Causal Inference and Statistical Designs

Example Sheet 1

1. **Neyman's variance formula.** Use the notation in Lectures 1 and 2. Assume the Neyman–Rubin causal model and complete randomization ($n_0, n_1 \geq 2$). Prove Neyman's formula for randomization variance:

$$\mathrm{Var}\{\hat{\beta} \mid Y(\cdot)\} = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{01}^2}{n},$$

where S_1^2, S_0^2, S_{01}^2 are the sample variances of $Y_i(1)$, $Y_i(0)$, and $Y_i(1) - Y_i(0)$, $i = 1, \ldots, n$, respectively, by following the steps below:

   (a) By using the fact that $\mathrm{Var}\{\hat{\beta} \mid Y(\cdot)\}$ is a quadratic form in $(Y(0), Y(1))$ that is invariant to simultaneously relabelling the units $1, \ldots, n$ in both vectors and to adding a constant to either $Y(0)$ or $Y(1)$, show that $\mathrm{Var}\{\hat{\beta} \mid Y(\cdot)\} = c_1 S_1^2 + c_0 S_0^2 + c_{01} S_{01}^2$ for constants c_1, c_0, c_{01} depending only on n_0, n_1.

   (b) Determine the constants c_1, c_0, c_{01} by computing $\mathrm{Var}(\hat{\beta})$ when $(Y_i(0), Y_i(1))$ are i.i.d. bivariate normal, independent of Z, with mean 0, variances σ_0^2 and σ_1^2, and correlation ρ. [*Hint:* apply the law of total variance to $\hat{\beta}$ in two ways, conditioning on $Y(\cdot)$ and on Z.]

2. **Kidney-stone treatments.** The 2×2 table below reports the outcomes of 350 patients who underwent open surgery during 1972–1980 and 350 who underwent a small puncture procedure called percutaneous nephrolithotomy during 1980–1985. Success meant that stones were absent or reduced to particles smaller than 2 mm at three months. More information can be found in the original study by Charig et al. (1986), *BMJ* 292, 879–882.

	Success	Failure
Open surgery	273	77
Small puncture	289	61

   Apply a one-sided Fisher exact test to assess evidence for higher success probability with the small puncture procedure. In R, use `alternative = "less"` in `fisher.test()` with the table in the displayed order. How would you interpret the result of this test causally?

   The original study further broke down the numbers into patient groups with small (left table below) and large (right) stones. Compare the success proportions within each group. How does your conclusion change with this additional information?

	Success	Failure
Open surgery	81	6
Small puncture	234	36

	Success	Failure
Open surgery	192	71
Small puncture	55	25

3. **Barnard's test.** Let $X_1 \sim \mathrm{Bin}(4, p_1)$ and $X_0 \sim \mathrm{Bin}(4, p_0)$ independently. Consider testing

$$H_0 : p_1 = p_0 \qquad \text{against} \qquad H_1 : p_1 > p_0.$$

Fisher's test conditions on $S = X_1 + X_0$, under which X_1 has a hypergeometric null distribution. To attain a prescribed conditional size, the test can randomize at its boundary: given $S = s$, reject if $X_1 > c_s$, reject with probability r_s if $X_1 = c_s$, and otherwise do not reject. The decision at equality uses independent randomization. Tocher (1950) showed that this test is uniformly most powerful among all unbiased tests of the same (unconditional) size; see Lehmann and Romano (2022, p.143-144).

Barnard's pooled-score test uses

$$T = \frac{X_1/4 - X_0/4}{\sqrt{\hat{p}(1-\hat{p})(1/4 + 1/4)}}, \qquad \hat{p} = \frac{X_1 + X_0}{8},$$

taking $T = 0$ when $\hat{p} \in \{0, 1\}$. At nominal level 0.05, it chooses the smallest attainable cutoff t such that $\mathsf{P}_{p,p}(T \geq t) \leq 0.05$ for every common null probability p. Here $t = \sqrt{24/5}$, so it rejects at $(X_1, X_0) = (3, 0), (4, 0), (4, 1)$.

(a) A test's size is its maximum null rejection probability. Show that Barnard's test has actual size $\alpha = 9/256$.

(b) Verify that the randomized Fisher test with

$$(c_0, \ldots, c_8) = (0, 1, 2, 3, 3, 4, 4, 4, 4),$$
$$(r_0, \ldots, r_8) = \frac{1}{2048}(72, 144, 336, 1008, 187, 1008, 336, 144, 72)$$

has conditional size α for every total $s = 0, \ldots, 8$. What is its unconditional size?

(c) Compare the powers of the two tests at $(p_1, p_0) = (0.30, 0.05)$, $(0.80, 0.20)$ and $(0.11, 0.10)$. Do these results contradict the theorem that the randomized Fisher test is uniformly most powerful among unbiased tests of this size?

4. **Probability integral transform.** Let $F(t) = \mathsf{P}(T \leq t)$ be the cumulative distribution function of a real random variable T. Prove that $\mathsf{P}\{F(T) \leq \alpha\} \leq \alpha$ for every $0 \leq \alpha \leq 1$, allowing atoms in the distribution of T.

5. **Regression adjustment.** Let $(X_i, Y_i(0), Y_i(1)) \in \mathbb{R}^3$ be i.i.d. with finite fourth moments, $\mathsf{E}\, X_i = 0$ and $\mathrm{Var}(X_i) > 0$. Use this known population mean rather than sample-centring X_i. Independently of these variables, assign $A_i \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\pi)$, $0 < \pi < 1$, and observe $Y_i = Y_i(A_i)$. Let $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ be the least-squares estimators of b from minimizing $\sum_{i=1}^n (Y_i - \alpha - bA_i)^2$, $\sum_{i=1}^n (Y_i - \alpha - bA_i - \gamma X_i)^2$ and $\sum_{i=1}^n (Y_i - \alpha - bA_i - \gamma X_i - \delta A_i X_i)^2$, respectively. Let $\beta = \mathsf{E}\{Y_i(1) - Y_i(0)\}$.

(a) Show that, for $j = 1, 2, 3$,

$$\sqrt{n}(\hat{\beta}_j - \beta) \xrightarrow{d} \mathrm{N}(0, V_j), \qquad V_j = \frac{\mathsf{E}\{(A_i - \pi)^2 \epsilon_{ij}^2\}}{\pi^2 (1-\pi)^2},$$

where ϵ_{ij} is the residual from the jth population least-squares regression (minimizing $\mathsf{E}\{(Y_i - \alpha - bA_i - \cdots)^2\}$). Do not assume any of the linear models is correctly specified.

(b) Show that $V_3 \leq \min\{V_1, V_2\}$.

6. **Stepped-wedge design and quasi-randomization tests.** In the first Dandenong trial reported by Haines et al. (2017), six wards withdrew weekend allied health services in a uniformly random order, one ward per month during months 2–7. The realized withdrawal months were $Z = (3, 5, 7, 6, 2, 4)$.

   Run the R starter on the course website to download the original data and create the clean data frame `data` and schedule matrix `S`. Each row of `data` is one admission. Let Y_i be log(patient length of stay), and let W_i, T_i be the ward and admission month. Under a withdrawal schedule z, exposure is

$$A_i(z) = \mathbf{1}\{T_i \geq z_{W_i}\},$$

   so $A_i(z) = 1$ means that services had been withdrawn when the patient was admitted.

   (a) Run these regressions, both with an intercept. Compare the exposure coefficients, standard errors and (two-sided) p-values.

```r
fit1 <- lm(log_los ~ exposure, data = data)
fit2 <- lm(log_los ~ exposure + factor(ward), data = data)
summary(fit1)
summary(fit2)
```

   (b) Test the sharp null that the withdrawal has no causal effect whatsoever. For each of the $6! = 720$ schedules in `S`, recompute exposure and refit both models, using their exposure coefficients as test statistics. Reproduce the upper-tail randomization p-values $60/720$ and $3/720$ in the lectures. Compare them with the p-values in (a). Why are these randomization tests valid even if the regression models are incorrect?

   (c) Now shuffle the `month` column across all hospital admissions, leaving each patient's ward and outcome and the observed schedule Z in place. Using the same exposure mapping and Monte Carlo simulations, find (approximate) upper-tail p-values using the same test statistics as in (b). Do these "quasi-randomization tests" give stronger causal evidence than (b)? Why or why not?

7. **Walks, acyclicity and the trek rule.** Let $G \in \mathbb{G}(V)$ be a directed mixed graph with vertex set V.

   (a) Let A be the adjacency matrix of the directed edges of G. Show that G is acyclic if and only if $A^{|V|} = 0$.

   (b) Suppose $\rho(\beta) < 1$, where $\rho(\beta)$ is the largest modulus of an eigenvalue of β. Prove that

$$\sum_{r=0}^{\infty} \beta^r = (I - \beta)^{-1}.$$

   [*Hint:* you may use Gelfand's formula $\rho(\beta) = \lim_{m\to\infty} \|\beta^m\|_2^{1/m}$, where $\|\cdot\|_2$ is the Euclidean operator norm.]

   (c) For a linear system on G with $I - \beta$ nonsingular, prove that

$$W[V_j \overset{\mathrm{t}}{\leftrightsquigarrow} V_k] = \emptyset \quad \Longrightarrow \quad \Sigma_{jk} = 0.$$

   If, in addition, $\rho(\beta) < 1$, show that

$$\Sigma = \sigma\big(W[V \overset{\mathrm{t}}{\leftrightsquigarrow} V]\big).$$

8. **Wright's path-analysis formula.** Suppose $G \in \mathbb{G}_A(V)$ is acyclic.

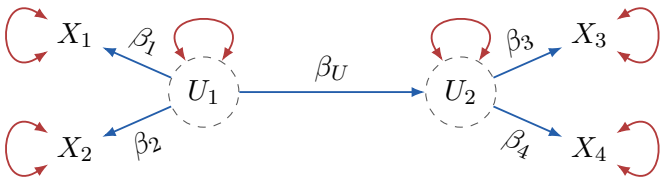
   (a) Prove Wright's formula from Lecture 5 by induction on $|V|$, assuming $\mathrm{Var}(V_j) = 1$ for every j:

   $$\mathrm{Cov}(V_j, V_k) = \sigma\big(P[V_j \rightsquigarrow V_k]\big), \qquad j \neq k.$$

   (b) How would you modify this formula if we no longer assume unit variances?

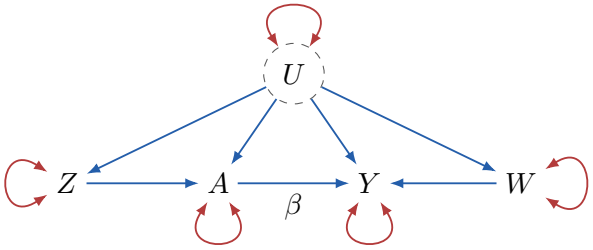
9. **Latent variables and identification.** Consider the following Gaussian linear systems.

   (a) In the first graph, only $X_1, \ldots, X_4$ are observed. Assume $\mathrm{Var}(U_1) = \mathrm{Var}(U_2) = 1$, $\beta_1, \ldots, \beta_4 \neq 0$ and $0 < |\beta_U| < 1$.



   Recover the coefficients and noise variances from $\mathrm{Cov}(X_1, \ldots, X_4)$ up to sign changes, and describe the sign ambiguity.

   (b) In the second graph, only U is unobserved. Let β be the coefficient of $A \longrightarrow Y$.



   Remove $A \longrightarrow Y$ (assuming $\beta = 0$) and prove the tetrad constraint $\Sigma_{ZY}\Sigma_{AW} - \Sigma_{AY}\Sigma_{ZW} = 0$. Use this to derive an identification formula for β in the original system.

10. **Marginalizing the noises.** Complete the noise-covariance calculation left open in the marginalization theorem in Lecture 6 (click here). Let $U = V \setminus \tilde{V}$ and suppose $\rho(\beta_{UU}) < 1$. For a nonsingular Gaussian linear system on G, express $\tilde{E}$ in the marginal system in terms of the original noises and expand its covariance into four terms. Match these to a decomposition of unblocked treks to prove

    $$\tilde{\Lambda}_{jk} = \sigma\big(W[V_j \overset{\mathrm{t}}{\leftrightsquigarrow} V_k \mid \tilde{V} \text{ in } G]\big), \qquad V_j, V_k \in \tilde{V}.$$

    Sum over treks whose two directed legs each have length at most n, then let $n \to \infty$.

11. **Regression and conditioning.**

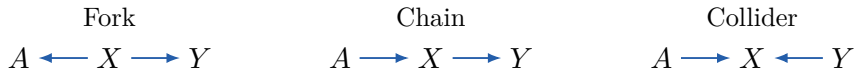
    (a) Let $(A, X, Y) \in \mathbb{R}^3$ be jointly Gaussian, with mean 0, covariance matrix $\Sigma \succ 0$ and precision matrix $\Omega = \Sigma^{-1}$. Let b_A and $b_{A|X}$ be the coefficients of A in the population regressions of Y on A and on (A, X). Derive

    $$b_A = \frac{\Sigma_{AY}}{\Sigma_{AA}},$$
    $$b_{A|X} = \frac{\Sigma_{AY} - \Sigma_{AX}\Sigma_{XY}/\Sigma_{XX}}{\Sigma_{AA} - \Sigma_{AX}^2/\Sigma_{XX}} = -\frac{\Omega_{AY}}{\Omega_{YY}}.$$

Show also that

$$\mathrm{Cor}(A, Y \mid X) = -\frac{\Omega_{AY}}{\sqrt{\Omega_{AA}\Omega_{YY}}}, \qquad A \perp\!\!\!\perp Y \mid X \iff \Omega_{AY} = 0.$$

(b) In each graph below, all directed coefficients are 1 and the intrinsic noises are independent $\mathcal{N}(0, 1)$ variables. Bidirected self-loops are omitted.



Using (a), calculate $\mathrm{Cov}(A, Y)$, $\mathrm{Cov}(A, Y \mid X)$, b_A and $b_{A|X}$. Compare the regression coefficients with the total causal effect of A on Y, and explain graphically.

(c) Add a child $D = X + E_D$ to the collider, where $E_D \sim \mathcal{N}(0, 1)$ is independent of the other noises. Calculate $\mathrm{Cov}(A, Y \mid D)$ and the coefficient of A in the regression of Y on (A, D). Can you tell from the graph why $\mathrm{Cov}(A, Y \mid D)$ is generally not zero?

12. **Academic self-concept and achievement.** Read the abstract and "Tests of Initial a Priori Model" (pp. 649–651) in Marsh (1990) (click here).

(a) What are the research question and main conclusion?

(b) In Figure 1, which variables are observed and which are latent? What is the meaning of the double-headed arrows?

(c) How do Models 1, 2 and 3 differ?

(d) Each latent grade has only one measurement. Is this measurement model identifiable without further assumptions? Where does Marsh discuss this?

(e) Suppose the latent grades are accurately observed. Is Model 3 generically identifiable, up to latent-factor signs and scales, from observed covariances? Justify your answer for the stipulated causal order and error assumptions. [Figure 1 suppresses nonsignificant arrows for display; use the full fitted forward model.]