Stat 425 HW8 Solutions

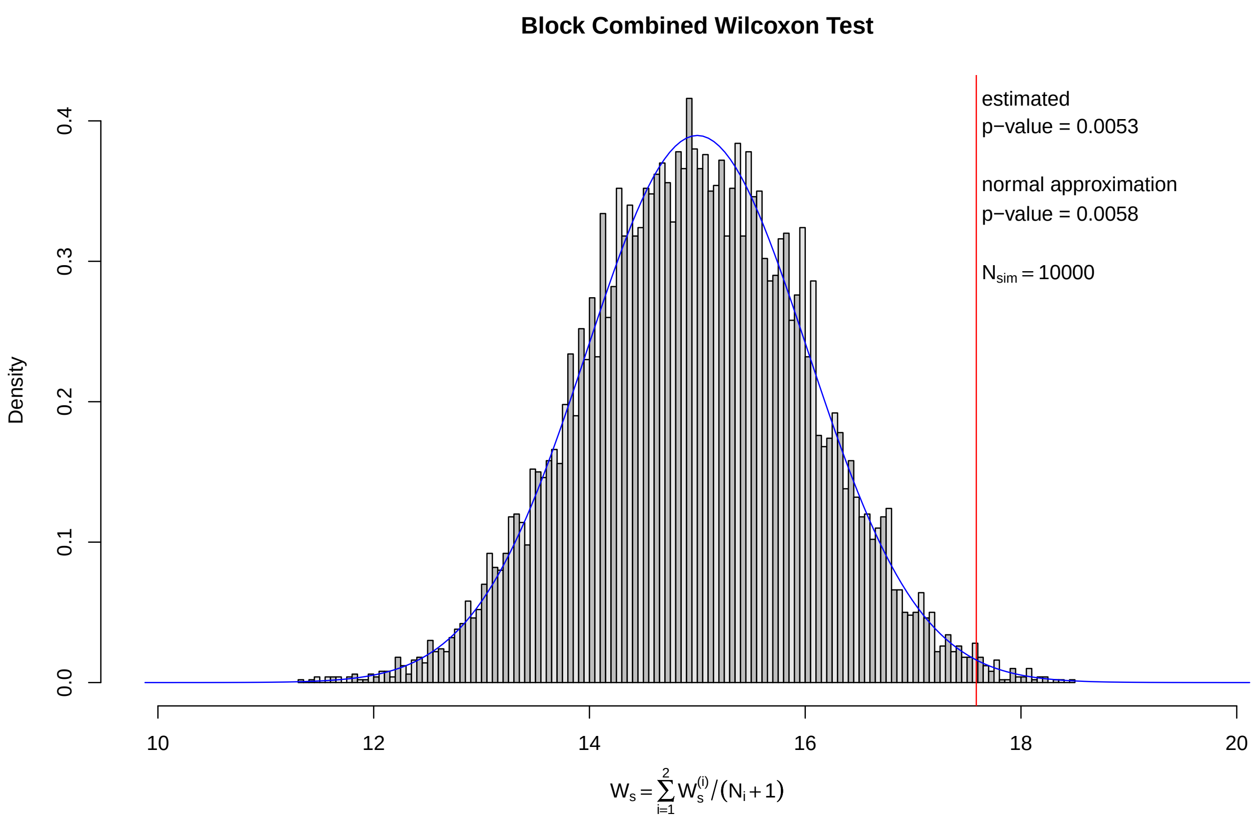
Fritz Scholz

For the first two problems you may/should use appropriate R functions introduced in class and posted on the class web site.

1. Problem 49 (p. 150 of Textbook). While the ordered data categories allow ranking (using midranks where appropriate) you may want to use scores of -2, -1, 0, 1, 2 for the five categories, as suggested in the next problem. Such scores will not change the rankings but they facilitate processing in R. Give your coding steps (or function) and the results (p-values and plots).

After the coding the data are as follows for the two studies and they were put in list form and then processed by the function `BlockedWilcoxon` with resulting plot shown below.

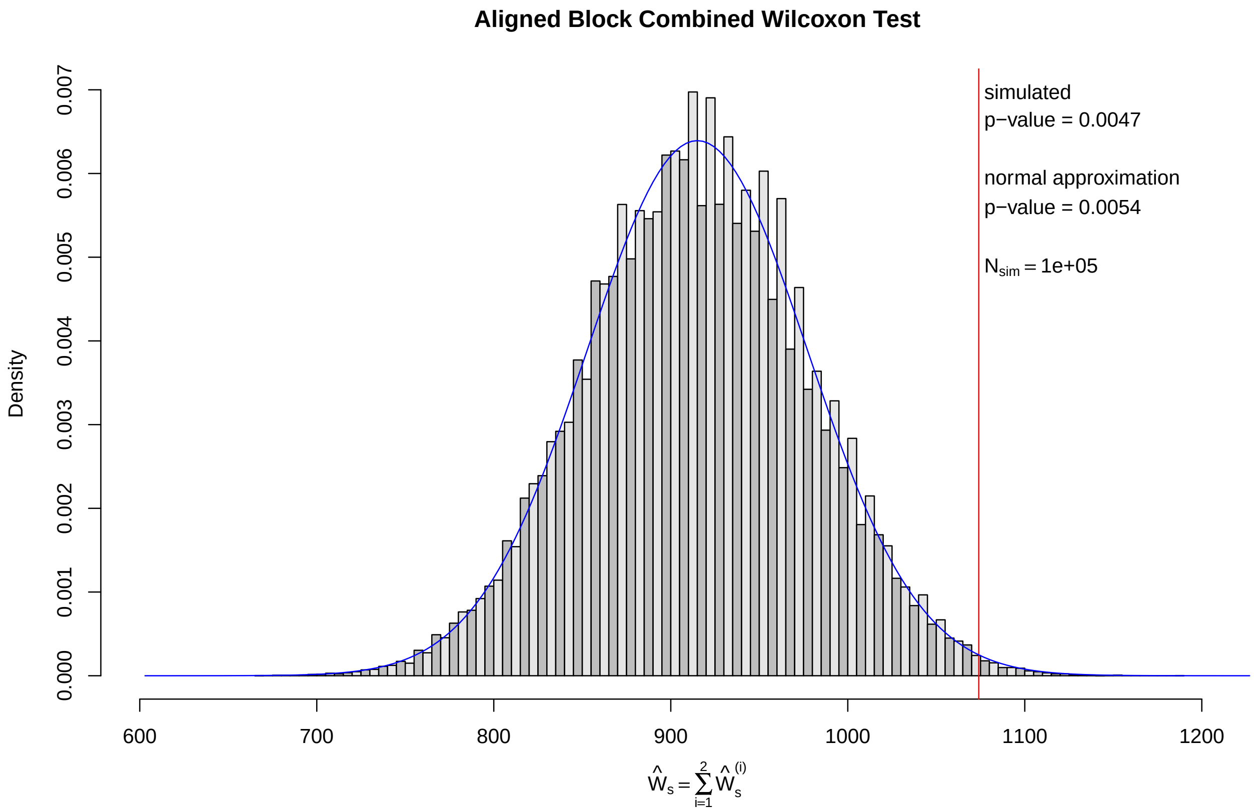
```
x1<-c(2,2,11,4,1); y1<-c(0,1,9,7,3); x2<-c(2,2,5,1,0); y2<-c(0,2,4,3,1)
z<-c(-2,-1,0,1,2)
x1 <- rep(z,x1); y1 <- rep(z,y1); x2 <- rep(z,x2); y2 <- rep(z,y2)
datlist49=list(list(x1,y1),list(x2,y2))
BlockedWilcoxon(datlist49,PDF=T)
```



2. Problem 55 (p. 151 of Textbook). Here the scores are needed for the alignment process. Give your coding steps (or function) and the results (p-values and plots).

```
AlignedBlockedWilcoxon(datlist49,align="median",PDF=T,Nsim=100000)
```

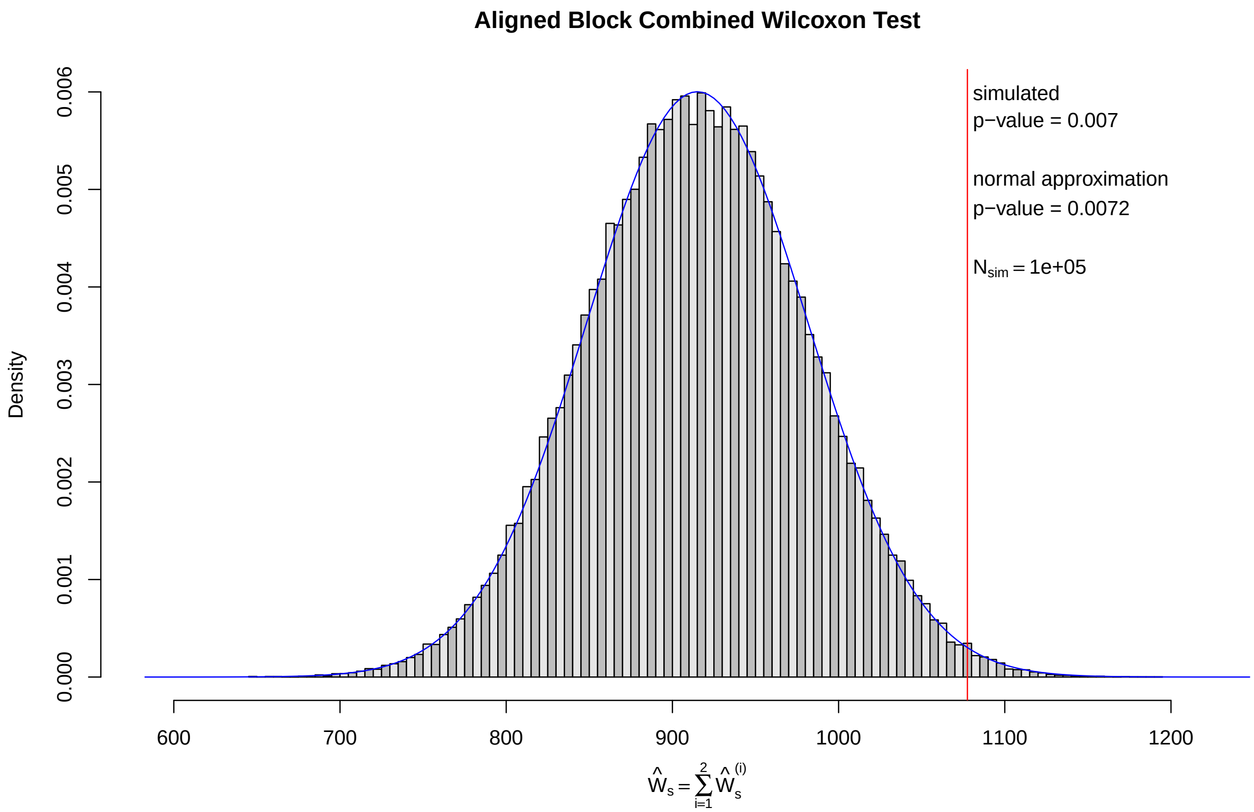
produced the following plot with median block alignment.



while

```
AlignedBlockedWilcoxon(datlist49,align="mean",PDF=T,Nsim=100000)
```

produced the following plot with mean block alignment. Note how much smoother the histogram looks under mean alignment. The median shows much more jerky behavior, especially with so many ties.



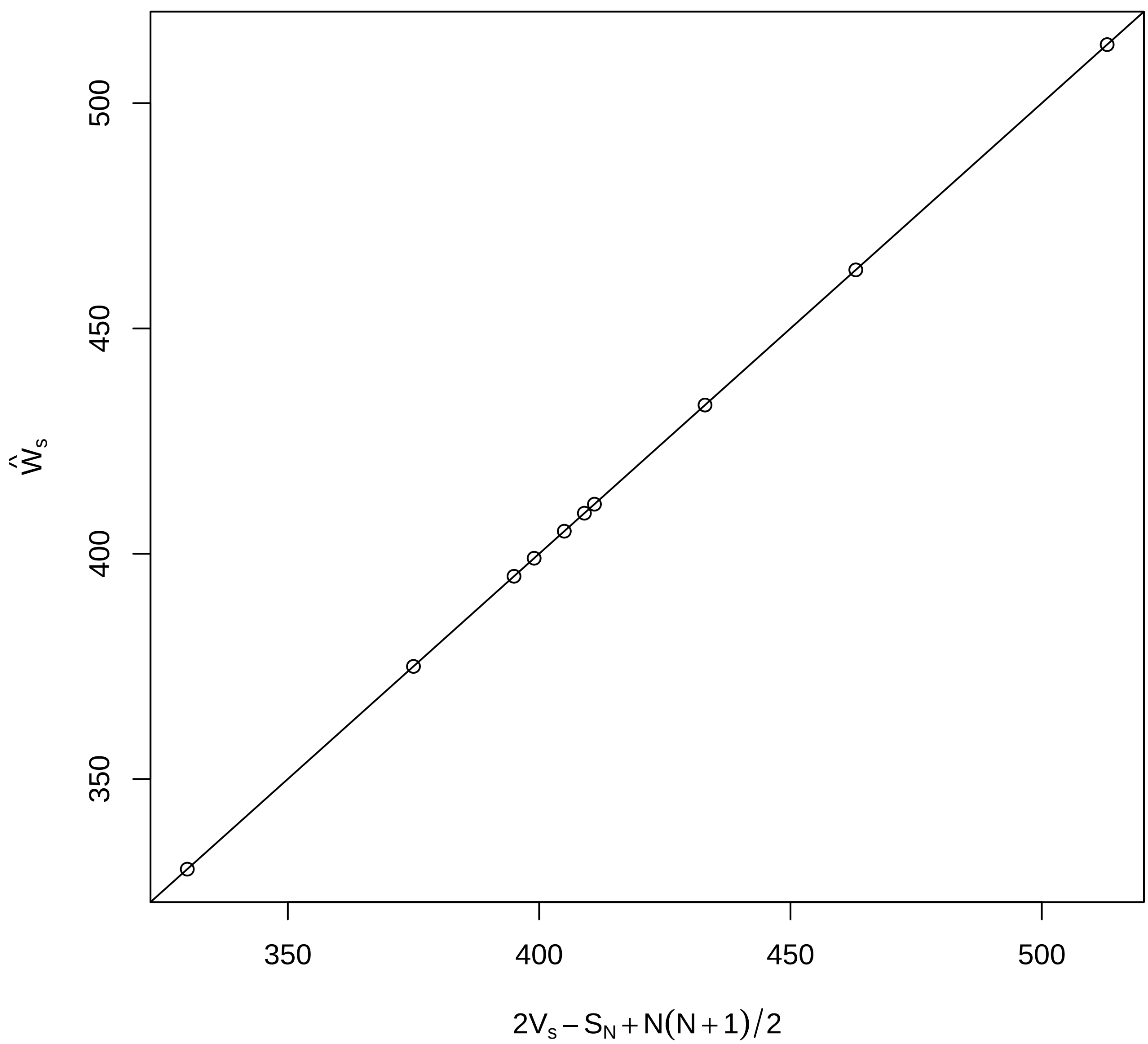
3. Explore the formula 3.32 of the Textbook (valid in the case of block size 2) for $N = 20$ blocks. Write a function that generates such block data pairs (x_i, y_i) from $y_i \sim \mathcal{N}(i \times .1 + \Delta, 1)$ (e.g. `y=rnorm(N,(1:N)*.1+Delta,1)`), and $x_i \sim \mathcal{N}(i \times .1, 1)$, $i = 1, \ldots, N$ and which computes $\hat{W}_s$, V_s and S_N according to their respective definitions. For V_s you may want to use $V_s = \texttt{wilcox.test(y,x,paired = T)\$statistic}$. Do this $N_{\text{sim}} = 10$ times and accumulate vectors of length $N_{\text{sim}} = 10$ for the three statistics $\hat{W}_s$, V_s and S_N. Check the formula relationship by plotting $\hat{W}_s$ against the appropriate linear relationship of V_s and S_N and superimpose the main diagonal. Do this for $\Delta = 0$ and $\Delta = 1$. Describe any change that you perceive.

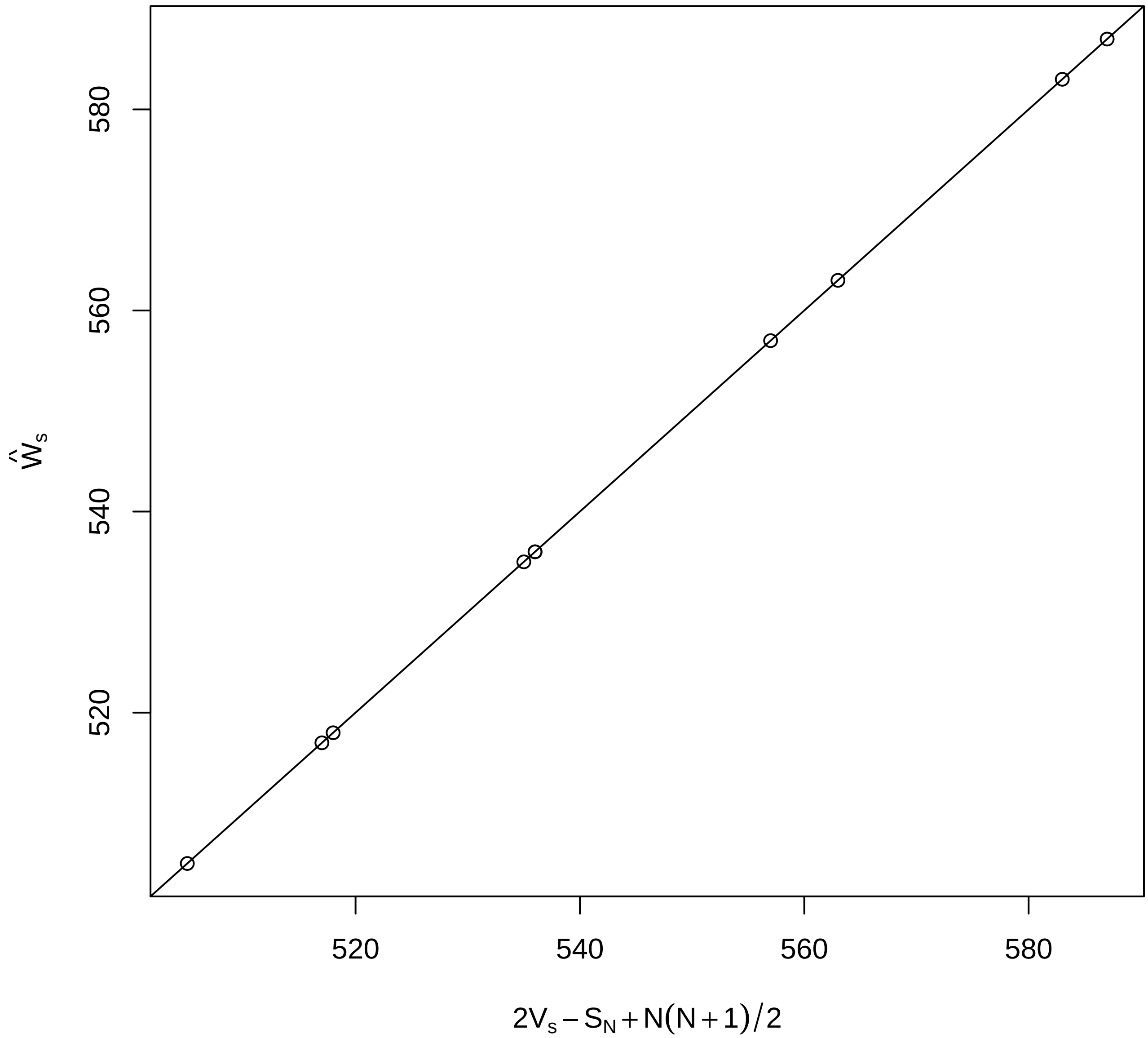
Write a separate function that simulates $\hat{W}_s$, V_s and S_N (with $N = 20$) $N_{\text{sim}} = 100$ times using the same types of data sets of $N = 20$ pairs $x_i \sim \mathcal{N}(i \times .1, 1)$, and $y_i \sim \mathcal{N}(i \times .1 + \Delta, 1)$ for $i = 1, \ldots, N$ to calculate each of the three statistics. Plot the $N_{\text{sim}} = 100$ values of $\hat{W}_s$ against the corresponding values of V_s. Plot a line to this data pattern via `abline(lsfit(Vs,Whats))`. What does to plot suggest or confirm?

Give your function codes and the resulting plots.

```
formula3.32a=function(Nsim=10,N=20,Delta=.3,PDF=F){
if(PDF==T) pdf(file="formula332.pdf",width=7)
What=rep(0,Nsim)
SN=rep(0,Nsim)
Vs=rep(0,Nsim)
for(i in 1:Nsim){
  x=rnorm(N,.1*(1:N),1); y=rnorm(N,.1*(1:N)+Delta,1)
  mxy=(x+y)/2; xt=x-mxy; yt=y-mxy; rt=rank(c(yt,xt)); What[i]=sum(rt[1:N])
  SN[i]=sum(y>x); Vs[i]=wilcox.test(y,x,paired=T)[[1]]
  }
```

```
plot(2*Vs-SN+N*(N+1)/2,What,xlab=expression(2*V[s]-S[N]+N*(N+1)/2),
   ylab=expression(hat(W)[s]))
abline(0,1)
if(PDF==T) dev.off()
}
```



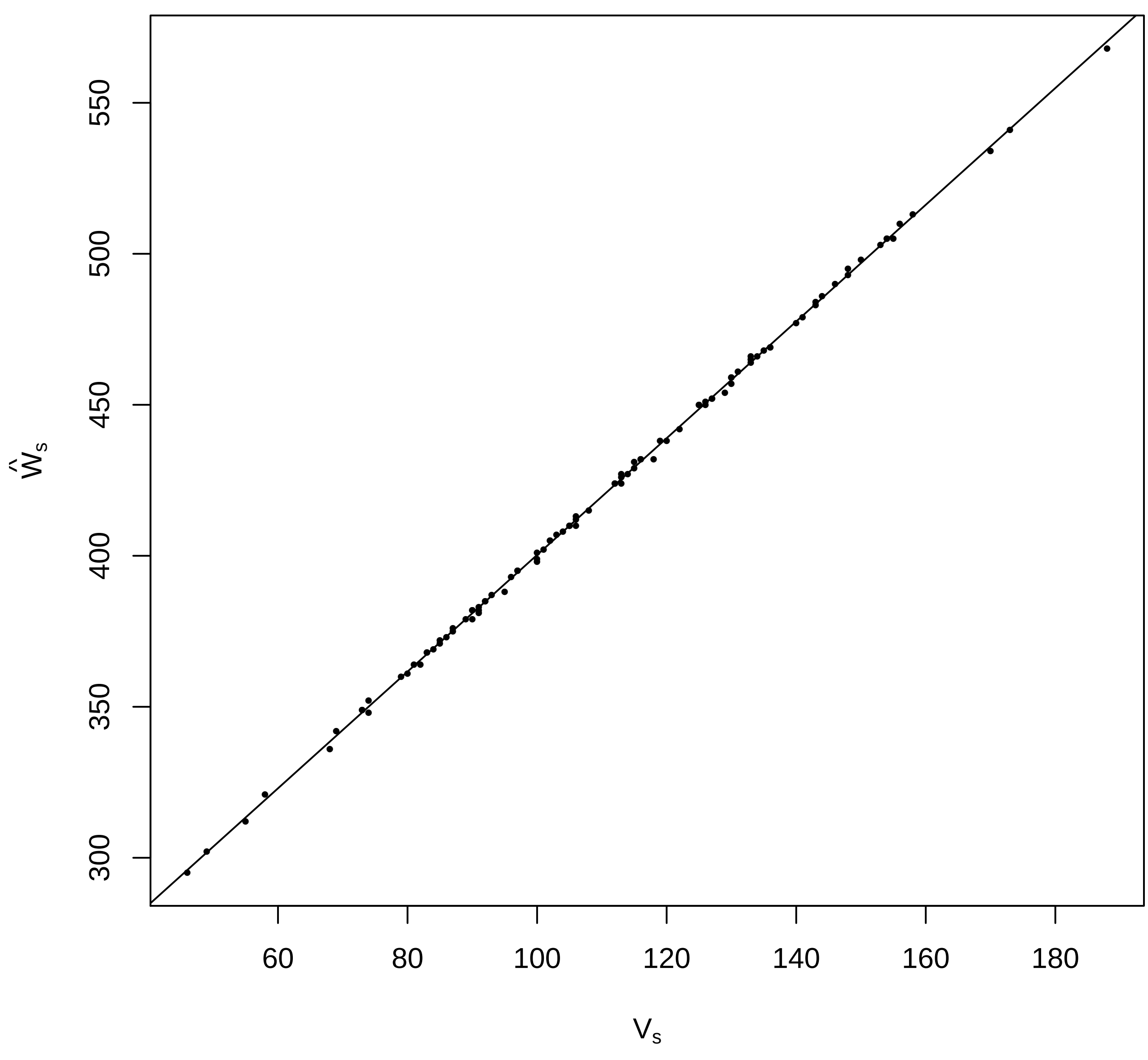


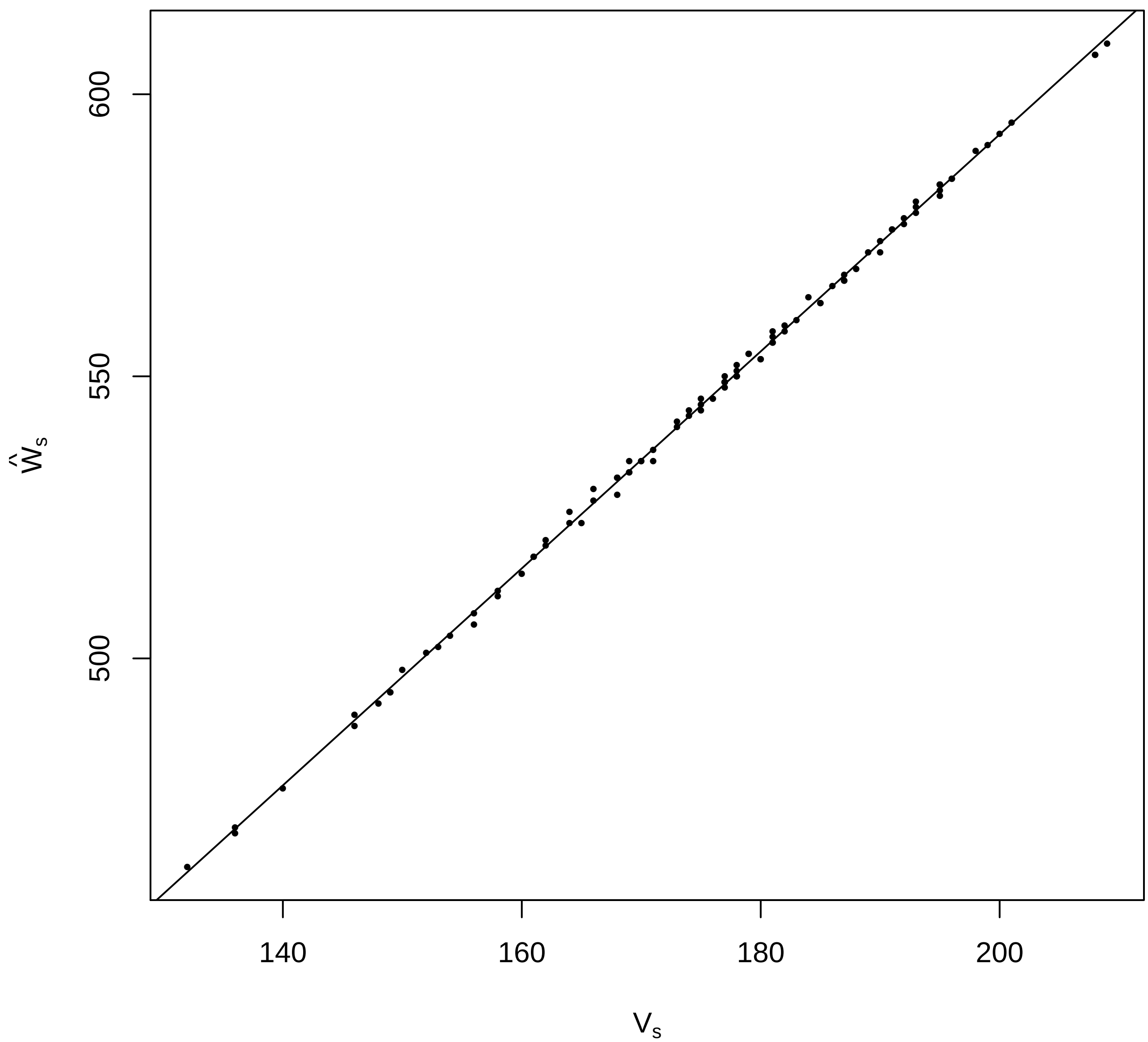
In both plots the points lie on the main diagonal as they should according to formula (3.32). In the second plot corresponding to $\Delta = 1$ the points are generally higher in their coordinate values than in the first plot. In the first plot we are dealing with values in the range of the null distribution of $\hat{W}_s$ while in the latter case we should see higher values beause we sampled data under the alternative that y values are generally increased by $\Delta = 1$ relative to x values.

The code for the simulation study comparing $\hat{W}_s$ with V_s values follows

```
formula3.32b=function(Nsim=100,N=20,Delta=0,PDF=F){
if(PDF==T) pdf(file="formula332.pdf",width=7)
What=rep(0,Nsim); SN=rep(0,Nsim); Vs=rep(0,Nsim)
for(i in 1:Nsim){
  x=rnorm(N,.1*(1:N),1); y=rnorm(N,.1*(1:N)+Delta,1)
  mxy=(x+y)/2; xt=x-mxy; yt=y-mxy; rt=rank(c(yt,xt)); What[i]=sum(rt[1:N])
  SN[i]=sum(y>x); Vs[i]=wilcox.test(y,x,paired=T)[[1]]
  }
plot(What,Vs,pch=16,cex=.5,xlab=expression(hat(W)[s]),ylab=expression(V[s]))
abline(lsfit(What,Vs))
if(PDF==T) dev.off()}
```

resulting in respective plots for $\Delta = 0$ and $\Delta = 1$





Both plots suggest that there is an almost linear relationship between $\hat{W}_s$ and V_s, i.e., they can serve as almost equivalent test statistic. This was pointed out in the slides and in the Text and was attributed to the relatively small variability of S_N in relation to the variability of V_s.