Lead Absorption in Children

Lanqin Zhao

Abstract: This study used graphical methods and simulation methods: permutation tests and bootstrap to analyze the data Morton et al. (1982) collected on lead absorption in children. The results indicated that the blood lead levels of the children in Exposed group and Control group are significantly different, and that the true mean differences of lead levels between the two groups is between 10.5 and 21.2 μg/dl. Moreover, the results illustrated lead exposure and personal hygienic practices of the workers directly affected the blood lead levels in their children.

Key words: graphical methods; permutation test; bootstrap; simulation

1. Introduction

Lead is an environmental pollutant especially worthy of attention because of its damaging effects on the neurological and intellectual development of children. It is generally agreed that children with blood lead levels above 40 micrograms per deciliter need medical treatment. Children above 60 on this scale should be immediately hospitalized for treatment (Miller and Kean, 1987). Morton et al. (1982) collected data on lead absorption by children whose parents worked at a factory in Oklahoma where lead was used in the manufacture of batteries. The concern was that children might be exposed to lead inadvertently brought home on the bodies or clothing of their parents. Levels of lead (in micrograms per deciliter) were measured in blood samples taken from 33 children who might have been exposed in this way. These children are the Exposed group. The researchers formed a Control group by making matched pairs. For each of the 33 children in the Exposed group, they selected a matching child—of the same age, living in the same neighborhood, and with parents employed at a place where lead is not used.

This study used graphical methods and simulation methods to analyze the data sets, LEADKIDS Morton et al collected in order to determine whether children of workers in a lead-related industry have significantly different blood lead levels from neighborhood children of the same age whose parents are not employed in a lead-related industry, and to estimate the confidence interval for mean difference. In addition, this study visualized the data set LEADEXP Morton et al collected to illustrate how lead exposure and personal hygienic practices of the workers affected the blood lead levels in their children.

2. Comparing the lead levels

2.1 Summarizing the data

```
summ.lead=lapply(leadkids, function(x) rbind(mean = mean(x) ,
                                              sd = sd(x) ,
                                              median = median(x) ,
                                              minimum = min(x) ,
                                              maximum = max(x) ,
                                              size = length(x) ) )
> data.frame(summ.lead )
         Exposed       Control       Diff
mean     31.84848      15.878788     15.96970
sd        14.40729      4.539807     15.86365
median    34.00000      16.000000    15.00000
minimum   10.00000      7.000000     -9.00000
maximum   73.00000      25.000000    60.00000
size     33.00000      33.000000     33.00000
```

TABLE 1: Summary of Data Set LEADKIDS (unit: µg/dl)

From table 1, median blood lead level for Exposed group is 34µg/dl, which is even greater than the maximum 25µg/dl for Control group. This means that over half of children in Exposed group have more lead than do children in Control group. The mean difference is 15.97, about 16, and the mean of the Control group is also about 16. Therefore, on average, children of workers in the battery plant have about 16 µg/dl more lead than their peers whose parents do not work in a lead-related industry. Then, on average, the effect of having a parent who works in the battery factory is to double the lead levels. The lead levels of the Exposed group are much more diverse, with range=73-10=63 and standard deviation =14.4, than those of the Control group, with range= 25- 7= 18 and standard deviation = 4.5.

2.2 Visualizing the data

```
plot(lead$Pair, lead$Exposed, col="red", pch=19, ylab="Lead", xlab="Pair",
main="Scatter Plot of Exposed(red) VS Control(green)")
points(lead$Pair, lead$Control, col="green", pch=21)
abline(h=c(40,60), col=c("yellow","red"))
```

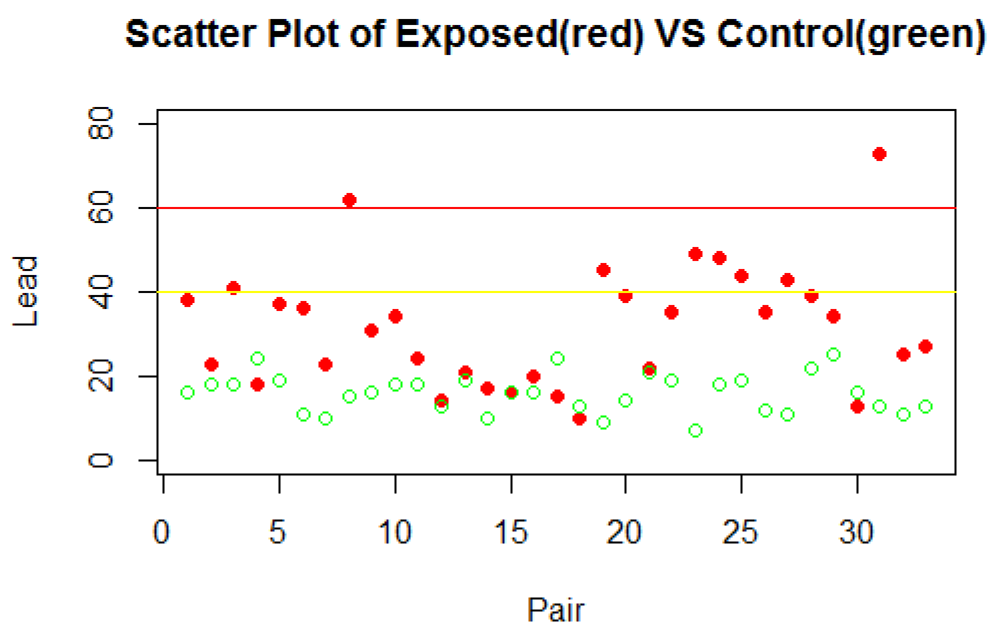

FIGURE 1. Blood lead levels plot for children in Exposed group and their matched pairs in Control group.

Figure 1 shows that Over 80% of the children in the Exposed group have higher blood lead levels than do their matcher pairs in the Control group. Lead levels of Exposed group are much more diverse than those of the Control group. Furthermore, some serious cases of lead poisoning, children needing medical treatment (with blood lead level above 40 ug/dl) and needing to be immediately hospitalized for treatment (with blood lead level above 60 ug/dl)are all in the Exposed group.

In order to have a clear picture how serious the children in the Exposed group have lead poisoning, see the below FIGURE 2.

```
h=hist(lead$Exposed, breaks=c(0,25,40,60,80))
h$density = h$counts/sum(h$counts)*100
plot(h,freq=FALSE,ylab="Relative Frequency", labels = T, ylim=c(0,60),
     col=c("light green","wheat","yellow", "red"))
```

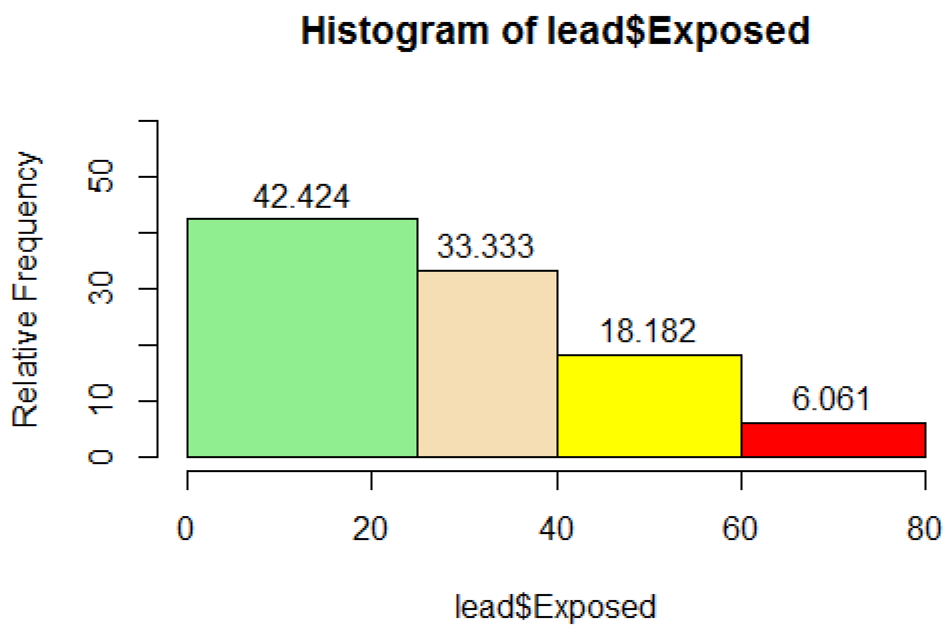

FIGURE 2. Blood lead levels relative frequency distribution plot for children in Exposed group

Figure 2 shows that 42.4% of children in the Exposed group have comparable lead levels, less than 25 µg/dl(maximum in Control group), to those in the Control group, and 57.6% have lead more than 25 µg/dl. Shockingly, 18.2% of children needs

medical treatment, and 6% of children needs to be immediately hospitalized for treatment.

2.3 Analyzing Data

2.1 and 2.2 results all show that lead levels between Exposed group and Control group are different. We know if there is no significant difference, a permutation within any pair of lead levels is as likely as the reverse, so I use permutation test to confirm it.

```
# Permutation Test on Matched Samples
# simulated permutation dist of mean differences
obs.mean=mean(lead$Diff)
b= 10000
set.seed(1234)
perm.mean = numeric(b)
for (i in 1:b) {
signs = sample( c(1,-1),length(lead$Diff), replace = T) #permute each pair
perm = lead$Diff*signs  # permute lead levels in each pair
```

```
perm.mean[i]=mean(perm)  # mean difference
}
# permutation test (Rejects at 5% level if obs. green line is outside red #lines)
hist(perm.mean, breaks = 30, prob=T, col="wheat")
abline(v=obs.mean, lwd=2,col="green")
abline(v=quantile(perm.mean, c(0.975, 0.025)), lty="dashed", col="red")
```

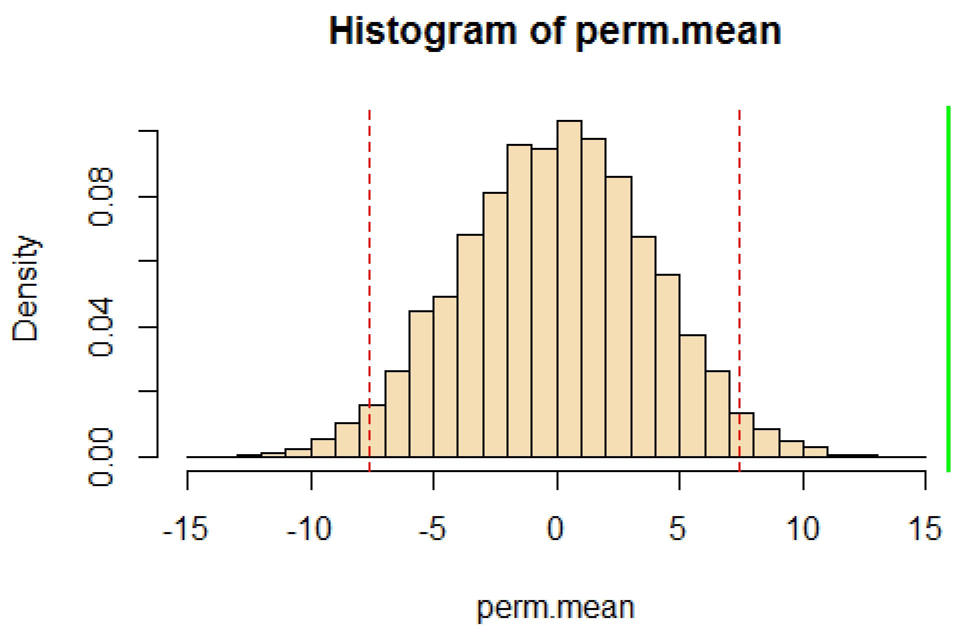

FIGURE 3 Permutation distribution of mean difference

```
mean( perm.mean > abs(obs.mean) | perm.mean < -abs(obs.mean) )
                                                  # p-value of two-sided perm test
> mean( perm.mean >abs(obs.mean) | perm.mean < -abs(obs.mean)  )
[1] 0
```

From FIGURE 3, the observed mean difference 15.97 (green line) is far outside of 95% confidence limits (red line), and the p-value is 0. Thus, it is extremely unlikely that we would see data as extreme as those actually collected unless workers at the battery factory were contaminating their children. We conclude that the difference between the lead levels of children in the Exposed and Control groups is large enough to be statistically significant. I compare this result with those of parametric t test and nonparametric Wilcoxon signed rank test as follow.

Compare to t test:

```
> t.test(lead$Diff)

        One Sample t-test
```

```
data:  lead$Diff
t = 5.783, df = 32, p-value = 2.036e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 10.34469 21.59470
sample estimates:
mean of x
  15.9697
```

Compare to Wilcoxon signed rank test:

```
> wilcox.test(lead$Diff)
        Wilcoxon signed rank test with continuity correction
data:  lead$Diff
V = 499, p-value = 1.155e-05
alternative hypothesis: true location is not equal to 0
```

Based on the outputs of the tests, these tests give the same result as that we get from the permutation test . The difference between the lead levels of children in the Exposed and Control groups is large enough to be statistically significant.

3 Estimate the difference

From above comparison, we know there is statistically significant difference between the lead levels of children in the Exposed and Control groups. we must want to know how big the difference is so that lead poisoning experts are able to evaluate whether the difference is of practical importance. I use bootstrap method to estimate the confidence interval for the mean difference.

```
# CI for mu of Diff from bootstrap distribution
n = length(lead$Diff); obs.mean = mean(lead$Diff)
# Estimated Distribution of V=xbar - mu
set.seed(1234)
B = 10000
re.x = sample(lead$Diff, B*n, repl=T)
RDTA = matrix(re.x, nrow=B)  # B x n matrix of resamples
re.mean = rowMeans(RDTA) # vector of B `xbar-star's
re.v = re.mean - obs.mean # vector of v-star's
v.UL = quantile(re.v, c(.975,.025)) # estimated Confidence limits of V
obs.mean - v.UL   # Bootstrap Confidence Interval
> obs.mean - v.UL
```

```
   97.5%     2.5%
10.48485 21.21212
```

I get the confidence interval for the mean difference is (10.5, 21.2). Then we are safe in believing that the true average difference is between 10.5 and 21.2 µg/dl. I visualize the result as follow.

```
# visualize the confidence interval
hist(lead$Diff, main="Bootstrap Confidence Interval")
abline(v=obs.mean - v.UL, lty="dashed", col="red" )
text(x=15.97,  y=1, expression(bar(x)==15.97),col="blue", cex=0.6)
```

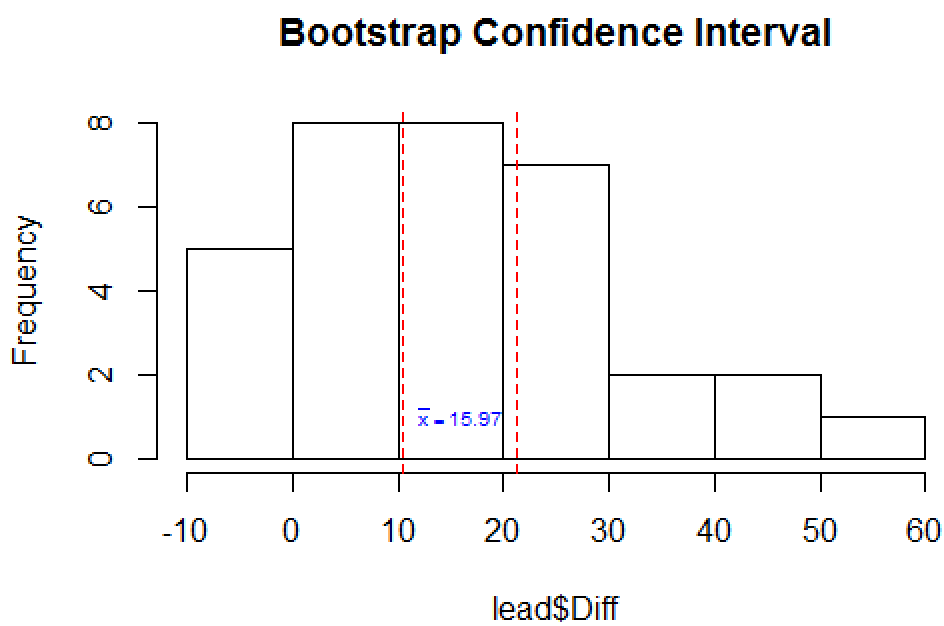

FIGURE 4   distribution of differences with confidence interval

Figure 4 is a histogram of the 33 lead level differences between Exposed children and their matched pairs. Also shown is the sample mean of the differences and 95% confidence interval. The sample mean is the best estimate of the population mean. It is very hard to imagine that this histogram could have resulted by sampling at random from a population with mean 0. This graphic is an effective summary of the data and of our previous conclusions.

I also compare this confidence interval from bootstrap with that from t test as follow.

t.test(lead$Diff)

```
> t.test(lead$Diff)
        One Sample t-test
data:  lead$Diff
```

```
t = 5.783, df = 32, p-value = 2.036e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 10.34469 21.59470
sample estimates:
mean of x
  15.9697
```

The confidence interval for the mean difference from t test is (10.3, 21.6), which is pretty close to the one from bootstrap.

4 Exploring the association between lead levels in children with lead exposure and hygienic of their parents

Notice that the lead levels of the Exposed group are much more diverse than those of the Control group. This suggests that some children in the Exposed group are getting a lot more lead, presumably from their working parents, than are others in this group. Perhaps some parents at the battery factory don't work in areas where they come into direct contact with lead. Perhaps some parents have good hygienic habits, and some don't. The following plot may reveal these guesses.

```
plot(leadexp$Lead, pch=leadexp$JobExp, col=leadexp$JobHyg, main="Lead vs JobEx
p and JobH yg")
abline(h=c(40,60), col=c("yellow","red"))
legend("topleft", title="JobExp", legend=c("H", "M", "L"),pch=c(3,2,1), cex=0.
6)                              # insert legend for JobExp
legend("topright", title="JobHyg", legend=c("Good", "Fair", "Poor"),
fill = 1:3, cex=0.6)                      # insert legend for JobHyg
```

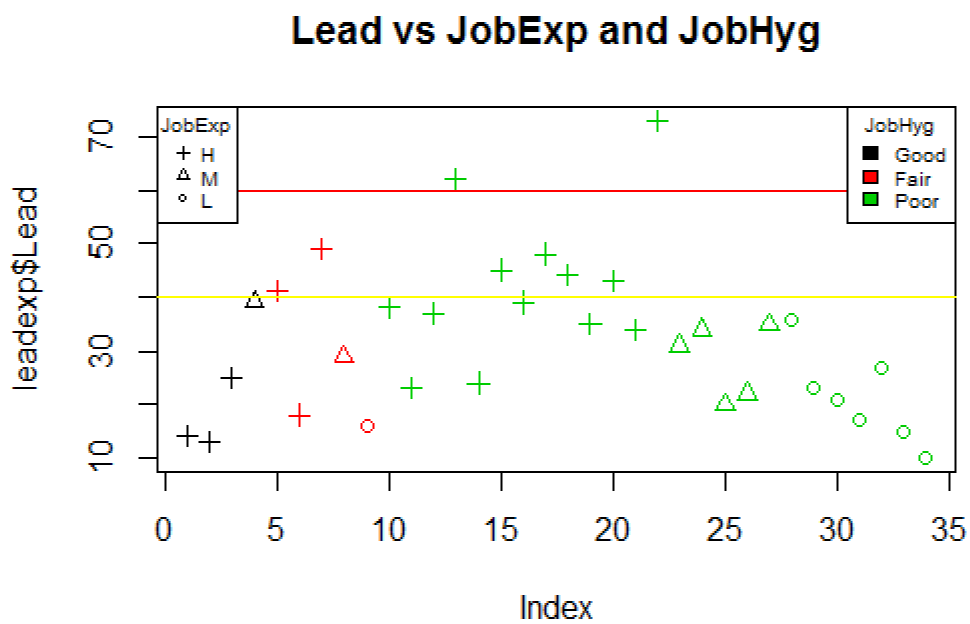

FIGURE 5 Lead levels in children vs lead exposure and hygiene of their parents

Figure 5 indicates that the children whose parents had higher lead exposure at work also have higher blood lead levels. However, the children whose parents have good personal hygiene have blood lead levels comparable to the control children. It appeared that only good parental personal hygiene, i.e., showering, shampooin g and changing clothes and shoes before leaving work, was effective for preventing from lead poisoning in children.

5 Summary

The previous results indicated that the blood lead levels of the children of employees in a lead-related industry are significantly different (p-value=0) from neighborhood children of the same age whose parents are not employed in a lead-related industry. About 24% of children of workers in the battery plant need medical treatment or hospital treatment. On average, children of workers in the battery plant have about 16 µg/dl more lead than their peers whose parents do not work in a lead-related industry, that means the effect of having a parent who works in the battery factory is to double the lead levels. And the true average difference of lead levels between the Exposed group and Control group is between 10.5 and 21.2 μg/dl. Moreover, there are distinct trends indicating that the risk of increased lead absorption is directly related to the lead exposure at work of the children's parents, and, more importantly, to personal hygienic habits.

6 Reference

D. Morton et al. (1982), “Lead absorption in children of employees in a lead-related industry,” American Journal of Epidemiology, Vol. 115, pages 549-555.

Bruce E. Trumbo(2002), *Learning Statistics with Real Data*.

7 Appendix

1. The data set LEADKIDS

Contains three variables, each with 33 cases. All involve measurements of lead in micrograms per deciliter of blood as follow:

Exposed: lead(μg/dl of whole blood) for children of workers in the battery factory

Control: lead(μg/dl of whole blood) for matched controls

Diff : the differences: 'Exposed' - 'Control'.

Complete listing of dataset LEADKIDS

	Exposed	Control	Diff
1	38	16	22
2	23	18	5
3	41	18	23
4	18	24	-6
5	37	19	18
6	36	11	25
7	23	10	13
8	62	15	47
9	31	16	15
10	34	18	16
11	24	18	6
12	14	13	1
13	21	19	2
14	17	10	7
15	16	16	0
16	20	16	4

17	15	24	-9
18	10	13	-3
19	45	9	36
20	39	14	25
21	22	21	1
22	35	19	16
23	49	7	42
24	48	18	30
25	44	19	25
26	35	12	23
27	43	11	32
28	39	22	17
29	34	25	9
30	13	16	-3
31	73	13	60
32	25	11	14
33	27	13	14

2. The dataset LEADEXP

Contains 34 subjects, including one child who had to be left out of LEADKIDS for lack of a valid control partner. It reports summary results as follows:

Lead: Lead (μg/dl of whole blood) for children of battery factory workers (same as Exposed in LEADKIDS, but not in the same order)

JobExp: Categorical variable rating parent's level of exposure (**1** = Low, **2** = Moderate, **3** = High)

JobHyg: Categorical variable rating parent's level of hygiene (**1** = Good, **2** = Moderately good, **3** = Poor)

Complete listing of dataset LEADEXP

	Lead	JobExp	JobHyg
1	14	3	1
2	13	3	1
3	25	3	1
4	39	2	1
5	41	3	2
6	18	3	2
7	49	3	2

8	29	2	2
9	16	1	2
10	38	3	3
11	23	3	3
12	37	3	3
13	62	3	3
14	24	3	3
15	45	3	3
16	39	3	3
17	48	3	3
18	44	3	3
19	35	3	3
20	43	3	3
21	34	3	3
22	73	3	3
23	31	2	3
24	34	2	3
25	20	2	3
26	22	2	3
27	35	2	3
28	36	1	3
29	23	1	3
30	21	1	3
31	17	1	3
32	27	1	3
33	15	1	3
34	10	1	3