STAT 140: Design of Experiments — Midterm 1 Mock Exam #2

Time: 75 minutes | Total: 100 points Covers: Modules 1–6 (Potential Outcomes, Assignment Mechanisms, Fisher, Neyman, Bayesian Inference)

Part A: True or False (40 points, 2 points each)

For each statement, state TRUE or FALSE and provide a brief justification (1–2 sentences). No credit without justification.

Potential Outcomes & Science Table

A1. The individual treatment effect $\tau_i = Y_i(1) - Y_i(0)$ can be directly computed from observed data for at least some units in a randomized experiment.

A2. In a science table with $N = 10$ units, there are exactly 10 missing potential outcomes (assuming binary treatment with no missing data).

A3. SUTVA (Stable Unit Treatment Value Assumption) states that a unit's potential outcome depends only on its own treatment assignment and not on the treatment assignments of other units.

A4. The observed outcome can be written as $Y_i^{\text{obs}} = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$. This equation requires SUTVA to hold.

A5. If the ATE (Average Treatment Effect) is zero, it is possible that some individual treatment effects τ_i are non-zero.

Assignment Mechanisms

A6. In a Bernoulli trial with $p = 0.5$ and $N = 10$, it is possible that all 10 units are assigned to treatment.

A7. Under a completely randomized design, the assignment mechanism is both individualistic and unconfounded.

A8. Block randomization always produces a more precise estimate of the ATE than a completely randomized design.

A9. In a paired randomized experiment with 10 pairs (20 units total), the number of possible assignment vectors is $\binom{20}{10} = 184{,}756$.

A10. Rerandomization restricts the set of acceptable allocations, but each accepted allocation still has equal probability of being selected.

Fisher Inference

A11. The Fisher randomization test requires the assumption that potential outcomes are normally distributed.

A12. Under the sharp null hypothesis, the test statistic T is a random variable whose randomness comes from the randomization of treatment assignment, not from sampling from a population.

A13. If the Fisher exact p-value is 0.03, it means there is a 3% probability that the treatment has no effect.

A14. The Fisher randomization test can use any test statistic, not just the difference in means $\bar{Y}_t - \bar{Y}_c$.

A15. The Hodges-Lehmann estimator $\hat{\tau}_{HL}$ is the value of τ_0 that maximizes the Fisher p-value when testing $H_0 : \tau_i = \tau_0$ for all i.

Neymanian Inference

A16. The difference-in-means estimator $\hat{\tau} = \bar{Y}_t - \bar{Y}_c$ is unbiased for the ATE under CRD, regardless of whether the treatment effect is constant or heterogeneous.

A17. If the individual treatment effects are constant ($\tau_i = \tau$ for all i), then Neyman's conservative variance estimator is exact (not conservative).

A18. Neyman's confidence interval is exact for any sample size.

Bayesian Inference

A19. In the Bayesian approach, single imputation correctly captures the uncertainty about the missing potential outcomes because it uses the best available estimate (the group mean).

A20. The Bayesian Bootstrap imputes missing potential outcomes by drawing from a parametric normal distribution fitted to the observed data.

Part B: Concept Explanations (30 points)

Answer each question in 3–5 sentences. Be precise and use correct terminology.

B1. (5 points) Explain the **Fundamental Problem of Causal Inference**. Why does this problem exist, and how does randomization help address it (even though it does not solve it completely)?

B2. (5 points) What is SUTVA? Give one realistic example where SUTVA would be violated and explain why the violation matters for causal inference.

B3. (5 points) A student claims: "Randomization is just about making the two groups look similar in terms of covariates (like age, gender, etc.)." Is this a complete explanation? What deeper role does randomization play in the potential outcomes framework?

B4. (5 points) Explain the difference between the **sharp null hypothesis** and the **weak null hypothesis** (also called the Neyman null). Why does the sharp null allow us to fill in the entire science table, while the weak null does not?

B5. (5 points) What does it mean for Neyman's variance estimator to be "conservative"? Under what specific condition is it no longer conservative but exact? Explain intuitively why the $S^2(\tau)/N$ term disappears in that case.

B6. (5 points) Compare the Fisher confidence interval (via test inversion) and the Neyman confidence interval. What are the key differences in how each is constructed, and when would you prefer one over the other?

Part C: Calculation Problems (30 points)

Problem C1: Complete Fisher Analysis (15 points)

A researcher tests whether **listening to classical music** during study improves exam scores. She conducts a CRD with **4 students**, assigning **2 to music** (treatment) and **2 to silence** (control).

Observed data:

Student	W_i	Y_i^{obs}
1	1 (Music)	88
2	1 (Music)	92
3	0 (Silence)	80
4	0 (Silence)	84

(a) (2 points) Compute $T^{\text{obs}} = \bar{Y}_t - \bar{Y}_c$.

(b) (2 points) How many possible allocations exist under CRD? List them all.

(c) (2 points) Write the sharp null hypothesis. Under this hypothesis, fill in the complete science table.

(d) (4 points) For each possible allocation, compute T. Present your results in a table.

(e) (2 points) Compute the two-sided Fisher exact p-value ($|T| \geq |T^{\text{obs}}|$).

(f) (3 points) Now compute the Hodges-Lehmann estimate. List all pairwise differences between treated and control units, and find their median.

Problem C2: Neymanian Analysis (10 points)

A large-scale experiment examines whether a **new teaching method** improves standardized test scores. 200 students participate: 100 in the new method (treatment), 100 in the traditional method (control).

Results:

Group	N	$\bar{Y}$	s^2
New Method	100	78.4	156.25
Traditional	100	73.1	182.49

(a) (2 points) Compute $\hat{\tau}$.

(b) (2 points) Compute $\hat{V}_{\text{Neyman}}$ and the standard error SE.

(c) (2 points) Construct a 95% confidence interval.

(d) (2 points) Construct a 99% confidence interval (use $z_{0.005} = 2.576$).

(e) (2 points) A skeptic claims: "The new method doesn't help at all." Based on your 95% CI, how do you respond? What if they made this claim against your 99% CI?

Problem C3: Bayesian Bootstrap (5 points)

Using the same data from Problem C1 (the music experiment with 4 students):

Student	W_i	Y_i^{obs}
1	1	88
2	1	92
3	0	80
4	0	84

(a) (1 point) What are the two sampling pools for the Bayesian Bootstrap?

(b) (2 points) Enumerate **all** possible completed science tables under the Bayesian Bootstrap (hint: each missing value can take one of 2 values, so there are $2^4 = 16$ combinations). For each, compute the ATE.

(c) (2 points) What is the mean and range of the ATE across all 16 imputations? How does this compare to the simple difference-in-means T^{obs} from C1(a)?

END OF EXAM