ECMT 676 Assignment #5

Joseph Janko

April 3, 2020

Part A

Question 1

Show the distribution of birth weights around the 1500-gram cutoff in three ways. Plot thefrequencies using 1-gram bins and 10-gram bins. [Be careful not pool together observations acrossthe threshold.] Is the distribution of birth weights smooth?

* Please note that Figure 1 is the histogram and Figure 2 is the scatter plot for the case of 1-gram bins. Figure 3 is the histogram and Figure 4 is the scatter plot for the results of 10-gram bin.

Conclusion: In both the cases of the 1-gram and 10-gram bins, we observe clustered peaks corresponding to observations with birth weight recorded at the exact ounce. Obviously this phenomena results in the distribution is not smooth.

Table 1: Birth Weights Summary Statistics

	Obs	Mean	Std.Dev.	Min	Max
bweight_normalized	376408	11.57565	89.01614	-150	150

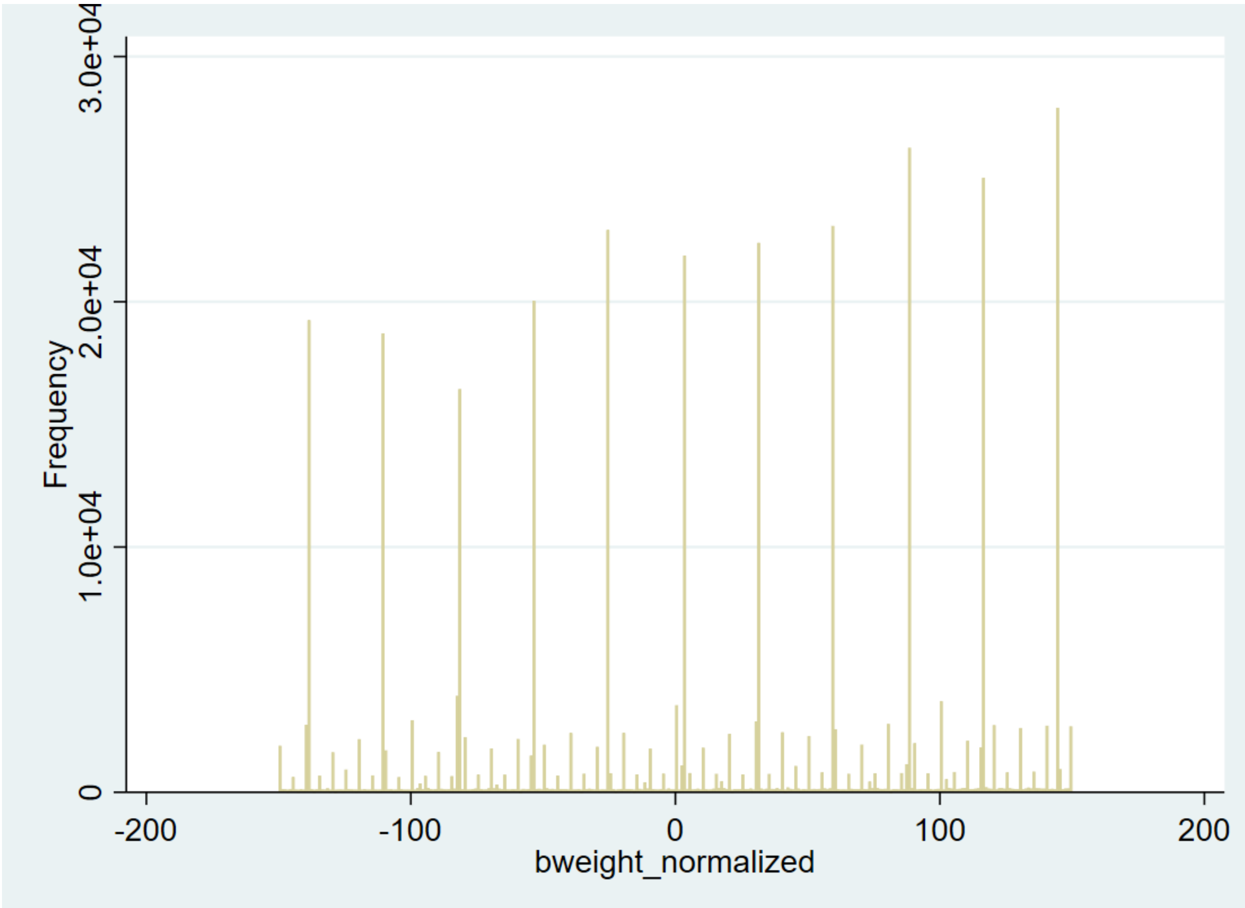

Figure 1: 1-gram Bin Histogram

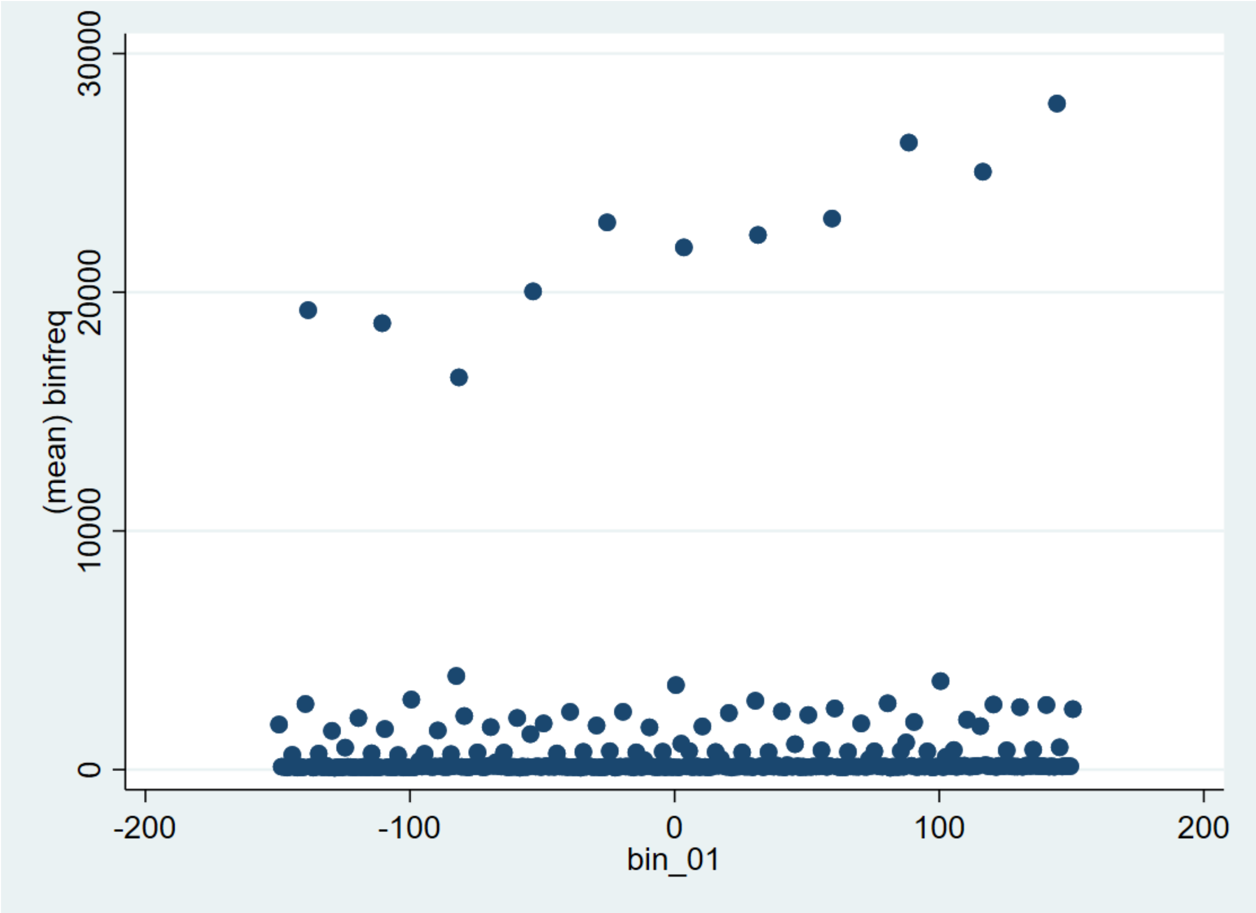

Figure 2: 1-gram Bin Scatter

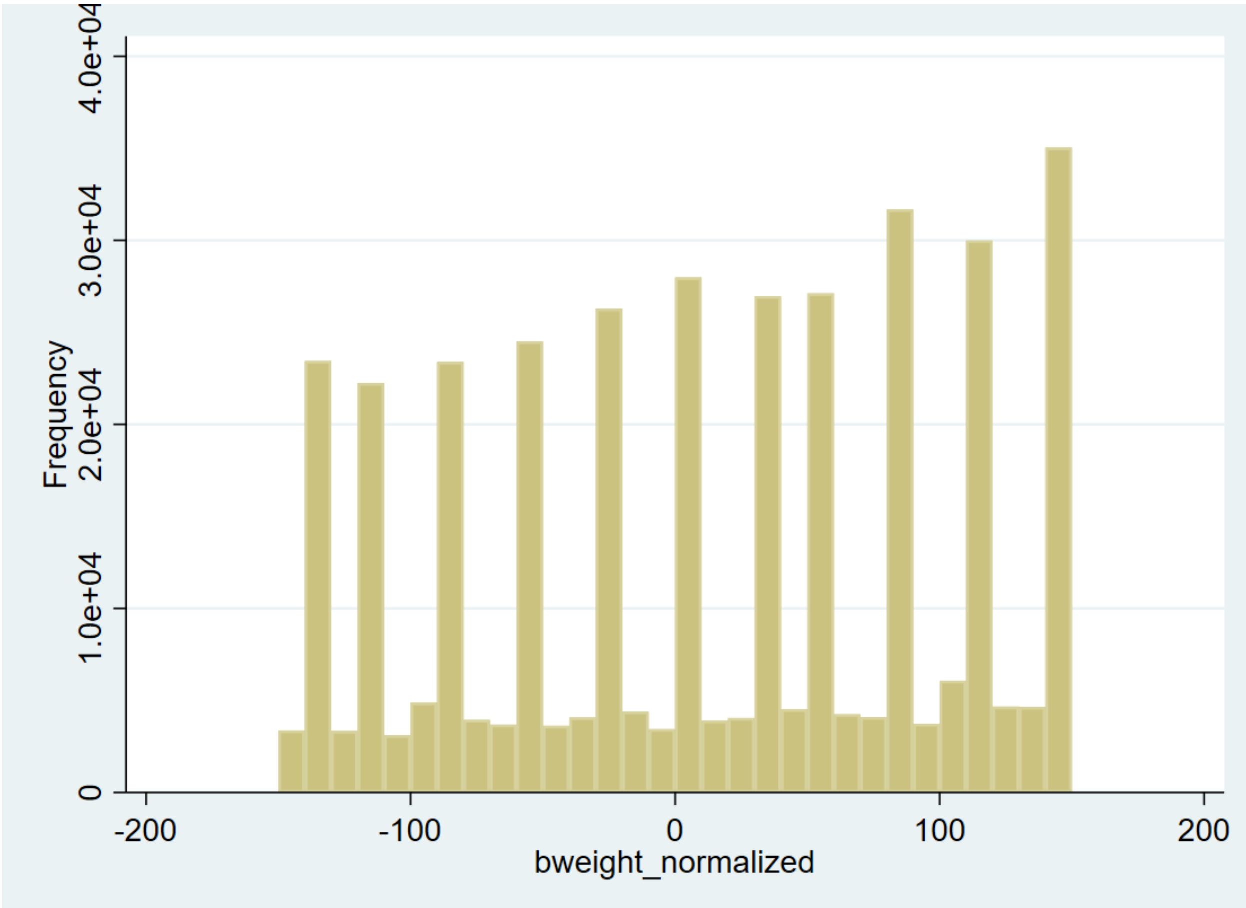

Figure 3: 10-gram Bin Histogram

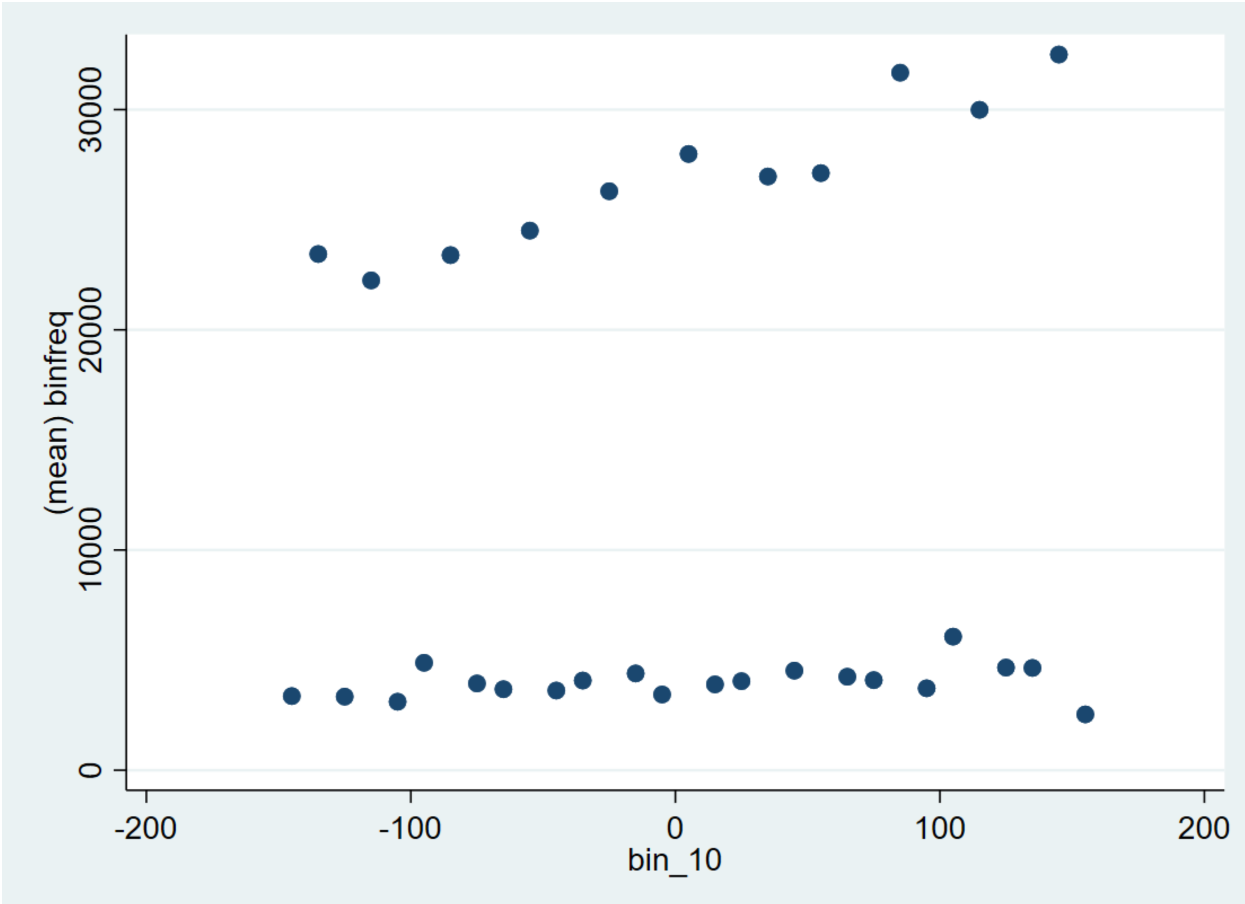

Figure 4: 10-gram Bin Scatter

Question 2

Using the 1-gram bins and associated frequencies from part 1a as your observations, estimate whether the distribution is discontinuous at the 1500-gram threshold. Estimate the discontinuity using a regression that is linear in birthweights, allowing the slope to be different on each side of the cutoff, using bandwidths of 150 grams and 50 grams. We will estimate the RDD using the formula below, which is taken from the class notes.

from 2020-04-02 13-48-13.png

Normalize running variable, $\tilde{r} = r - c$

$$E[y|\tilde{r}] = \alpha + \theta 1(\tilde{r} > 0) + \beta\tilde{r} + \gamma\tilde{r}1(\tilde{r} > 0) \quad for\ \tilde{r} < h\ and\ \tilde{r} > -h$$

Please note that α is represented by the constant term. θ is represented by bwt_indicator. β is represented by bin_01. The interaction_term is represented by gamma in the above equation. Since the coefficient of bwt_indicator nor interaction term does not have a significant p-value less than .05, The running variable is smooth across the threshold for both bandwidths.

Table 2: Results for Q2

	(1) bin=1-gram Bandwidth=150	(2) bin=1-gram Bandwidth=50
bwt_indicator	433.9	1538.0
	(970.8)	(1555.9)
bin_01	-1.894	-5.125
	(6.499)	(7.425)
interaction_term	3.282	-29.37
	(11.83)	(44.20)
Constant	909.5	708.3
	(558.3)	(465.3)
Observations	301	101
R^2	0.003	0.014
Adjusted R^2	-0.008	-0.016

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 3

Are the results sensitive to the choice of bandwidth? Does the RD design pass the balanced covariates test?

Please note that white in the tables below indicate the mother's race. In addition LessHS_Edu is indcation of the mother having less than high school education. For both white mothers and mothers with less than high school education the bwt_indicator and bweight_normalized is insignifcant for every choice of bandwith. The interaction term is insignificant for every iteration but a bandwith of 90g and when the mother has less than high school education. Furthermore the covariates estimates are all within the standard error estimation of each other at each bandwidth. The RD design passes the balanced covariate test".

Table 3: Results for Q3 Triangle Kernel Weights

	(1) Bwth=90 white	(2) Bwth=90 LessHS_Edu
bwt_indicator	0.00482	0.00474
	(0.00479)	(0.00473)
bweight_normalized	-0.000110	0.0000313
	(0.0000826)	(0.0000816)
interaction_term	0.000188	-0.000202*
	(0.000103)	(0.000102)
Constant	0.654***	0.276***
	(0.00404)	(0.00398)
Observations	230248	209819
R^2	0.000	0.000
Adjusted R^2	0.000	0.000
Pseudo R^2		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Results for Q3 Triangle Kernel Weights

	(1) Bwth=60 white	(2) Bwth=60 LessHS_Edu
bwt_indicator	0.00843	0.00199
	(0.00633)	(0.00626)
bweight_normalized	-0.000220	0.000153
	(0.000163)	(0.000161)
interaction_term	0.000240	-0.000332
	(0.000198)	(0.000197)
Constant	0.651***	0.279***
	(0.00562)	(0.00555)
Observations	158693	144608
R^2	0.000	0.000
Adjusted R^2	0.000	0.000
Pseudo R^2		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Results for Q3 Triangle Kernel Weights

	(1) Bwth=30 white	(2) Bwth=30 LessHS_Edu
bwt_indicator	0.00712	-0.000952
	(0.0111)	(0.0109)
bweight_normalized	-0.000179	0.000395
	(0.000459)	(0.000452)
interaction_term	0.000397	-0.00115
	(0.000628)	(0.000619)
Constant	0.652***	0.284***
	(0.0105)	(0.0104)
Observations	68207	62089
R^2	0.000	0.000
Adjusted R^2	-0.000	0.000
Pseudo R^2		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Part B

Question 1

Estimate the effect of very low birth weight classification on one-year mortality (agedth5) usingthe same specifications as those in part A.3 but use robust standard errors. Are the estimates sensitive to the bandwidth?

The coefficient of bwt_indicator is positive and significant at the 0.05 significance level for every bandwidth.

The bweight_normalized coefficient is negative and significant at every bandwdith. We see in the 60−g and 90−g cases the interaction term is not significant, however, the 30−g case is signficant. This gives evidence that the slope between one-year mortality and birth weight for children below 1500 grams and above 1500 grams are very similar.

Table 6: Output for Part B Q1, Note the Dependent Var, One-year Mortality

	(1) Bwth=90 TriKernelWght	(2) Bwth=60 TriKernelWght	(3) Bwth=30 TriKernelWght
bwt_indicator	0.0114***	0.0161***	0.0263***
	(0.00234)	(0.00302)	(0.00507)
bweight_normalized	-0.000166***	-0.000279***	-0.000554**
	(0.0000397)	(0.0000756)	(0.000207)
interaction_term	-0.0000858	-0.000128	-0.00112***
	(0.0000506)	(0.0000952)	(0.000299)
Constant	0.0525***	0.0497***	0.0445***
	(0.00193)	(0.00260)	(0.00472)
Observations	230248	158693	68207
R^2	0.000	0.001	0.001
Adjusted R^2	0.000	0.000	0.001
Pseudo R^2			

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 2

Repeat B.1 after dropping observations that fall exactly at the 1500-gram cutoff. Have the results changed? Should RD estimates ever be sensitive to dropping observations exactly at the cutoff? Explain.

There appears to be change in the ouput from removing observations at the 1500-gram cutoff. We observe that the coefficient of bwt_indicator is still significant at the 5 percent level, however, the magnitudes of the coefficient become smaller in comparison to Part B Question 1 for every chosen bandwith. The interaction terms for all bandwiths are not signficant at the 5 percent level. RD estimates compare the mean expected outcomes below and above the cutoff, thereby the estimate should not be changed by removing the values at the 1500-g.

Table 7: Output for Part B Q2, Note the Dependent Var, One-year Mortality

	(1) Bwth=90 TriKernelWght	(2) Bwth=60 TriKernelWght	(3) Bwth=30 TriKernelWght
bwt_indicator	0.00650**	0.0101***	0.0175***
	(0.00235)	(0.00303)	(0.00508)
bweight_normalized	-0.000166***	-0.000279***	-0.000554**
	(0.0000397)	(0.0000756)	(0.000207)
interaction_term	0.0000226	0.0000771	-0.000153
	(0.0000506)	(0.0000954)	(0.000291)
Constant	0.0525***	0.0497***	0.0445***
	(0.00193)	(0.00260)	(0.00472)
Observations	226704	155149	64663
R^2	0.000	0.000	0.000
Adjusted R^2	0.000	0.000	0.000
Pseudo R^2			

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 3

Which set of estimates are you most inclined to believe? Why? What do you conclude about the effect of very low birth weight classification on infant mortality?

The estimate should not be sensitive to the observations that fall exactly at the 1500-g cutoff, therefore the output of Part B Question 2 should used in analysis instead of Part B Question 1. This part essentially removes the clustering peak effects around the exact ounce. The results from this section is indcation that there is not a significant effect of very low birth weight classification on one-year mortality rate.