**Module 1 lingo sheet**

1. Terms from the potential outcomes framework
   * **Lingo 1**: In the potential outcomes framework, what are the names for the symbols and
   * **Answer:** Decision or treatment, realized outcome, potential outcome with treatment, potential outcome without treatment
   * **Lingo 2:** In the potential outcome framework, what is the causal effect of a decision?
   * **Answer**: , or the difference in potential outcomes
   * **Lingo 3:** What is the fundamental problem of causal inference?
   * **Answer:** We only observe one of and , not both
2. Prediction questions vs. causal questions
   * **Lingo 4**: What’s the difference between a prediction question and a causal question?
   * **Answer**: for a prediction question, we only need one potential outcome. For a causal question, we need both.
3. Confounding
   * **Lingo** **5**: What’s a confound?
   * **Answer**: A factor which is correlated with decision or treatment, correlated with outcome, and not caused by decision/ treatment
   * **Lingo 6**: What’s Simpson’s paradox?
   * **Answer**: A statistical phenomenon in which a relationship that appears in several groups of data disappears or reverses when the groups are combined.
   * **Lingo 7**: What’s an intermediate outcome?
   * **Answer**: A factor which is correlated with decision or treatment, correlated with outcome, and caused by decision/ treatment. (note difference vs. confound)

**Module 2 lingo sheet**

1. Potential outcomes related terms
   * **Lingo 8:** What is balance (in the context of an experiment)?
   * **Answer:** Balance is the idea that each treatment arm should be similar in terms of pretreatment variables
   * **Lingo 9:** What are pretreatment variables?
   * **Answer:** Pretreatment variables are ones measured or determined prior to treatment assignment
2. Basic elements of experiments
   * **Lingo 10**: What are “units of randomization”?
   * **Answer:** Units that you can both (1) sort randomly into treatment and control, (2) measure outcomes for
   * **Lingo 11:** What are “treatments”?
   * **Answer:** The treatment is the experience that you can control that differs between the units randomized to treatment and those randomized to control
3. Outcome metrics
   * **Lingo 12:** What is an Overall evaluation criteria?
   * **Answer:** The overall evaluation criteria is an outcome or combination of outcomes that is sufficient for decision making
   * **Lingo 13:** What is a proxy outcome?
   * **Answer:** A proxy outcome is an outcome which is not itself directly relevant, but highly correlated with the desired outcome.
4. Types of experiments
   * **Lingo 14:** What is a policy evaluation experiment?
   * **Answer:** Policy evaluation experiments are experiments where the treatment is a decision the business might implement
   * **Lingo 15:** What is a mechanism test experiment?
   * **Answer:** Mechanism tests experiments are experiments to investigate a theory about customer behavior or other business variables

**Module 3 lingo sheet**

1. Subgroups – Improving Precision
   * **Lingo 16:** What is demeaning (in the context of subgroup analysis)?
   * **Answer:** We take the outcome metric for each unit in our data and subtract off its subgroup mean.
2. Subgroups - Balance
   * **Lingo 18:** What is stratified random sampling?
   * **Answer:** It’s the process that we organize data into separate strata (mutually exclusive groupings) and **randomly** select for treatment within each of the strata.
3. Subgroups - Heterogeneous Treatment Effects
   * **Lingo 19:** What are “heterogeneous treatment effects (HTEs)” (in the context of subgroup analysis)?
   * **Answer:** Difference in average outcome metric, by subgroup

**Module 4 lingo sheet**

1. Terms related to hypothesis testing
   * **Lingo 18:** Null hypothesis
   * **Answer:** The hypothesis under consideration.
   * **Lingo 19:** Significance level
   * **Answer:** The lowest probability of a divergence we would be willing to tolerate. The false positive rate (or % of true nulls we will reject).
   * **Lingo 20:** Test statistic
   * **Answer:** Transformation of the data, with higher values less if H0 is true.
   * **Lingo 21:** p-value
   * **Answer:** the probability of observing test statistic at least as large as our actual statistic if H0 is true.
   * **Lingo 22:** Decision rule
   * **Answer:** The rule that determines whether we reject H0 or not under certain circumstances.
   * **Lingo 23:** Confidence interval
   * **Answer:** If p-value is less than , then confidence intervals contain .
2. Terms related to power
   * **Longo 24:** Type I error
   * **Answer:** A false positive.
   * **Lingo 25:** Type II error
   * **Answer:** A false negative.
   * **Lingo 26:** power
   * **Answer:** The true positive rate (or % of false nulls we will reject)
   * **Lingo 27:** When is an effect size practically significant?
   * **Answer:** When it is large enough to justify scaling up treatment**.**
   * **Lingo 28:** What is a noisy estimate?
   * **Answer:** An estimate that is **not** statistically significant but does not rule out a practically significant effect size.
   * **Lingo 29:** When is a result a precise zero?
   * **Answer:** When it rules out a practically significant effect size.
3. Multiple testing and p-hacking
   * **Longo 30:** p-hacking
   * **Answer:** Data analysis practices that inflate statistical significance. Usually a combination of conviction and carelessness.

**Formulas Needed**

* Formula for confidence interval given standard error:
  + CI = x̄ ± (1.96 × SE)
  + CI = ate +- (1.96 \* SE)
* How does standard error change (increase/decrease/remain the same) if the sample size increases? if the standard deviation increases? :
* SE= STD / sqrt(n) 🡪 As sample size increases, the standard error decreases
* SE for difference in means: sqrt (SE(Mt)^2 + SE(Mc)^2)

**Module 5 lingo sheet**

1. Legal and ethical principles
   * **Lingo 31:** What is respect for persons (in the context of experimental ethics)?
   * **Answer:** The idea that experimental subjects should be treated as fully autonomous, with their own rights and decision-making capacity.
   * **Lingo 32:** What is beneficence (in the context of experimental ethics)?
   * **Answer:** The principle that experiments should strive for maximal benefit from the experiment against minimal cost.
   * **Lingo 33:** What is justice (in the context of experimental ethics)?
   * **Answer:** The idea that the procedures should be fair。
   * **Lingo 34:** What is the principle of informed consent?
   * **Answer:** The idea that study participants should be fully informed about the experiment, including all its risks, and should have the decision to opt in or opt out.

**Module 6 lingo sheet**

1. Clustering
   * **Lingo 35:** What is aggregate unit?
   * Answer: An aggregate unit is the level at which randomization takes place.
   * **Lingo 36**: What is clustering (in the context of an experiment)?
   * **Answer:** when the treatment is delivered simultaneously to many individual units who are connected in some way.
2. Imperfect take-up, non-compliance, and excludability
   * **Lingo 37**: What is non-compliance?
   * **Answer:**  When you offer or expose units to treatment, but for a variety of reasons only a subset of people you offered the treatment to actually take-up the treatment.
   * **Lingo 38**: What is the “intention to treat (ITT)” comparison?
   * **Answer:** The entire group randomized into the offer of treatment compared to the control group.
   * **Lingo 39:** What is the “treatment on the treated (ToT)” effect?
   * **Answer:** The effect on those individuals who actually received or complied with the treatment, rather than on everyone who was assigned to the treatment group.
   * **Lingo 40:** What is the excludability condition, in general?
   * **Answer:** Only way assignment to treatment group (i.e., offer of treatment) can affect average outcomes is through the people who take treatment.
   * **Lingo 41:** What is attrition problem?
   * **Answer:** We don’t observe outcome for everyone.
   * **Lingo 42:** What is differential attrition?
   * **Answer:** More people drop out of control than treatment, or vice versa.

**Module 7 lingo sheet**

**1. Natural experiments, event study, and DID**

* **Lingo 43:** What is an observational study
* **Answer:** When the researcher doesn’t control the assignment process.
* **Lingo 44:** What is a natural experiment?
* **Answer:** An observational study where we understand the assignment process and the assignment is as-if random.
* **Lingo 45:** What is an event study?
* **Answer:** A research strategy that exploits abrupt changes in X at a moment in time
* **Lingo 46:** What is a difference in differences design?
* **Answer:** A research strategy that compares how the outcome changed over time for a group of treated units to how it changed over time for a group of comparison units.
* **Lingo 47:** What is the parallel trends assumption?
* **Answer:** The assumption that comparison units and treatment units would have moved in parallel, had there been no treatment.

2. **Regression discontinuity designs**

* **Lingo 48:** What is the regression discontinuity design (RDD)?
* **Answer:** a quasi-experimental research design used to estimate causal effects when a treatment or intervention is assigned based on whether a continuous variable crosses a specific threshold or cutoff.
* **Lingo 49:** What is the scoring variable?
* **Answer:** the continuous variable that determines the treatment assignment in a RDD.