**Name: Adrian Dsouza**

**Email:** [**ad7628@nyu.edu**](mailto:ad7628@nyu.edu)

**1. How do I assess the statistical significance of an insight?**

* I use hypothesis testing to check if what I observed is likely due to chance.
* I start with a null hypothesis (usually "no effect" or "no difference").
* Then I calculate the **p-value** — if it's below 0.05, I take the result seriously.
* Basically, if the data looks too unusual under the assumption that nothing’s going on, I reject that assumption.

**2. What is the Central Limit Theorem, and why is it important?**

* The Central Limit Theorem says that **sample means become normally distributed** if the sample is large enough—even if the original data isn’t normal.
* This is super useful because it lets me apply z-tests and t-tests on real-world data.
* It also explains why things like confidence intervals and standard error work the way they do.

**3. What is statistical power?**

* Power is the chance that I’ll detect an effect if there *really is* one.
* High power means fewer false negatives (Type II errors).
* I’ve learned that power depends on sample size, effect size, and how much noise (variance) there is in the data.

**4. How do I control for biases?**

* I use **random assignment** in experiments to cancel out confounding effects.
* Blinding helps avoid biased behavior or measurements.
* I also try to **standardize** conditions and sometimes use **statistical controls** like regression when I can't randomize.
* This is especially important because humans naturally look for patterns that confirm what they already believe (confirmation bias).

**5. What are confounding variables?**

* These are variables that mess with my ability to see a clear cause-effect relationship.
* For example, if I'm studying the effect of exercise on health, **diet** could be a confounder.
* That’s why randomizing or adjusting for them in analysis is key.

**6. What is A/B testing?**

* A/B testing is when I compare two versions of something (like two website designs) to see which one performs better.
* I randomly assign users to version A or B and track a metric like click-through rate.
* Then I use statistical tests to see if any difference is just by chance or actually meaningful.

**7. What are confidence intervals?**

* A confidence interval is a range that likely contains the true value I'm trying to estimate (like a mean).
* For example, a 95% CI means I’m 95% sure the true value is somewhere in that range.
* It’s a better way to report results than just saying "here’s the average."