ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

HYPOTHESIS TESTING FOR A PROPORTION

PART 2

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Decision Errors

- Hypothesis testing is not perfect: We can make errors based on the data.
- Analogy: Innocent people wrongfully going to jail, and guilty people going free.

		Test Conclusion	
		Do Not Reject H0	Reject H0 in Favor of HA
The Truth	H0 is True	✓	**Type 1 error**
	HA is True	**Type 2 error**	✓

Hypothesis Testing: Significance Levels

- Hypothesis testing entails either rejecting or failing to reject H_0
- We only reject H_0 if we have strong evidence.
- “Strong evidence” is measured by a **significance level** α.
- If the null hypothesis is true, the significance level indicates how often the data lead us to incorrectly reject H_0.
- **Mathematically:** $\mathbb{P}(\textbf{rejecting } H_0 \mid H_0 \textbf{ is true}) = \alpha$
- **Confidence level =** $1 - \alpha$**, so a** $95\,\%$ **confidence interval implies** $\alpha = 0.05$**.**

Formal Testing Using P-Values

- **p-values** are a way of quantifying the strength of the evidence **against** the null hypothesis and in favor of the alternative hypothesis.
- Using p-values is **more common** than using confidence intervals in hypothesis testing.
- **Definition:** A p-value is the probability of observing data at least as favorable to H_A as our current data set, if H_0 were true . We usually use a **summary statistic** to compute p.
- **Let's see how this works on an example:** A polling firm asks a random sample of 1000 adults whether they support Policy A. Let's set up a hypothesis test, at a significance level of $\alpha = 0.05$, to determine whether a majority of adults support Policy A.

H_0: $p = 0.5$ (i.e., 50% of adults support the measure). $\implies p_0 = 0.5$

H_1: $p \neq 0.5$

Formal Testing Using P-Values

- **Example (cont'd):** We know from the CLT that the distribution of $\hat{p}$ is approximately normal when the independence and success failure probabilities both hold. (In our example, they do!)
- The distribution of $\hat{p}$ assuming that H_0 is true is called the **null distribution.**
- Now suppose the result of the poll is that 42% of the sample support the measure.
- Since we are assuming that H_0 is true, we **use p_0 to calculate the standard error** $SE = \sqrt{p_0(1-p_0)/n} = \sqrt{0.5^2/1000} = 0.016.$
- **Important:** This is different from how we calculate confidence intervals (there we use $\hat{p}$ to compute SE). The difference is because when using p-values we are *conditioning on* H_0.

Formal Testing Using P-Values

- **Example (cont'd):** In a **two-sided** test, we ask ourselves:
- "What is the probability that $\hat{p}$ is at least as far in the tails as 0.42 under the null distribution?"
- In our example the null distribution is normal with mean $\mu = 0.5$ and SE = 0.016.
- Under this distribution, we want to compute $\mathbb{P}(\hat{p} \leq 0.42$ or $\hat{p} \geq 0.58)$. This can be done by integration (using software), or by using tables.
- Turns out $\mathbb{P}(\hat{p} \leq 0.42$ or $\hat{p} \geq 0.58) \approx 9.64 \times 10^{-8}$
- This is our p-value!

Formal Testing Using P-Values

Let's interpret this:

- We just calculated that if H_0 were true, the probability of taking a random sample as extreme as the observed proportion is **TINY** $(p - value = 9.64 \times 10^{-8})$.
- So, either we just lucked on an extremely improbable event, or the alternative hypothesis is true!
- More formally, we compare the p-value to α, and since it is less than $\alpha = 0.05$, we reject the null hypothesis.
- Moreover, we notice that the data strongly indicates that a majority of adults do not support Policy A.

Summary: Hypothesis Testing for a Single Proportion

- **There are 4 steps to conducting a hypothesis test for a single proportion**
- **Step 1 (Prepare):** Identify the parameter of interest, list hypotheses, identify the significance level α, and identify $\hat{p}, n$.
- **Step 2 (Check):** Check the conditions for using the CLT, assuming H_0. **Use the null-value** to check the Success/Failure Condition.
- **Step 3 (Calculate):** If the conditions hold, calculate SE (using p_0), and calculate the p-value (either using Z-scores and tables, or integration/software)
- **Step 4 (Conclude):** Compare the p-value to α, and provide your conclusions in the context of the problem at hand.

Final Thoughts: Picking α

- Typically, it is common to see $\alpha = 0.05$, but it is important to adjust it based on the application.
- If making Type 1 errors is particularly bad, we can use a smaller significance level, e.g., $\alpha = 0.01$. In effect, you are demanding stronger evidence before rejecting the null-hypothesis.
- Conversely, if making a Type 2 error is particularly bad, you may want to pick a larger α, e.g., $\alpha = 0.1$.