1. **Morton et al. (1982) studied lead in the blood of children whose parents worked in a factory where lead was used in making batteries. They were concerned that children were exposed to lead inadvertently brought home by their parents. Their study included 33 such children from different families -- they are the exposed or treated children. The outcome R was the level of lead found in a child's blood in mcg/dl of whole blood. The covariate x was two-dimensional, recording age and neighborhood of residence. They matched each exposed child to one control child of the same age and neighborhood whose parents were employed in other industries not using lead. The data file data_lead.csv in the class directory shows the levels of lead found in the children's blood in mcg/dl of whole blood. If this study were free of hidden bias, which may or may not be the case, we would be justified in analyzing these data using methods for a uniform randomized experiment with 33 matched pairs.**

   Perform the following:
   (i) a one-sided test of the null of no treatment effect using the Wilcoxon signed rank test, do not use a shortcut function,
   (ii) the Hodges-Lehmann estimate of the size of an additive effect,
   (iii) the associated 95% confidence interval, and
   (iv) confirm i-iii using wilcoxon.test() or similar function. Be sure to interpret your results.

Let's start with the data set:

pair	exposed	control	diff	|diff|	rank	q
1	38	16	22	22	21	21
2	23	18	5	5	7	7
3	41	18	23	23	22.5	22.5
4	18	24	-6	6	8.5	0
5	37	19	18	18	20	20
6	36	11	25	25	25	25
7	23	10	13	13	13	13
8	62	15	47	47	31	31
9	31	16	15	15	16	16
10	34	18	16	16	17.5	17.5
11	24	18	6	6	8.5	8.5
12	14	13	1	1	1.5	1.5
13	21	19	2	2	3	3
14	17	10	7	7	10	10
15	16	16	0	0	0	0
16	20	16	4	4	6	6
17	15	24	-9	9	11.5	0
18	10	13	-3	3	4.5	0
19	45	9	36	36	29	29
20	39	14	25	25	25	25
21	22	21	1	1	1.5	1.5
22	35	19	16	16	17.5	17.5
23	49	7	42	42	30	30
24	48	18	30	30	27	27
25	44	19	25	25	25	25
26	35	12	23	23	22.5	22.5
27	43	11	32	32	28	28
28	39	22	17	17	19	19
29	34	25	9	9	11.5	11.5
30	13	16	-3	3	4.5	0
31	73	13	60	60	32	32
32	25	11	14	14	14.5	14.5
33	27	13	14	14	14.5	14.5

It's easy to create the rank column if you sort by *|diff|* and then sort back by *pair*. If you sum the rank column you get 499; that's the Wilcoxon signed rank statistic $v = \sum_{i=1}^{I} q_i * I_{[diff_i>0]}$.

If you search you'll find a table of critical values for the Wilcoxon signed rank statistic. For example, I grabbed this table:

Critical Values for the Wilcoxon Signed Ranks test

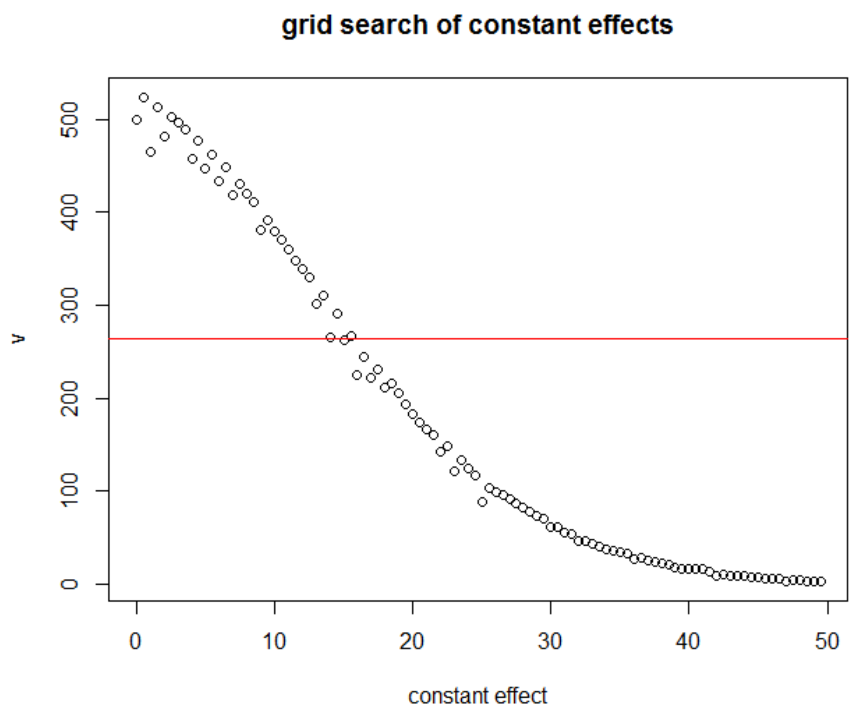
n	$\alpha1$ = $\alpha2$ =	5% 10%	2.5% 5%	1% 2%	0.5% 1%	n	$\alpha1$ = $\alpha2$ =	5% 10%	2.5% 5%	1% 2%	0.5% 1%
1		—	—	—	—	26		110	98	84	75
2		—	—	—	—	27		119	107	92	83
3		—	—	—	—	28		130	116	101	91
4		—	—	—	—	29		140	126	110	100
5		0	0	—	—	30		151	137	120	109
6		2	0	—	—	31		163	147	130	118
7		3	2	0	—	32		175	159	140	128
8		5	3	1	0	33		187	170	151	138
9		8	5	3	1	34		200	182	162	148
10		10	8	5	3	35		213	195	173	159
11		13	10	7	5	36		227	208	185	171
12		17	13	9	7	37		241	221	198	182
13		21	17	12	9	38		256	235	211	194
14		25	21	15	12	39		271	249	224	207
15		30	25	19	15	40		286	264	238	220
16		35	29	23	19	41		302	279	252	233
17		41	34	27	23	42		319	294	266	247
18		47	40	32	27	43		336	310	28	261
19		53	46	37	32	44		353	327	296	276
20		60	52	43	37	45		371	343	312	291
21		67	58	49	42	46		389	361	328	307
22		75	65	55	48	47		407	378	345	322
23		83	73	62	54	48		426	396	362	339
24		91	81	69	61	49		446	415	379	355
25		100	89	76	68	50		466	434	397	373

To save space, this table uses a slightly different way of coding the statistic. Note that sum of all of the ranks 1 to 32 (we dropped 33 because one of the pairs was 0) is 528. We can think of calculating the statistic using either (exposed-control) or (control-exposed). The sum of these statistics will equal 528. We might designate the way we calculated $v_+ = \sum_{i=1}^{I} q_i * I_{[diff_i>0]}$ and the other option as $v_- = \sum_{i=1}^{I} q_i * I_{[diff_i<0]}$. Thus 499+$v_-$=528 implies $v_- = 29$. This table that I found uses the minimum of v_- and v_+ and tests if $\min(v_-, v_+) < v_{crit}$. From this we see that our p-value is less than 0.005.

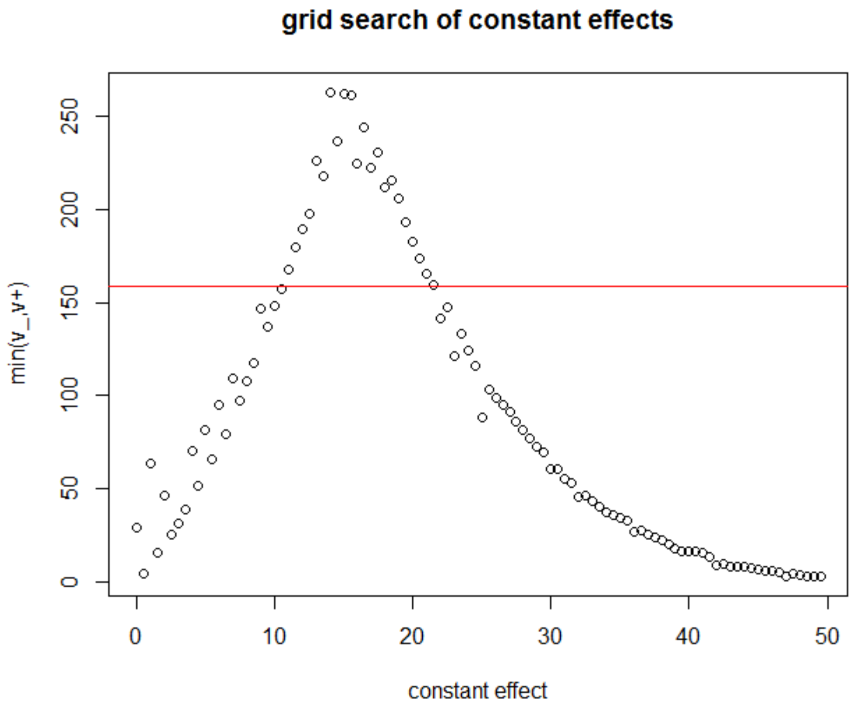
The exposed appear to be quite large compared to the control. To find the Hodges-Lehman we need to find the constant, additive effect that returns the expected value of the test statistic. The expected value of the test statistic is $\frac{I(I+1)}{4} = \frac{32(32+1)}{4} = 264$. There are a number of ways we can do this.

Looking ahead, we can do a grid search of values for the shift that produces the expected value of the test statistic. The grid search will be handy so we can get the 95% confidence interval. (There are wwwwaaaaayyyy better ways to do this.)

I put the corresponding code at the end of this. The red line is the 280 to give us a sense of where it crosses. You can eye-ball it at around 15. You then check more carefully in that range.



You can probably see things better on this plot which looks at the constant effect and the definition used by the table, $\min(v_-, v_+)$.



If we use the critical value for 32 pairs and two-side alpha of 0.05, which is 159. You can eyeball that it crosses the critical value somewhere around 10 and around 21. You can also see here that it peaks around 15.

```
> wilcox.test(d$exposed,d$control,paired=TRUE,conf.int = TRUE)

	Wilcoxon signed rank test with continuity correction

data:  d$exposed and d$control
V = 499, p-value = 1.155e-05
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
  9.999943 21.499970
sample estimates:
(pseudo)median
      15.49996
```

The shift parameter is 15.5 and its 95% CI is (10, 21.5). This is calculated under the constant, additive model.

sample code:

```
d <- read.csv("data_lead.csv",header=TRUE)

m <- function(x){wilcox.test(d$exposed-x,d$control,paired=TRUE)$statistic} #g
et the WSR stat out

iters_mat <- matrix(NA,nrow=100,ncol=2) #use to store grid
for(i in 1:100){
  iters_mat[i,1] <-  (i/2-.5)
  iters_mat[i,2] <- m((i/2-.5))
}
plot(iters_mat,
     xlab="constant effect",
     ylab="v",
     main="grid search of constant effects")
abline(h=264,col="red")


w <- ifelse(iters_mat[,2]<(528-iters_mat[,2]),
            iters_mat[,2],
            528-iters_mat[,2])
plot(iters_mat[,1],w,
     xlab="constant effect",
     ylab="min(v_,v+)",
     main="grid search of constant effects")
abline(h=159,col="red")


wilcox.test(d$exposed,d$control,paired=TRUE,conf.int = TRUE)
```