1) A scientist has a large population of $4n$ people; $2n$ of the people are men and $2n$ of the people are women. The scientist assigns treatment to exactly half of the men and half of the women, and assigns the rest of the people to control. Let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes of person $i$ given treatment and control respectively. The scientist wants to estimate the average treatment effect ($ATE$):

$$ATE = \frac{1}{4n}\sum_{i=1}^{4n} Y_i(1) - \sum_{i=1}^{4n} Y_i(0)$$

The scientist assumes the following model:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 M_i + \epsilon_i$$

Here, $T_i$ is a treatment indicator, and $M_i$ is an indicator variable for whether subject $i$ is a male. The scientists estimates the $ATE$ by computing $\widehat{ATE} = \hat{\beta}$ using OLS. Under these assumptions, the OLS estimate is unbiased for the $ATE$.

**TRUE:** Under this setup, the treatment indicators $T_i$ and the male indicators $M_i$ are uncorrelated. From page 121 of Professor Sekhon's notes, it follows that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{4n}(T_i - \bar{T})(Y_i - \bar{Y})}{\sum_{i=1}^{4n}(T_i - \bar{T})^2} = \frac{\frac{1}{2}\sum_{i=1}^{4n}(Y_i(1)T_i - \bar{Y}) - \frac{1}{2}\sum_{i=1}^{4n}(Y_i(0)(1 - T_i) - \bar{Y})}{\sum_{i=1}^{4n}\frac{1}{4}}$$

$$= \frac{\sum_{i=1}^{4n} Y_i(1)T_i}{2n} - \frac{\sum_{i=1}^{4n} Y_i(0)(1 - T_i)}{2n}$$

which is an unbiased estimate of the ATE.

2) Consider a large medical trial for a new weight loss drug. Before the trial, each patient has their weight, height, and body fat percentage measured. A goodness-of-health score is calculated for each patient based on these measurements (higher scores are a proxy for worse health). Assume that patients do not have time to manipulate their weight or body fat once selected to participate in the trial. Historically, a histogram of patient goodness-of-health scores closely follows a normal distribution with mean $c$. It is thought that the effect of the drug varies with the value of this score.

   a) Suppose that the trial is conducted so that people with a goodness-of-health score of $c$ or above are given the drug and that people with a score below $c$ are not given the drug. Under this setup, what meaningful inferences can be made about the effect of the drug on weight loss? Discuss the parameter of interest and the methods used to estimate this parameter. What assumptions are required to estimate this parameter?

   **ANSWER:** This is a textbook example of when regression discontinuity is useful. We estimate the local average treatment effect (LATE)

$$\mathbb{E}(Y_i(1) - Y_i(0)|score = c)$$

   This parameter can be estimated by, say, fitting two lines within a small neighborhood $(c-\epsilon, c+\epsilon)$; one line is fitted to points within $(c-\epsilon, c)$, and another line is fitted to points within $[c, c+\epsilon)$; and taking the difference between the estimates at $c$ for both of these lines. This framework requires continuity of potential outcomes at $c$, and smoothness of covariates within the neighboorhood.

   b) Consider the following mechanism for treatment assignment: Before being assigned to treatment or control, each patient rolls a 6-sided die. If the die comes up as 1, 2, 3, or 4 and the patient has a score of $c$ or above, that patient takes the weight loss drug, otherwise they receive a placebo. If the die comes up as a 5 or 6 and the patient has a score below $c$, that patient takes the weight loss drug, otherwise they receive a placebo. Suppose that the die roll is unknown to the experimenter. What does the estimate in a) measure now? Can the parameter of interest in part a) still be estimated? If so, how? If not, why not?

**ANSWER:** We still want to estimate $\mathbb{E}(Y_i(1) - Y_i(0)|score = c)$. However, by fitting a regression within $(c - \epsilon, c)$, and obtaining the estimate at $c$, we no longer estimate $\mathbb{E}(Y_i(0)|score = c)$, but instead we estimate $1/3\mathbb{E}(Y_i(1)|score = c) + 2/3\mathbb{E}(Y_i(0)|score = c)$. Similarly, fitting a regression within $[c, c + \epsilon)$ and predicting at $score = c$ gives an estimate of $2/3\mathbb{E}(Y_i(1)|score = c) + 1/3\mathbb{E}(Y_i(0)|score = c)$.

It follows that the process in a) gives an estimate of

$$
\begin{aligned}
& 2/3\mathbb{E}(Y_i(1)|score = c) + 1/3\mathbb{E}(Y_i(0)|score = c) \\
& -(1/3\mathbb{E}(Y_i(1)|score = c) + 2/3\mathbb{E}(Y_i(0)|score = c)) \\
= \; & \frac{\mathbb{E}(Y_i(1)|score = c) - \mathbb{E}(Y_i(0)|score = c)}{3}
\end{aligned}
$$

Thus, one way we can estimate the quantity $\mathbb{E}(Y_i(1) - Y_i(0)|score = c)$ is to use the estimation procedure in a) and multiply the that answer by 3.

c) Suppose that the effect of the drug is thought to be the same for all patients with goodness-of-health scores within the interval $(c - 5, c + 5)$. Suppose that patients with scores below $c$ are ineligible to receive the weight loss drug. Patients with scores of $c$ or above are given an appointment to receive the drug. The drug is administered only once during the trial, and only at this appointment. Some patients fail to arrive at their appointment. Under this setup, discuss at least two types of inference possible for measuring the effect of the drug on weight loss? Discuss the parameters of interest and the methods used to estimate these parameters. What assumptions are required to estimate these parameters? Which estimate will be larger (in absolute value)?

**ANSWER:** Let $A_i$ denote an "assigned-to-treatment" indicator, and let $\#A$ denote the number of units that were assigned to treatment. Let $\hat{\alpha}$ denote the proportion of treated units that took treatment.

We can estimate either the intention to treat effect (ITT) or the effect of the treatment on the treated (ETT).

The ITT parameter is:
$$\mathbb{E}(Y_i(1) - Y_i(0))$$

The ETT parameter is:
$$\mathbb{E}(Y_i(1) - Y_i(0)|compliance)$$

The estimate for the ITT is:
$$\frac{\sum_{i=1}^n Y_i A_i}{\#A} - \frac{\sum_{i=1}^n Y_i(1 - A_i)}{n - \#A}$$

The estimate for ETT, which is an IV estimator, is:
$$\frac{\frac{\sum_{i=1}^n Y_i A_i}{\#A} - \frac{\sum_{i=1}^n Y_i(1 - A_i)}{n - \#A}}{\hat{\alpha}}$$

Since $\hat{\alpha} \leq 1$, the ETT estimator will be greater (in absolute value).

There is a warning with this analysis; assignment to treatment is not explicitly random. It may be argued that a plus-or-minus 5 point fluctuation in score is essentially random. Additionally, it may be argued that, aside from the difference in score, these groups are identical in every way. Some assumption like this is necessary in order to properly carry out the analysis; if some observed or unobserved covariate affects response, the estimates will be biased. This assumption can be falsified by checking balance on pretreatment covariates, if some were measured. Randomization is the only real way to ensure estimation without bias.

The ITT estimate requires one of the above assumptions to hold. The ETT estimate requires one of the above assumptions, and also requires the proportion of noncompliers in the treatment group to be the same as the number of noncompliers in the control group.

3) Suppose that there is a study with a total of $2n$ subjects. Exactly $n$ of these people are smokers. A height and weight are measured for each subject. Suppose that there are enough people so that the joint distribution of the heights and weights is extremely close to a multivariate normal distribution. The researcher wants to test whether smoking affects 40-yard dash times.

a) A statistician notices some imbalance in the average weight and height between the smokers and the non-smokers. To fix the imbalance, the statistician matches smokers to non-smokers by matching on the Mahalanobis distance with height and weight covariates (with replacement, nearest neighbor). Will the differences in average height and average weight between the smokers and matched non-smokers be as small or smaller as they were before matching? Why or why not?

**ANSWER:** Yes. Matching with the mahalanobis distance is EPBR when the data is ellipsoidal. The multivariate normal is an ellipsoidal distribution. See page 368 in Professor Sekhon's notes for more detail.

b) Suppose instead that all subjects in the study are twins. For each set of twins, one twin is a smoker and one twin is a non-smoker, and both twins in each set have the same height and weight. In his analysis, the statistician believes that the smoking sibling in a twin pair is essentially random, though he concedes that some unobserved trait may help explain a twin's propensity for being a smoker.

For each set of twins $s$, let $(1, s)$ denote the twin that smokes, and let $(2, s)$ denote the non-smoking twin. Let $T_{is}$ denote random smoking indicators; $T_{is} = 1$ if the $i$th unit in the $s$th twin pair smokes, $i = 1, 2$. For this study, for each pair $s$, the smoking indicators are observed to be $T_{1s} = 1$ and $T_{2s} = 0$. The statistician models the probability that a subject smokes in the following way:

$$\log\left(\frac{P(T_{is} = 1)}{1 - P(T_{is} = 1)}\right) = \alpha + \kappa_1 h_{is} + \kappa_2 w_{is} + \gamma u_{is} \tag{1}$$

where $h_{is}$ and $w_{is}$ are the height and weight of twin $(i, s)$, and $u_{is}$ is the value of an unobserved covariate for that twin. The statistician also assumes that any subject cannot influence any other subject to smoke or not smoke (smoking is independent across all subjects).

Show that, under this model, the probability that subject $(1, s)$ is a smoker is:

$$P(T_{1s} = 1 | T_{1s} + T_{2s} = 1) = \frac{e^{\gamma u_{1s}}}{e^{\gamma u_{1s}} + e^{\gamma u_{2s}}} \tag{2}$$

Hint: Use $P(A|B) = P(A \cap B)/P(B)$, and find an expression for

$$\frac{P(T_{1s} = 1 \cap T_{2s} = 0)}{P(T_{1s} = 0 \cap T_{2s} = 1)} = \left(\frac{P(T_{1s} = 1)}{1 - P(T_{1s} = 1)}\right)\left(\frac{P(T_{2s} = 1)}{1 - P(T_{2s} = 0)}\right)$$

**ANSWER:** Note that, under the assumption that twins in the same pair have the same height and weight:

$$\left(\frac{P(T_{1s} = 1)}{1 - P(T_{1s} = 1)}\right)\left(\frac{P(T_{2s} = 0)}{1 - P(T_{2s} = 0)}\right)$$
$$= \left(\frac{P(T_{1s} = 1)}{1 - P(T_{1s} = 1)}\right)\left(\frac{1 - P(T_{2s} = 1)}{P(T_{2s} = 1)}\right)$$
$$= \frac{\exp(\alpha + \kappa_1 h_{1s} + \kappa_2 w_{1s} + \gamma u_{1s})}{\exp(\alpha + \kappa_1 h_{2s} + \kappa_2 w_{2s} + \gamma u_{2s})}$$
$$= \exp(\alpha - \alpha + \kappa_1 h_{1s} + \kappa_2 w_{1s} - (\kappa_1 h_{2s} + \kappa_2 w_{2s}) + \gamma u_{1s} - \gamma u_{2s})$$
$$= \exp(\gamma u_{1s} - \gamma u_{2s})$$

3

It follows that:

$$P(T_{1s} = 1|T_{1s} + T_{2s} = 1) = \frac{P(T_{1s} = 1 \cap T_{1s} + T_{2s} = 1)}{P(T_{1s} + T_{2s} = 1)}$$

$$= \frac{P(T_{1s} = 1 \cap T_{2s} = 0)}{P(T_{1s} = 1 \cap T_{2s} = 0) + P(T_{1s} = 0 \cap T_{2s} = 1)}$$

$$= \frac{1}{1 + \dfrac{P(T_{1s} = 0 \cap T_{2s} = 1)}{P(T_{1s} = 1 \cap T_{2s} = 0)}}$$

$$= \frac{1}{1 + \dfrac{1}{\exp(\gamma u_{1s} - \gamma u_{2s})}}$$

$$= \frac{1}{1 + \dfrac{\exp(\gamma u_{2s})}{\exp(\gamma u_{1s})}}$$

$$= \frac{\exp(\gamma u_{1s})}{\exp(\gamma u_{1s}) + \exp(\gamma u_{2s})}$$

c) Suppose that $0 \le u_{is} \le 1$ and that $\gamma > 0$. Find an upper and lower bound (sharper than just 1 and 0) for the probability $P(T_{1s} = 1|T_{1s} + T_{2s} = 1)$. Denote these bounds by $p_s^+$ and $p_s^-$ respectively. Do the same for $P(T_{1s} = 0|T_{1s} + T_{2s} = 1)$. Comment, in one sentence, on how these bounds change if $\gamma < 0$.

**ANSWER:** For $P(T_{1s} = 1|T_{1s} + T_{2s} = 1)$:
Biggest value when $u_{1s} = 1$ and $u_{2s} = 0$:

$$P(T_{1s} = 1|T_{1s} + T_{2s} = 1) = \frac{e^\gamma}{e^\gamma + 1}$$

Smallest value when $u_{1s} = 0$ and $u_{2s} = 1$:

$$P(T_{1s} = 1|T_{1s} + T_{2s} = 1) = \frac{1}{e^\gamma + 1}$$

For $P(T_{1s} = 0|T_{1s} + T_{2s} = 1)$:
Note that

$$P(T_{1s} = 0|T_{1s} + T_{2s} = 1) = \frac{\exp(\gamma u_{2s})}{\exp(\gamma u_{1s}) + \exp(\gamma u_{2s})}$$

Biggest value when $u_{1s} = 0$ and $u_{2s} = 1$:

$$P(T_{1s} = 1|T_{1s} + T_{2s} = 1) = \frac{e^\gamma}{e^\gamma + 1}$$

Smallest value when $u_{1s} = 1$ and $u_{2s} = 0$:

$$P(T_{1s} = 1|T_{1s} + T_{2s} = 1) = \frac{1}{e^\gamma + 1}$$

If $\gamma < 0$, the minimums become maximums and vice versa.

d) Let $y_{is}$ denote the 40-yard dash time of subject $(i, s)$ in milliseconds. Let $Z_s$ denote an indicator variable for the smoker having the faster 40-yard dash time: $Z_s = 1$ if and only if the smoking twin has a faster 40-yard dash time than the non-smoking twin. Let $d_s$ denote the rank of $|y_{1s} - y_{2s}|$; higher ranks denote larger absolute values. Assume there are no ties between $y_{1s}$ and $y_{2s}$ within any twin pair $s$, and that $|y_{1s} - y_{2s}| \ne |y_{1t} - y_{2t}|$ for all distinct twin pairs $s, t$.

The Wilcoxon signed rank statistic is:

$$W = \sum_{s=1}^{n} d_s Z_s.$$

Let $Z_s^+$ and $Z^-$ be independent and identically distributed bernoulli random variables (or indicator variables) with $P(Z_s^+ = 1) = p_s^+$ and $P(Z_s^- = 1) = p_s^-$. Consider the following random variables:

$$
\begin{aligned}
W^+ &= \sum_{s=1}^{n} d_s Z_s^+ \\
W^- &= \sum_{s=1}^{n} d_s Z_s^-
\end{aligned}
$$

Show that, under the null hypothesis that smoking does not effect 40-yard dash times, the following property holds:

$$\mathbb{E}(W^-) \leq \mathbb{E}(W | T_{1s} + T_{2s} = 1) \leq \mathbb{E}(W^+)$$

**ANSWER:** Note that:

$$
\begin{aligned}
\mathbb{E}(W^-) &= \sum_{s=1}^{n} d_s \mathbb{E}(Z_s^-) = \sum_{s=1}^{n} d_s p_s^- \\
\mathbb{E}(W | T_{1s} + T_{2s} = 1) &= \sum_{s=1}^{n} d_s \mathbb{E}(Z_s | T_{1s} + T_{2s} = 1) \\
&= \sum_{s=1}^{n} d_s \left[ P(T_{1s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{1s} > y_{2s}) + P(T_{2s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{2s} > y_{1s}) \right] \\
\mathbb{E}(W^+) &= \sum_{s=1}^{n} d_s \mathbb{E}(Z_s^+) = \sum_{s=1}^{n} d_s p_s^+
\end{aligned}
$$

Since

$$
\begin{aligned}
& P(T_{1s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{1s} > y_{2s}) + P(T_{2s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{2s} > y_{1s}) \\
\geq \; & p_s^- \mathbf{1}(y_{1s} > y_{2s}) + p_s^- \mathbf{1}(y_{2s} > y_{1s}) = p_s^-
\end{aligned}
$$

it follows that $\mathbb{E}(W | T_{1s} + T_{2s} = 1) \geq \mathbb{E}(W^-)$. The same steps can be followed to show that $\mathbb{E}(W | T_{1s} + T_{2s} = 1) \leq \mathbb{E}(W^+)$

e) In fact, it can be shown that under this null hypothesis, for any $a$:

$$P(T^- \geq a) \leq P(T \geq a | Z_{1s} + Z_{2s} = 1) \leq P(T^+ \geq a) \tag{3}$$

Discuss, in about 3 -5 sentences or so, how property (3) can be exploited to test the exact null of no treatment effect.

**ANSWER:** Suppose our Wilcoxon rank sum test gave a test statistic equal to $w$. The $p$-value for this test statistic is $P(W \geq w | T_{1s} + T_{2s} = 1)$; if unobserved variables affect the treatment assignment, this $p$-value can be hard to obtain, even by simulation. However, we can bound the $p$-value by $P(W^- \geq w)$ and $P(W^+ \geq w)$; these probabilities are easy to simulate. By adjusting $\gamma$—a measure of weight that unobserved covariates have on treatment assignment—we can demonstrate that a result is still significant even if treatment assignment can be partially explained (up to a point) by unobserved covariates.

f) [BONUS QUESTION] Prove property (3).

**ANSWER:** We show $P(W \geq a | T_{1s} + T_{2s} = 1) \leq P(W^+ \geq a)$. The other inequality is proven the same way.

For the ease of writing, denote:

$$p_s^* = P(T_{1s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{1s} > y_{2s}) + P(T_{2s} = 1 | T_{1s} + T_{2s} = 1)\mathbf{1}(y_{2s} > y_{1s})$$

Define the random variables $Z_s^{+*}$ where

1. If $Z_s = 1$ then $Z_s^{+*} = 1$.

2. If $Z_s = 0$ then $P(Z_s^{+*} = 1) = \frac{p^+ - p_s^*}{1 - p_s^*}$

Note that $P(Z_s^{+*} = 1) = p^+$. Moreover, since the original $Z_s$ are independent of each other, it follows that the joint distribution of $(Z_1^+, \ldots, Z_n^+)$ is the same as $(Z_1^{+*}, \ldots, Z_n^{+*})$. Thus, if we define $W^{+*} = \sum_{s=1}^{n} d_s Z_s^{+*}$, then $W^{+*}$ has the same distribution as $W^+$.

Now, by how we defined $Z_s^{+*}$, it immediately follows that $W^{+*} \geq W$. Thus $P(W^{+*} \geq a) \geq P(W \geq a | T_{1s} + T_{2s} = 1)$. And since $W^{+*}$ has the same distribution as $W^+$, it follows that $P(W^+ \geq a) \geq P(W \geq a | T_{1s} + T_{2s} = 1)$.

For an intuition for this problem, see the "Biased Coins" section on the Wikipedia page: http://en.wikipedia.org/wiki/Coupling_(probability)