

MGTECON 603 - Problem Set 1

(Instructor: Guido Imbens)

Wooyong Park

Collaborators: Cem Kozanoglu, Roberto Gonzalez Tellez

September 27, 2025

1 Part I

Descriptive Statistics

The summary statistics of the data are shown in tables [A.1](#), [A.2](#), and [A.3](#) in the appendix.

The histogram of the outcome variable `earnings1yr` (Figure 1.1) displays a heavily zero-inflated distribution, and overall we have more treated units than control units.

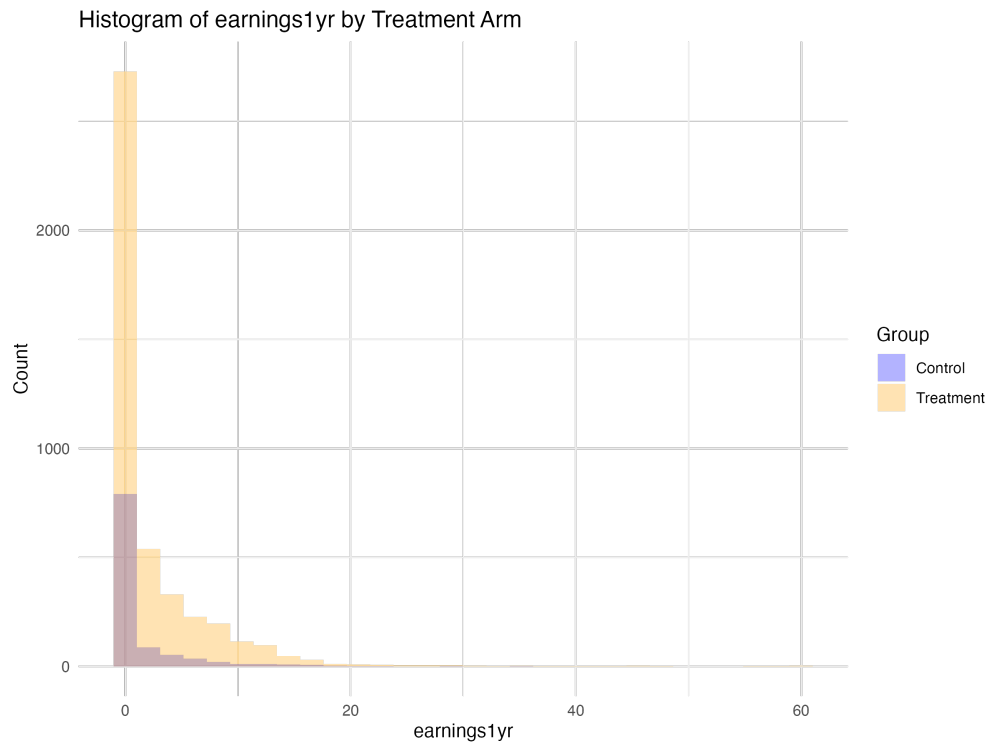


Figure 1.1: Histogram of earnings1yr by treatment arm

(a) Difference-in-Means

Assuming our data is not a random sample but the population itself, the number M given in the problem must be the actual number treated in our dataset, which is 4,383. I create 99, 999, 9999, and 99999 random draws of treatment assignments with exactly 4,383 treated units to create the same number of test statistics that would be comparable to the actual test statistic. – In principle, I must have $\binom{n}{M}$ possible treatment assignments, but since this number is extremely large, I will only create at maximum 99,999 random draws, assuming that it is a large enough number so that the difference of the exact population p-value and the p-value that I have from the 99,999 random draws is negligible – The reason I have multiple number of draws is to see if the p-value changes significantly. If I see histograms and p-values converging at some point, we might not need to increase the numbers of draws.

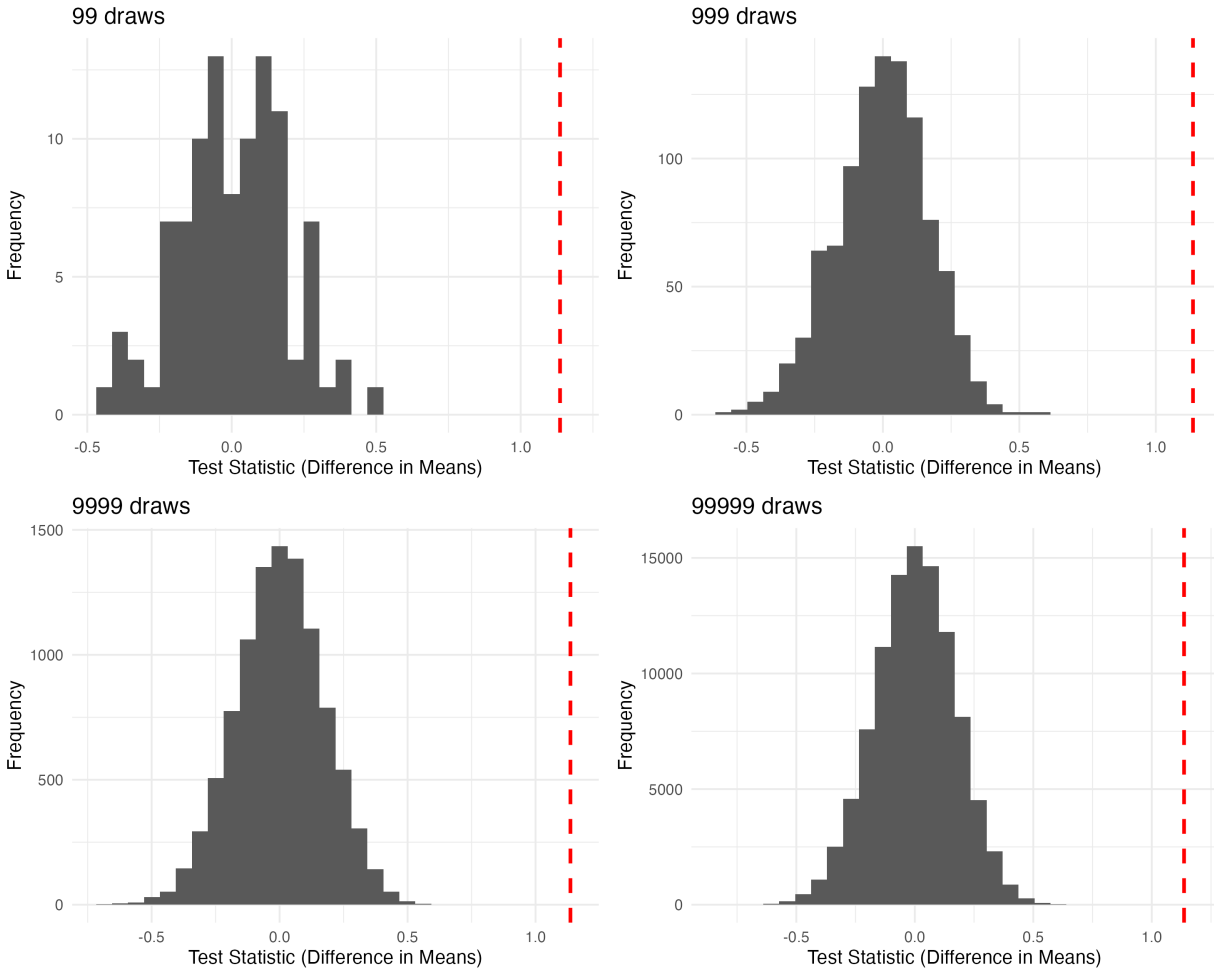


Figure 1.2: Difference-in-means by number of draws

Figure 1.3 shows that the histograms of the test statistics under the null hypothesis of no treatment effect, generated from different numbers of random draws outside the actual draw we

have. The red dashed line indicates the actual test statistic calculated from the observed data. Here, we can see that even 999 draws seem to be enough to get a reasonable approximation of the null distribution, and the p-value does not change much as we increase the number of draws. I remain trying different number of draws for problem I-(b) and (c) as well; however, if I had to choose a number, 999 seems to be sufficient. The actual test statistic is 1.1362, and the p-value under 999 draws is 0.0000.¹

(b) Difference-in-Medians

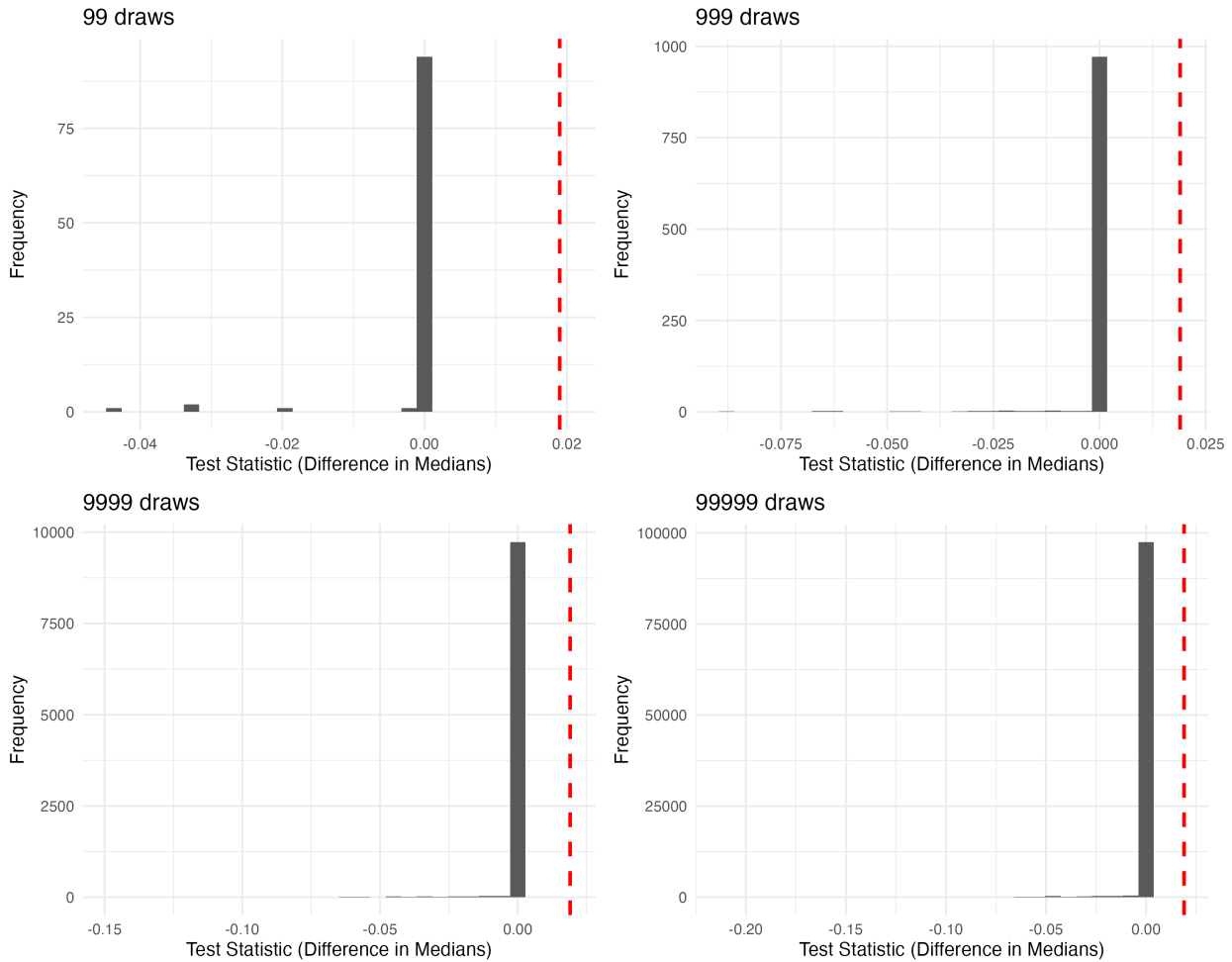


Figure 1.3: Difference-in-medians by number of draws

With the difference-in-medians, we obtain the actual test statistic 0.019 and the p-value 0.017 under 999 draws (table A.4). This might not be the ideal test statistic to use, given that we have many zeros in the data (`earnings1yr` has 52% zeros). Chen and Roth (2024) explains how we can separate the treatment effects at the extensive margin and the intensive margin when the outcome

¹The p-values for each test statistic and number of draws are in table A.4.

is a weakly positive random variable, and Feng (2021) compares zero-inflated count models and hurdle models. In this case, if the treatment does not strongly affect the extensive margin, the median of the two groups will not capture the treatment effect under a zero-inflated distribution. To account for the zero-inflated nature of the data and the distinction between the extensive and intensive margins, we can use a hurdle model, which is demonstrated in the next section.

(c) Hurdle Estimates

As for another statistic that we can use, we can use a hurdle model to account for the zero-inflated nature of the data. The model I consider is as follows:

$$P(Y_i > 0|W_i) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 W_i + \beta_2 X_i + \beta_3 W_i \tilde{X}_i))} \quad (1.1)$$

$$\mathbb{E}[\log(Y_i)|W_i] = \gamma_0 + \gamma_1 W_i + \gamma_2 X_i + \gamma_3 W_i \tilde{X}_i \quad (1.2)$$

where \tilde{X}_i is the vector of demeaned covariates (high school diploma, female, age, single) excluding the treatment variable W_i . Here, I allow the heterogeneity in the treatment effect by the observed covariates, and the covariates for the interaction term are demeaned so that β_1 and γ_1 capture the treatment effect for the individual with average characteristics.

The test statistic I consider from equations 1.1 and 1.2 is as follows:

$$\hat{T} = \beta_1 + \gamma_1 \quad (1.3)$$

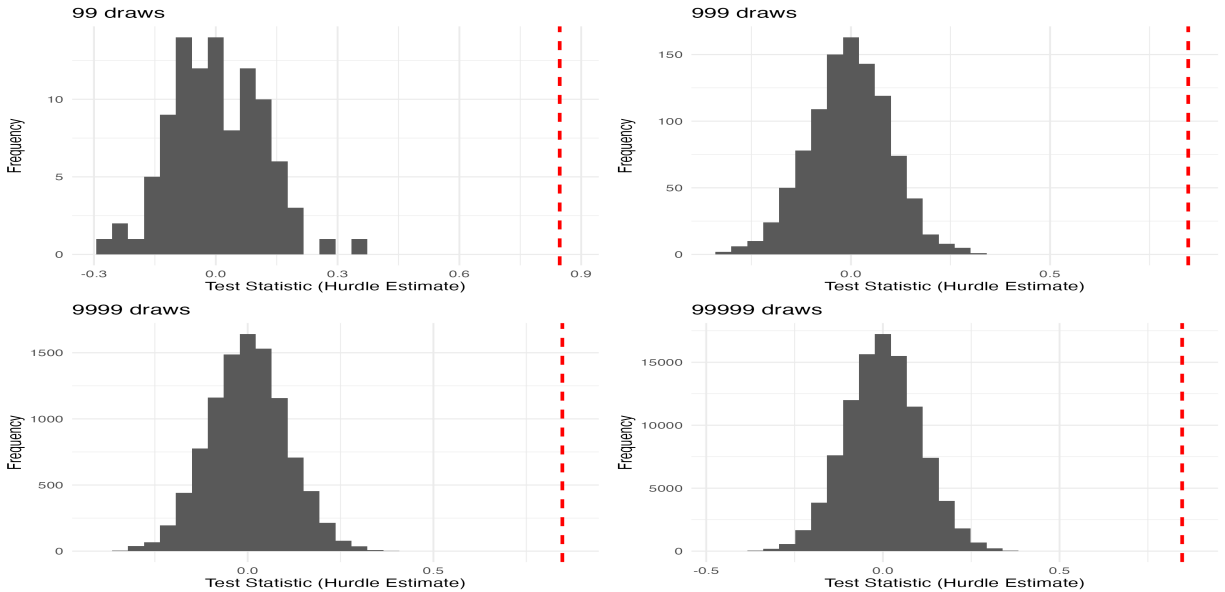


Figure 1.4: Hurdle model estimate by number of draws

Figure 1.4 shows that the histograms of 1.3 under the null hypothesis of no treatment effect, generated from different numbers of random draws outside the actual draw we have. The red dashed line indicates the actual test statistic calculated from the observed data. The actual test statistic is 0.8470, and the p-value under 999 draws is 0.0000(table A.4).

(d) Bernoulli Process

Figure 1.5 shows the histograms of the three test statistics(mean, median, and hurdle) under the null hypothesis of no treatment effect from 999 draws, assuming a coin toss experiment with $p = 0.2$. The red dashed line indicates the actual test statistic calculated from the observed data.

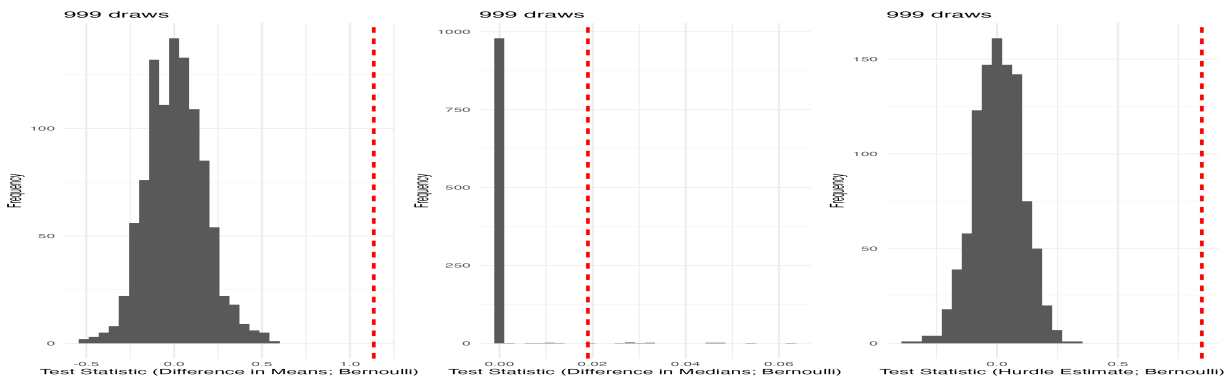


Figure 1.5: Bernoulli process

The p-values under 999 draws for all three test statistics are in table 1, compared to the permutation p-values in the previous sections. In this case, we have smaller p-values for the median statistic in the coin toss case. However, we cannot compare the other p-values directly since they are all zero. I believe the p-values of these two approaches can be very different if we have a large imbalance in the treatment and control groups or if the number of observations is small. The reasons are discussed in Part 2.

Test Statistic	Bernoulli	Permutation
Mean	0.0000	0.0000
Median	0.0140	0.0170
Hurdle	0.0000	0.0000

Table 1: P-values of the test statistics under different approaches(999 draws)

2 Part II

(a) Theoretical Calculations

In the current situation, with 49 treatment units and 51 control units, it is better to consider the permutation of the treatment status rather than considering the Bernoulli process that the

experiment design actually followed. This is not a universal rule, but in this case, since we have a balanced treatment and control groups, it is more appropriate to consider the permutation of the treatment status. Assuming we use the usual test statistics such as difference-in-means or difference-in-medians, their variances usually increase with the imbalance of the treatment assignment. If we were to consider the Bernoulli process, we also consider the test statistics under imbalanced treatment and control groups, and the specific values of these statistics will be more affected by the noise. In other words, we will have more test statistics since we are considering 2^{100} possible treatment assignments instead of $\binom{100}{49}$ possible treatment assignments, but the extra test statistics will not have the same noise-to-signal ratio as the statistics under balanced treatment assignment. However, if we limit to the case where we have 49 and 51 treated and control units, the variance of the test statistics are less likely to be affected by the noise. Therefore, we should consider the permutation of the treatment status in this case.

simulations	permutation	bernoulli
100	0.0099	0.0099
1,000	0.0010	0.0020
10,000	0.0010	0.0003
50,000	0.0009	0.0009

Table 2: p-values: Permutation vs. Bernoulli

Table 2 presents the p-values of the test statistics under the null hypothesis of no treatment effect based on the two different approaches: permutation and Bernoulli. The DGP of this simulated data is as follows:

```
Y0 <- rnorm(100, mean=50, sd=10)
Y1 <- 1.1*Y0 + rnorm(100, mean=0, sd=1)
```

where $Y_i(1)$ and $Y_i(0)$ are the potential outcomes for the i -th unit under the treatment and control scenarios, respectively.

We see that the p-values of the test statistics under the null hypothesis of no treatment effect are very close to each other, regardless of the number of random draws.

However, consider the situation where we ended up with 30 treated units and 70 control units.

In this case, our test statistic will be more likely to be affected by the noise. Therefore, we should consider the Bernoulli process in this case. Table 3 is from the same DGP of the potential outcomes and Bernoulli process with $p = 0.5$, but with 30 treated units and 70 control units. That is,

```
Y0 <- rnorm(100, mean=50, sd=10)
Y1 <- 1.1*Y0 + rnorm(100, mean=0, sd=1)
W <- c(repeat(1, 30), repeat(0, 70))
W <- shuffle(W)
```

```

for (i in 1:num_simulations) {
  Wmock <- rbinom(100, 1, 0.5) # random assignment with p=0.5
  diff_in_means <- mean(data$Y_obs[Wmock == 1]) - mean(data$Y_obs[Wmock == 0])
  test_statistics <- c(test_statistics, diff_in_means)
}

```

simulations	permutation	bernoulli
100	0.0198	0.0396
1,000	0.0539	0.0200
10,000	0.0439	0.0305
50,000	0.0447	0.0286

Table 3: p-values: Permutation vs. Bernoulli(30 T vs 70 C)

In table 3, we see that the p-values of the test statistics under the null hypothesis of no treatment effect are significantly lower when we consider the Bernoulli process. Therefore, it is better to consider the Bernoulli process when the resulted imbalance in treatment and control groups is large.

References

- Chen, Jiafeng and Jonathan Roth (2024). “Logs with zeros? Some problems and solutions”. In: *The Quarterly Journal of Economics* 139.2, pp. 891–936.
- Feng, Cindy Xin (2021). “A comparison of zero-inflated and hurdle models for modeling zero-inflated count data”. In: *Journal of Statistical Distributions and Applications* 8.1, p. 8.

variable	numNA	numZeros	fracZeros	mean	sd	min	max	10%	20%	30%	40%	50%	90%	95%	99%	99.9%
W	0	1,036	0.1912	0.8088	0.3933	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
earnings1yr	0	2,858	0.5274	2.3683	4.9519	0.0000	60.0000	0.0000	0.0000	0.0000	0.0000	0.0000	8.0403	11.4974	21.3039	53.1307
earnings4yr	0	3,252	0.6001	4.0321	8.1327	0.0000	60.0000	0.0000	0.0000	0.0000	0.0000	0.0000	14.8926	21.7532	37.5919	59.7337
hs.diploma	0	2,584	0.4768	0.5232	0.4995	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
female	0	653	0.1205	0.8795	0.3256	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
age	0	0	0.0000	33.6402	8.1958	15.0000	70.0000	24.0000	27.0000	29.0000	31.0000	33.0000	45.0000	49.0000	57.0000	62.5820
child	0	4,532	0.8363	0.1637	0.3700	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000
single	0	727	0.1342	0.8658	0.3409	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
hs.diploma_dm	0	0	0.0000	0.0000	0.4995	-0.5232	0.4768	-0.5232	-0.5232	-0.5232	-0.5232	-0.4768	0.4768	0.4768	0.4768	0.4768
female_dm	0	0	0.0000	0.0000	0.3256	-0.8795	0.1205	-0.8795	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205
age_dm	0	0	0.0000	0.0000	8.1958	-18.6402	36.3598	-9.6402	-6.6402	-4.6402	-2.6402	-0.6402	11.3598	15.3598	23.3598	28.9418
single_dm	0	0	0.0000	0.0000	0.3409	-0.8658	0.1342	-0.8658	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342

Table A.1: Summary Statistics

variable	numNA	numZeros	fracZeros	mean	sd	min	max	10%	20%	30%	40%	50%	90%	95%	99%	99.9%
W	0	1,036	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
earnings1yr	0	679	0.6554	1.4493	3.4950	0.0000	34.5994	0.0000	0.0000	0.0000	0.0000	0.0000	5.1112	8.6316	16.3443	28.6786
earnings4yr	0	678	0.6544	3.0354	6.8941	0.0000	51.5277	0.0000	0.0000	0.0000	0.0000	0.0000	11.7650	17.7781	31.8940	51.4125
hs.diploma	0	494	0.4768	0.5232	0.4997	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000
female	0	130	0.1255	0.8745	0.3314	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
age	0	0	0.0000	33.8205	8.3839	15.0000	66.0000	25.0000	28.0000	29.0000	31.0000	33.0000	45.0000	49.0000	58.0000	62.9650
child	0	865	0.8349	0.1651	0.3714	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000
single	0	126	0.1216	0.8784	0.3270	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
hs.diploma_dm	0	0	0.0000	0.0000	0.4997	-0.5232	0.4768	-0.5232	-0.5232	-0.5232	-0.5232	-0.4768	0.4768	0.4768	0.4768	0.4768
female_dm	0	0	0.0000	-0.0050	0.3314	-0.8795	0.1205	-0.8795	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205
age_dm	0	0	0.0000	0.1803	8.3839	-18.6402	32.3598	-8.6402	-5.6402	-4.6402	-2.6402	-0.6402	11.3598	15.3598	24.3598	29.3248
single_dm	0	0	0.0000	0.0125	0.3270	-0.8658	0.1342	-0.8658	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342

Table A.2: Summary Statistics

variable	numNA	numZeros	fracZeros	mean	sd	min	max	10%	20%	30%	40%	50%	90%	95%	99%	99.9%
W	0	0	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
earnings1yr	0	2,179	0.4971	2.5855	5.2141	0.0000	60.0000	0.0000	0.0000	0.0000	0.0000	0.0190	8.4014	11.7693	23.5991	56.8758
earnings4yr	0	2,574	0.5873	4.2677	8.3822	0.0000	60.0000	0.0000	0.0000	0.0000	0.0000	0.0000	15.3113	22.4685	38.1191	60.0000
hs.diploma	0	2,090	0.4768	0.5232	0.4995	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
female	0	523	0.1193	0.8807	0.3242	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
age	0	0	0.0000	33.5975	8.1511	16.0000	70.0000	24.0000	27.0000	29.0000	31.0000	33.0000	44.0000	49.0000	57.0000	61.6180
child	0	3,667	0.8366	0.1634	0.3697	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000
single	0	601	0.1371	0.8629	0.3440	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
hs.diploma_dm	0	0	0.0000	0.0000	0.4995	-0.5232	0.4768	-0.5232	-0.5232	-0.5232	-0.5232	-0.4768	0.4768	0.4768	0.4768	0.4768
female_dm	0	0	0.0000	0.0012	0.3242	-0.8795	0.1205	-0.8795	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205	0.1205
age_dm	0	0	0.0000	-0.0426	8.1511	-17.6402	36.3598	-9.6402	-6.6402	-4.6402	-2.6402	-0.6402	10.3598	15.3598	23.3598	27.9778
single_dm	0	0	0.0000	-0.0030	0.3440	-0.8658	0.1342	-0.8658	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342	0.1342

Table A.3: Summary Statistics

Test Statistic	99 Draws	999 Draws	9999 Draws	99999 Draws
Difference in Means	0.0000	0.000	0.0000	0.0000
Difference in Medians	0.0404	0.017	0.0166	0.0162
Hurdle Estimate	0.0000	0.000	0.0000	0.0000

Table A.4: P-values of the test statistics under the null hypothesis of no treatment effect by number of draws