Student names:

Jurriën Boogert
Snr. 2085383
Anr. 872949

Giorgos Vermaak
Snr. 2108346
Anr. 482782

Ka Wai Tang
Snr. 2064845
Anr. 870245

Exercise 1.

a. Null hypothesis states there's no difference between average suicide rates in 2011 and 2005 or the suicide rate in 2011 is smaller than the suicide rate in 2005. H_0: $\mu_{2011} \leq \mu_{2005}$

Alternative hypothesis states that the average suicide rate in 2011 is larger than in 2005. H_1: $\mu_{2011} > \mu_{2005}$

You could also reformulate the above hypothesis as difference in means so the null hypothesis would be that the difference between mean suicide rates from 2011 and 2005 is zero or smaller. The alternative hypothesis is that the difference between mean suicide rates from 2011 and 2005 is larger than zero:

H_0: $\mu_{2011} - \mu_{2005} \leq 0$
H_1: $\mu_{2011} - \mu_{2005} > 0$

b. Paired t-tests are used when the same item or group is tested twice or are related, which is the case here. The veterans are tested in 2005 and in 2011 to see how many of the population of veterans committed suicide. Every paired-samples t-test can be reduced to a one-sample t-test. Compute a new variable, a difference score. Then test the null hypothesis using this new variable, the difference in means. So a one-sample t-test is suitable as well. An independent-samples t-test compares the averages/means of two independent or unrelated groups to determine if there is a significant difference between the two. Therefore the independent-samples t-test is not suitable in this case, because the groups are not independent, but related. A z-test is not suitable, because the population variance per state is not known.

c. Figure 1 shows the histogram and Figure 2 the qq-plot of the variable vet-rate for the years 2005 and 2011 and their difference. Looking at the figures the data is not unambiguously normally distributed, but has the tendency to be normal only in the center. The values in the tails are especially deviating from the normal distribution, it has

some large outliers. Also the sample size is small, so with large outliers and a small sample size it is plausible it is not normally distributed. When running the Shapiro-Wilk test on the separate groups in R the results suggest the data is not significantly different from normal, but if we run the test on the difference in means between the two groups the test suggests the distribution is significantly different from Normal with p-value = 0.027. We therefore think the assumption of normality does not hold, because the groups are paired (so we look at the difference in means), the distribution has outliers (seen in qq-plot and histogram of difference in means).

$$\overline{X}_{vet\_rate,\ 2005} = 30.27 \qquad \overline{X}_{vet\_rate,\ 2011} = 35.69$$

The mean suicide rate for the veterans in 2005 was 30.27, while the rate in 2011 was 5.42 higher than in 2005. The suicide rate in 2011 was 35.69 per 100.000 veterans. We expect a significant difference between mean suicide rates between 2005 and 2011, because the difference in means, 5.42, is large in comparison with the null hypothesis of zero difference in means, but when the distribution is not normal, inferences are not to be trusted with full confidence.

Histogram of 2005 veteran suicides

Histogram of 2011 veteran suicides
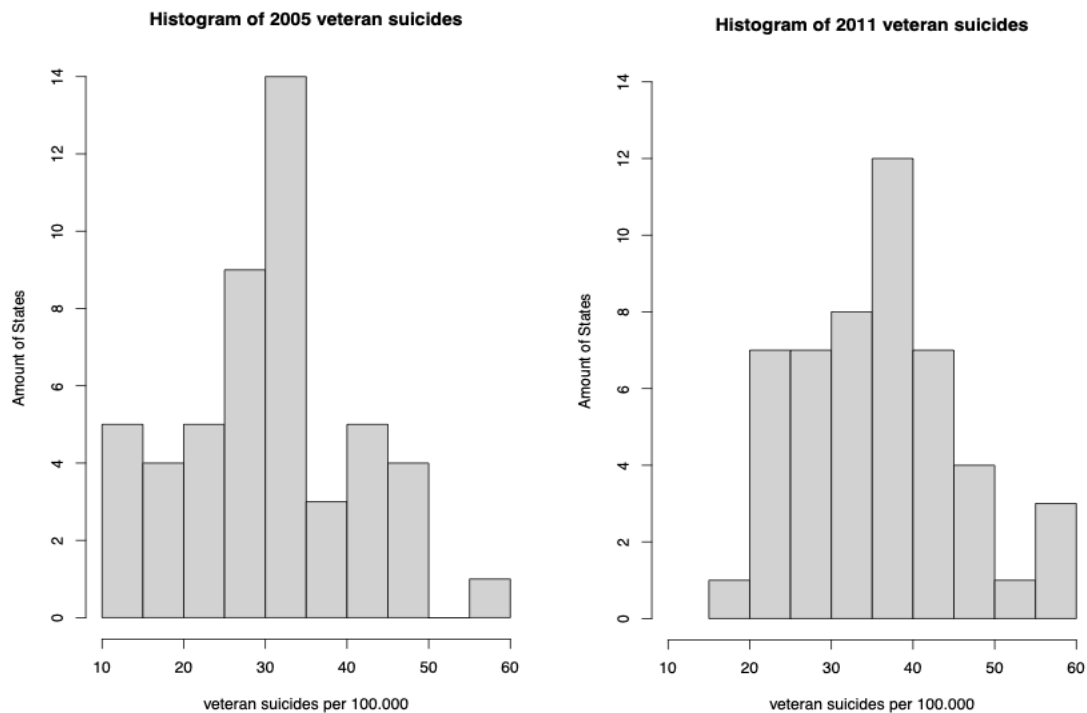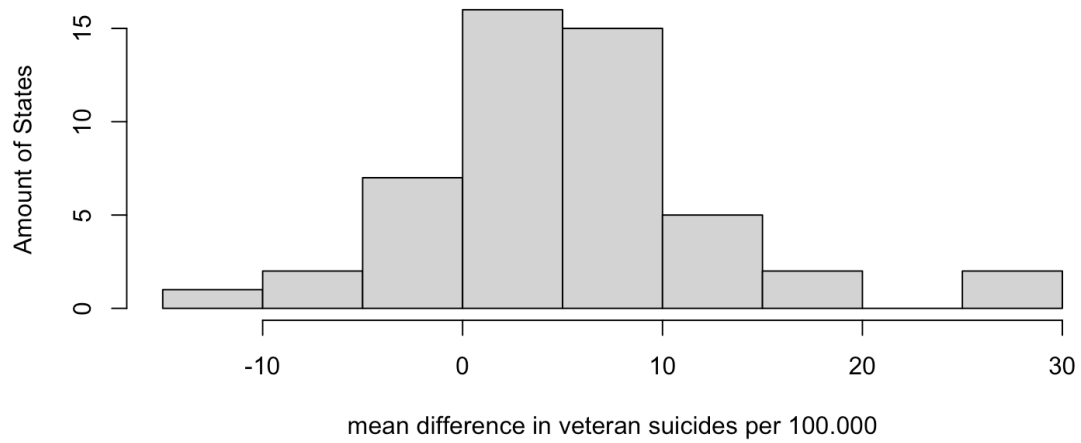
Histogram of the difference in veteran suicides between 2005-2011
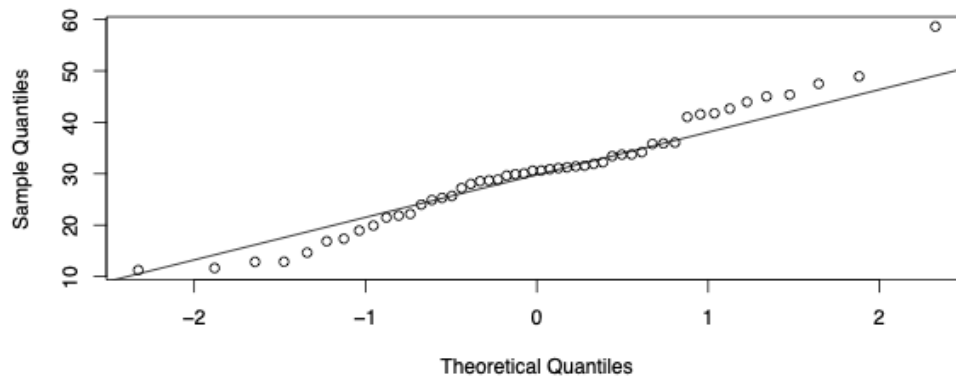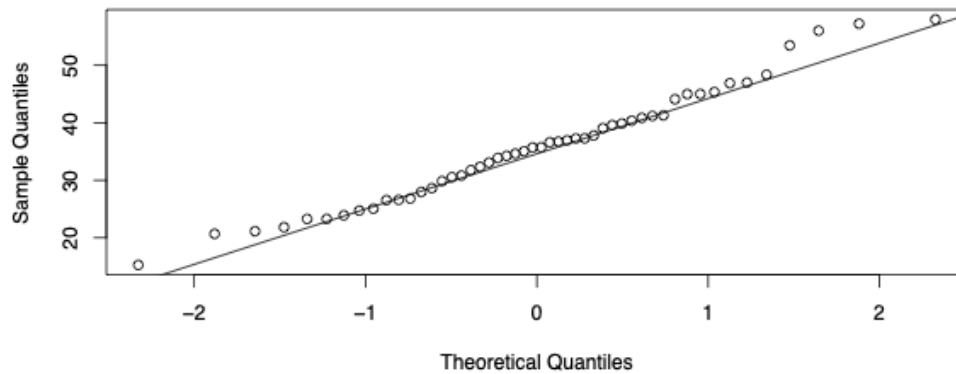

Figure 1. Histogram of the suicide rates for the years 2005 and 2011 and their difference

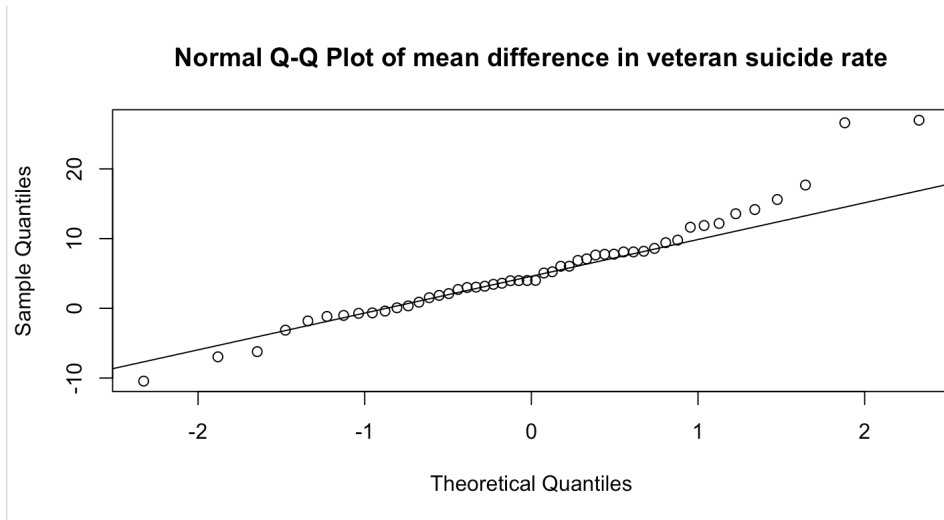**Normal Q-Q Plot of mean difference in veteran suicide rate**

Figure 2. QQ-plots of the actual quantiles of the distribution of the data plotted against the theoretical quantiles of a normal distribution

d. Results of a paired t-test and the one samples t-test indicate a significant increase in veteran suicides in 2011 ($\overline{X}$ = 35.69, SD = 9.91) in relation to 2005 ($\overline{X}$ = 30.27, SD = 10.36), $t(49)$ = 5.33, p = 1.25e-06. At the $\alpha = 0.05$ level of significance, the H_0 that the difference in means between suicide rate of 2005 and 2011 is the same or that the rate for 2011 is smaller than for 2005, is rejected in favor of H_1, that there is a significant increase in suicides in 2011 compared to 2005, because $p < 0.05$.

e. Figure 3. Shows the histogram of bootstrap differences in means with the dashed red line indication the difference between the two sample means and the blue line indication the mean of the bootstrap distribution

   The shape of the bootstrap distribution looks normal based on face value.

Bootstrap distribution of the difference in mean suicide rate 2011 vs 2005
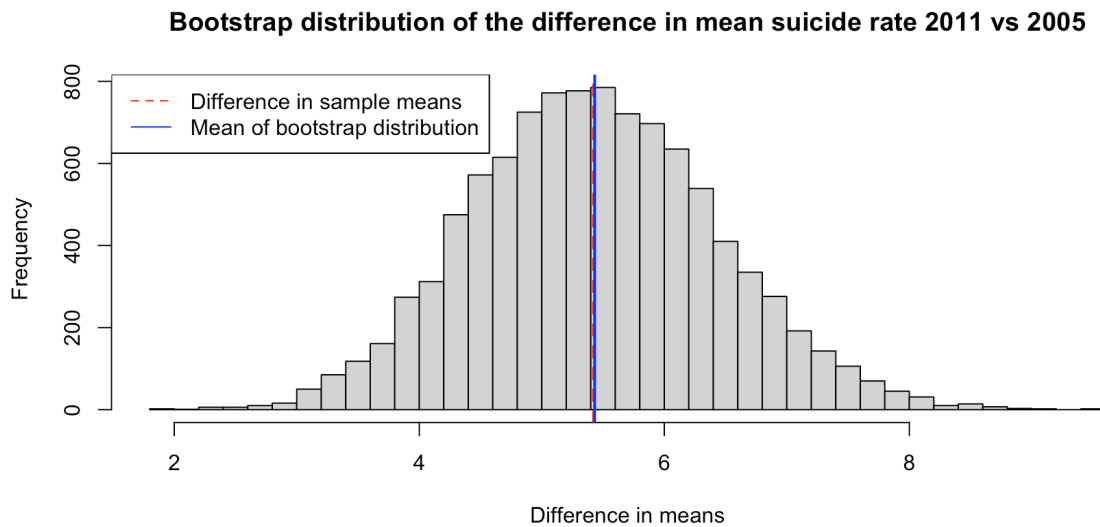
Figure 3. Shows the bootstrap distribution of the differences in sample means. The blue and red lines represent the mean of the bootstrap distribution and difference in sample means respectively.

$$\overline{X}_{\text{bootstrap, difference in means}} = 5.40$$

$$\overline{X}_{\text{vet\_rate, 2011}} - \overline{X}_{\text{vet\_rate, 2005}} = 35.69 - 30.27 = 5.42$$

Looking at figure 3 we see the mean of the bootstrap distribution and difference in sample means are very close. The bias is $5.40 - 5.42 = -0.02$, which is very small in comparison with the values of the data, so we can say the mean of our bootstrap distribution is unbiased.

The 95% bootstrap percentile interval with lower bound of 3.79 does closely match the outcome of the 95% bootstrap t confidence interval with lower bound 3.76.

Percentile interval: [3.79;Inf]
T confidence interval: [3.76;Inf]

The bias of the bootstrap distribution is small and the distribution is close to Normal, therefore the bootstrap t and percentile confidence intervals will agree closely, which they do. Because zero is not contained in the intervals the null hypothesis of no difference between groups or that the suicide rate in 2005 was higher than in 2011 can be rejected in favor of the alternative hypothesis that the suicide rate in 2011 was higher than in 2005.

f. With a permutation test we tested the assumption that -on average- the suicide rate in 2011 is larger than in 2005. The resulting permutation distribution is shown in figure 4. The permutation test estimate of the P-value is 0.0001.

Permutation distribution of the difference in mean suicide rate 2011 vs 2005
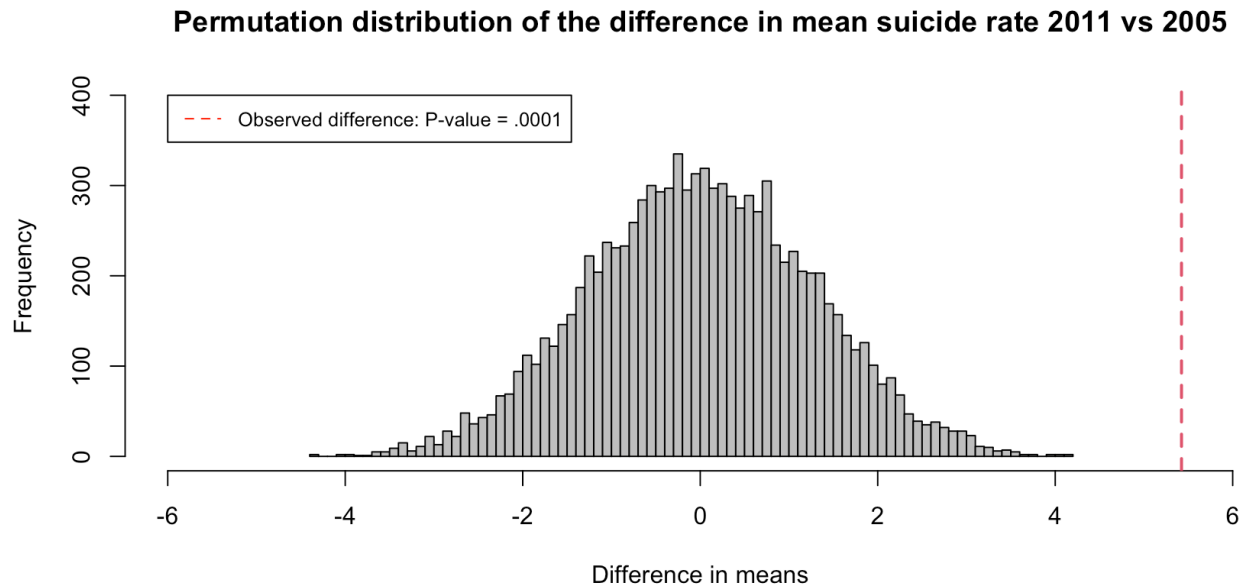
Figure 4. Permutation distribution of the two samples from 2005 and 2011 suicide rates.

g. The observed difference in mean suicide rate is larger than expected under the null hypothesis where we assume no difference. The probability of observing a difference of 5.42 or larger while the null hypothesis is true is 0.0001, which is smaller than the cut-off value of 0.05, therefore we reject the null hypothesis in favor of the alternative hypothesis that the average suicide rate in 2011 is larger than in 2005 based on the resulting p-value for the bootstrap and permutation test.

Exercise 2.

1. The null hypothesis states that there is no difference in means between the two groups n_1 and n_2:

   H_0: $\mu_{n1} = \mu_{n2} \Rightarrow \mu_{n1} - \mu_{n2} = 0$

   The alternative hypothesis states that there is a significant difference in means between the two groups n_1 and n_2:

   H_1: $\mu_{n1} \neq \mu_{n2} \Rightarrow \mu_{n1} - \mu_{n2} \neq 0$

   For the permutation test the formulation of the null hypothesis is that two different groups come from the same distribution. The alternative hypothesis is that the two groups come from a different distribution.

   H_0: $Distribution_{n1} = Distribution_{n2} \Rightarrow Distribution_{n1} - Distribution_{n2} = 0$
   H_1: $Distribution_{n1} \neq Distribution_{n2} \Rightarrow Distribution_{n1} - Distribution_{n2} \neq 0$

2. Tables 1.1 to 1.4 show the results of the four different testing procedures. For the Welch's t-test the rejection rate is for every condition under the null hypothesis controlled for. All values are closely around $\alpha = 0.05$. The power for the smaller sample size of $n_1 = 10$ is smaller than the power for the larger sample sizes (n_1=100,n_1=200). The power is somewhat lower for the largest variance groups ($\sigma^2_2 = 10$).

   For the Student's t-test the only conditions for which the rejection rate is controlled for are the ones where n_1=100 (same as n_2). When the sample size of $n_1 = 10$, the smaller the variance of group2 the higher the rejection rate. When the sample size of $n_1 = 200$, the smaller the variance of group2 the lower the rejection rate. For all conditions under the null hypothesis where the variance of the two groups are equal the rejection rate is around $\alpha = 0.05$ and thus controlled for. The power of the Student's t-test for the smaller sample size of $n_1 = 10$ is smaller than the power for the larger sample sizes (n_1=100,n_1=200). The power is somewhat lower for the largest variance groups ($\sigma^2_2 = 10$) and especially low for the smallest sample size with the largest variance.

   For the Permutation test we see that the $\alpha = 0.05$ is controlled for for sample size n_1=10 and $\sigma^2_2 = 2$, and sample size n_1=100 and $\sigma^2_2 = 10$. However, the outcome may be influenced due to some random variation, because the differences are small. In the case of sample size n_1=10 and $\sigma^2_2 = 1$, and sample size n_1=200 and $\sigma^2_2 = 10$, the rejection rates are too high. In all the other cases, the rejection rates are too low. This doesn't make the permutation test a good test for controlling the alpha level/Type-1 error when

sample sizes are different and variances are different. Looking at the power, we see that there is a high power for sample sizes of n_1=100 and n_1=200. Even though the power decreases for larger variances, they are still fairly high, assuming a minimum power of 0.8. For a sample size of 10, the power levels are not as high, even below 0.8 for all variances, and in the case of σ^2_2 = 10, the power is only 0.034. So, a larger sample size will give a higher power.

For the Non-parametric Bootstrap procedure the rejection rate is inflated for a smaller sample size of n_1=10 for all variance levels. For n_1=100 and n_1=200 the rejection rate is closer to $\alpha = 0.05$. The only case where the rejection rate is underestimated is for the smallest variance of σ^2_2 = 1. For the power almost the same applies as for Welch's t-test. The power for the smaller sample size of n_1 = 10 is smaller than the power for the larger sample sizes (n_1=100,n_1=200). The power is somewhat lower for the largest variance groups (σ^2_2 = 10).

Conclusion
The permutation test seems to perform well when the sample sizes of both groups are the same and/or when the variances of both groups are the same. In the case when the sample size of the first group is small the rejection rate seems inflated for a small variance and too low for a larger variance than the second group. When the sample size for the first group is larger than the sample size for the second group the inverse holds. The power is highly variable when the sample size of one group is small and is very low when the variance is high. When the group means are truly different, when the alternative hypothesis is true, the power of all tests is high for larger sample size (n = +100).

Our advice would be: only use the Permutation test when you assume equal distributions, have a high sample size (+100) or when you know variances are the same for both tested groups. The best performing test is in all cases is Welch's t-test and this would be the preference to use, but the power is not as high for the smaller sample size as for the groups with higher sample sizes. The Student's t-test assumes equal variances so performs well when the variances are indeed equal for both groups. It actually has higher power when variances of both groups are equal and is the preferred method to choose, als when sample sizes are different. When sample sizes are the same for both groups it doesn't make a difference if variances are different for both groups for the rejection rate, they are all around $\alpha = 0.05$. As an extra remark about power: When the rejection rate under the null hypothesis is not controlled for then there is no technically valid reason to talk about the power.

Welch's t-test					
Type I error (rejection rate)			Power $(1 - \beta)$		
$\mu_1 = \mu_2 = 0$			$\mu_1 = 0; \mu_2 = 1$		
Sample size (n_1)	Variance (σ^2_2)	Rejection rate	Sample size (n_1)	Variance (σ^2_2)	Rejection rate
10	1	0.050	10	1	0.502
	2	0.049		2	0.496
	10	0.045		10	0.418
100	1	0.052	100	1	0.9999
	2	0.051		2	0.9986
	10	0.050		10	0.820
200	1	0.050	200	1	1
	2	0.048		2	0.9998
	10	0.052		10	0.849

Table 1.1 Rejection rates of the null hypothesis obtained for the Welch's test for data generated 1) under the null hypothesis and 2) under the alternative hypothesis, for each of the nine conditions

Student's T-test					
Type I error (rejection rate)			Power (1 - β)		
$\mu_1 = \mu_2 = 0$			$\mu_1 = 0; \mu_2 = 1$		
Sample size (n_1)	Variance (σ^2_2)	Rejection rate	Sample size (n_1)	Variance (σ^2_2)	Rejection rate
10	1	0.139	10	1	0.756
	2	0.049		2	0.569
	10	0		10	0.038
100	1	0.052	100	1	0.9999
	2	0.051		2	0.9986
	10	0.050		10	0.822
200	1	0.028	200	1	1
	2	0.048		2	0.9998
	10	0.119		10	0.928

Table 1.2 Rejection rates of the null hypothesis obtained for the Student's t-test test for data generated 1) under the null hypothesis and 2) under the alternative hypothesis, for each of the nine conditions

Permutation test					
Type I error (rejection rate)			Power (1 - β)		
$\mu_1 = \mu_2 = 0$			$\mu_1 = 0; \mu_2 = 1$		
Sample size (n_1)	Variance (σ^2_2)	Rejection rate	Sample size (n_1)	Variance (σ^2_2)	Rejection rate
10	1	0.128	10	1	0.747
	2	0.051		2	0.565
	10	0.000		10	0.034
100	1	0.041	100	1	1
	2	0.039		2	0.998
	10	0.051		10	0.796
200	1	0.028	200	1	1
	2	0.039		2	0.999
	10	0.133		10	0.907

Table 1.3 Rejection rates of the null hypothesis obtained for the Permutation test for data generated 1) under the null hypothesis and 2) under the alternative hypothesis, for each of the nine conditions

Non-Parametric Bootstrap test					
Type I error (rejection rate)			Power (1 - β)		
$\mu_1 = \mu_2 = 0$			$\mu_1 = 0; \mu_2 = 1$		
Sample size (n_1)	Variance (σ^2_2)	Rejection rate	Sample size (n_1)	Variance (σ^2_2)	Rejection rate
10	1	0.086	10	1	0.638
	2	0.096		2	0.614
	10	0.067		10	0.474
100	1	0.032	100	1	1
	2	0.048		2	0.994
	10	0.055		10	0.818
200	1	0.046	200	1	1
	2	0.056		2	1
	10	0.045		10	0.851

Table 1.4 Rejection rates of the null hypothesis obtained for the Non-Parametric Bootstrap test for data generated 1) under the null hypothesis and 2) under the alternative hypothesis, for each of the nine condition

R code

Exercise 1.

```
setwd("Documents/DATA_SCIENCE_AND_SOCIETY/Computational
Statistics/assignments/assignment2")
library(readr)

# Load data
Vet_2005 <- read_csv("VetSuicides_2005.csv")
Vet_2011 <- read_csv("VetSuicides_2011.csv")

# Plots histograms
hist(Vet_2005$vet_rate,xlab="veteran suicides per 100.000",ylab="Amount of States",
main="Histogram of 2005 veteran suicides", xlim = c(10,60), ylim = c(0,14))
hist(Vet_2011$vet_rate,xlab="veteran suicides per 100.000",ylab="Amount of
States",main="Histogram of 2011 veteran suicides", xlim = c(10,60), ylim = c(0,14))

# Plot difference in means
diff_mean <- Vet_2011$vet_rate - Vet_2005$vet_rate
hist(diff_mean,xlab="veteran suicides per 100.000",ylab="Amount of
States",main="Histogram of the difference in veteran suicides between 2005-2011")

# Shapiro Wilk tests
shapiro.test(Vet_2005$vet_rate)
shapiro.test(Vet_2011$vet_rate)
shapiro.test(Vet_2011$vet_rate - Vet_2005$vet_rate)

# QQplots
qqnorm(Vet_2005$vet_rate, main="Normal Q-Q Plot of 2005 veteran suicide rate");
qqline(Vet_2005$vet_rate)
qqnorm(Vet_2011$vet_rate, main="Normal Q-Q Plot of 2011 veteran suicide rate");
qqline(Vet_2011$vet_rate)
qqnorm(diff_mean, main="Normal Q-Q Plot of difference in veteran suicide rate between
2011 and 2005"); qqline(diff_mean)

# Means of vet rates
mean(Vet_2005$vet_rate)
mean(Vet_2011$vet_rate)

# Difference in mean vet rates
mean(Vet_2011$vet_rate) - mean(Vet_2005$vet_rate)

# SD for 2005 and 2011
```

```r
sd(Vet_2005$vet_rate)
sd(Vet_2011$vet_rate)

# t.tests (student and welch)
t.test(Vet_2011$vet_rate,Vet_2005$vet_rate, alternative = "greater", var.equal = FALSE,
paired = TRUE)
t.test(Vet_2011$vet_rate - Vet_2005$vet_rate, alternative = "greater")



#NON-PARAMETRIC BOOTSTRAP PROCEDURE FOR PAIRED SAMPLES
set.seed(420)
B <- 10000000
paired_boot_vet <- function(B,data1,data2){
  diff_rates <- vector(mode = "numeric", length = B)
  diff_mean <- Vet_2011$vet_rate - Vet_2005$vet_rate
  for (i in 1:B){
    paired_shuffle <- sample(diff_mean, 50, replace = TRUE)

    diff_rates[i] <- mean(paired_shuffle)
  }
  return(diff_rates)
}

diff_rates <- paired_boot_vet(10000,Vet_2005$vet_rate,Vet_2011$vet_rate)

#plotting bootstrap distribution
hist(diff_rates,breaks=30,xlab="Difference in means",ylab="Frequency",
main="Bootstrap distribution of the difference in mean suicide rate 2011 vs 2005")
abline(v = mean(diff_mean), col="red", lwd=2, lty=2)
abline(v = mean(diff_rates), col="blue", lwd=2, lty=1)
legend("topleft", legend = c("Difference in sample means","Mean of bootstrap
distribution"),
    lty = c('dashed',"solid"),
    col = c("red","blue"))

#bootstrap difference in means
mean(diff_rates)

#bias bootstrap distribution
mean(diff_rates) - (mean(Vet_2011$vet_rate) - mean(Vet_2005$vet_rate))

#95% bootstrap percentile interval
quantile(diff_rates, 0.05)
```

```r
#95% bootstrap t confidence interval
t <- qt(0.05,9999, lower.tail = FALSE)

"
#upperbound
mean(diff_rates) + t*sd(diff_rates)
"

#lowerbound
mean(diff_rates) - t*sd(diff_rates)

#PERMUTATION PROCEDURE V1
set.seed(420)
B <- 9999

diff <- Vet_2011$vet_rate - Vet_2005$vet_rate
diff_perm_means <- vector(mode = "numeric", length=B)
for (i in 1:B){
  scale <- sample(c(-1,1),50,replace = TRUE)
  diff_perm_means[i] <- mean(diff * scale)
}

#plotting permutation distribution
a=hist(diff_perm_means,100,xlim = c(-6,6),ylim = c(0,400))
barcols <- a$breaks

barcols[a$breaks<round(mean(diff_mean),0)] = "grey"
barcols[a$breaks>=round(mean(diff_mean),0)] = 2

a <- hist(diff_perm_means,100,xlim = c(-6,6),ylim = c(0,400),col=barcols,xlab="Difference
in means",ylab="Frequency", main="Permutation distribution of the difference in mean
suicide rate 2011 vs 2005")

abline(v=mean(diff_mean), col =2, lwd = 2, lty=2)
legend(-6.000, 400, legend = "Observed difference: P-value = .0001",
    cex = .8,
    lty = 'dashed',
    col = "red")
'legend(-6.0, 345, legend = "0.05% Significance level",
    fill = "red",
    cex = .8,
    col = "red")
'
```

```
#calculate p-value
(1+sum(diff_perm_means>(mean(diff_mean))))/(B+1)
```

Exercise 2

```
# MONTE CARLO SIMULATION OF PERMUTATION PROCEDURE
set.seed(420)
S <- 1000
B <- 9999

n1 <- c(10,100,200)
mu1 <- 0
var1 <- 2

n2 <- 100
mu2 <- c(0,1)
var2 <- c(1,2,10)

diff_perm_means <- vector(mode = 'numeric', length = B)

permutation_test <- function(){
  for (i in mu2){
    for (j in n1){
      for (k in var2){
        rejection <- 0
        for (l in 1:S){
          group1 <- rnorm(n = j, mean = mu1, sd = sqrt(var1))
          group2 <- rnorm(n = n2, mean = i, sd = sqrt(k))
          observed_stat <- mean(group2) - mean(group1)
          for (m in 1:B){
            shuffle <- sample(c(group1,group2), size = j+100, replace = FALSE)
            perm1 <- shuffle[1:j]
            perm2 <- shuffle[(1+j):(j+100)]
            diff_perm_means[m] <<- mean(perm2) - mean(perm1)
          }
          if (((1+sum(diff_perm_means > observed_stat)) / (B+1)) <= 0.025){
            rejection <- rejection + 1
          }
          if (((1+sum(diff_perm_means < observed_stat)) / (B+1)) <= 0.025){
            rejection <- rejection + 1
          }
```

```r
        #print(paste0('MC_sample:',counter <- counter + 1))
      }
      print(paste0('n1:',j, ' mu2:',i, ' var2:',k, '    ', rejection/S))
    }
  }
 }
}
permutation_test()

#----------------------------------------------------------------------------
# MONTE CARLO SIMULATION OF WELCH T-TEST
set.seed(420)
S <- 10000

n1 <- c(10,100,200)
mu1 <- 0
var1 <- 2

n2 <- 100
mu2 <- c(0,1)
var2 <- c(1,2,10)


welch_test <- function(){
  for (i in mu2){
    for (j in n1){
      for (k in var2){
        counter <- 0
        for (l in 1:S){
          group1 <- rnorm(n = j, mean = mu1, sd = sqrt(var1))
          group2 <- rnorm(n = n2, mean = i, sd = sqrt(k))
          t <- t.test(group2,group1)
          if (t$p.value < 0.05){
            counter <- counter + 1
          }
        }
        print(paste0('n1:',j, ' mu2:',i, ' var2:',k, '    ', counter/S))
      }
    }
  }
}
welch_test()

#----------------------------------------------------------------------------
```

```
# MONTE CARLO SIMULATION OF STUDENT-T TEST
set.seed(420)
S <- 10000

n1 <- c(10,100,200)
mu1 <- 0
var1 <- 2

n2 <- 100
mu2 <- c(0,1)
var2 <- c(1,2,10)

studentsT_test <- function(){
  for (i in mu2){
    for (j in n1){
      for (k in var2){
        counter <- 0
        for (l in 1:S){
          group1 <- rnorm(n = j, mean = mu1, sd = sqrt(var1))
          group2 <- rnorm(n = n2, mean = i, sd = sqrt(k))
          t <- t.test(group2,group1, var.equal = TRUE)
          if (t$p.value < 0.05){
            counter <- counter + 1
          }
        }
        #rejection_rates[paste0('n1:',j, ' mu2:',i, ' var2:',k)] <- counter / S
        print(paste0('n1:',j, ' mu2:',i, ' var2:',k, '    ', counter/S))
      }
    }
  }
}
studentsT_test()

#--------------------------------------------------------------------------------
# MONTE CARLO SIMULATION OF NON-PARAMETRIC BOOTSTRAP PROCEDURE
set.seed(420)
S <- 1000
B <- 10000

n1 <- c(10,100,200)
mu1 <- 0
var1 <- 2

n2 <- 100
```

```r
mu2 <- c(0,1)
var2 <- c(1,2,10)

diff_boot_means <- vector(mode = 'numeric', length = B)

bootstrap_test <- function(){
  for (i in mu2){
    for (j in n1){
      for (k in var2){
        rejection <- 0
        for (l in 1:S){
          group1 <- rnorm(n = j, mean = mu1, sd = sqrt(var1))
          group2 <- rnorm(n = n2, mean = i, sd = sqrt(k))
          for (m in 1:B){
            boot1 <- sample(group1,j,replace = TRUE)
            boot2 <- sample(group2,100,replace = TRUE)
            diff_boot_means[m] <<- mean(boot2) - mean(boot1)
          }

          low <- quantile(diff_boot_means, 0.025)[[1]]
          high <- quantile(diff_boot_means, 0.975)[[1]]
          if (low > 0){
              rejection <- rejection + 1
          }
          if (high < 0){
            rejection <- rejection + 1
          }
        }
        print(paste0('n1:',j, ' mu2:',i, ' var2:',k, '    ', rejection/S))
      }
    }
  }
}
bootstrap_test()
```