

# T-tests, differences, and pooled variance

*Dan Yurovsky*

*11/9/2016*

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
theme_set(theme_bw())
```

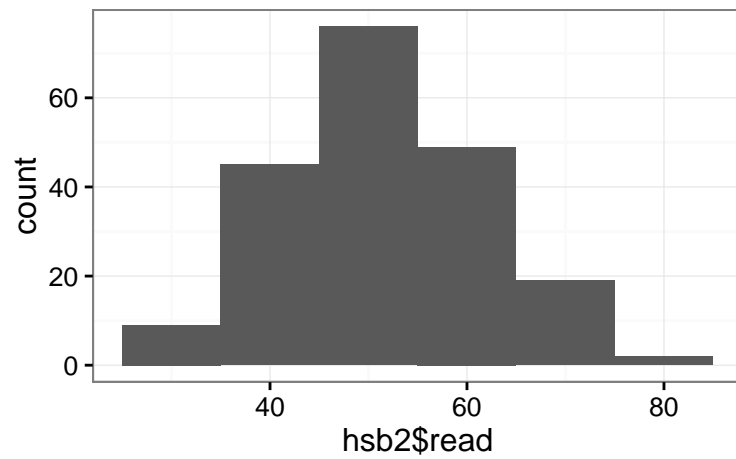
```
hsb2 <- read.delim("hsb2.txt") %>%
```

```
  mutate(diff = read - write)
```

```
head(hsb2)
```

```
##   id gender  race   ses schtyp      prog read write math science socst
## 1  70   male white   low public  general   57   52   41     47    57
## 2 121 female white middle public vocational 68   59   53     63    61
## 3  86   male white   high public  general   44   33   54     58    31
## 4 141   male white   high public vocational 63   44   47     53    56
## 5 172   male white middle public  academic 47   52   57     53    61
## 6 113   male white middle public  academic 44   52   51     63    61
##   diff
## 1     5
## 2     9
## 3    11
## 4    19
## 5    -5
## 6    -8
```

```
qplot(hsb2$read, binwidth = 10)
```



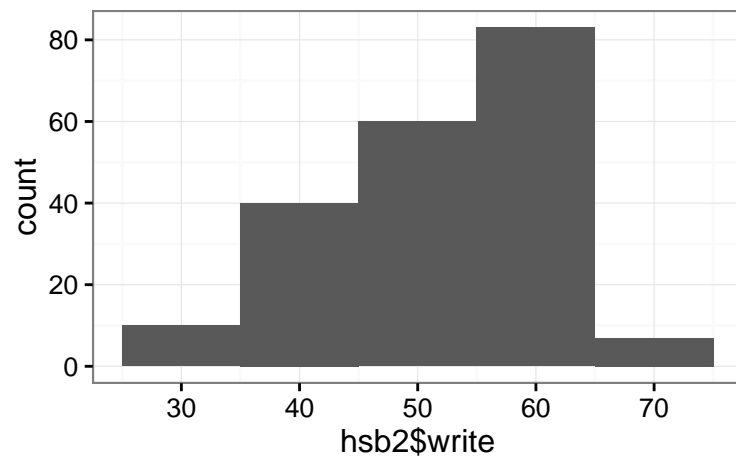
```
mean(hsb2$read)
```

```
## [1] 52.23
```

```
sd(hsb2$read)
```

```
## [1] 10.25294
```

```
qplot(hsb2$write, binwidth = 10)
```



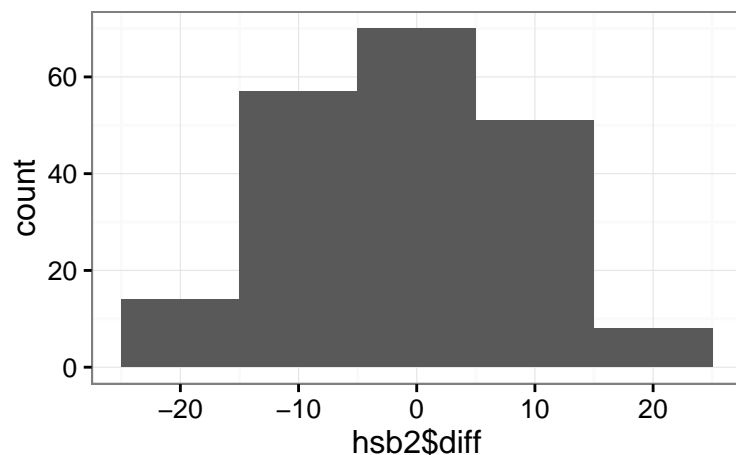
```
mean(hsb2$write)
```

```
## [1] 52.775
```

```
sd(hsb2$write)
```

```
## [1] 9.478586
```

```
qplot(hsb2$diff, binwidth = 10)
```



```
mean(hsb2$diff)
```

```
## [1] -0.545
```

```
sd(hsb2$diff)
```

```
## [1] 8.886666
```

```
# Test if the reading scores's mean is drawn from a population who's mean  
# is the mean of the writing scores samples
```

```
t.test(hsb2$read, mu = mean(hsb2$write))
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: hsb2$read
```

```
## t = -0.75173, df = 199, p-value = 0.4531
```

```
## alternative hypothesis: true mean is not equal to 52.775
```

```
## 95 percent confidence interval:
```

```
## 50.80035 53.65965
```

```
## sample estimates:
```

```
## mean of x
```

```
## 52.23
```

```
# Test if the difference between reading and writing scores is drawn from  
# a population with a mean of 0
```

```
t.test(hsb2$read - hsb2$write, mu = 0)
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: hsb2$read - hsb2$write
```

```
## t = -0.86731, df = 199, p-value = 0.3868
```

```
## alternative hypothesis: true mean is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -1.7841424 0.6941424
```

```
## sample estimates:
```

```
## mean of x
```

```
## -0.545
```

```
# Test if the reading scores's mean is drawn from the  
# same population as the writing scores mean
```

```
t.test(hsb2$read - sample(hsb2$write), mu = 0)
```

```
##
## One Sample t-test
##
## data: hsb2$read - sample(hsb2$write)
## t = -0.56691, df = 199, p-value = 0.5714
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.440758 1.350758
## sample estimates:
## mean of x
## -0.545
```