

ward_abigail_HW6

Abigail Ward

##Question 1

```
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

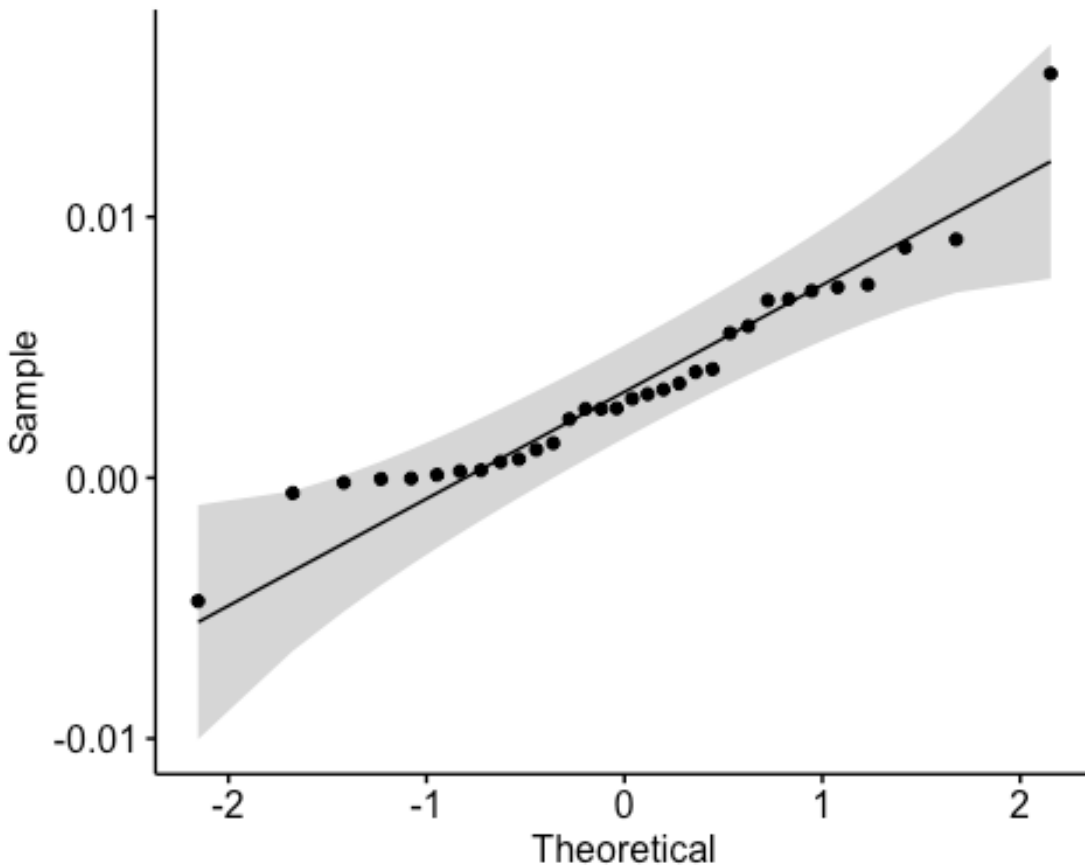
dat<-USSeatBelts
dat<-pivot_wider(dat,id_cols=state,names_from = year,values_from =
c(fatalities,drinkage))
```

1A

```
dat$drinkage_1983=="no"

## [1] FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## [13]  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE
## [25] FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE
## [37] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## [49]  TRUE  TRUE  TRUE

fatal.diff<-dat$fatalities_1983[dat$drinkage_1983=="no"]-
  dat$fatalities_1988[dat$drinkage_1983=="no"]
ggqqplot(fatal.diff)
```



The qq plot shows a fairly linear and even distribution of the data. There are the two end point which do not seem to follow the exact curve but are still within the shaded region. It is a fairly safe assumption to consider this data to be normally distributed.

1B

```
t.test(fatal.diff)

##
##  One Sample t-test
##
## data:  fatal.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
```

The p-value for the student's t test is 0.00001839 which is less than the 0.05 and therefore it is reasonable to reject the null hypothesis. This means that the mean of the differences in "fatalities" between 1983 and 1988 for jurisdictions that went from "no" to "yes" in "drinkage" during this period is not zero. Yes the one-sample t-test is a valid test of the

location of the mean at zero since the data was fairly normal. Here we are comparing the mean change in fatalities to zero, which will indicate whether any change occurred corresponding to the change in the drinking age.

1C

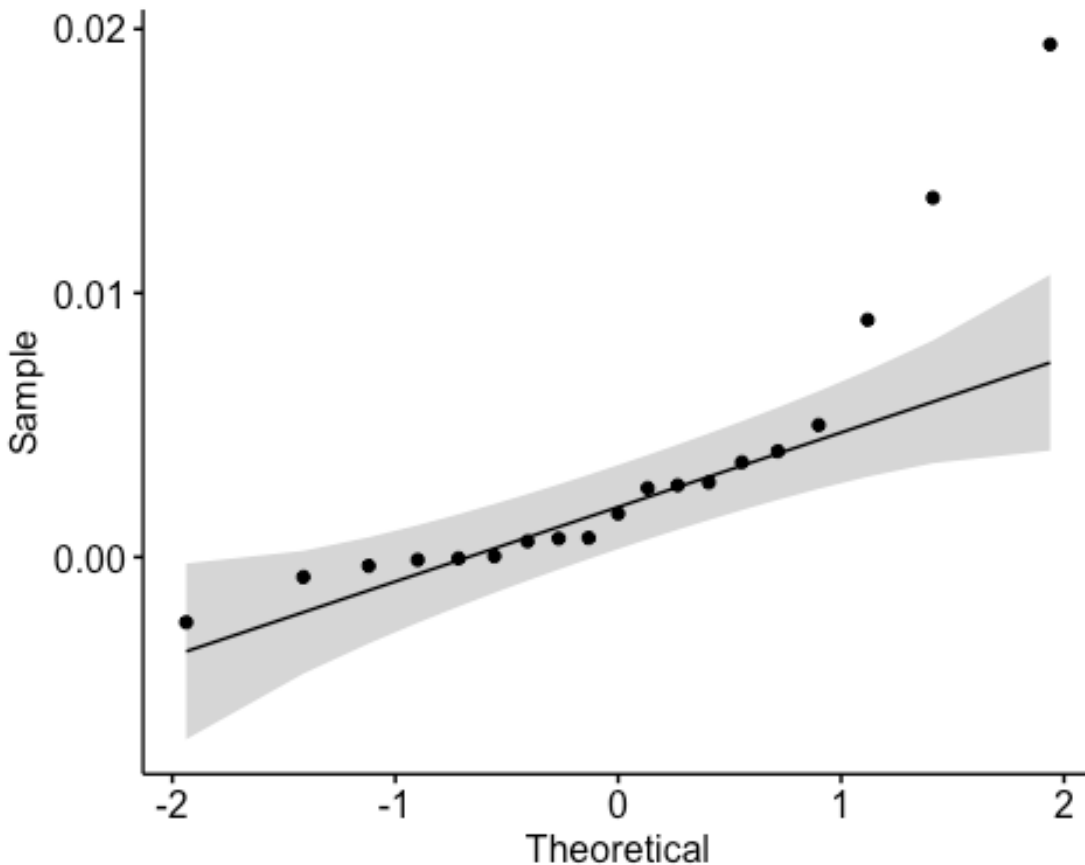
```
t.test(fatal.diff, conf.level = 0.99)

##
##  One Sample t-test
##
## data:  fatal.diff
## t = 5.0541, df = 31, p-value = 1.839e-05
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  0.001583076 0.005344006
## sample estimates:
##  mean of x
## 0.003463541
```

The 99% confidence interval is from 0.001583076 to 0.005344006 which does not include 0.00. This would indicate that there were more fatalities in 1983 than in 1988 because the 99% confidence interval only contains positive values and therefore consistent with a drop in the fatality rate between 1983 and 1988.

1D

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



```
mean(fatal.diff.yes)
## [1] 0.003302584
mean(fatal.diff)
## [1] 0.003463541
```

No you cannot conclude that the increased drinking age caused a reduction in the fatality rate. We did see a decline in the fatality rate after the drinking age was imposed, but there was also a decline when there was already a drinking age imposed in 1983. This would indicate that there were additional factors such as speed limits, car safety features and more that could have also impacted the fatality rate. We cannot conclude causation from this study.

Question 2

Population is $Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ and the null hypothesis is that the sample is from a $Normal(\mu = 2\sqrt{2} - 1, \sigma^2 = 4)$ population

Gamma distribution with mean $2\sqrt{2}$ and variance 4 given the null hypothesis that the sample is from a $Normal(\mu = 2\sqrt{2} - 1, \sigma^2 = 4)$ population

$Normal(\mu = 2\sqrt{2}, \sigma^2 = 4)$ distribution but the values are rounded using the null hypothesis is that the sample is from a $Normal(\mu = 2\sqrt{2} - 1, \sigma^2 = 4)$ population.

2A

```
qnorm(.995)
## [1] 2.575829
pnorm(2.575829)
## [1] 0.995

shp<-2
scl<-sqrt(4/shp)
sig<-sqrt(shp*scl^2) # sigma in 2a
mu<-shp*scl # 2*sqrt(2), mu in 2a
n<-50

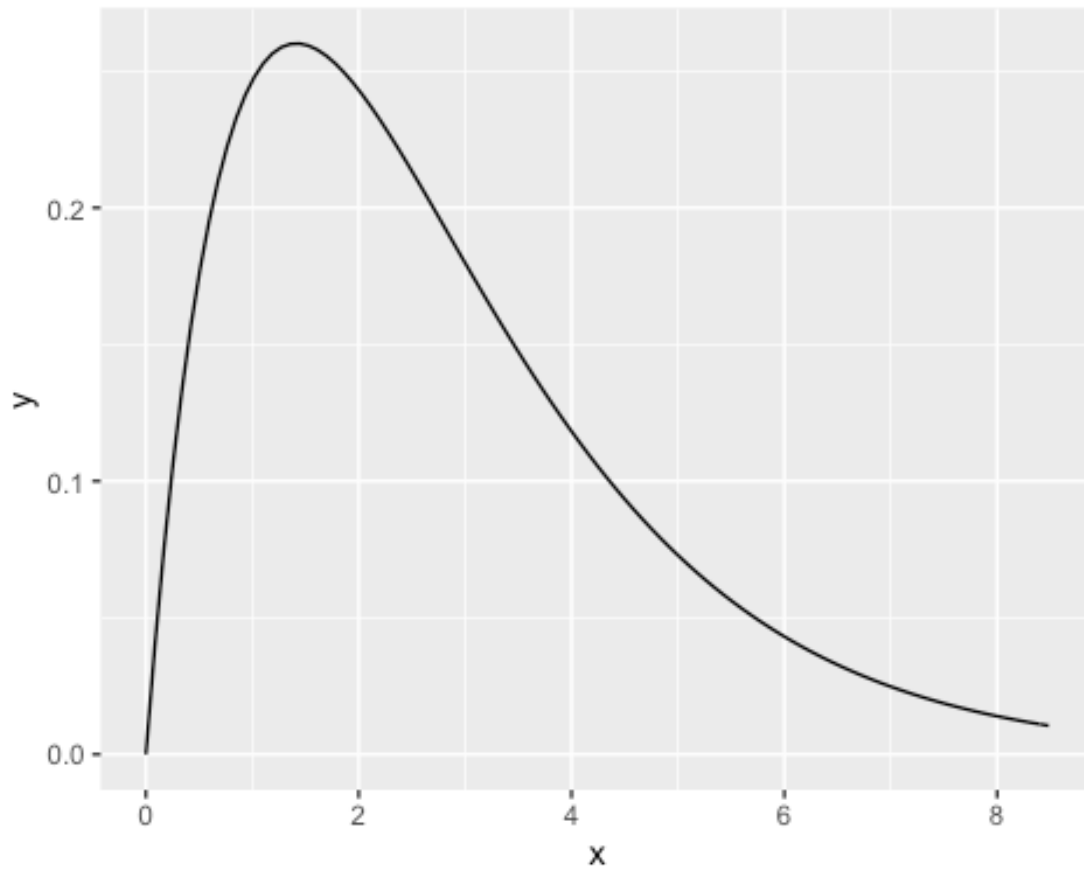
a<-qnorm(.995)
a
## [1] 2.575829
a*sig/sqrt(n)
## [1] 0.7285545
sig/sqrt(n)
## [1] 0.2828427
sig
## [1] 2
sqrt(50)
## [1] 7.071068
b.lower<-mu-1-a*sig/sqrt(n)
b.upper<-mu-1+a*sig/sqrt(n)

(pnorm(b.upper,mu,sig/sqrt(n),lower.tail=FALSE)+
  pnorm(b.lower,mu,sig/sqrt(n)))
## [1] 0.831398
```

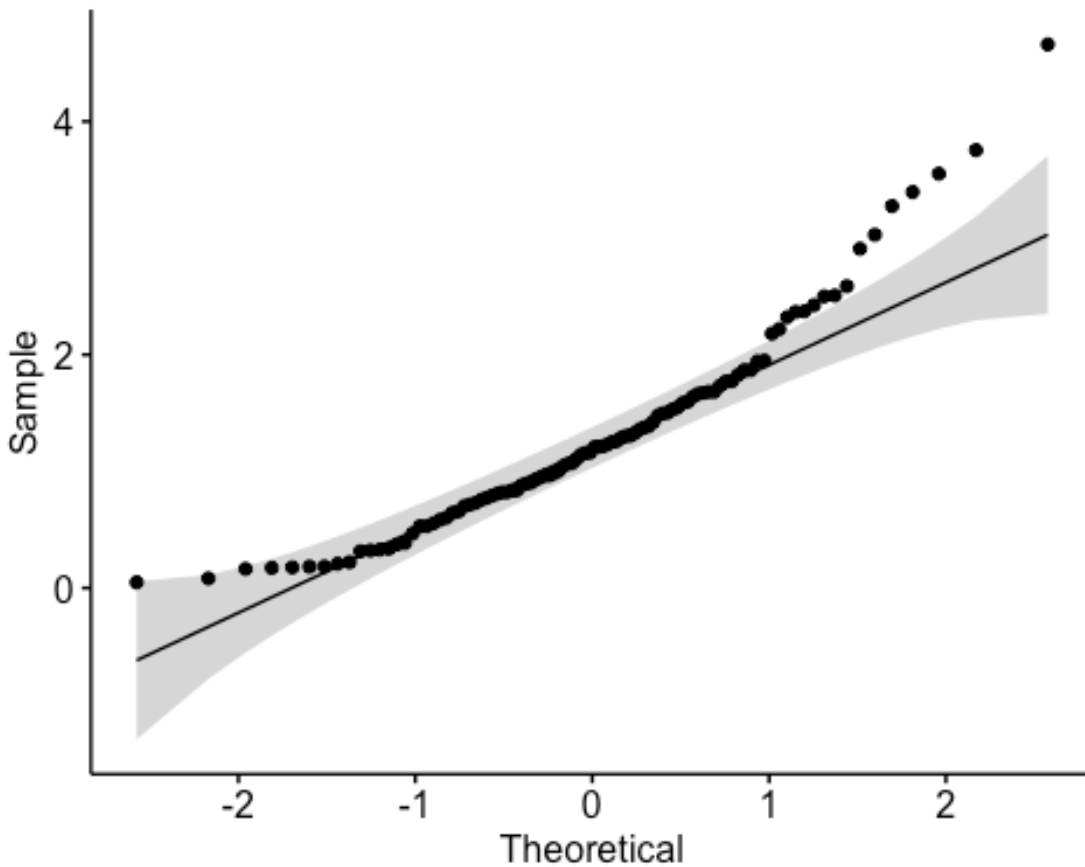
The probability that a two-sided z-test of performed on the sample will have a p-value that is less than or equal to 0.01 is about 83%.

2B

```
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))
```



```
ggqqplot(rgamma(100,shp,scl))
```

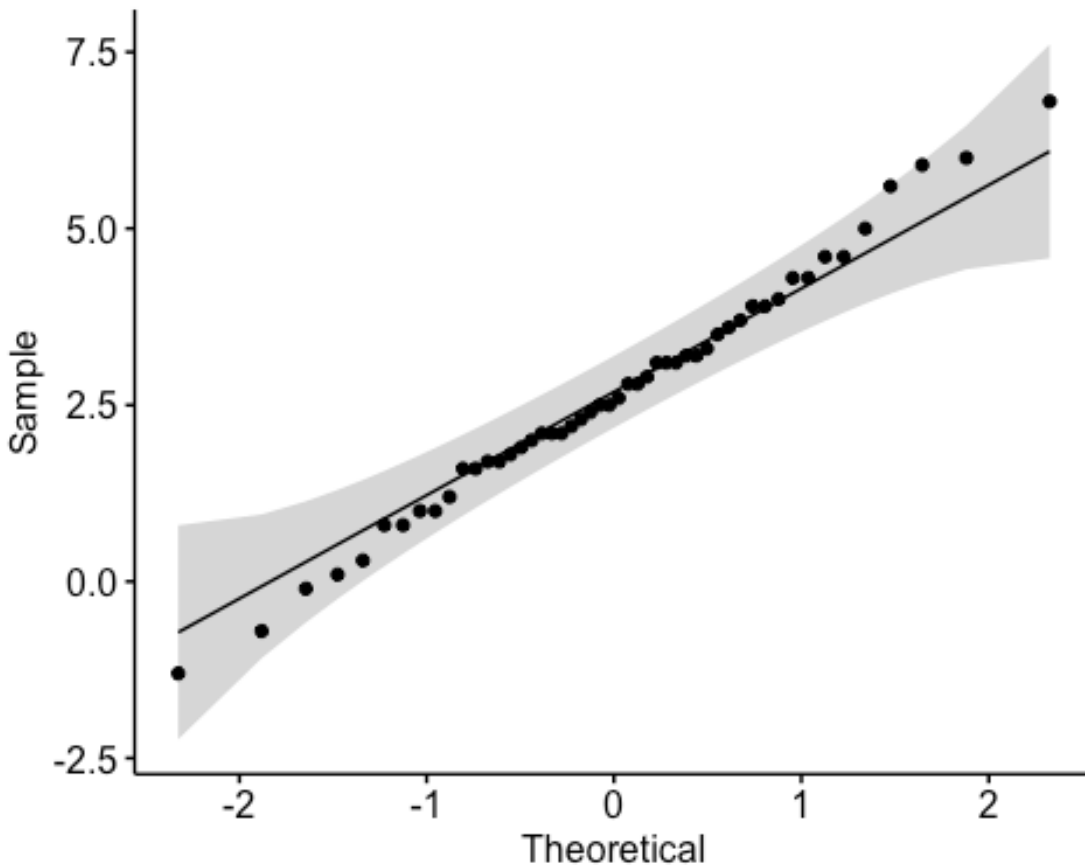


```
z.p.val<-function(a=shp,s=sc1){
  samp<-rgamma(n,shape=a,scale=s)
  return(2*pnorm(-abs((mean(samp)-(mu-1))/(sig/sqrt(n)))))
}
set.seed(1234567)
ps<-replicate(100000,z.p.val())
mean(ps<.01)
## [1] 0.83074
```

The probability that a two-sided z-test of performed on the sample will have a p-value that is less than or equal to 0.01 is 83%.

2C

```
set.seed(1234567)
ggqqplot(round(rnorm(n,mu,sig),1))
```



```
z.p.val.round<-function(a=shp,s=sc1){
  samp<-round(rnorm(n,mu,sig),1)
  return(2*pnorm(-abs((mean(samp)-(mu-1))/(sig/sqrt(n)))))
}

ps.round<-replicate(100000,z.p.val.round())
mean(ps.round<.01)

## [1] 0.83037
```

The probability that a two-sided z-test of performed on the sample from the normal distribution with rounding will have a p-value that is less than or equal to 0.01 is about 83%.

2D

```
z.p.val.true<-function(a=shp,s=sc1){
  samp<-rgamma(n,shape=a,scale=s)
  return(2*pnorm(-abs((mean(samp)-mu)/(sig/sqrt(n)))))
}

set.seed(1234567)
ps.true<-replicate(100000,z.p.val.true())
mean(ps.true<.01)
```



```
## [1] 0.01002
```

The probability that a two-sided z-test performed on the sample will have a p-value that is less than or equal to 0.01 is 1.002% meaning the gamma distribution here resembles the normal distribution.

2E

```
set.seed(123456)
z.p.val.round.true<-function(a=shp,s=scl){
  samp<-round(rnorm(n,mu,sig),1)
  return(2*pnorm(-abs((mean(samp)-mu)/(sig/sqrt(n)))))
}

ps.round.true<-replicate(100000,z.p.val.round.true())
mean(ps.round.true<.01)

## [1] 0.00993
```

The probability that a two-sided z-test performed on the sample will have a p-value that is less than or equal to 0.01 is 0.993% meaning the rounded normal distribution here resembles the normal distribution.

2F

Given the sample size, and the p value, the correctness of the p-value and the power of the test does not seem to distinguish between the normal and gamma distribution. This can be seen in the similar values for the power of the test in parts a and b.

The correctness of the p-value and the power of the test does not seem to be strongly affected by rounding the normal data as seen in part a and c.