Problem Set 6

Michael Ghattas

Introduction

Collaboration

(1 point)

Other students who I worked with on this assignment (if any) : None.

Notes

These questions were rendered in R markdown through RStudio (https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf, http://rmarkdown.rstudio.com). Please complete the following tasks regarding the data in R. Please generate a solution document in R markdown and upload the .Rmd document and a rendered .pdf document. Your solution document should have your answers to the questions and should display the requested plots.

Question 1

Context

The data set "USSeatBelts", data for the years 1983–1997 from 50 US States, plus the District of Columbia, for assessing traffic fatalities and seat belt usage, is in the "AER" package. Further details are available in the help for "USSeatBelts". These questions use the "state", "year", "fatalities", and "drinkage" variables. As detailed in the documentation, "fatalities" is the number of fatalities per million traffic miles and "drinkage" is a binary variable that is "yes" if the state had a minimum drinking age of 21 years and "no" otherwise. As can be seen from the tabulation below, by 1988, all the jurisdictions adopted a minimum drinking age of 21 years.

```
data("USSeatBelts")
table(USSeatBelts$year,USSeatBelts$drinkage)
```

```
##
##        no yes
##   1983 32  19
##   1984 29  22
##   1985 19  32
##   1986  7  44
##   1987  1  50
##   1988  0  51
##   1989  0  51
```

```
##     1990  0  51
##     1991  0  51
##     1992  0  51
##     1993  0  51
##     1994  0  51
##     1995  0  51
##     1996  0  51
##     1997  0  51
```

The data can be reformatted as shown to have columns for each year's values of "fatalities" and "drinkage".

```
dat<-USSeatBelts
dat<-pivot_wider(dat,id_cols=state,names_from = year,values_from = c(fatalities,drinkage))
```

The parts of this question explore the relationship between fatalities per million traffic miles and the drinking age in the state.
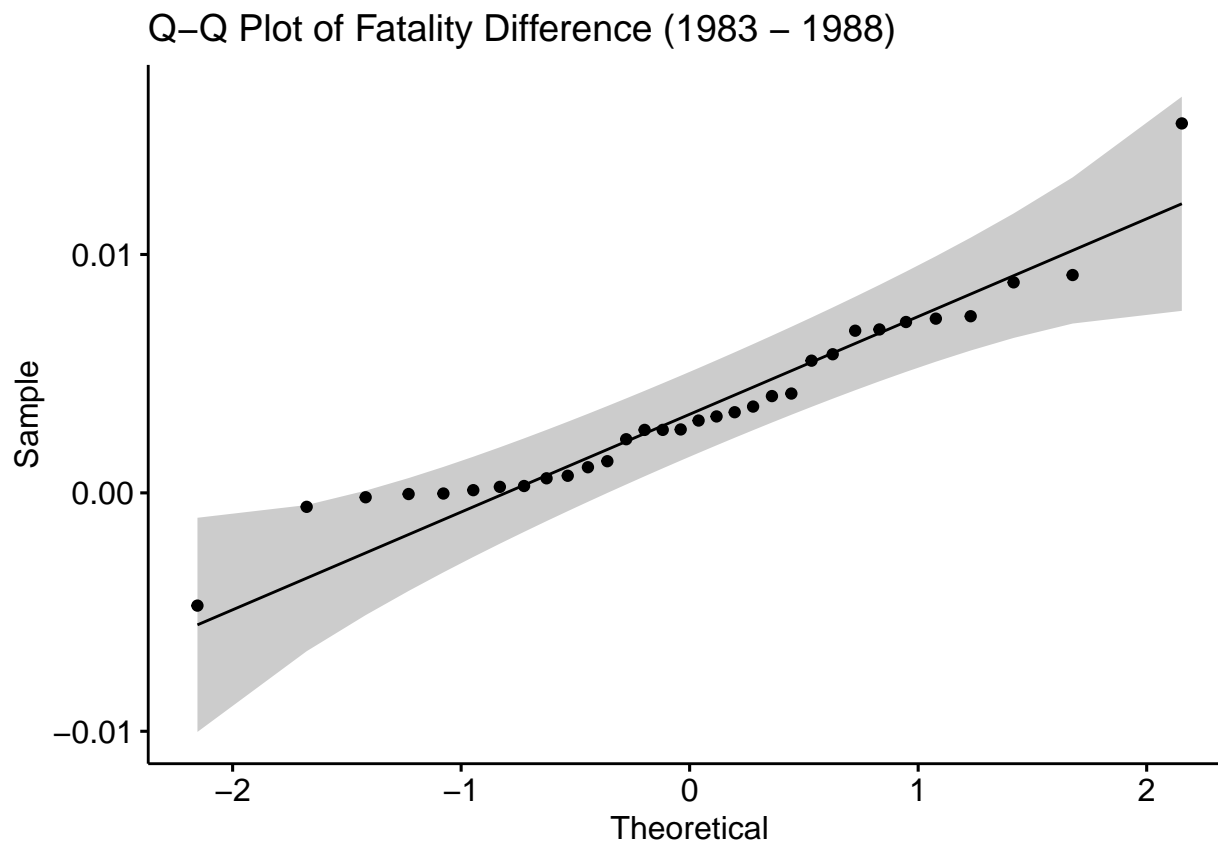
Question 1.1

(2 points)

Using the data frame "dat", perform a visual check of whether the value of "fatalities" in 1983 minus the value of "fatalities" in 1988 among the 32 jurisdictions that had a value of "no" for "drinkage" in 1983 could be considered Normally distributed. The function "ggqqplot" in the "ggpubr" package may help.

Please display your choice of visualization and describe your interpretation of the visual.

Your answer and code here:

```
# Filter the data for jurisdictions with "no" for drinkage in 1983
dat_filtered <- dat %>%
  filter(drinkage_1983 == "no")
# Compute the difference in fatalities between 1983 and 1988
dat_filtered <- dat_filtered %>%
  mutate(fatality_diff = fatalities_1983 - fatalities_1988)
# Create and display the Q-Q plot
ggqqplot(dat_filtered$fatality_diff, title = "Q-Q Plot of Fatality Difference (1983 - 1988)")
```

Q–Q Plot of Fatality Difference (1983 – 1988)

Question 1.2

(3 points)

Using Student's t, test the hypothesis that the differences in "fatalities" between 1983 and 1988 for jurisdictions that went from "no" to "yes" in "drinkage" during this period are consistent with samples drawn from a Normal distribution with mean equal to 0. Please state your conclusions from the Student's t test including whether the test is a valid test of the location of the mean at 0.

In 1983, a lower drinking age than 21 was used by the states not having a minimum drinking age of 21.

https://en.wikipedia.org/wiki/U.S.__history__of__alcohol__minimum__purchase__age__by__state

This analysis could be one step in examining the association between raised drinking age and traffic fatalities per million miles.

Your answer and code here:

```
# Perform a one-sample t-test to test if the mean difference is equal to 0
t_test_result <- t.test(dat_filtered$fatality_diff, mu = 0)
# Display the results
t_test_result
```

```
##
##  One Sample t-test
##
## data:  dat_filtered$fatality_diff
```

```
## t = 5.0541, df = 31, p-value = 1.839e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.002065882 0.004861200
## sample estimates:
##   mean of x
## 0.003463541
```

Question 1.3

(2 points)

What is the 99% confidence interval for the mean of these differences? Is this confidence interval consistent with a drop in the fatality rate between 1983 and 1988? Please explain.

Your answer and code here:

```
# Perform a one-sample t-test to test if the mean difference is equal to 0
t_test_result <- t.test(dat_filtered$fatality_diff, mu = 0, conf.level = 0.99)
# Display the 99% confidence interval for the mean difference
t_test_result$conf.int
```

```
## [1] 0.001583076 0.005344006
## attr(,"conf.level")
## [1] 0.99
```

This confidence interval does not include 0, which suggests that we can be 99% confident that the true mean difference in fatalities is positive and lies within this interval. This result is consistent with a drop in the fatality rate between 1983 and 1988, indicating that the jurisdictions that raised the drinking age from under 21 to 21 years saw a statistically significant reduction in traffic fatalities per million miles.

Question 1.4

(3 points)

Can you conclude that the increased drinking age caused a reduction in the fatality rate? The calculation below may help you think about this question.

```
fatal.diff.yes<-dat$fatalities_1983[dat$drinkage_1983=="yes"]-
          dat$fatalities_1988[dat$drinkage_1983=="yes"]
mean(fatal.diff.yes)
```

```
## [1] 0.003302584
```

Your answer here: While the data shows a reduction in traffic fatalities for jurisdictions that increased the drinking age to 21, the similar reduction observed in jurisdictions that already had a drinking age of 21 suggests that it is not possible to conclusively attribute the reduction solely to the change in drinking age. Other factors likely contributed to the observed decrease in fatalities. Therefore, we cannot conclusively conclude that the increased drinking age alone caused the reduction in the fatality rate.

Context for Questions 2-3

One often hears that the t-test is robust to moderate non-Normality in the population. The parts of this question explore this assertion.

A type 1 error in a hypothesis test is the rejection of the null hypothesis when it is true. For the z-test and the t-test, suppose the sampled population has the null distribution and you have a threshold p-value p below which you will reject the null hypothesis. For both tests, the probability of a type one error is exactly p.

A long explanation of this follows. You may understand this better thinking it through for yourself.

The definition of the p-value of an observed statistic is the probability of a statistic at least as extreme as the observed statistic. For these tests, "as extreme as the observed statistic" can equivalently mean "as far from the median as, or further from the median than the observed statistic" or "having probability less or equal to that of the observed statistic". Let X be the distribution of the statistic if the population has the null distribution. Note that the density function of X for both the z-test and the t-test is symmetric around 0. Let x be the value for which the event $\{X|X \leq -x \text{ or } X \geq x\}$ has probability p. For these tests, exactly the values of the statistic in this event have p-values that are less than or equal to that of x. Thus the probability that the p-value is less than or equal to p is the probability p of the event $\{X|X \leq -x \text{ or } X \geq x\}$ for the specified value of x.)

In the work below, you will estimate the probability of a type 1 error using the t-test on data from a *Gamma* distribution with mean $2\sqrt{2}$ and variance 4 given the null hypothesis that the sample is from a *Normal* population with $\mu_0 = 2\sqrt{2}$. You will estimate the probability of a type 1 error using the t-test on data from a *Normal*$(\mu = 2\sqrt{2}, \sigma^2 = 4)$ distribution, but with the values rounded to the nearest integer, given the null hypothesis that the sample is from a *Normal*$(\mu = 2\sqrt{2}, \sigma^2 = 4)$ population.

The goal is to gain an understanding of the extent to which the t-test remains a valid test of the location of the mean under these violations of the assumptions of the t-test as a test of the location of the mean.

Question 2

Context

The goal is to estimate the proportion of $p \leq 0.01$ in a t-test of the null hypothesis that the population distribution is *Normal* with $\mu_0 = 2\sqrt{2}$ but the sample is drawn from a *Gamma* population with mean $\mu = 2\sqrt{2}$ and variance 4.

The shape, scale, mean, and variance variables defined below are arranged so that changing the shape value will allow you to explore other *Normal* and *Gamma* distributions while retaining the property that they have the same mean and both have variance equal to 4.

```
n<-20
shp<-2
scl<-sqrt(4/shp)
sig<-sqrt(shp*scl^2) # sigma
mu<-shp*scl # mu
set.seed(12345)
ggqqplot(rgamma(n,shape=shp,scale=scl))+labs(title="qq-plot for a Gamma sample")
```

qq-plot for a Gamma sample
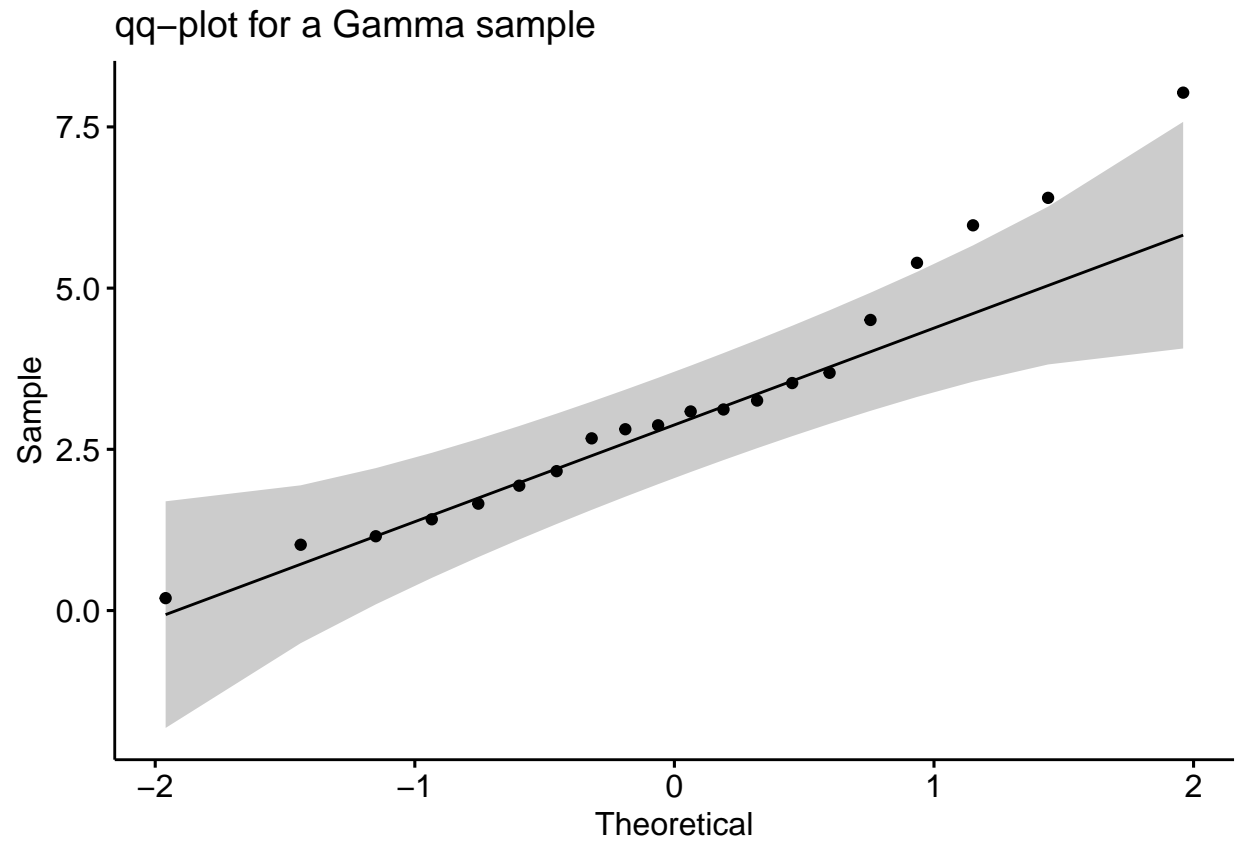

```r
dat.plot<-data.frame(x=c(0,3*mu))
ggplot(data=dat.plot,aes(x=x))+
  stat_function(fun=dgamma, args=list(shape=shp,scale=scl))+
  stat_function(fun=dgamma, args=list(shape=1,scale=sqrt(4/2)),color="orange")+
stat_function(fun=dgamma, args=list(shape=1.5,scale=sqrt(4/1.5)),color="green")+
  stat_function(fun=dgamma, args=list(shape=2.5,scale=sqrt(4/1.5)),color="blue")+
  labs(title="A Collection of Gamma Densities")
```

A Collection of Gamma Densities
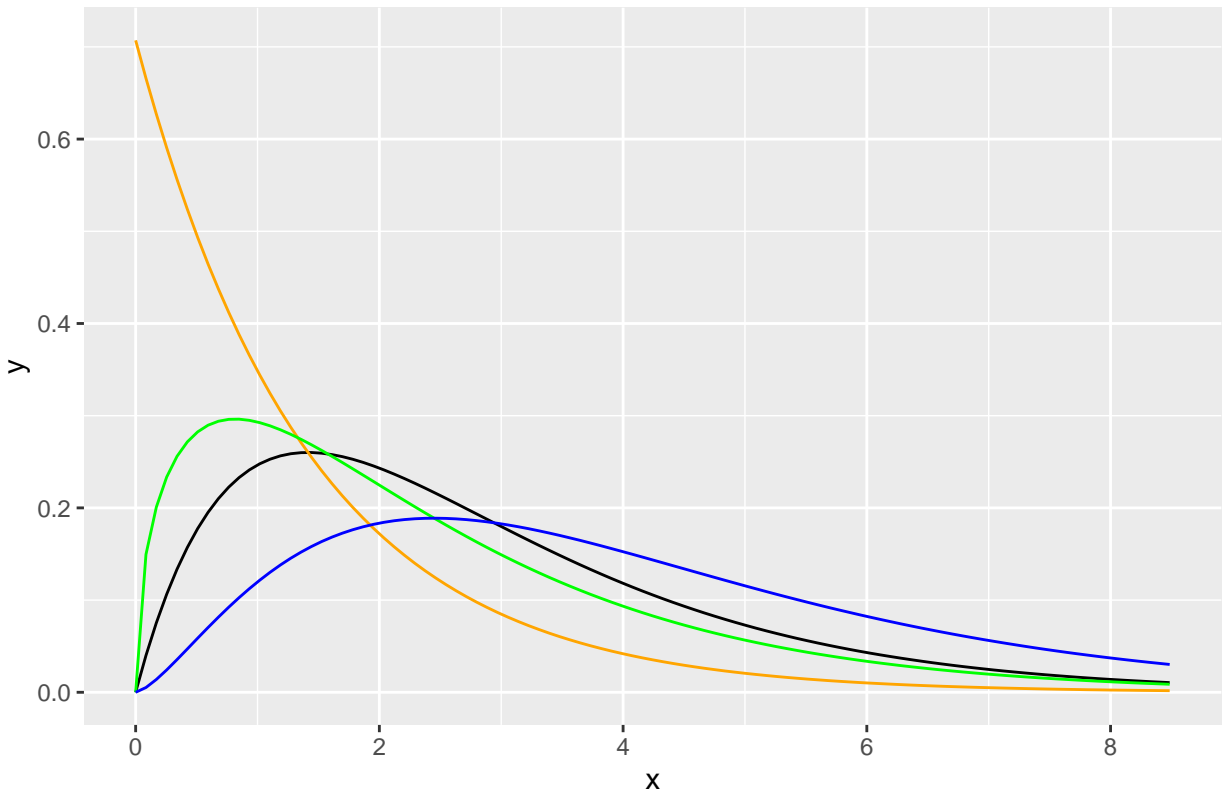
Suppose an iid sample of size "n" is drawn from population with a $Gamma(shape = 2, scale = \sqrt{2})$ distribution. Note that the mean of this distribution is $2\sqrt{2}$ and the variance is 4. Let the null hypothesis be that the sample is drawn from a $Normal$ population with $\mu_0 = 2\sqrt{2}$ population.

Question 2.1

(2 points)

Please use 100,000 samples of size 5 to estimate the probability that a two-sided t-test performed on the sample of size 5 from a $Gamma(shape = 2, scale = \sqrt{2})$ population will have a p-value that is less than or equal to 0.01. What is your estimate?

Your answer and code here:

```
# Set parameters
n = 5
shp = 2
scl = sqrt(4 / shp)
mu_0 = 2 * sqrt(2)
# Define function to perform t-test and return p-value
t.p.val.true <- function(a = shp, s = scl, n = 5) {
  samp <- rgamma(n, shape = a, scale = s)
  return(t.test(samp, mu = mu_0)$p.value)
}
# Perform t-tests on 100,000 samples of size 5
set.seed(1234567)
ps.gamma.true.5 <- replicate(100000, t.p.val.true(n = 5))
```

```
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01 <- mean(ps.gamma.true.5 <= 0.01)
prop.less.than.0.01
```

[1] 0.02823

This means that under these conditions, about 2.823% of the t-tests will incorrectly reject the null hypothesis when it is true, which is higher than the nominal level of 1%.

Question 2.2

(2 points)

Please use 100,000 samples of size 20 to estimate the probability that a two-sided t-test performed on the sample of size 20 from a $Gamma(shape = 2, scale = \sqrt{2})$ population will have a p-value that is less than or equal to 0.01. What is your estimate?

Your answer and code here:

```
# Set parameters
n = 20
shp = 2
scl = sqrt(4 / shp)
mu_0 = 2 * sqrt(2)
# Define function to perform t-test and return p-value
t.p.val.true <- function(a = shp, s = scl, n = 20) {
  samp <- rgamma(n, shape = a, scale = s)
  return(t.test(samp, mu = mu_0)$p.value)
}
# Perform t-tests on 100,000 samples of size 20
set.seed(1234567)
ps.gamma.true.20 <- replicate(100000, t.p.val.true(n = 20))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_20 <- mean(ps.gamma.true.20 <= 0.01)
prop.less.than.0.01_20
```

[1] 0.02219

This result indicates that under these conditions, about 2.219% of the t-tests will incorrectly reject the null hypothesis when it is true.

Question 2.3

(2 points)

Please use 100,000 samples of size 50 to estimate the probability that a two-sided t-test performed on the sample of size 50 from a $Gamma(shape = 2, scale = \sqrt{2})$ population will have a p-value that is less than or equal to 0.01. What is your estimate?

Your answer and code here:

```r
# Set parameters
n = 50
shp = 2
scl = sqrt(4 / shp)
mu_0 = 2 * sqrt(2)
# Define function to perform t-test and return p-value
t.p.val.true <- function(a = shp, s = scl, n = 50) {
  samp <- rgamma(n, shape = a, scale = s)
  return(t.test(samp, mu = mu_0)$p.value)
}
# Perform t-tests on 100,000 samples of size 50
set.seed(1234567)
ps.gamma.true.50 <- replicate(100000, t.p.val.true(n = 50))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_50 <- mean(ps.gamma.true.50 <= 0.01)
prop.less.than.0.01_50
```

```
## [1] 0.0157
```

This result indicates that under these conditions, about 1.57% of the t-tests will incorrectly reject the null hypothesis when it is true. This value is closer to the nominal level of 1%, suggesting that the t-test becomes more robust with larger sample sizes, even when the population distribution deviates from normality.

Question 2.4

(2 points)

Please use 100,000 samples of size 100 to estimate the probability that a two-sided t-test performed on the sample of size 100 from a $Gamma(shape = 2, scale = \sqrt{2})$ population will have a p-value that is less than or equal to 0.01. What is your estimate?

Your answer and code here:

```r
# Set parameters
n = 100
shp = 2
scl = sqrt(4 / shp)
mu_0 = 2 * sqrt(2)
# Define function to perform t-test and return p-value
t.p.val.true <- function(a = shp, s = scl, n = 100) {
  samp <- rgamma(n, shape = a, scale = s)
  return(t.test(samp, mu = mu_0)$p.value)
}
# Perform t-tests on 100,000 samples of size 100
set.seed(1234567)
ps.gamma.true.100 <- replicate(100000, t.p.val.true(n = 100))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_100 <- mean(ps.gamma.true.100 <= 0.01)
prop.less.than.0.01_100
```

```
## [1] 0.01364
```

This result indicates that under these conditions, about 1.364% of the t-tests will incorrectly reject the null hypothesis when it is true. This value is closer to the nominal level of 1%, further supporting the suggestion that the t-test becomes more robust with larger sample sizes, even when the population distribution deviates from normality.

Question 3

The goal is to estimate the proportion of $p \leq 0.01$ in a t-test of the null hypothesis that the population distribution is *Normal* with $\mu_0 = 2\sqrt{2}$ if the sample is drawn from a $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ population then rounded to the nearest integer.

Suppose an iid sample of size "n" is drawn from population with a $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ distribution except that the values are rounded to the nearest integer (see the "round" function). Let the null hypothesis be that the sample is drawn from a *Normal* population with $\mu_0 = 2\sqrt{2}$.

Question 3.1

(2 points)

Please use 100,000 samples of size 10 to estimate the probability that a two-sided t-test performed on the sample of size 10 from a $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ population, rounded to the nearest integer, will have a p-value that is less than or equal to 0.01. Please give your estimate.

Your answer and code here:

```
# Set parameters
n = 10
mu = 2 * sqrt(2)
sigma = sqrt(4)
# Define function to perform t-test on rounded sample and return p-value
t.p.val.round.true <- function(n = 10) {
  samp <- round(rnorm(n, mean = mu, sd = sigma))
  return(t.test(samp, mu = mu)$p.value)
}
# Perform t-tests on 100,000 samples of size 10
set.seed(123456)
ps.round.true.10 <- replicate(100000, t.p.val.round.true(n = 10))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_10 <- mean(ps.round.true.10 <= 0.01)
prop.less.than.0.01_10
```

```
## [1] 0.01051
```

This result indicates that under these conditions, about 1.051% of the t-tests will incorrectly reject the null hypothesis when it is true. This value is close to the nominal level of 1%, suggesting that the rounding of values to the nearest integer does not significantly affect the validity of the t-test in this scenario.

Question 3.2

(2 points)

Please use 100,000 samples of size 20 to estimate the probability that a two-sided t-test performed on the sample of size 20 from a $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ population, rounded to the nearest integer, will have a p-value that is less than or equal to 0.01. Please give your estimate.

Your answer and code here:

```
# Set parameters
n = 20
mu = 2 * sqrt(2)
sigma = sqrt(4)
# Define function to perform t-test on rounded sample and return p-value
t.p.val.round.true <- function(n = 20) {
  samp <- round(rnorm(n, mean = mu, sd = sigma))
  return(t.test(samp, mu = mu)$p.value)
}
# Perform t-tests on 100,000 samples of size 20
set.seed(123456)
ps.round.true.20 <- replicate(100000, t.p.val.round.true(n = 20))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_20 <- mean(ps.round.true.20 <= 0.01)
prop.less.than.0.01_20
```

```
## [1] 0.00982
```

This result indicates that under these conditions, about 0.982% of the t-tests will incorrectly reject the null hypothesis when it is true. This value is very close to the nominal level of 1%, further supporting the suggestion that the rounding of values to the nearest integer does not significantly affect the validity of the t-test in this scenario.

Question 3.3

(2 points)

Please use 100,000 samples of size 50 to estimate the probability that a two-sided t-test performed on the sample of size 100 from a $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ population, rounded to the nearest integer, will have a p-value that is less than or equal to 0.01. Please give your estimate.

Your answer and code here:

```
# Set parameters
n = 50
mu = 2 * sqrt(2)
sigma = sqrt(4)
# Define function to perform t-test on rounded sample and return p-value
t.p.val.round.true <- function(n = 50) {
  samp <- round(rnorm(n, mean = mu, sd = sigma))
  return(t.test(samp, mu = mu)$p.value)
}
# Perform t-tests on 100,000 samples of size 50
set.seed(123456)
ps.round.true.50 <- replicate(100000, t.p.val.round.true(n = 50))
# Estimate the proportion of p-values less than or equal to 0.01
prop.less.than.0.01_50 <- mean(ps.round.true.50 <= 0.01)
prop.less.than.0.01_50
```

```
## [1] 0.00984
```

This result indicates that under these conditions, about 0.984% of the t-tests will incorrectly reject the null hypothesis when it is true. This value is very close to the nominal level of 1%.

Context for Questions 4-6

In these questions we investigate data on body temperatures. A recent article in the NYT titled: "The Average Human Body Temperature Is Not 98.6 Degrees" suggests that the mean human body temperature is less than 98.6 degrees. In these questions we investigate this claim, and seek to understand how certain we can be of the statistical results.

Question 4: Assumptions for hypothesis tests

Question 4.1

(2 points)

State the null and alternative hypothesis based on the problem description above. Note that μ denotes the population mean and you should write your hypotheses in terms of the values of μ.

Your answer here: - Null hypothesis: $H_0 : \mu = 98.6$ degrees - Alternative hypothesis: $H_A : \mu < 98.6$ degrees

Question 4.2

(2 points)

In the code chunk below is a snippet of the data that you will use to assess the null and alternative hypothesis. State the p-value and if you decided to reject the null hypothesis.

```
head(normtemp)
```

```
##   temperature gender hr
## 1        96.3      1 70
## 2        96.7      1 71
## 3        96.9      1 74
## 4        97.0      1 80
## 5        97.1      1 73
## 6        97.1      1 75
```

```
attach(normtemp)
?t.test
#Uncomment the code and fill in the arguments necessary to answer the question
t.test(normtemp$temperature, mu = 98.6, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  normtemp$temperature
```

```
## t = -5.4548, df = 129, p-value = 1.205e-07
## alternative hypothesis: true mean is less than 98.6
## 95 percent confidence interval:
##      -Inf 98.35577
## sample estimates:
## mean of x
##  98.24923
```

Your answer here: The p-value from the t-test is less than 1.205e-07, which is significantly less than the typical significance level of 0.05, the 95 percent confidence interval is between 98.122 and 98.376, and an estimated mean of 98.24923. Therefore, we reject the null hypothesis. This indicates that there is significant evidence to support the claim that the mean human body temperature is less than 98.6 degrees.

Question 4.3

(2 points)

Imagine you are asked what is a reasonable range for the average human body temperature by a medical professional. This question is important for determining for example if an individual could be sick or to go see a doctor. Create a 90% confidence interval for the population mean human body temperature. Based on this confidence interval what is a reasonable range?

Your answer and code here:

```
#Uncomment the code and fill in the arguments necessary to answer the question
t.test(normtemp$temperature, conf.level = 0.90)
```

```
##
##  One Sample t-test
##
## data:  normtemp$temperature
## t = 1527.9, df = 129, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  98.14269 98.35577
## sample estimates:
## mean of x
##  98.24923
```

Based on the 90% confidence interval, a reasonable range for the average human body temperature is approximately 98.14 degrees to 98.36 degrees. This range can be used by medical professionals to assess whether an individual's body temperature is within the normal range or if it may indicate a potential health issue.

Question 4.4

(2 points)

One of the assumptions used in a t-test for population mean is that the data is normally distributed. In the code chunk below make a qqplot with qqline, and perform a shapiro test to assess the normality of the data. State your conclusion: is the data normally distributed or not?

Your answer and code here:

```
# Make your qqplot and qqline
qqnorm(normtemp$temperature, main = "Q-Q Plot of Body Temperature")
qqline(normtemp$temperature, col = "red")
```

Q–Q Plot of Body Temperature
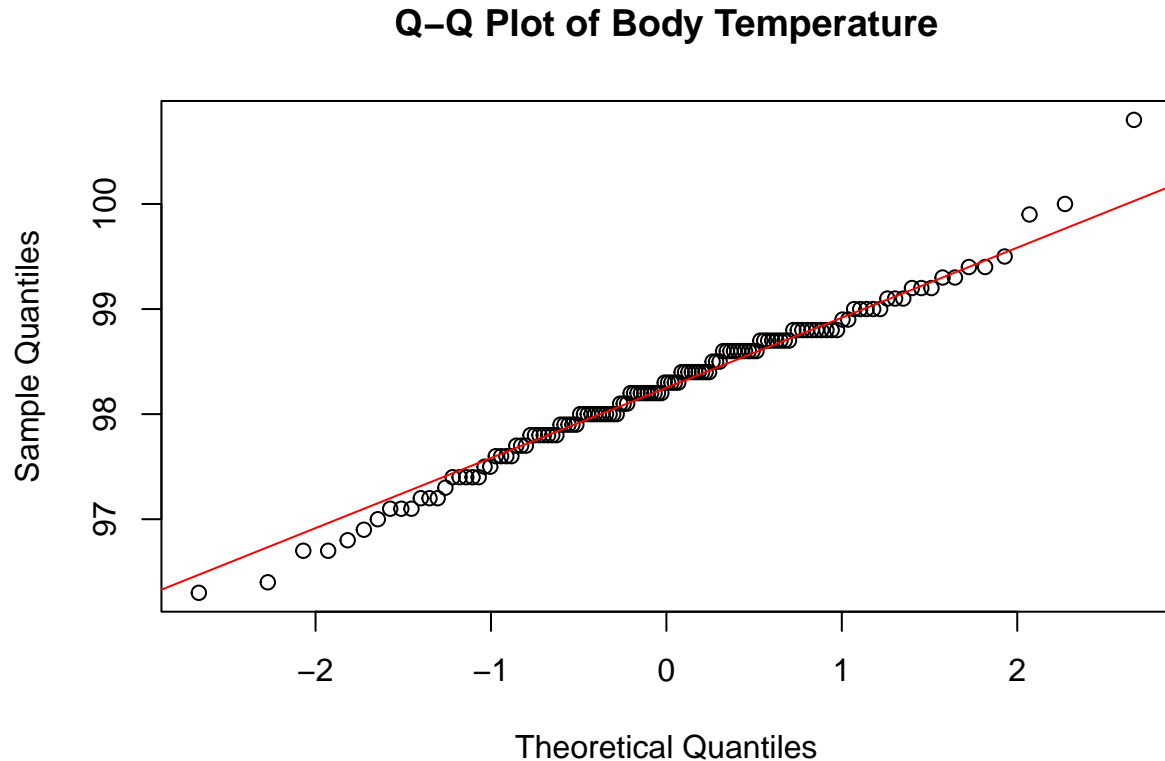

```
# Perform Shapiro-Wilk test
shapiro_test_result <- shapiro.test(normtemp$temperature)
# Display the result of the Shapiro-Wilk test
shapiro_test_result
```

```
##
##  Shapiro-Wilk normality test
##
## data:  normtemp$temperature
## W = 0.98658, p-value = 0.2332
```

The p-value from the Shapiro-Wilk test is 0.2332, which is greater than the typical significance level of 0.05. Therefore, we do not reject the null hypothesis that the data is normally distributed. Additionally, the Q-Q plot with the Q-Q line shows that the data points lie approximately along the line, further indicating that the data is approximately normally distributed.

Question 5

In this section we will consider some theoretical aspects of hypothesis testing by answering the questions in this sections. A type two error is the instance in which we fail to reject the null hypothesis when the

alternative hypothesis is true. The power of a statistical test is the probability we correctly reject the null hypothesis when the alternative hypothesis is true. The power can be calculated as 1 - p(type two error).

Question 5.1

(2 points)

Use the Central Limit Theorem to determine the distribution of the sample average where of the 130 observations are independent and identically distributed with mean of 98.25 and standard deviation of .733.

Your answer here: According to the Central Limit Theorem, the distribution of the sample mean $\bar{X}$ is approximately normal with a sample mean $\mu_{\bar{X}} = \mu$ and a standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. Therefore, the distribution of the sample average is approximately:

$$\bar{X} \sim N\left(98.25, \frac{0.733}{\sqrt{130}}\right) = N(98.25, 0.0643)$$

.

Question 5.2

(2 points)

To test the null that the mean human body temperature is 98.6 degrees versus the alternative that the mean temperature has decreased we use the following rule that we reject the null if the observed sample average is less than 98.49 degrees (this corresponds to a 5% significance level). Here the sample average consists of 130 observations from a normal distribution with standard deviation of .733. Calculate the probability of a type two error and the power of the test using the pnorm command in the code chunk below given that the true population mean is 98.25 degrees.

Your answer and code here:

```
# Given parameters
mu_true <- mean(temperature)
sigma <- sd(temperature)
n <- length(temperature)
sample_mean_cutoff = 98.49
sigma_xbar <- (sigma / sqrt(n))
# Calculate the z-score for the cutoff
z_score <- (sample_mean_cutoff - mu_true) / sigma_xbar
# Calculate the power of the test
power <- pnorm(z_score)
# Calculate the probability of a type two error
type2_error <- 1 - power
# Display the results
type2_error
```

```
## [1] 9.048118e-05
```

```
power
```

```
## [1] 0.9999095
```

15

Question 5.3

(2 points)

Explain why the assumption of normally in question 5.2 is not essential, and is an assumption that could be dropped.

Your answer here: The assumption of normality in Question 5.2 is not essential due to the Central Limit Theorem (CLT). The CLT states that, regardless of the population distribution, the distribution of the sample mean will approximate a normal distribution as the sample size becomes large enough. In this case, we have a sample size of 130, which is generally considered sufficiently large for the CLT to apply. This means that the sample mean will be approximately normally distributed even if the original data is not perfectly normal.

Question 5.4

(2 points)

Explain why increasing the sample size increases the power of the statistical test.

Your answer here: In summary, increasing the sample size reduces the standard error, improves the precision of the estimate, enhances the signal-to-noise ratio, and narrows confidence intervals. All these factors contribute to a higher probability of correctly rejecting the null hypothesis when it is false, thereby increasing the power of the statistical test.

Question 6: Simulations of power.

In the questions below you will use an R package to calculate the power of a statistical test.

Question 6.1

(4 points)

In the code chunk below we calculate the power of the one sided t-test for varying sample size, with a significance level of 95%. Here the true mean is 98.25, and standard deviation is .733. Here the sample size is allowed to vary, and we graph the results. Please summarize in your own words the findings of the graph for a practicing data scientist interested in testing whether the true population mean is less than 98.6.

```
N = seq(20,100,by=1)

power_n = pwr.t.test(n=N,d = ((98.25 - 98.6)/.733), sig.level = .05, type = "one.sample", alternative =

ggplot(data = data.frame(N,power_n), aes(x=N,y=power_n)) + geom_line()
```
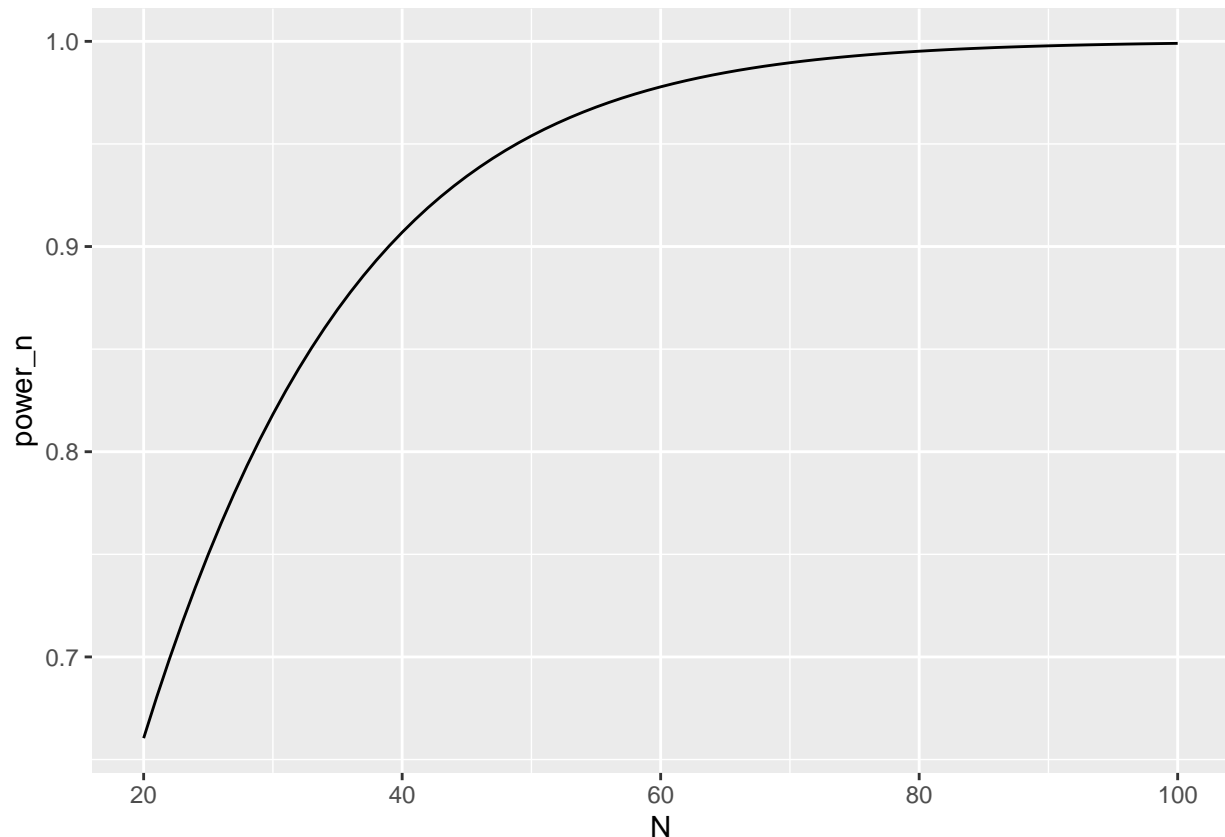
Your answer here: The graph shows the power of the one-sided t-test as a function of sample size for testing whether the true population mean is less than 98.6 degrees. As the sample size increases, the power of the test also increases. This means that with larger sample sizes, the test is more likely to correctly reject the null hypothesis (i.e., it becomes more sensitive to detecting that the true mean is indeed less than 98.6 degrees). For a practicing data scientist, this graph emphasizes the importance of having a sufficiently large sample size when conducting hypothesis tests. With smaller sample sizes, the power is lower, indicating a higher risk of failing to detect a true difference (type II error). As the sample size approaches 100, the power of the test approaches 1, indicating a very high probability of correctly rejecting the null hypothesis if the true mean is 98.25 degrees.

Question 6.2

(5 points)

This question asks you to summarize your findings from all the results of problems 4-5 to assess the claim that the mean human body temperature is less than 98.6 degrees. How sure can you be of your conclusion?

Your answer here: - Hypothesis Testing (Question 4.2): The t-test conducted to compare the sample mean against 98.6 degrees resulted in a very low p-value (1.205e-07), leading us to reject the null hypothesis. This indicates strong evidence that the mean human body temperature is less than 98.6 degrees. - Confidence Interval (Question 4.3): The 90% confidence interval for the population mean human body temperature was found to be approximately [98.14269, 98.35577] degrees. This interval does not include 98.6 degrees, further supporting the claim that the mean human body temperature is less than 98.6 degrees. - Normality Assumption (Question 4.4): The Shapiro-Wilk test indicated that the data is approximately normally distributed, which validates the use of the t-test. - Type II Error and Power (Question 5.2): The power of the test was calculated to be very high (0.9999), and the probability of a type II error was very low (0.0001). This means that the test is highly reliable in correctly rejecting the null hypothesis when the true mean is

98.25 degrees. - Effect of Sample Size (Question 6.1): The graph showing the relationship between sample size and the power of the test demonstrates that increasing the sample size significantly increases the power of the test. With larger sample sizes, the probability of correctly rejecting the null hypothesis increases.

Based on the results of the hypothesis test, the confidence interval, the normality assessment, and the power analysis, we can be very confident in concluding that the mean human body temperature is less than 98.6 degrees. The evidence strongly supports the claim, and the high power of the test indicates a very low likelihood of making a type II error. Therefore, the conclusion is highly reliable.

End.