Homework #8

Derek Sonderegger

5.

In the `Lock5Data` package, the dataset `Wetsuit` describes an experiment with the goal of quantifying the effect of wearing a wetsuit on the speed of swimming. (It is often debated among triathletes whether or not to wear a wetsuit when it is optional.) A set of $n = 12$ swimmers and triathletes did a 1500 m swim in both the wetsuit and again in regular swimwear. The order in which they swam (wetsuit first or non-wetsuit first) was randomized for each participant. Reported is the maximum velocity during each swim.

```
data('Wetsuits', package='Lock5Data')
Wetsuits.Long <- Wetsuits %>%
  mutate(Participant = factor(1:12) ) %>%
  gather('Suit', 'MaxVelocity', Wetsuit,NoWetsuit) %>%
  arrange( Participant, Suit) %>%
  mutate(Suit = factor(Suit))
```

(a) Why did the researcher randomize which suit was worn first?

Assumibly the swimmers will be tired after their first 1.5 k swim and we would expect their second swim to be slower. So to account for that, we'll randomly assign the wetsuit to be either the first or second swim so that on average, those differences of first or second swim should balance out.

(b) Plot the velocities for the wetsuit and non-wetsuit for each participant. Comment on if you detect any difference in the means of these two distributions.

```
ggplot(Wetsuits.Long, aes(y=MaxVelocity)) +
  geom_point(aes(x=Suit, color=Suit)) +
  facet_wrap( ~ Participant ) +
  geom_line( aes(x=as.numeric(Suit)) )
```
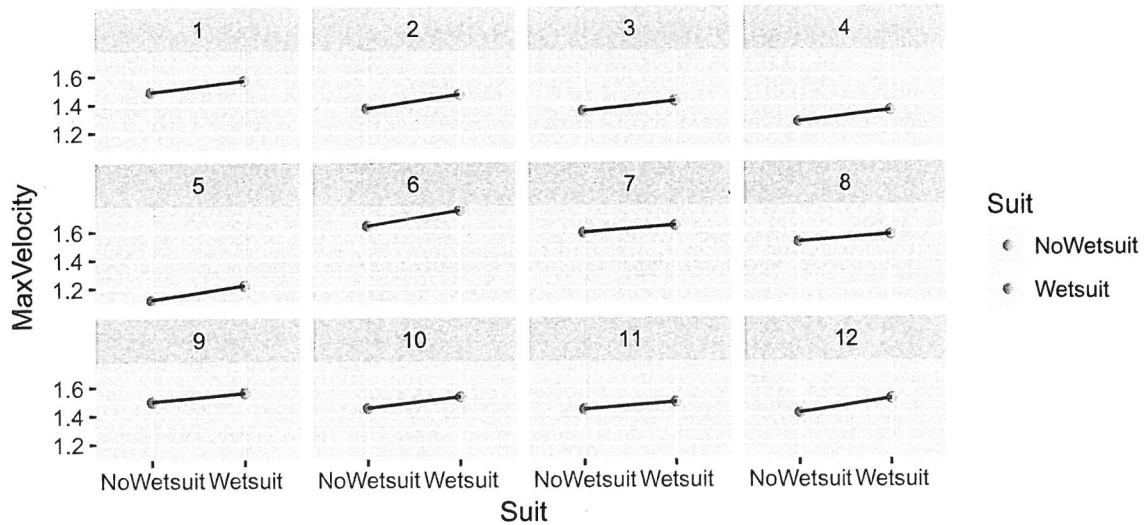
For every participant, they had a higher maximum velocity in the wetsuit than they did in the nonwetsuit. This certainly must be evidence to that the wetsuit will be faster.

The actual mean observed difference is

```
Wetsuits.Long %>%
  group_by(Participant)                %>%
  arrange(Suit)                        %>%   # NoWetsuit, Wetsuit
  summarise( d = diff(MaxVelocity)) %>%      # Wetsuit - NoWetsuit
  summarise( dbar = mean(d))

## # A tibble: 1 × 1
##     dbar
##    <dbl>
## 1 0.0775
```

(c) Ignore the pairing and do a two-sample t-test using the asymptotic method. What do you conclude?

```
t.test(MaxVelocity ~ Suit, data=Wetsuits.Long)

##
##  Welch Two Sample t-test
##
## data:  MaxVelocity by Suit
## t = -1.3688, df = 21.974, p-value = 0.1849
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -0.19492937  0.03992937
## sample estimates:
## mean in group NoWetsuit   mean in group Wetsuit
##                1.429167                1.506667
```

Ignoring the pairing, we see insufficient evidence to reject H_0 and we cannot conclude that the average maximum velocity is different between the wetsuit group and the non-wetsuit group. The reason for this result is that the observed increase is pretty small relative to the variability amongst all the swimmers. Without doing something clever to account for the swimmer-to-swimmer variabilty we don't have enough statistical power to detect a difference of this magnitude.

(d) Plot the difference in velocity for each swimmer. Comment on if the observed difference in velocity seems to indicate that which should be preferred (wetsuit or non-wetsuit).

```
Wetsuits.Diff <- Wetsuits.Long %>%
  group_by(Participant)  %>%
    arrange( Suit)  %>%               # NoWetsuit, Wetsuit
    summarise(d = diff(MaxVelocity))  # Wetsuit - NoWetsuit
```

```
ggplot(Wetsuits.Diff, aes(x=d)) + geom_histogram(binwidth=.005)
```
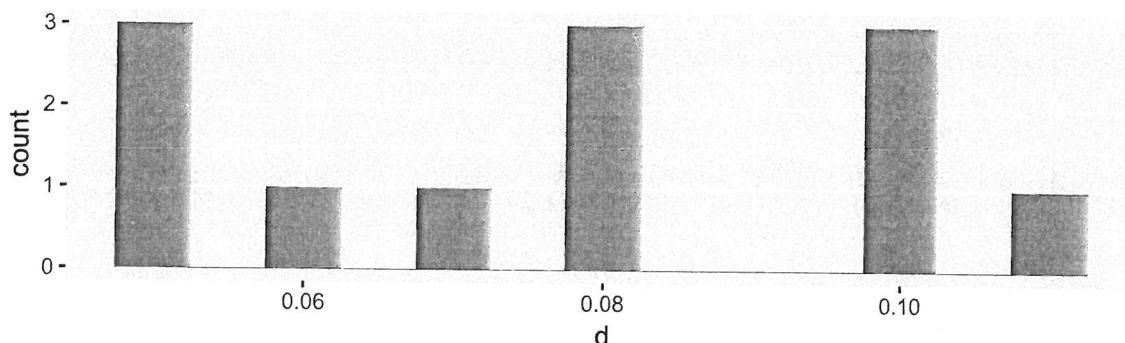

This view makes it clear that all of our differences (Wetsuit - No Wetsuit) are positive, implying that every single swimmer had a higher maximum velocity with the wetsuit than without it.

(e) Do a paired t-test using the asymptotic method. Also calculate the 95% confidence interval. What do you conclude?

```
t.test( Wetsuits.Diff$d )

##
##   One Sample t-test
##
## data:  Wetsuits.Diff$d
## t = 12.318, df = 11, p-value = 8.885e-08
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   0.06365244 0.09134756
## sample estimates:
## mean of x
##     0.0775
```

These data are highly significant (t=12.3, df=11, p=8.9e-8; 95% CI=(0.0637, 0.0913)) evidence to reject the null hypothesis of no difference in max velocity between the wetsuit and no-wetsuit groups and we conclude that the wetsuit does increase the maximum velocity.

(f) Do a paired t-test using the permutation method. Also calculate the 95% confidence interval using the bootstrap method. What do you conclude?

```
PermDist <- mosaic::do(10000)*{ Wetsuits.Long %>%
    group_by(Participant) %>%
    mutate( Suit = mosaic::shuffle(Suit) ) %>%
    arrange( Suit ) %>%                      # order is NoWetsuit, Wetsuit
    summarise( d = diff(MaxVelocity))  %>%  # d = Wetsuit - NoWetsuit
    summarise(dbar.star = mean(d)) }
BootDist <- mosaic::do(10000)*{ Wetsuits.Long %>%
    group_by(Participant) %>%
    arrange( Suit ) %>%                      # order is NoWetSuit, WetSeuit
    summarise( d = diff(MaxVelocity)) %>%  # d = Wetsuit - NoWetSuit
    mosaic::resample() %>%
```

```
        summarise(dbar.star = mean(d)) }
CI <- quantile(BootDist$dbar.star, probs=c(0.025, 0.975))
p.value <- 2 * mean( PermDist$dbar.star >= 0.0775 )
CI

##       2.5%      97.5%
## 0.06583333 0.08916667

p.value

## [1] 4e-04
```

Our p.value is quite small compared to any reasonable α so we conclude this is statistically significant evidence to conclude that the mean difference in maximum velocity is not zero, but rather that the wetsuit increases the maximum velocity of a swimmer. A 95% confidence for that increase is $(0.066, 0.089)$ meters per second.

To put this increase in context, the typical max velocity was about 1.5 m/sec, so this is about a 5% increase in max velocity. For a competetive swimmer, a 5% increase would be massive.

1.

To find the probabilities for the F-distribution, we will use the function `pf(f, df1, df2)` where the f is the value for which we want to find the probability of finding a value less than. That is $P(F_{2,10} < 4.2) =$
`pf(4.2, df1=2, df2=10)`

(a) Using the probability function for the F-distribution in R, find the following probabilities:

  i. $P\left(F_{5,5} < \frac{1}{2}\right)$
  `pf(1/2, df1=5, df2=5)`

  `## [1] 0.2325113`

  ii. $P\left(F_{5,5} > \frac{2}{1}\right)$
  `1 - pf(2/1, df1=5, df2=5)`

  `## [1] 0.2325113`

  iii. $P\left(F_{4,10} > \frac{6}{1}\right)$
  `1 - pf(6, df1=4, df2=10)`

  `## [1] 0.0099689`

  iv. $P\left(F_{10,4} < \frac{1}{6}\right)$
  `pf(1/6, df1=10, df2=4)`

  `## [1] 0.0099689`

(b) From what you calculated in problem 1, comment on the reciprocal symmetry of the F-distribution.

  The F-distribution is centered at 1 (ie the median is 1) and the area of probability from some point (call it a/b) to the tail (either towards 0 or towards ∞) is the same as for b/a so long as we switch the order of the degrees of freedom as well.

2.

In this exercise we will examine the variability of samples from various distributions and how easily departures from normality are detected using qqplots and the Shapiro-Wilks test. Under no circumstances should you turn in page after page of output or graphs. Produce a table that summarizes how often the test rejects the null hypotheses and include at most one figure of QQ-plots. To receive credit, you must comment on the table and graph and describe what you observe and why you observed what you did.

(a) The following code will create a random sample from a normal distribution and draw the qqplot. Also notice the results of the Shapiro-Wilks test. Investigate the behavior of repeated samples (ie run this code at least 10 times). Repeat with increased sample sizes (do this for n=5,25,100,400). Describe your results.

```
par(mfrow=c(1,2)) # 1 row of 2: graphs side-by-side
n <- 5                     # sample size is 5
x <- rnorm(n, mean=25, sd=5) # draw random sample
hist(x)                    # histogram
qqnorm(x)                  # qqplot for normality
qqline(x)                  # add a line to the above plot
shapiro.test(x)            # do the test for normality
```

For all values of n, the Shapiro-Wilks test rejected the null hypothesis very rarely, which is good because, in this case, the null hypothesis is true and the data are drawn from a normal distribution. So using an α level of 0.05, we should only reject the null hypothesis 5% of the time.

Most students executed the above code 10-20 times and saw 0,1, or 2 rejections, but to really understand how often we reject the null hypothesis, we should do this thousands of times. If we want to go a little crazy with R, we could write some looping code to look at how often we reject the null hypothesis for each sample size. However, this large number of replicate trick is purely optional.

```
# make a data frame for where my results will be placed
Results <- NULL

for(n in c(5,25,100,400)){
  for( i in 1:1000 ){
    x <- rnorm( n, mean=25, sd=5)              # generate data
    Results <- rbind(Results, data.frame(
        n = n,
        p.value = shapiro.test(x)$p.value))  # calculate a p.value
  }
}

Results %>%
  group_by(n) %>%
  summarize(proportion.reject = mean(p.value < 0.05))

## # A tibble: 4 × 2
##        n proportion.reject
##    <dbl>         <dbl>
## # 1     5         0.045
## # 2    25         0.054
## # 3   100         0.038
## # 4   400         0.053
```

Here the proportion.reject is the type I error rate, and so should be very close to $\alpha = 0.05$.

(b) Repeat problem (a) but consider samples drawn from a distribution that is not normal, in this case, the gamma distribution with parameters **shape=3**, and **rate=2**.

```
# make a data frame for where my results will be placed
Results <- NULL

for(n in c(5,25,100,400)){
  for( i in 1:1000 ){
    x <- rgamma( n, shape=3, rate=2)              # generate data
    Results <- rbind(Results, data.frame(
```

```
        n = n,
        p.value = shapiro.test(x)$p.value))   # calculate a p.value
  }
}

Results %>%
  group_by(n) %>%
  summarize(proportion.reject = mean(p.value < 0.05))

## # A tibble: 4 × 2
##        n proportion.reject
##    <dbl>            <dbl>
## 1      5            0.077
## 2     25            0.510
## 3    100            0.991
## 4    400            1.000
```

Here the proportion.reject is the power to detect the difference between a gamma distribution and a normal.

For small values of n, the Shapiro-Wilks test does not reject the null hypothesis of normality, but as our sample size increases, we more consistently reject H_0. This is exactly what we expect as our power to detect a difference increases with increasing sample sizes.

4

We are interested in testing if the variance is equal among two populations that are known to be normal. A sample of size $n_1 = 15$ from the first population resulted in a sample mean and standard deviation of $\bar{x}_1 = 52$ and $s_1 = 7$ while the sample of size $n_2 = 20$ from the second population had a sample mean and standard deviation of $\bar{x}_2 = 42$ and $s_2 = 4$. Perform an F-test with $\alpha = 0.05$ to test if the variances are different. Because the data is not given, all calculations must be done by-hand (except the usual probability look up).

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

The appropriate test statistic is a f-statistic

$$f_{n_1-1,n_2-1} = \frac{s_1^2}{s_2^2} = \frac{7^2}{4^2} = \frac{49}{16} = 3.06$$

Because this value is larger than 1, the p-value is the 2 times the probability of observing 3.06 or larger.

$$p - value = 2 \cdot P\left(F_{14,19} > 3.06\right)$$

```
2* (1-pf(3.06, df1=14, df2=19))
```

```
## [1] 0.02490965
```

Given this small p-value we can reject the null hypothesis and conclude that the group variances are not equal.

5.

The life span of an electrical component was studied under two operating voltages (110 and 220). Ten components were randomly assigned to operate at 110 volts and 16 were assigned to 220 volts. The time to failure (in hundreds of hours) for the 26 components were obtained:

110	19.25	19.7	19.75	19.9	19.95	20.05	20.13	20.2	20.4	20.6
220	9.7	9.75	9.8	9.82	9.85	9.90	9.92	9.96		
	10.01	10.02	10.10	10.11	10.13	10.19	10.28	10.31		

(a) Calculate the mean and variance of each sample group

```
LifeSpan <- data.frame(
  life = c(19.25,19.70,19.75,19.90,19.95,20.05,20.13,20.20, 20.4, 20.6,
           9.7,  9.75, 9.80, 9.82, 9.85, 9.90, 9.92, 9.96,
           10.01,10.02,10.10,10.11,10.13,10.19,10.28,10.31),
  voltage = factor(c( rep(110, 10), rep(220, 16) )) )

LifeSpan %>%
  group_by(voltage) %>%
  summarize(Mean = mean(life),
            Var  = var(life) )

## # A tibble: 2 × 3
##   voltage      Mean        Var
##   <fctr>      <dbl>      <dbl>
## 1     110 19.993000 0.14515667
## 2     220  9.990625 0.03427292
```
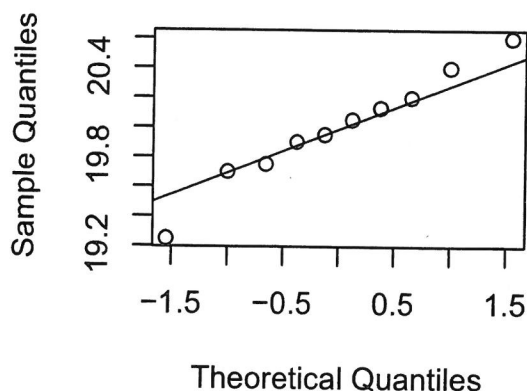
(b) Test the assumption that the data in each group is normally distributed.

(i) Create the QQplots first and comment on their fit.
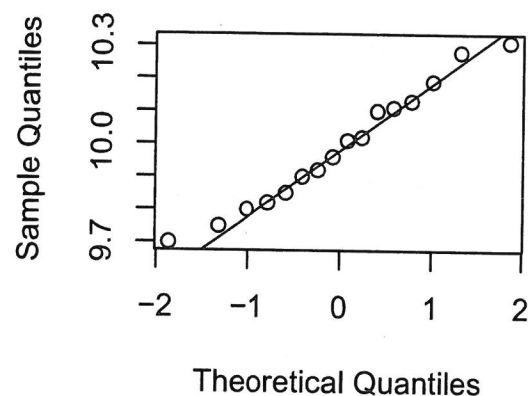
```
Volt110 <- LifeSpan %>% filter(voltage == 110)
Volt220 <- LifeSpan %>% filter(voltage == 220)

par(mfrow=c(1,2)) # 1 row, 2 columns of graphs
qqnorm( Volt110$life ); qqline( Volt110$life )
qqnorm( Volt220$life ); qqline( Volt220$life )
```

The 110 group looks a bit odd with the smallest value being far from the normality line, but with only 10 observations, it is hard to really say it isn't normal. The 220 group looks good, with perhaps only the smallest value substantially off of the normality line.

(ii) Perform the Shapiro-Wilks test to assess normality.

```
shapiro.test( Volt110$life )
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Volt110$life
## W = 0.98348, p-value = 0.981
```

```
shapiro.test( Volt220$life )
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Volt220$life
## W = 0.96927, p-value = 0.8267
```

The Shapiro-Wilks test also fails to detect non-normality in the data, possibly due to the small sample sizes.

(c) Test the assumption that the variances in each group are equal

   i) By hand, perform a two-side hypothesis test that variances in each group are equal. (Here, "by hand" means to calculate the f-statistic by hand and then form the probability statement that defines the p-value. Then use the pf() function to calculate the actual p-value.)

$$f = \frac{s_{110}^2}{s_{220}^2} = \frac{0.145}{0.0343} = 4.23$$

And therefore the p-value is $2 \cdot P\left(F_{9,15} > 4.23\right) =$

```
2 * (1 - pf( 4.23, 9,15 ))
```

```
## [1] 0.01382366
```

Given such a small p-value, we can reject the null hypothesis of equal variance.

   ii) Using the R function var.test() confirm your calculations in part (ii).

```
var.test(life ~ voltage, data=LifeSpan)
```

```
##
##  F test to compare two variances
##
## data:  life by voltage
## F = 4.2353, num df = 9, denom df = 15, p-value = 0.01374
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##    1.356295 15.964423
## sample estimates:
## ratio of variances
##            4.235317
```