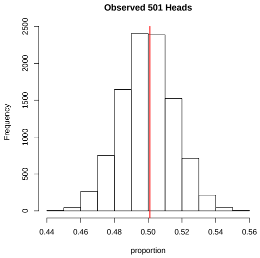
- The point estimate $\widehat{p}$ of p is $\dfrac{\#\text{ of successes}}{\#\text{ of trials}}$.
- Since it is based on a sample, $\widehat{p}$ has a sampling distribution
- The standard error is $\sqrt{\frac{p(1-p)}{n}}$ (Chapter 6).
- Furthermore, since conditions hold, the sampling distribution is Normal (CLT)

Sampling Distribution of $\widehat{p}$

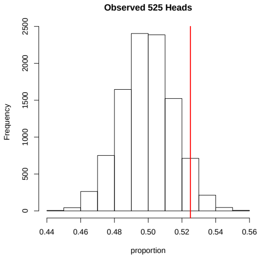
Under H_0 that the coin is fair i.e. $p = p_0 = 0.5$, the sampling distribution of $\widehat{p}$ when $n = 1000$ is:



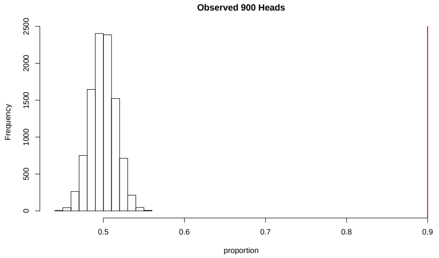
Say we observe...



Say we observe...



Say we observe...



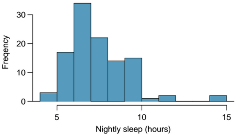
p-Value Definition

The p-value or *observed* significance level is the probability of observing a test statistic as extreme or more extreme (in favor of the alternative) as the one observed, assuming H_0 is true.

It is NOT the probability of H_0 being true. This is the most common misinterpretation of the p-value.

Example about Sleep Habits

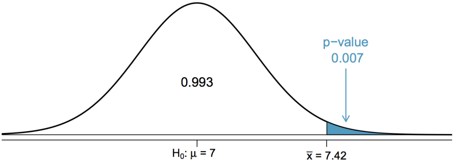
A poll found that college students sleep about 7 hours a night. Researchers suspect that Reedies sleep more. They want to investigate this claim at a pre-specified $\alpha = 0.05$ level. They sample $n = 110$ Reedies and find that $\overline{x} = 7.42$ and $s = 1.75$ and the histogram looks like:



Example about Sleep Habits

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.

Next Time

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