Mike Higgins
Stat 239A
Midterm

2a) Let $p_{j,k}$ denote the proportion of candidates with $D_i = j$ and $R_i = k$. We can write

$$\mathbb{E}(Y_i(1) - Y_i(0))$$
$$= \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 1, R_i = 1)p_{1,1} + \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 1, R_i = 0)p_{1,0}$$
$$+ \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 0, R_i = 1)p_{0,1} + \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 0, R_i = 0)p_{0,0}. \quad (1)$$

It follows from (1) that

$$\mathbb{E}(\hat{\delta}_{naive}) = \mathbb{E}(Y_i(1)|D_i = 1, R_i = 1) - \mathbb{E}(Y_i(0)|D_i = 0, R_i = 1)$$
$$= \mathbb{E}(Y_i(1) - Y_i(0)) - \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 1, R_i = 0)p_{1,0}$$
$$- \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 0, R_i = 0)p_{0,0} + \mathbb{E}(Y_i(0)|D_i = 1, R_i = 1)p_{1,1}$$
$$- \mathbb{E}(Y_i(1)|D_i = 0, R_i = 1)p_{0,1} + \mathbb{E}(Y_i(1)|D_i = 1, R_i = 1)(1 - p_{1,1})$$
$$- \mathbb{E}(Y_i(0)|D_i = 0, R_i = 1)(1 - p_{0,1}). \quad (2)$$

Thus, the bias in the estimator is

$$Bias(\hat{\delta}_{naive}) = - \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 1, R_i = 0)p_{1,0} - \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 0, R_i = 0)p_{0,0}$$
$$+ \mathbb{E}(Y_i(0)|D_i = 1, R_i = 1)p_{1,1} - \mathbb{E}(Y_i(1)|D_i = 0, R_i = 1)p_{0,1}$$
$$+ \mathbb{E}(Y_i(1)|D_i = 1, R_i = 1)(1 - p_{1,1}) - \mathbb{E}(Y_i(0)|D_i = 0, R_i = 1)(1 - p_{0,1}). \quad (3)$$

2b) If $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$, then (3) simplifies to

$$- \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0)p_{1,0} - \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0)p_{0,0}$$
$$+ \mathbb{E}(Y_i(0)|R_i = 1)p_{1,1} - \mathbb{E}(Y_i(1)|R_i = 1)p_{0,1}$$
$$+ \mathbb{E}(Y_i(1)|R_i = 1)(1 - p_{1,1}) - \mathbb{E}(Y_i(0)|R_i = 1)(1 - p_{0,1})$$
$$= - \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0)(p_{1,0} + p_{0,0})$$
$$+ \mathbb{E}(Y_i(1)|R_i = 1)(1 - p_{1,1} - p_{0,1})$$
$$- \mathbb{E}(Y_i(0)|R_i = 1)(1 - p_{1,1} - p_{0,1}). \quad (4)$$

Note that $p_{1,1} + p_{0,1} = \lambda$ and $p_{1,0} + p_{0,0} = 1 - \lambda$. Thus, we can write the bias in (4) as

$$Bias(\hat{\delta}_{naive}) = [\mathbb{E}(Y_i(1) - Y_i(0)|R_i = 1) - \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0)](1 - \lambda). \quad (5)$$

The bias is zero if and only if the incumbency advantage among candidates that run in the next election is the same as the advantage for those that do not run in the next election.

2c) If the incumbency advantage is at least as big among candidates that re-run, and if the independence assumption holds, then $Bias(\hat{\delta}_{naive}) \geq 0$. That is, the smallest possible value of $\mathbb{E}(\delta)$ is $\mathbb{E}(\hat{\delta}_{naive})$.

Since $0 \leq Y_i(j) \leq 1$, it follows that $-1 \leq \mathbb{E}(Y_i(1) - Y_i(0)|R_i) \leq 1$. Thus, if the incumbency advantage is at least as big among candidates that re-run, it follows that

$Bias(\hat{\delta}_{naive}) \leq 2(1 - \lambda)$. (If most of the candidates run in the next election, this bias is small).

More precisely, under the assumption that $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$, it follows that $\mathbb{E}(\hat{\delta}_{naive}) = \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 1)$. Thus,

$$
\begin{aligned}
Bias(\hat{\delta}_{naive}) &= \mathbb{E}(\hat{\delta}_{naive}) - \mathbb{E}(\delta) \\
&= \left[\mathbb{E}(\hat{\delta}_{naive}) - \mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0) \right](1 - \lambda)
\end{aligned}
\tag{6}
$$

And so,

$$
\mathbb{E}(\delta) = \lambda \mathbb{E}(\hat{\delta}_{naive}) + (1 - \lambda)\mathbb{E}(Y_i(1) - Y_i(0)|R_i = 0) \leq \lambda \mathbb{E}(\hat{\delta}_{naive}) + (1 - \lambda)
\tag{7}
$$

3a) In this scenario, we have a binomial-randomized controlled experiment ($p = 0.5$); by design, we are guaranteed that treatment assignment is independent of potential outcome.

A natural quantity to estimate given this design is the ATE of the training program on status and income. We can make inferences about the ATE over all workers in our study.

Suppose there are N workers in total. Let $T_i = 1$ if worker i participates in the training program, and $T_i = 0$ if the worker does not. Let $Y_i(1)$ and $Y_i(0)$ denote the value of the response variable for worker i (either status or income) under treatment and control respectively. Estimates of the ATE include:

$$
\sum_{i=1}^{N} \frac{Y_i(1)T_i}{\sum T_i} - \sum_{i=1}^{N} \frac{Y_i(0)(1 - T_i)}{\sum(1 - T_i)}.
\tag{8}
$$

and

$$
\frac{1}{N/2}\sum_{i=1}^{N} Y_i(1)T_i - \frac{1}{N/2}\sum_{i=1}^{N} Y_i(0)(1 - T_i).
\tag{9}
$$

Both estimates can be shown to be unbiased. Inferences assume the SUTVA assumption. Through this assumption, we can test the (sharp) null hypothesis of no treatment effect on all workers using a permutation test.

3b) Despite randomization into treatment and control, we should not use the approach as in part a); workers may respond differently depending on the amount of compensation they receive before beginning the program.

Instead, this may be a more reasonable assumption: the compensation for units that were assigned a number $X \approx 2$ is similar enough that the effect of treatment on units that had X slightly greater than 2 would be the same as the effect of treatment with for units with X slightly less than 2 had they been assigned treatment. Because assignment to treatment is determined by a fixed threshold of X, this suggests making inference using a (sharp) regression discontinuity design. We would estimate the local average treatment effect (LATE) at $X = 2$;
$$
\mathbb{E}(Y_i(1) - Y_i(0)|X_i = 2).
$$

We can obtain an estimate of the local average treatment effect (LATE) $\hat{\beta}$ at $X = 2$ by fitting two OLS models (or LOESS, or whatever your favorite method is) over data within

some small bandwidth δ around $X = 2$; one model fitted to data before the cutpoint and one fitted to data after the cutpoint. The estimate of the LATE is the difference of the model estimates at the cutpoint.

3c) Let $Z_i = 1$ if the coin flips heads and let $Z_i = 0$ if the coin flips tails. Let $P(Z_i = 1) = p$
Recall, the LATE is
$$\mathbb{E}(Y_i(1) - Y_i(0)|X_i = 2)$$

What we estimate to the left of the cutpoint is:

$$\mathbb{E}(Y_i(0)|X_i = 2)$$

and what we estimate to the right of the cutpoint is

$$p\mathbb{E}(Y_i(1)|X_i = 2, Z_i = 1) + (1 - p)\mathbb{E}(Y_i(0)|X_i = 2, Z_i = 0)$$

If smoothness conditions necessary for RD hold, they will still hold in this model, as long as response is independent of the coin flip (if people can self-select to be in treatment or control, this property will no longer hold). The quantity estimated using RD will then be

$$p\mathbb{E}(Y_i(1)|X_i = 2) + (1-p)\mathbb{E}(Y_i(0)|X_i = 2) - \mathbb{E}(Y_i(0)|X_i = 2) = p(\mathbb{E}(Y_i(1)|X_i = 2) - \mathbb{E}(Y_i(0)|X_i = 2))$$

This expectation scales with the probability of heads of the coin flip. If p is known, an estimate of the LATE can be obtained by estimating the LATE without factoring in the coin flip, and then dividing by $1/p$.