Math 3070 § 1. **Oxygen Consumption Example: Hodges-Lehman** Name: Example
Treibergs **Estimator for Center. Bootstrap & Simulation.** July 8, 2011

The Hodges-Lehman estimator is introduced in Problem 7.35 of Devore, *Probability and Statistics for Engineering and the Sciences,* 8*th ed.,* Brooks Cole, 2012. Devore applies it to data from the article "Oxygen Consumption During Fire Supression: Error of Heart Rate Estimation," *Ergonomics,* 1991. The sample gives oxygen consumption in (mL/kg/min) for ten firefighters.

The Hodges-Lehmann Estimator is the median of pairwise averages of all observations. It is known to be a good estimator for the center of a symmetric distribution. We apply it to the fireman data, and do simulations to assess its standard error by a bootstrap procedure. We take $B = 10,000$ samples of size $n = 10$ with replacement from the data (as a proxy for the background distribution) and compute the Hodges-Lehman for each. The bootstrapped standard error is $S_{\hat{\mu}} = 1.715$.

We also take $B = 10,000$ samples of size $n = 10$ from t-distribution with 5 degrees of freedom. Its variance is $\sigma^2 = \alpha/(\alpha - 2) = 5/(5 - 2) = 1.667$ so $\sigma = 1.291$. The simulated standard error of the Hodges-Lehman estimator is in this example $S_{\hat{\mu}} = .376$.

We use **R**'s vector arithmetic. Let a sample $X_,.X_2, \ldots, X_n$, be taken from a pdf symmetric about μ. For each $1 \le i \le j \le n$ we compute

$$\bar{X}_{i,j} = \frac{X_i + X_j}{2},$$

the mean of two observations. Then the Hodges-Lehman estimator is

$$\hat{\mu} = \operatorname*{median}_{i \le j} \bar{X}_{i,j}.$$

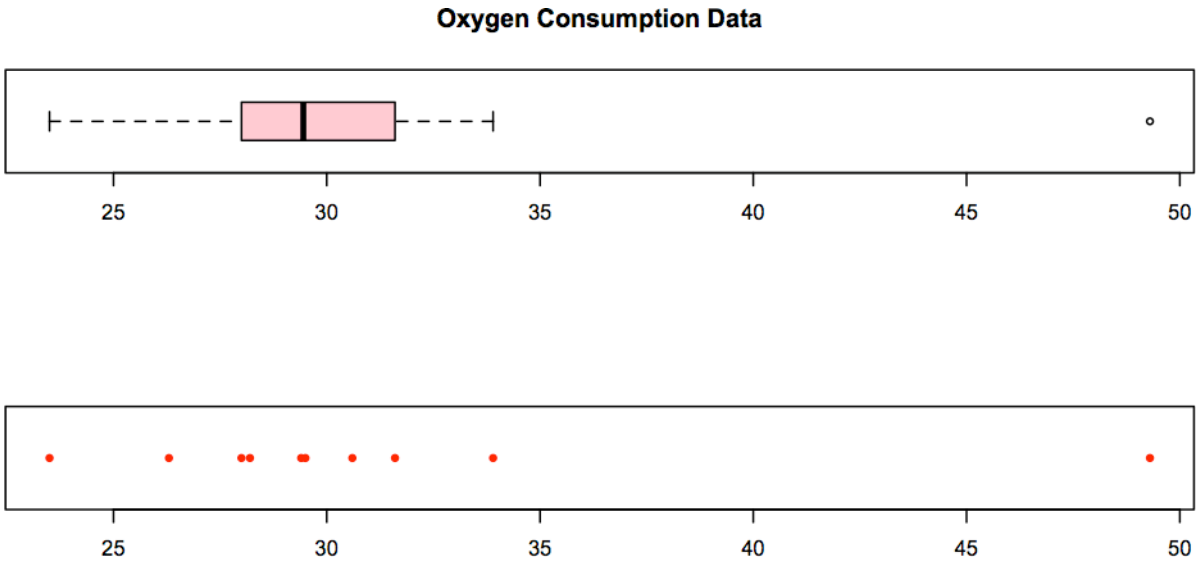
To compute it, `.5*outer(x,x,"+")` makes an array of pairwise averages. The statistic is the median of the upper triangle including the diagonal. The median of the whole matrix double counts off-diagonal entries. By throwing in another diagonal, we get adouble count of all entries, thus has the same median

$$\hat{\mu} = \text{median}\left(\left\{\bar{X}_{i,j} : 1 \le i, j \le n\right\} \cup \left\{\bar{X}_{i,i} : 1 \le i \le n\right\}\right).$$

R Session:

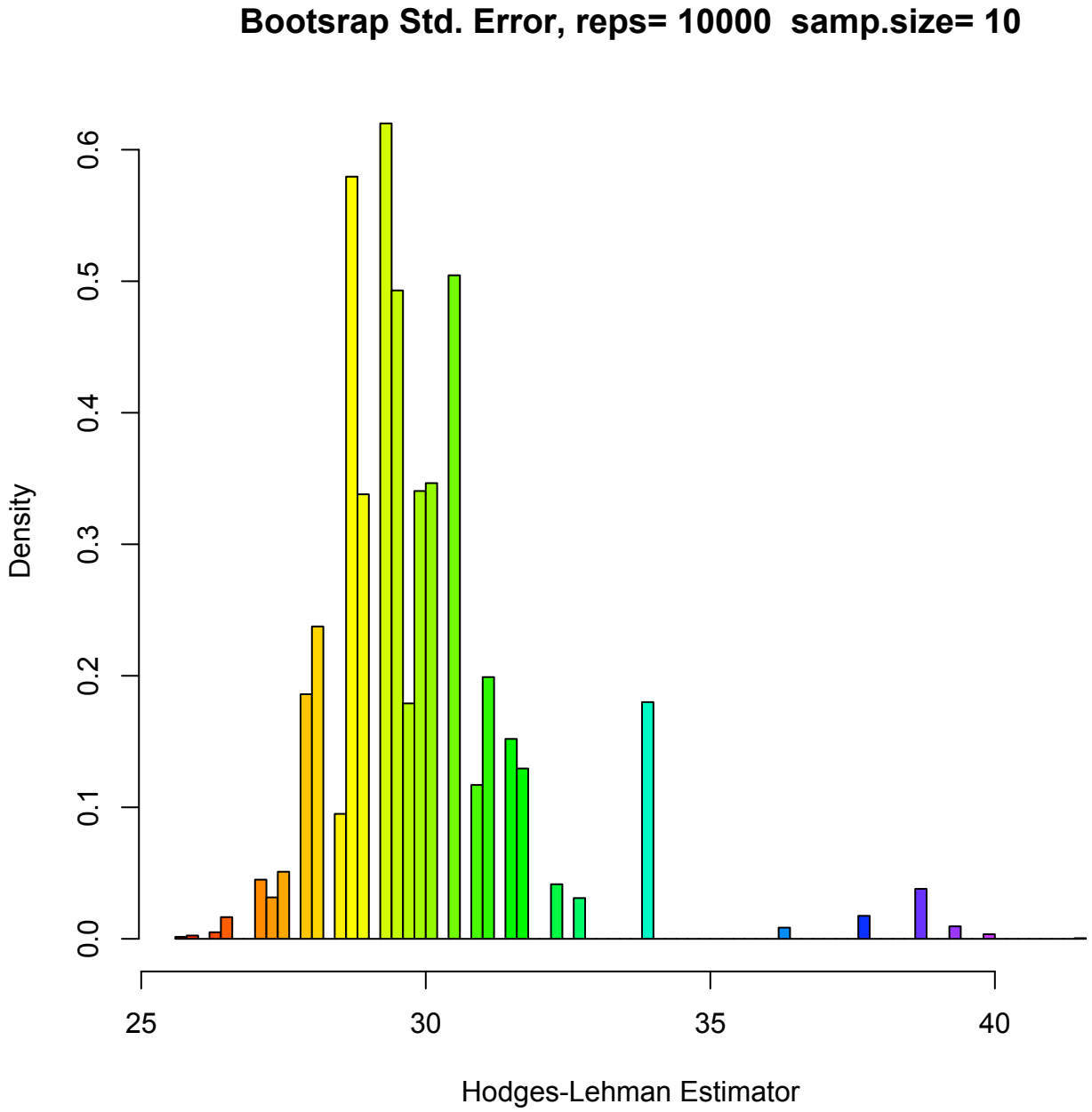
```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
  Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]
[Workspace restored from /Users/andrejstreibergs/.RData]
```

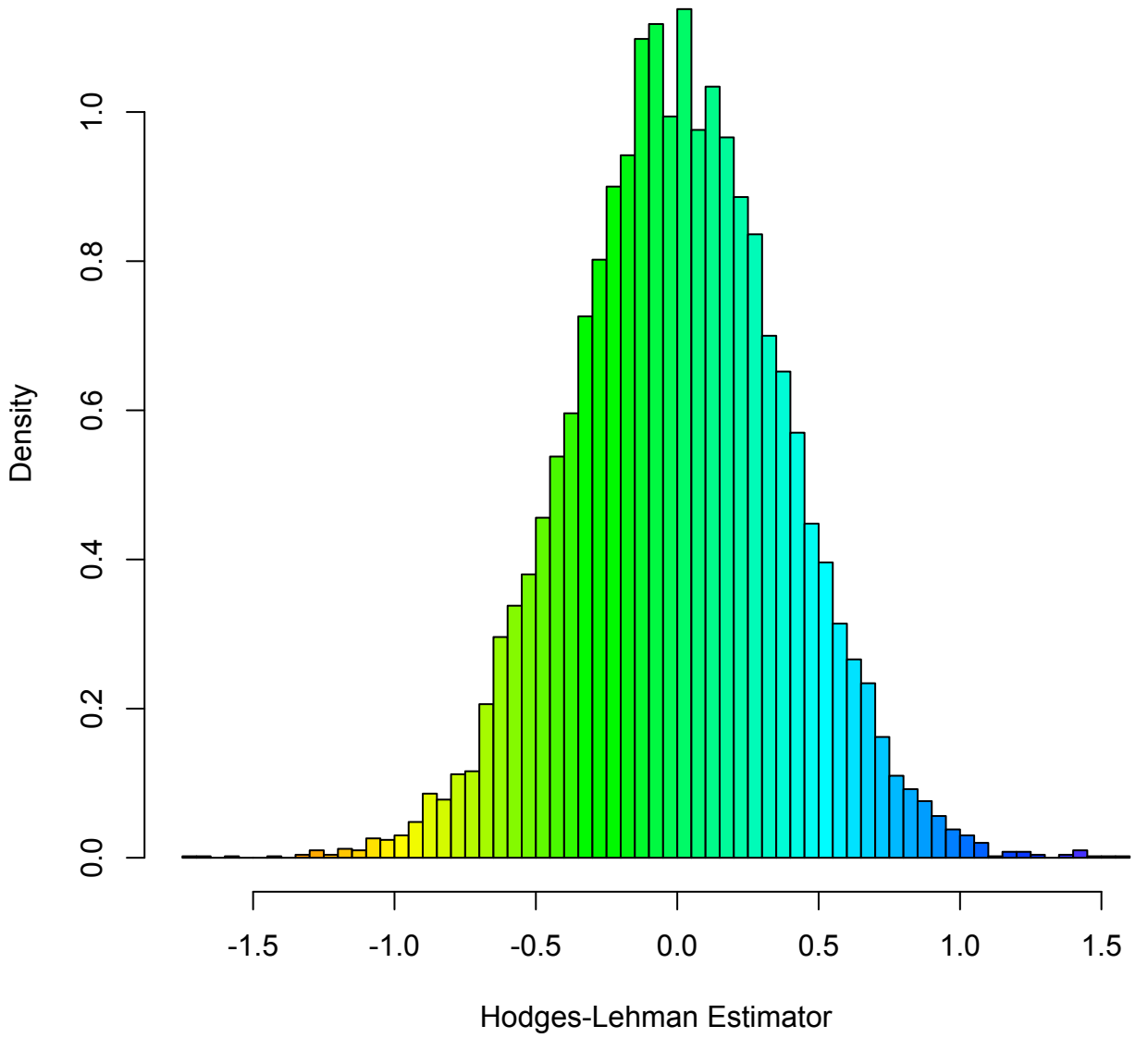
```
> ###################### ENTER OXY CONS DATA ###################################
> # Devore 1.44
> x <- scan ()
1: 29.5 30.6 49.3 28.2 28.0 26.3 33.9 29.4 23.5 31.6
11:
Read 10 items
> x
 [1] 29.5 30.6 49.3 28.2 28.0 26.3 33.9 29.4 23.5 31.6
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.50   28.05   29.45   31.03   31.35   49.30
> range(x)
[1] 23.5 49.3
> # Find sample range.
> max(x)-min(x)
[1] 25.8

> # Find variance and s.d.
> var(x); sd(x)
[1] 49.31122
[1] 7.022195

> # plot boxplot sand stripchart of this data.
> boxplot(x,horizontal=T,col="pink")
> title("Oxygen Consumption Data")
> stripchart(x,method="stack",pch=19,col=2)
> # M3074OxyCons1.pdf
```



```
> ################### FUNCTION TO COMPUTE HODGES-LEHMAN STATISTIC ############
> # devore 6.35  Hodges lehman estimator
> hl <- function(x){ median(c(.5*outer(x, x, "+"), x)) }
> hl(x)
[1] 29.5
> ############### TO UNDERSTAND THE CODE, WE DISSECT IT: ######################
> outer(x,x,"+")
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,] 59.0 60.1 78.8 57.7 57.5 55.8 63.4 58.9 53.0  61.1
 [2,] 60.1 61.2 79.9 58.8 58.6 56.9 64.5 60.0 54.1  62.2
 [3,] 78.8 79.9 98.6 77.5 77.3 75.6 83.2 78.7 72.8  80.9
 [4,] 57.7 58.8 77.5 56.4 56.2 54.5 62.1 57.6 51.7  59.8
 [5,] 57.5 58.6 77.3 56.2 56.0 54.3 61.9 57.4 51.5  59.6
 [6,] 55.8 56.9 75.6 54.5 54.3 52.6 60.2 55.7 49.8  57.9
 [7,] 63.4 64.5 83.2 62.1 61.9 60.2 67.8 63.3 57.4  65.5
 [8,] 58.9 60.0 78.7 57.6 57.4 55.7 63.3 58.8 52.9  61.0
 [9,] 53.0 54.1 72.8 51.7 51.5 49.8 57.4 52.9 47.0  55.1
[10,] 61.1 62.2 80.9 59.8 59.6 57.9 65.5 61.0 55.1  63.2
> .5*outer(x,x,"+")
       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
 [1,] 29.50 30.05 39.40 28.85 28.75 27.90 31.70 29.45 26.50 30.55
 [2,] 30.05 30.60 39.95 29.40 29.30 28.45 32.25 30.00 27.05 31.10
 [3,] 39.40 39.95 49.30 38.75 38.65 37.80 41.60 39.35 36.40 40.45
 [4,] 28.85 29.40 38.75 28.20 28.10 27.25 31.05 28.80 25.85 29.90
 [5,] 28.75 29.30 38.65 28.10 28.00 27.15 30.95 28.70 25.75 29.80
 [6,] 27.90 28.45 37.80 27.25 27.15 26.30 30.10 27.85 24.90 28.95
 [7,] 31.70 32.25 41.60 31.05 30.95 30.10 33.90 31.65 28.70 32.75
 [8,] 29.45 30.00 39.35 28.80 28.70 27.85 31.65 29.40 26.45 30.50
 [9,] 26.50 27.05 36.40 25.85 25.75 24.90 28.70 26.45 23.50 27.55
[10,] 30.55 31.10 40.45 29.90 29.80 28.95 32.75 30.50 27.55 31.60
> c(.5*outer(x,x,"+"),x)
  [1] 29.50 30.05 39.40 28.85 28.75 27.90 31.70 29.45 26.50 30.55 30.05 30.60
 [13] 39.95 29.40 29.30 28.45 32.25 30.00 27.05 31.10 39.40 39.95 49.30 38.75
 [25] 38.65 37.80 41.60 39.35 36.40 40.45 28.85 29.40 38.75 28.20 28.10 27.25
 [37] 31.05 28.80 25.85 29.90 28.75 29.30 38.65 28.10 28.00 27.15 30.95 28.70
 [49] 25.75 29.80 27.90 28.45 37.80 27.25 27.15 26.30 30.10 27.85 24.90 28.95
 [61] 31.70 32.25 41.60 31.05 30.95 30.10 33.90 31.65 28.70 32.75 29.45 30.00
 [73] 39.35 28.80 28.70 27.85 31.65 29.40 26.45 30.50 26.50 27.05 36.40 25.85
 [85] 25.75 24.90 28.70 26.45 23.50 27.55 30.55 31.10 40.45 29.90 29.80 28.95
 [97] 32.75 30.50 27.55 31.60 29.50 30.60 49.30 28.20 28.00 26.30 33.90 29.40
[109] 23.50 31.60
> sort(c(.5*outer(x,x,"+"),x))
  [1] 23.50 23.50 24.90 24.90 25.75 25.75 25.85 25.85 26.30 26.30 26.45 26.45
 [13] 26.50 26.50 27.05 27.05 27.15 27.15 27.25 27.25 27.55 27.55 27.85 27.85
 [25] 27.90 27.90 28.00 28.00 28.10 28.10 28.20 28.20 28.45 28.45 28.70 28.70
 [37] 28.70 28.70 28.75 28.75 28.80 28.80 28.85 28.85 28.95 28.95 29.30 29.30
 [49] 29.40 29.40 29.40 29.40 29.45 29.45 29.50 29.50 29.80 29.80 29.90 29.90
 [61] 30.00 30.00 30.05 30.05 30.10 30.10 30.50 30.50 30.55 30.55 30.60 30.60
 [73] 30.95 30.95 31.05 31.05 31.10 31.10 31.60 31.60 31.65 31.65 31.70 31.70
 [85] 32.25 32.25 32.75 32.75 33.90 33.90 36.40 36.40 37.80 37.80 38.65 38.65
 [97] 38.75 38.75 39.35 39.35 39.40 39.40 39.95 39.95 40.45 40.45 41.60 41.60
[109] 49.30 49.30
```

```
> ###################### COMPUTATION OF THE MEDIAN ############################
> # in this case  x  has length  n=10. There are 10 x 10 in the matrix +
> # 10  more for the extra diagonal. the median entry is the average of the
> # 55th and 56th, namely  (29.5+29.5)/2=29.5.
> median(c(.5*outer(x,x,"+"),x))
[1] 29.5
>
> ############## BOOTSTRAP BY SAMPLING FROM THIS DATA ######################
> # simulate Sampling Distribution
> B <- 10000
> v <- replicate(B,hl(sample(x,n,replace=T)))
> summary(v)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  25.75   28.80   29.50   29.88   30.55   41.60
> sd(v)
[1] 1.715152

> hist(v, col = rainbow(90), xlab = "Hodges-Lehman Estimator", freq = F,
+ main = paste("Bootsrap Std. Error, reps=", B, " samp.size=", n), breaks ="fd")
>  # M3074OxyCons2.pdf
>
> ############## SIMULATE BY SAMPLING FROM T(5) ##############################
>
> # samle from dist symmetric about zero
> B <- 10000
> n <- 10
> v <- replicate(B,hl(rt(n,5)))
> summary(v)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.718000 -0.244800  0.001588  0.002746  0.252100  1.584000
> sd(v)
[1] 0.3762987
>
> hist(v, col = rainbow(90), xlab = "Hodges-Lehman Estimator", freq = F,
+ main = paste("Simulate Std. Error, reps=", B, " samp.size=", n,
+ " from T(5)"), breaks = "fd")
> # M3074OxyCons3.pdf
```


Bootsrap Std. Error, reps= 10000 samp.size= 10
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
25
30
35
40
Hodges-Lehman Estimator


Simulate Std. Error, reps= 10000 samp.size= 10 from T(5)
Density
0.0
0.2
0.4
0.6
0.8
1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Hodges-Lehman Estimator