

## Chapter 9

### 9.1

Suppose that the sample means and variances are  $\bar{X}_1 = 15$ ,  $\bar{X}_2 = 12$ ,  $s_1^2 = 8$ ,  $s_2^2 = 24$ ,  $n_1 = 20$ ,  $n_2 = 10$

Verify that  $s_p^2 = 13.14$ ,  $T = 2.14$  and that Student's T test rejects the hypothesis of equal means with  $\alpha = 0.05$ .

```
xbar1 <- 15; xbar2 <- 12; s1sq <- 8; s2sq <- 24
n1 <- 20; n2 <- 10
alpha <- 0.05

variance <- ( (n1 - 1)*s1sq + (n2 - 1)*s2sq ) / ( n1 + n2 - 2 )
variance
```

```
[1] 13.14286
```

```
t.stat <- (xbar1 - xbar2) / sqrt( variance*(1/n1 + 1/n2) )
t.stat
```

```
[1] 2.136637
```

```
crit <- qt(alpha/2, df = n1 + n2 - 2)
```

```
ifelse(abs(t.stat) >= crit, "Reject Null", "Fail to Reject")
```

```
[1] "Reject Null"
```

### 9.2

For two independent groups of subjects,  $\bar{X}_1 = 45$ ,  $\bar{X}_2 = 36$ ,  $s_1^2 = 4$ ,  $s_2^2 = 16$ ,  $n_1 = 20$ ,  $n_2 = 30$ . Assume that the population variances of the two groups are equal and verify that the estimate of this common variance is 11.25.

```
xbar1 <- 45; xbar2 <- 36; s1sq <- 4; s2sq <- 16
n1 <- 20; n2 <- 30

variance <- ( (n1 - 1)*s1sq + (n2 - 1)*s2sq ) / ( n1 + n2 - 2 )

stopifnot(variance == 11.25)
```

### 9.3

Still assuming equal variances, test the hypothesis of equal means using Student's T test and the data in the last exercise. Use  $\alpha = 0.05$

```
alpha <- 0.05
```

```
t.stat <- (xbar1 - xbar2) / sqrt( variance * (1/n1 + 1/n2) )

crit <- qt(alpha/2, df = (n1 + n2 - 2))

ifelse(abs(t.stat) >= crit, "Reject null", "Fail to Reject")

[1] "Reject null"
```

#### 9.4

Repeat the last problem, only use Welch's test for comparing means.

```
w.stat <- (xbar1 - xbar2) / sqrt(s1sq/n1 + s2sq/n2)

q1 <- s1sq/n1; q2 <- s2sq/n2
df <- (q1 + q2)^2 / ( q1^2/(n1 - 1) + q2^2/(n2 - 1) )

crit <- qt(1 - alpha/2, df)

ifelse(abs(w.stat) >= crit, "Reject Null", "Fail to Reject")

[1] "Reject Null"
```

#### 9.5

Comparing the results for the last two problems, what do they suggest regarding the power of Welch's test compared to Student's T test when the sample variances differ sufficiently.

```
t.stat

[1] 9.29516

w.stat

[1] 10.50974
```

*Roughly, the W statistic has more power because it is larger than the T statistic.*

#### 9.6

For two independent groups  $\bar{X}_1 = 86$ ,  $\bar{X}_2 = 80$ ,  $s_1^2 = s_2^2 = 25$ ,  $n_1 = n_2 = 20$ . Assume that the population variances of the two groups are equal and verify that Student's T rejects with  $\alpha = 0.01$ .

```
xbar1 <- 86; xbar2 <- 80; s1sq <- s2sq <- 25
n1 <- n2 <- 20; alpha <- 0.01
```

```
t.stat <- (xbar1 - xbar2) / sqrt(2 * s1sq/n1 )
crit <- qt(alpha/2, df = n1 + n2 - 2)
ifelse(abs(t.stat) >= crit, "Reject", "Fail to Reject")

[1] "Reject"
```

## 9.7

Repeat the last exercise using Welch's method.

```
w.stat <- (xbar1 - xbar2) / sqrt( 2 * s1sq/n1 )
q1 <- s1sq/n1; q2 <- s2sq/n2
df <- (q1 + q2)^2 / ( q1^2/(n1 - 1) + q2^2/(n2 - 1))
crit <- qt(alpha/2, df = 38)
ifelse(abs(w.stat) >= crit, "Reject", "Fail to Reject")

[1] "Reject"
```

## 9.8

Comparing the results of the last two problems, what do they suggest about using Student's T versus Welch's method when the sample variances are approximately equal?

```
t.stat
```

```
[1] 3.794733
```

```
w.stat
```

```
[1] 3.794733
```

*With equal sample variances, Student's T and Welch give exactly the same result, suggesting that when the sample variances are approximately equal, the choice between T and W makes little difference.*

## 9.9

For  $\bar{X}_1 = 10$ ,  $\bar{X}_2 = 5$ ,  $s_1^2 = 21$ ,  $s_2^2 = 29$ ,  $n_1 = n_2 = 16$ , compute a 0.95 confidence interval for the difference between the means using Welch's method and state whether a decision can be made about which group has the larger population mean.

```
xbar1 <- 10; xbar2 <- 5; s1sq <- 21; s2sq <- 29
n1 <- n2 <- 16; alpha <- 0.05

q1 <- s1sq/n1; q2 <- s2sq/n2
df <- (q1 + q2)^2 / (q1^2 / (n1 - 1) + q2^2 / (n2 - 1) )

variance <- sqrt(s1sq/n1 + s2sq/n2)

(xbar1 - xbar2) + qt(c(Lower = alpha/2, Upper = 1 - alpha/2), df)*variance
```

```
      Lower      Upper
1.385859 8.614141
```

### 9.10

Repeat the last problem, only use Student's T instead.

```
sqp <- (s1sq+s2sq) / 2

(xbar1 - xbar2) + qt(c(Lower = alpha/2, Upper = 1 - alpha/2), n1 + n2 - 2) * sqrt(sqp * (1/n1 + 1/n2))

      Lower      Upper
1.389738 8.610262
```

### 9.11

Two methods for training accountants are to be compared. Students are randomly assigned to one of the two methods. At the end of the course, each student is asked to prepare a tax return for the same individual. The amounts of the refund reported by the students are:

M1: 132, 204, 603, 50, 125, 90, 185, 134 M2: 92, -42, 121, 63, 182, 101, 294, 36

Using Welch's test would you conclude that the methods differ in terms of the average return? Use  $\alpha = 0.05$ .

```
dat <- data.table(
  M1 = c(132, 204, 603, 50, 125, 90, 185, 134),
  M2 = c(92, -42, 121, 63, 182, 101, 294, 36))

t.test(dat$M1, dat$M2, conf.level = .95)
```

Welch Two Sample t-test

```
data: dat$M1 and dat$M2
t = 1.1922, df = 11.193, p-value = 0.2579
alternative hypothesis: true difference in means is not equal to 0
```

95 percent confidence interval:

-71.17601 240.17601

sample estimates:

mean of x mean of y

190.375 105.875

The p-value is 0.258, so no, fail to reject and make no decision about which group has the larger population mean.

## 9.12

Responses to stress are governed by the hypothalamus. Imagine you have two randomly sampled groups of individuals between the ages of 60 and 65. The first shows signs of heart disease and the other does not. You want to determine whether the groups differ in terms of some measure associated with the hypothalamus. For the first group of subjects with no heart disease, the measures are:

11.1, 12.2, 15.5, 17.6, 13.0, 7.5, 9.1, 6.6, 9.5, 18.0, 12.6

For the other group with heart disease, the kmeasures are:

18.2, 14.1, 13.8, 12.1, 34.1, 12.0, 14.1, 14.5, 12.6, 12.5, 19.8, 13.4, 16.8, 14.1, 12.9

Determine whether the groups differ based on Welch's test. Use  $\alpha = 0.05$ .

```
dat <- data.table(
  M1 = c(11.1, 12.2, 15.5, 17.6, 13.0, 7.5, 9.1, 6.6, 9.5, 18.0, 12.6),
  M2 = c(18.2, 14.1, 13.8, 12.1, 34.1, 12.0, 14.1, 14.5, 12.6, 12.5, 19.8, 13.4, 16.8, 14.1, 12.9)
```

Warning in as.data.table.list(x, keep.rownames = keep.rownames, check.names = check.names, : Item 1 has 11 rows but longest item has 15; recycled with remainder.

```
t.test(dat$M1, dat$M2, conf.level = .95)
```

Welch Two Sample t-test

data: dat\$M1 and dat\$M2

t = -1.7809, df = 24.075, p-value = 0.08756

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-6.6057164 0.4857164

sample estimates:

mean of x mean of y

12.60667 15.66667

*Fail to reject because the p-value is greater than 0.05 and make no decision about which group has the larger population mean.*

**9.13**

The 0.95 confidence interval for the difference between the means, using Student's T, is (2.2, 20.5).

What are the practical concerns with this confidence interval?

*The actual probability coverage could differ substantially from 0.95.*

**9.14**

For the first of two binomial distributions, there are 15 successes among 24 observations. For the second, there are 23 successes among 42 observations.

Test  $H_0 : p_1 = p_2$  with a Type I error probability of 0.05 using the Storer-Kim method, followed by Beal's method.

Note the p-values and comment on what this suggests in terms of power.

```
# Storer-Kim
```

```
twobinom(15, 24, 23, 42)
```

```
$p.value
```

```
[1] 0.5594805
```

```
$p1
```

```
[1] 0.625
```

```
$p2
```

```
[1] 0.547619
```

```
$est.dif
```

```
[1] 0.07738095
```

```
# Beal's
```

```
twobicipv(15, 24, 23, 42)
```

```
$p.value
```

```
[1] 0.999
```

```
$ci
```

```
[1] -0.2018983 0.3377457
```

```
$p1
```

```
[1] 0.625
```

```
$p2
```

```
[1] 0.547619
```

*In this particular case, the p-value returned by Beal's method is substantially larger than the p-value returned by the Storer Kim method, suggesting that generally the Storer Kim method will have more power.*

### 9.15

A film producer wants to know which of two versions of a particular scene is more likely to be viewed as disturbing. A group of 98 individuals views the first version and 20 say that it is disturbing. Another group sees the second version and 30 of 70 people say that it is disturbing.

What is the probability these two probabilities are equal?

```
# Storer-Kim
```

```
twobinom(20, 98, 30, 70)
```

```
$p.value
```

```
[1] 0.00174666
```

```
$p1
```

```
[1] 0.2040816
```

```
$p2
```

```
[1] 0.4285714
```

```
$est.dif
```

```
[1] -0.2244898
```

```
# Beal's
```

```
twobicipv(20, 98, 30, 70)
```

```
$p.value
```

```
[1] 0.005
```

```
$ci
```

```
[1] -0.37821798 -0.06101773
```

```
$p1
```

```
[1] 0.2040816
```

```
$p2
```

```
[1] 0.4285714
```

Both reject, again Storer Kim method has more power.

## 9.16

It is found that among 121 individuals who take a training program on investing in commodities, 20 make money during the next year and the rest do not. With another training program, 15 of 80 make money.

Test the hypothesis that the probability of making money is the same for both training programs with a Type I error probability of 0.05.

```
p1 <- 20/121; p2 <- 15/80
Z <- (p1-p2)/sqrt(p1*(1-p1)/121 + p2*(1-p2)/80)
crit <- qnorm(0.05/2)
ifelse(abs(Z) >= crit, "Reject Null", "Fail To Reject")
```

```
[1] "Reject Null"
```

## 9.17

In a study dealing with violence between factions in the Middle East, one goal was to compare measures of depression for two groups of young males. In the first group, no family member was wounded or killed by someone belonging to the opposite faction, and the measures were:

- 22, 23, 12, 11, 30, 22, 7, 42, 24, 33, 28, 19, 4, 34, 15, 26, 50, 27, 20, 30, 14, 42

The second group consisted of young males who had a family member killed or wounded. The observed values were:

- 17, 22, 16, 16, 14, 29, 20, 20, 19, 14, 10, 8, 26, 9, 14, 17, 21, 16, 14, 11, 14, 11, 29, 13, 4, 16, 16, 7, 21

Test the hypothesis of equal means with Student's T test and  $\alpha = 0.05$ , and compute a .95 confidence interval.

```
dat <- data.table(
  X1 = c(22, 23, 12, 11, 30, 22, 7, 42, 24, 33, 28, 19, 4, 34, 15, 26, 50, 27, 20, 30, 14, 42),
  X2 = c(17, 22, 16, 16, 14, 29, 20, 20, 19, 14, 10, 8, 26, 9, 14, 17, 21, 16, 14, 11, 14, 11, 29, 13, 4, 16, 16, 7, 21),
)
```

Warning in as.data.table.list(x, keep.rownames = keep.rownames, check.names = check.names, : Item 1 has 22 rows but longest item has 29; recycled with remainder.

```
t.test(dat$X1, dat$X2)
```

Welch Two Sample t-test

data: dat\$X1 and dat\$X2

t = 2.9112, df = 43.24, p-value = 0.005671

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

2.09871 11.55646

sample estimates:



```
mean of x mean of y
22.82759 16.00000
```

```
t.test(dat$X1, dat$X2, var.equal = T)
```

### Two Sample t-test

```
data: dat$X1 and dat$X2
t = 2.9112, df = 56, p-value = 0.005159
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.129498 11.525675
sample estimates:
mean of x mean of y
22.82759 16.00000
```

*Student's T yields a shorter confidence interval, but the confidence interval based on Student's T can be less accurate compared to the confidence interval based on Welch's method.*

## 9.18

For the data in the last problem, the difference between the medians is -7.5 and the corresponding McKean-Schrader estimate of the standard error is  $\sqrt{(S_1^2 + S_2^2)} = 3.93$ .

Verify that you do not reject the hypothesis of equal medians with  $\alpha = 0.05$

```
msmed(dat$X1, dat$X2)
```

```
[1] "Warning: Group 1 has tied values. Might want to used medpb"
[1] "WARNING: tied values detected."
[1] "Estimate of standard error might be highly inaccurate, even with n large"
[1] "Warning: Group 2 has tied values. Might want to used medpb"
[1] "WARNING: tied values detected."
[1] "Estimate of standard error might be highly inaccurate, even with n large"
```

```
$test
      Group Group      test      crit      se    p.value
[1,]      1      2 1.769899 1.959964 3.390025 0.07704901
```

```
$psihat
      Group Group psihat  ci.lower ci.upper
[1,]      1      2      6 -0.644326 12.64433
```

**9.19**

Using the data in Exercise 17, use a percentile bootstrap method to compare the medians. Note that the p-value is 0.01. This is in contrast to the result in Exercise 18, where the hypothesis of equal medians is not rejected at the 0.05 level.

```
medpb2(dat$X1, dat$X2)
```

```
$n1
[1] 29

$n2
[1] 29

$p.value
[1] 0.0275

$ci
[1] 3 12

$est1
[1] 22

$est2
[1] 16

$est.dif
[1] 6
```

*The p-value indicates that the hypothesis of equal medians would be rejected at the 0.01 level, in contrast to the method based on the McKean-Schrader estimator, which has a p-value equal to 0.057.*

**9.20**

Repeat the analysis of the sexual attitude data, the second example in Section 9.10.2, only now use *twobinom* and *twobicipv*.

```
twobinom(49,105,101,156)
```

```
$p.value
[1] 0.003722579

$p1
[1] 0.4666667

$p2
[1] 0.6474359
```

```
$est.dif
[1] -0.1807692

twobicipv(49,105,101,156)

$p.value
[1] 0.008

$ci
[1] -0.31510569 -0.04021425

$p1
[1] 0.4666667

$p2
[1] 0.6474359
```

## 9.21

Does the consumption of alcohol limit our attention span? An article in the July 10, 2006 Los Angeles Times described a study conducted at the University of Washington where 23 people were given enough alcohol to reach a BAC of 0.04%. A second group of 23 people drank no alcohol. The researchers then showed members of both groups a 25-second video clip in which two teams passed a ball back and forth and asked them to count the number of times one team passed the ball. During the clip, a person in a gorilla suit walks through the crowd, thumps its chest, and walks off.

Researchers found that 11 of the 23 participants in the control group saw the gorilla, compared to 10 in the alcohol group.

Verify that you do not reject the null hypothesis that the two groups have the same probability of seeing the gorilla using the Storer-Kim method with  $\alpha = 0.05$ .

```
twobinom(11,23,10,23)
```

```
$p.value
[1] 0.7515369

$p1
[1] 0.4782609

$p2
[1] 0.4347826

$est.dif
[1] 0.04347826
```

**9.22**

Despite any problems it might have, summarize how you would justify using Student's T test to compare two independent groups.

*Student's T controls the Type I error probability when comparing identical distributions. It is possible that Student's T accurately detects a difference in the distributions that might be missed by other methods.*

**9.23**

Summarize any practical concerns about Student's T test and comment on how they might be addressed.

*In terms of understanding the nature of any difference that might exist, Student's T can be highly inaccurate in some situations. Confidence intervals for the difference between the means can be inaccurate when distributions differ. Problems with heteroscedasticity can be reduced by using a method that allows heteroscedasticity, such as Welch's test or a bootstrap-t method. But inaccurate confidence intervals, as well as relatively low power due to outliers, are always a concern when using means. Other methods might have substantially more power, such as techniques based on the median or 20% trimmed mean. Even with discrete distributions where outliers are rare, the methods in Section 9.11 might have more power. The only safe way to determine whether alternative methods make a difference is to try them.*

**9.24**

Summarize the relative merits of comparing groups with medians.

*The median better reflects the typical value for skewed distributions, it can have relatively high power when outliers are common. Accurate confidence intervals can be computed over a fairly broad range of situations, much broader than methods based on means. But for skewed distributions, it is possible that means have more power, even when the Type I error probability is controlled reasonably well, because the difference between the means might be larger. As noted in Chapter 2, there are situations where the mean can be argued to be a better measure of location. Also, when dealing with distributions that have light tails (outliers are relatively uncommon), the mean or some smaller amount of trimming can have a smaller standard error.*

**9.25**

For each of two independent groups, 1,000 bootstrap samples are generated and it is found that there are 10 instances where the bootstrap sample trimmed mean for the first group is less than the bootstrap sample trimmed mean for the second. There are two instances where the bootstrap sample trimmed means are equal. Determine a p-value when testing the hypothesis of equal population trimmed means.

$B = 1000$ ,  $A = 10$ ,  $C = 2$ ,  $Q = (A + 0.5C)/B = 0.01 + 0.5(0.002) = 0.011$ . Set  $P$  equal to  $Q$  or  $1 - Q$ , whichever is smaller, so  $P = 0.011$  and the p-value is  $2P = 2(0.011) = 0.022$

**9.26**

Repeat the analysis of Salk's data in section 9.1.6 using Welch's method.

```
salk <- data.table::fread(paste0(data.dir, "salk_dat.txt"), fill = T)
t.test(salk[,1],salk[,2])
```

Welch Two Sample t-test

```
data:  salk[, 1] and salk[, 2]
t = 3.5139, df = 51.639, p-value = 0.0009277
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 30.05544 110.11123
sample estimates:
mean of x mean of y
 18.00000 -52.08333
```

**9.27**

**9.28**

**9.29**

**9.30**