Statistically Speaking: *Difference Tests*

DAVID L. TABB, PH.D.

SEPTEMBER 21, 2017

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Overview

- Statistical history of beer and bullets
- Test statistics and p-values
- Examples of difference testing in R

Wikimedia Commons

William S. Gosset

- Born in 1876, he joined the Guinness Brewery after earning a chemistry degree at Oxford University. A leave of absence in 1906-1907 under Karl Pearson at UCL led to his publication of Student’s T distribution.
- He used a pen name because Guinness wanted to retain their statistical advantage.
- His mild personality helped bond the contentious British statistical community.

https://priceonomics.com/the-guinness-brewer-who-revolutionized-statistics/

Frank Wilcoxon

- An American born in Ireland (1882), Wilcoxon completed his Ph.D. in physical chemistry at Cornell in 1924.
- During World War II, his caution protected ordnance workers from accidents.
- His second statistics paper substituted ranks for original values in difference tests to create the rank-sum and signed-rank tests.

http://www.swlearning.com/quant/kohler/stat/biographical_sketches/bio21.1.html

A test statistic summarizes key information

STUDENT'S T TEST

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_p\sqrt{2/n}}$$

- $\overline{X}_i$ is mean of sample.
- s_p is pooled std. dev.
- n is number of samples.

MANN-WHITNEY U TEST

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

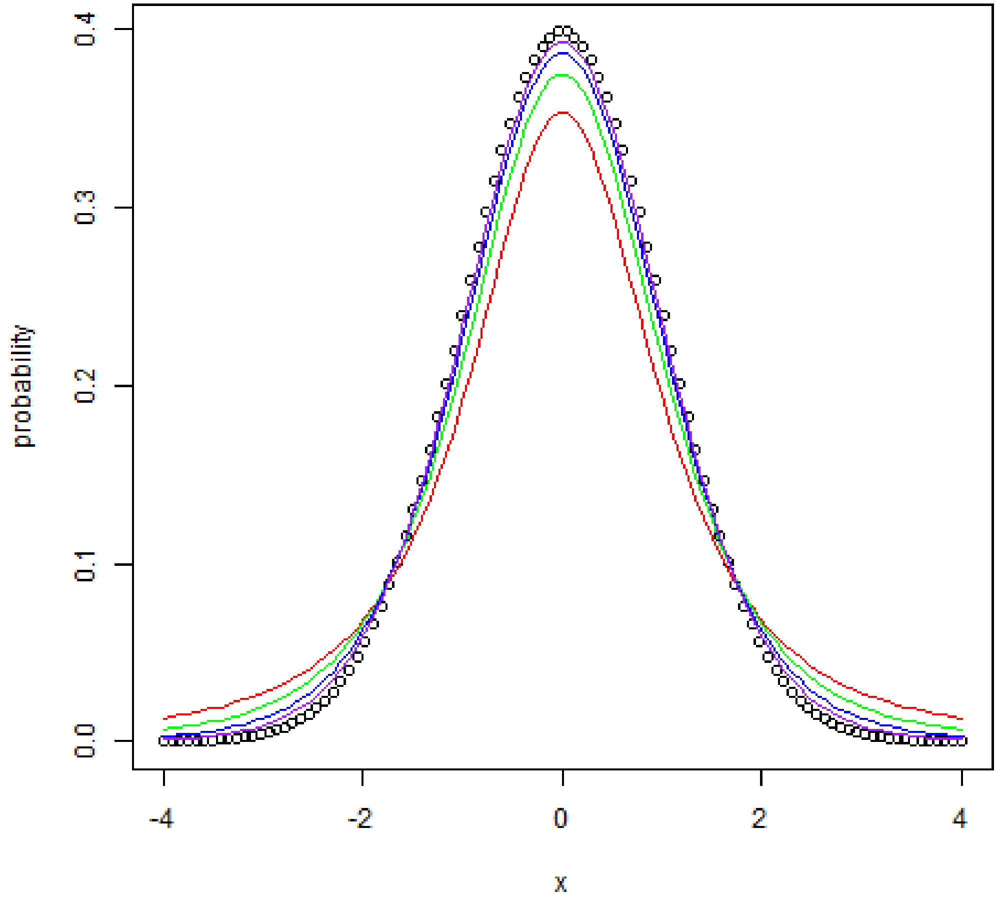
- R_1 is summed rank of samples in set 1.
- n_1 is number of samples in set 1.

Determine probability of a more extreme test statistic

- Under *null hypothesis* (H_0), we assume two distributions have the same mean.
- Under this assumption, we ask how probable a higher or lower test statistic would be *by random chance*.
- Gosset described the t-distribution in 1908; we can compute these probabilities! We call them *p-values*.

http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-t-tests-t-values-and-t-distributions

T-distribution depends upon *degrees of freedom*



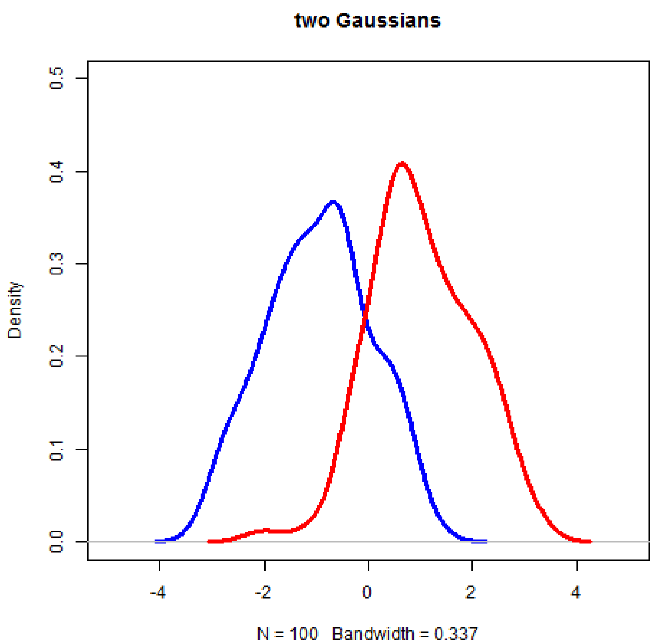
Red: t,df=2 Green: t,df=4 Blue: t,df=8 Purple: t,df=16 Dots: normal

Key concepts in difference testing

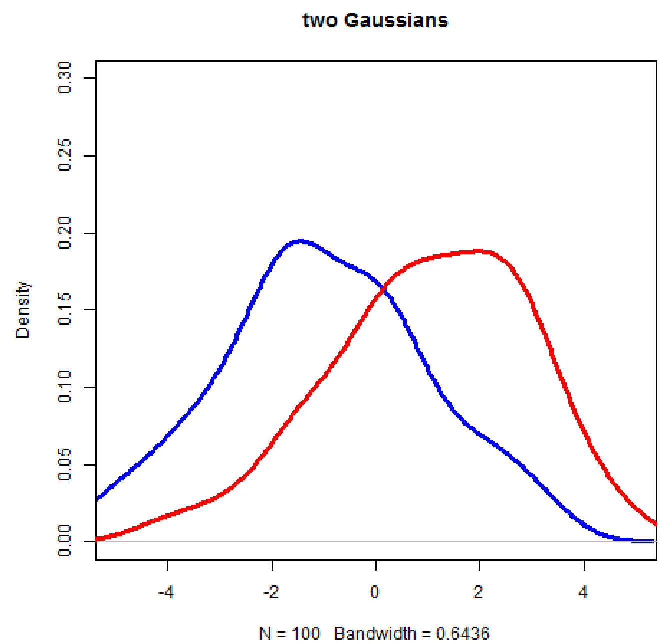
- Paired: If you have two snapshots of each sample (say, before and after), each value in one cohort pairs with one value in the other.
- One-Sided or Two-Sided: If you hypothesized that values will rise in B than in A rather than fall, use a one-sided test. A two-sided test thinks both increases and decreases are important changes.

Difference requires mean separation and low variance

STANDARD DEVIATION=1



STANDARD DEVIATION=2



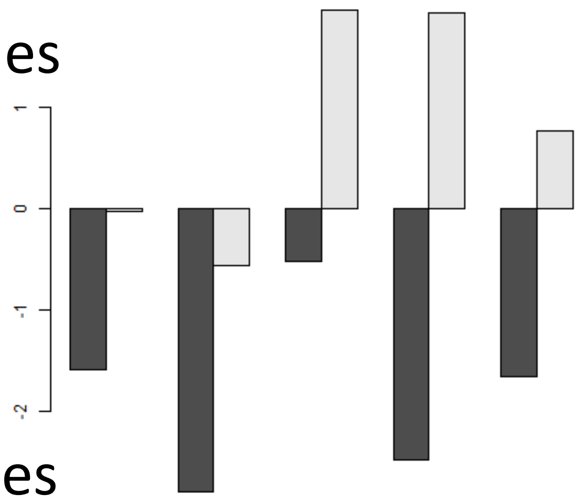
In both cases, mean of blue is -1 and mean of red is +1.

T-Test application in R

```
a <- rnorm(5,mean=-1, sd=1)
b <- rnorm(5,mean=+1, sd=1)
barplot(rbind(a,b),beside=T)
t.test(a,b)
```

Just for plotting

b values

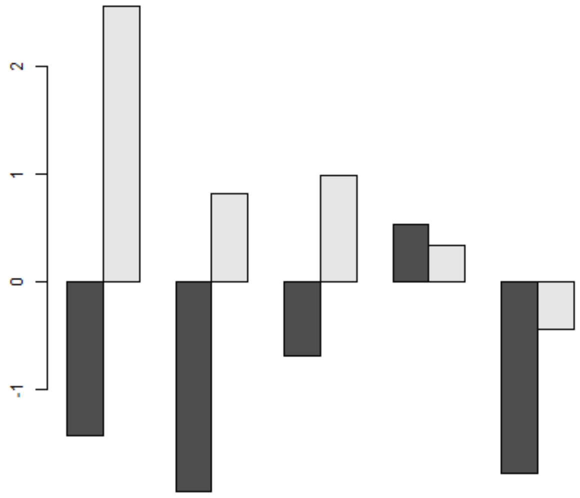


a values

```
Welch Two Sample t-test
data:  a and b
t = -4.0638, df = 7.5563,
p-value = 0.004075
alternative hypothesis: true
difference in means is not
equal to 0
95 percent confidence
interval:
 -4.114532 -1.115893
sample estimates:
 mean of x  mean of y
-1.8039078  0.8113044
```

U-Test application in R

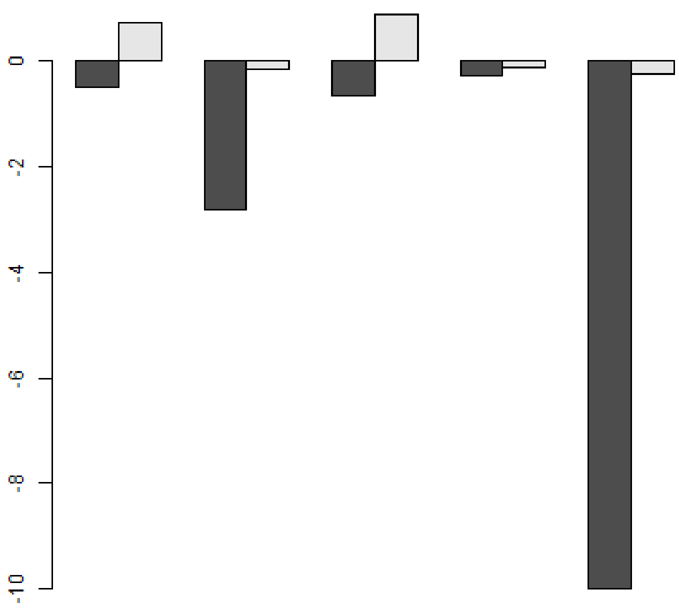
```
a <- rnorm(5,mean=-1, sd=1)
b <- rnorm(5,mean=+1, sd=1)
barplot(rbind(a,b),beside=T)
wilcox.test(a,b)
```




```
Wilcoxon rank sum test
data:  a and b
W = 2, p-value = 0.03175
alternative hypothesis:
true location shift is
not equal to 0
```

Now with an outlier!

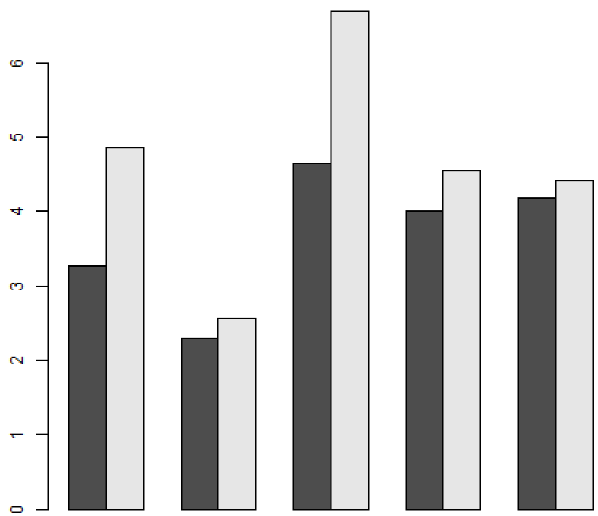
```
a <- c(rnorm(4,mean=-1,sd=1),-10)
b <- rnorm(5,mean=+1,sd=1)
```



- T-test p-value: 0.1718
- U-test p-value: 0.007937
- T-test dislikes high variance.
- U-test sees all As are lower than all Bs!

Paired, one-sided example

```
a <- 5*runif(5)
b <- a + abs(rnorm(5))
t.test(a,b, paired=T,
alternative="less")
```



- As written: p-value=0.032
- Paired,two-sided: p-value=0.063
- Unpaired, one-sided: p-value=0.134

Takeaways

- Difference testing is essential in biomedical research, but it may be misused.
- Student’s T-test has been with us for just over a century.
- Be sure you know whether or not your data are paired and if a sided test is appropriate.
- When your data are irregularly distributed, opt for a non-parametric test instead.