HYPOTHESIS TESTING

How do we draw conclusions about population parameters (or relationships between variables) based on sample statistics?

LESSON 32: HYPOTHESIS TESTING USING *P*-VALUES
What Does It Mean to be Unusual?

PART A : H_0 and H_1, NULL AND ALTERNATIVE HYPOTHESES

A null hypothesis (denoted by H_0) is an **assumption** that calculations are based on. In a hypothesis test, we **decide** to do one of two things:

- Reject H_0

OR

- Do not reject H_0.

WARNING 1: We never "accept H_0."

An alternative hypothesis (denoted by H_1 or H_a) serves as the **complement to the null**.

- H_1 is "everything that H_0 is not."
- Either H_0 or H_1 is true.

In a **hypothesis test**, we **decide** to do one of the following:

- Reject H_0, in which case we **accept** H_1.

WARNING 2: We **sometimes accept** H_1, even though we **never accept** H_0.

§ *Think: "The Null Never gets love."*

OR

- Do not reject H_0.

PART B: A HYPOTHESIS TEST and *P*-VALUES

Example 1 (Six Red Cards: One-Tailed Analysis)

We will look at the top six cards of a deck of cards.

WHAT ARE THE HYPOTHESES, H_0 and H_1?

• The **null hypothesis** H_0 is that the cards are **randomly ordered** in the deck.

• The **alternative hypothesis** H_1 is that the cards are **not randomly ordered** in the deck.

WHAT DO WE OBSERVE FROM THE SAMPLE DATA?

Let's say that we **observe** that the top six cards are **all red**.

HOW UNLIKELY ARE OUR SAMPLE RESULTS UNDER H_0?

The **probability** that the top six cards are all red **under** H_0 (that is, assuming that the null hypothesis is true) is about 0.0113 (or 1.13%).

• Using the notation of conditional probability and the methods of Lesson 13:

$$P\left(\text{top 6 cards are red} \mid H_0\right) = \frac{26}{52} \cdot \frac{25}{51} \cdot \frac{24}{50} \cdot \frac{23}{49} \cdot \frac{22}{48} \cdot \frac{21}{47}$$
$$\approx 0.0113, \text{ or } 1.13\%$$

• We may think of 0.0113 as a *P*-value, a probability that describes **how unusual** the sample results are under H_0. (A **lower** *P*-value means "**more unusual**.")

WHAT DO WE DECIDE TO DO ABOUT H_0?

The **lower** the ***P*-value** is, the **more unusual** the sample results are under H_0, and the more likely it is we will **reject** H_0 (and therefore accept H_1).

§ *Think: "If the P-value is **low**, the null must **go** [away]."*

(Source: Triola and others.)

WHAT DO WE CONCLUDE?

If 0.0113 is **too low** for H_0 to "survive," then we **reject** H_0. That is, we **reject the assumption** that the cards were **randomly ordered** in the deck.

• In later Lessons, we will refine the wording of conclusions.

§

PART C: ONE-TAILED vs. TWO-TAILED *P*-VALUES

Example 2 (Six Red Cards: Two-Tailed Analysis)

In Example 1, we calculated a one-tailed *P*-value because we focused on the probability of the top six cards being **all red**.

One could argue that it is more appropriate to calculate the probability of the top six cards being **all the same color** (all red or all black). This would be a two-tailed *P*-value.

$$\begin{aligned} & P(\text{top 6 cards are the same color} \mid H_0) \\ &= P(\text{top 6 cards are red} \mid H_0) + P(\text{top 6 cards are black} \mid H_0) \\ &\approx 0.0113 + 0.0113, \\ &\qquad \text{or } 2(0.0113) \quad [\text{by symmetry}] \\ &\approx 0.0226, \text{ or } 2.26\% \end{aligned}$$

§

PART D: WHY "TAILS"?

In Examples 1 and 2, we discussed **one-tailed** and **two-tailed** *P*-values.
Why do we discuss **tails** at all?

Example 3 (100 Coin Flips: One- and Two-Tailed Analyses)

A magician's coin will be flipped 100 times.

WHAT ARE THE HYPOTHESES, H_0 and H_1?

- The **null hypothesis** H_0 is that the coin is **fair**.
- The **alternative hypothesis** H_1 is that the coin is **not fair**.

WHAT DO WE OBSERVE FROM THE SAMPLE DATA?

Let's say that the coin comes up **heads 59 times** out of the 100 flips.

HOW UNLIKELY ARE OUR SAMPLE RESULTS UNDER H_0?

Let X = the number of heads in 100 flips of a fair coin.

Then, $X \sim \text{Bin}\left(n = 100,\ p = \frac{1}{2}\right)$, where p is the probability of heads on a single flip.

Let $P(x)$ = the probability of getting x heads out of 100 flips of a fair coin.

- It is **not** enough to look at $P(59)$, which is about 0.0159, or 1.59%.
- 50 is the **most likely** number of heads, and even $P(50)$ is not that high. It is only about 0.0796, or 7.96%.

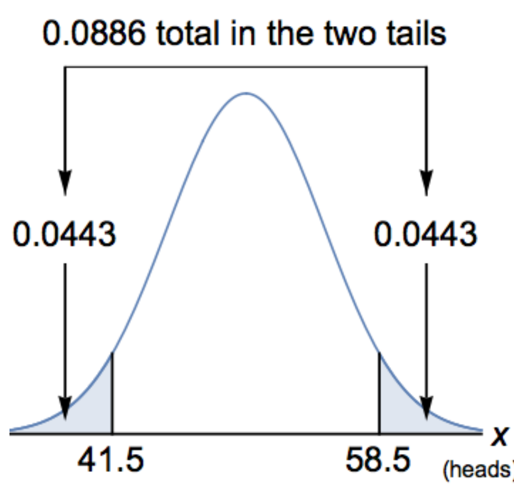
• It is more appropriate to find the probability of getting **at least 59 heads** if a fair coin is flipped 100 times. This is a **one-tailed *P*-value**.

$$P(X \geq 59) = P(59) + P(60) + P(61) + \ldots + P(100)$$
$$\approx 0.0443, \text{ or } 4.43\%$$

• We could also use a **two-tailed *P*-value**. By **symmetry**, we could **add on** the probability of getting **at least 59 tails** (that is, **at most 41 heads**) if a fair coin is flipped 100 times.

$$\begin{aligned} & P(\text{at least 59 heads } \textbf{or} \text{ at least 59 tails}) \\ &= P(\text{at least 59 heads } \textbf{or} \text{ at most 41 heads}) \\ &= P(X \geq 59 \text{ or } X \leq 41) \\ &= P(X \geq 59) + P(X \leq 41) \\ &= 2 \cdot P(X \geq 59) \quad [\text{by symmetry}] \\ &\approx 2(0.0443) \\ &\approx 0.0886, \text{ or } 8.86\% \end{aligned}$$

In the figure below, we use the fact that the distribution of X is binomial but can be **approximated by a normal distribution** (the approximation has about 0.0446, or 4.46%, in each tail; see Lesson 23). We are using **continuity corrections**, hence the 41.5 and the 58.5 (heads) on the x-axis.



WHAT DO WE DECIDE TO DO ABOUT H_0?

The decision to **reject** or to **not reject** H_0 depends on which *P*-value to take … and how low is "too low" for H_0 to survive.

WHAT DO WE CONCLUDE?

Our conclusion will depend on that decision.

§

P-Value

A *P*-value is a probability of obtaining sample results **at least as extreme** as what we observe, assuming that H_0 is true.

PART E : α, THE SIGNIFICANCE LEVEL OF A HYPOTHESIS TEST

How **low** must the ***P*-value** be for us to **reject** H_0? Bottom line: **lower than** α.

In Lesson 25, Part E, we introduced α, the Greek letter alpha.

- The **confidence level** of a confidence interval (CI) is denoted by $1-\alpha$.
- α is a **failure probability** that tells how likely it is that a population parameter (such as a population mean μ) is **not** contained within a CI for that parameter.

Here, α is the significance level of a hypothesis test.

- It represents the **threshold** for what it means for sample results to be **"too unusual"** under H_0 for H_0 to survive.
- It is typically assumed that $\alpha = 0.05$ (or 5%), but $\alpha = 0.01$ and $\alpha = 0.10$ are also used.
- The **higher** α is, the more likely it is that we will **reject** H_0.
We could say something like: "If α is high, the null tends to die."
Too many rhymes could be too confusing; we may be best off sticking with:

*"If the P-value is **low**, the null must **go** [away]."*

Remember: "**Low**" means "lower than α."

Example 4 (Coins and "Your Personal Significance Level")

A magician gives you a coin and tells you that the magician's favorite side is "heads." You flip the coin and **keep getting heads**. Depending on your "personal value" for α, when do you start saying that the results are **too unusual** for the coin to be fair (using a one-tailed analysis – only consider heads).

Let $P(x)$ = the probability of getting *x* heads in *x* flips of a fair coin.

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$P(1) = \frac{1}{2} = 0.5$			
$P(2) = \frac{1}{4} = 0.25$			
$P(3) = \frac{1}{8} = 0.125$			
$P(4) = \frac{1}{16} = 0.0625$	**too unusual**		
$P(5) = \frac{1}{32} = 0.03125$	**too unusual**	**too unusual**	
$P(6) = \frac{1}{64} = 0.015625$	**too unusual**	**too unusual**	
$P(7) = \frac{1}{128} = 0.0078125$	**too unusual**	**too unusual**	**too unusual**

§

PART F: DECISION RULES FOR HYPOTHESIS TESTS

Decision Rule for a Hypothesis Test: *P*-Value Method

- If *P*-value < α, then **reject** H_0.

  (In this case, the *P*-value is too low for H_0 to survive; it must "go [away].")

- If *P*-value > α, then **do not reject** H_0.

We will ignore the case where *P*-value = α. The probability of that happening is 0 anyway.

Decision Rule for a Hypothesis Test: Confidence Interval (CI) Method

Let's say H_0 states that the value of a population parameter is *c*.

- If a $1-\alpha$ confidence interval (CI) for the parameter **does not contain** *c*, then **reject** H_0.

- If a $1-\alpha$ confidence interval (CI) for the parameter **contains** *c*, then **do not reject** H_0.

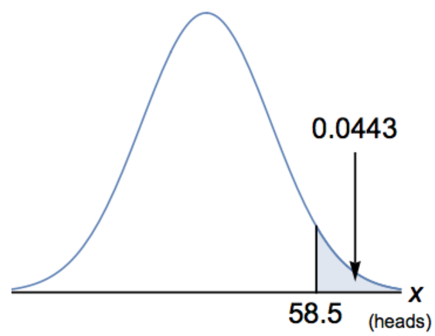
PART F: IS 59 HEADS UNUSUAL? FOUR METHODS

Example 5 (Is 59 Heads Unusual? Revisiting Example 3)

In Example 3, a magician's coin was flipped 100 times, and we observed 59 heads. H_0 states that the coin is **fair**. H_1 states that the coin is **not fair**. Use the significance level: $\alpha = 0.05$. Do we decide to **reject** or **not reject** H_0? What do we **conclude** about the assumption that the coin is fair?

We will use four methods to answer these questions. Different methods lead to different decisions and different conclusions.

Method 1: Using a One-Tailed P-Value



$$\text{One-tailed } P\text{-value} = P(X \geq 59)$$
$$= P(59) + P(60) + P(61) + ... + P(100)$$
$$\approx 0.0443, \text{ or } 4.43\%$$

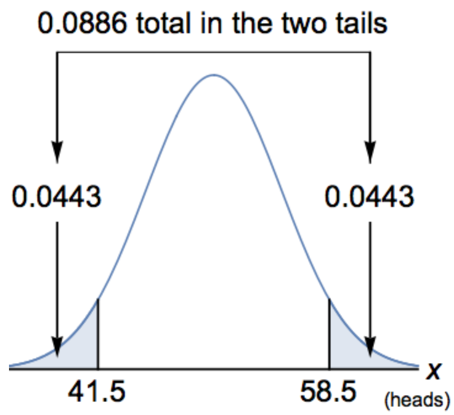
$0.0443 < 0.05$, so P-value $< \alpha$. The P-value is "**low**."

*"If the P-value is **low**, the null must **go** [away]."*

We **reject** H_0.

That is, we **reject the assumption** that the coin is fair.

Method 2: Using a Two-Tailed P-Value



$$\text{Two-tailed } P\text{-value} = P(X \geq 59 \text{ or } X \leq 41) \approx 0.0886, \text{ or } 8.86\%$$

$0.0886 > 0.05$, so *P*-value $> \alpha$. The *P*-value is "**not low**."

WARNING 3: "Not low" does **not** necessarily mean "high."

We **do not reject** H_0.

That is, we **do not reject the assumption** that the coin is fair.

Method 3: Using a Confidence Interval (CI)

H_0 states that the coin is **fair**, meaning that the probability of heads $p = 0.500$.

$\alpha = 0.05$, so $1 - \alpha = 0.95$. We can use a **95% confidence interval (CI)** for *p*. Using the methods of Lesson 30, a 95% CI for *p* is:

$$(0.494,\ 0.686)$$

0.500 **is contained in** this CI, so we **do not reject** H_0.

That is, we **do not reject the assumption** that the coin is fair.

Method 4: Using the "Two SD" Rule for Unusual z Scores

Remember from Lesson 10:

> The "Two SD" Rule for Unusual *z* Scores
>
> z scores that are either less than -2 or greater than 2 (that is, $z < -2$ or $z > 2$) are considered "unusual." The corresponding x-values are also considered unusual.

In Lesson 23, Example 1 we used a **normal approximation** to a **binomial distribution** to analyze 100 flips of a fair coin.

- The number of heads, $X \overset{\text{approx.}}{\sim} \text{N}(\mu = 50,\ \ \sigma = 5)$.
- By a continuity correction, $P(X \geq 59) \approx P(X_c > 58.5)$.
- $x_c = x = 58.5 \ \Rightarrow \ z = \dfrac{x - \mu}{\sigma} = \dfrac{58.5 - 50}{5} = 1.70$

1.70 is **not** an "unusual" *z* score, so we **do not reject** H_0.

That is, we **do not reject the assumption** that the coin is fair.

§