# HomeWork1

*Adon Rosen*

*9/9/2019*

### Load the library(s)

```
library(foreign) ## Will be used to load .sav file
library(ggplot2) ## Will be used for plotting
print("Done loading librarys")
```

```
## [1] "Done loading librarys"
```

```
all.dat <- read.spss("./salary.sav", to.data.frame=T)
```

```
## re-encoding from CP1252
```

```
## I am now going to write a csv, so I can upload this to github and have the data in a remote location
write.csv(all.dat, "./salary.csv", quote=F, row.names=F)
all.dat <- read.csv('./salary.csv')
```

## The following sections will be used to answer Problem #1

### Here is problem 1A

```
mean.salary <- mean(all.dat$salary)
median.salary <- median(all.dat$salary)
var.salary <- var(all.dat$salary)
print(paste("The mean of salary is: ", mean.salary))
```

```
## [1] "The mean of salary is:  72820.6666666667"
```

```
print(paste("The median of salary is: ", median.salary))
```

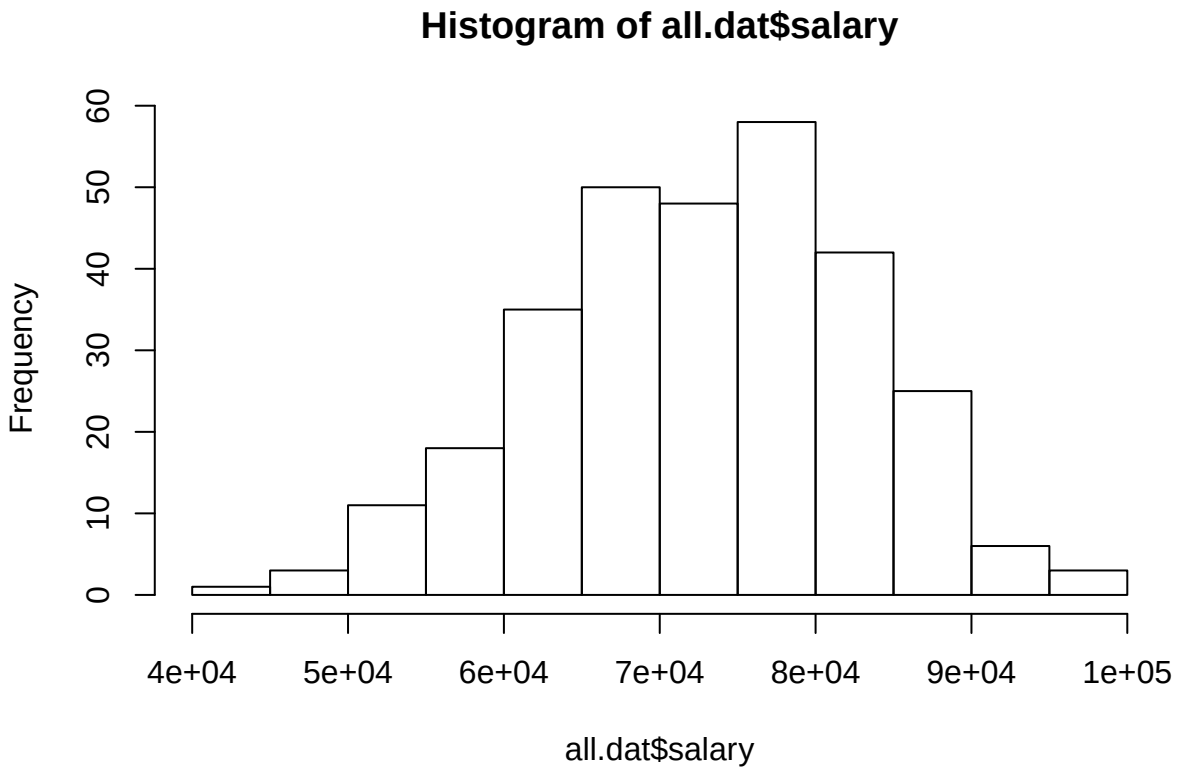
```
## [1] "The median of salary is:  73300"
```

```
print(paste("The variance of salary is: ", var.salary))
```

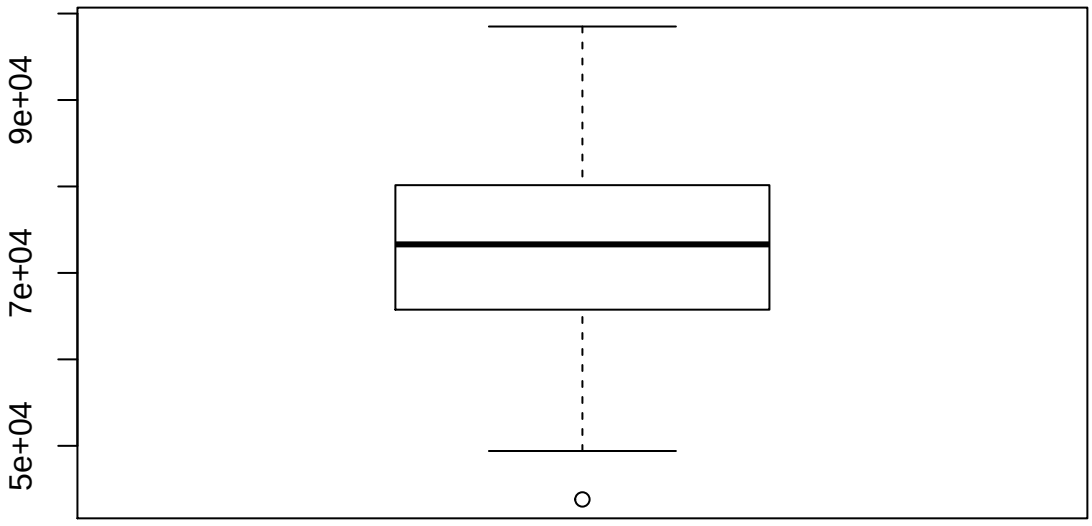
```
## [1] "The variance of salary is:  104096962.764771"
```

### Problem 1B

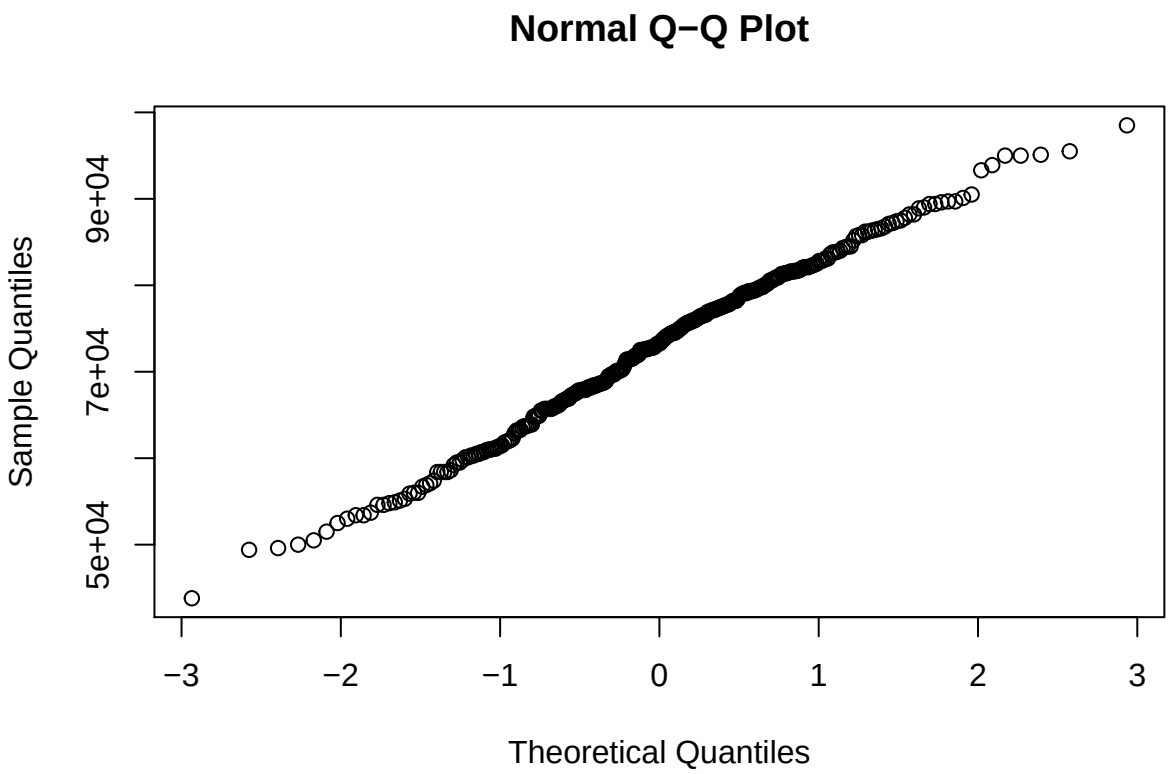
```
hist(all.dat$salary)
```




```
boxplot(all.dat$salary)
```




```
qqnorm(all.dat$salary)
```

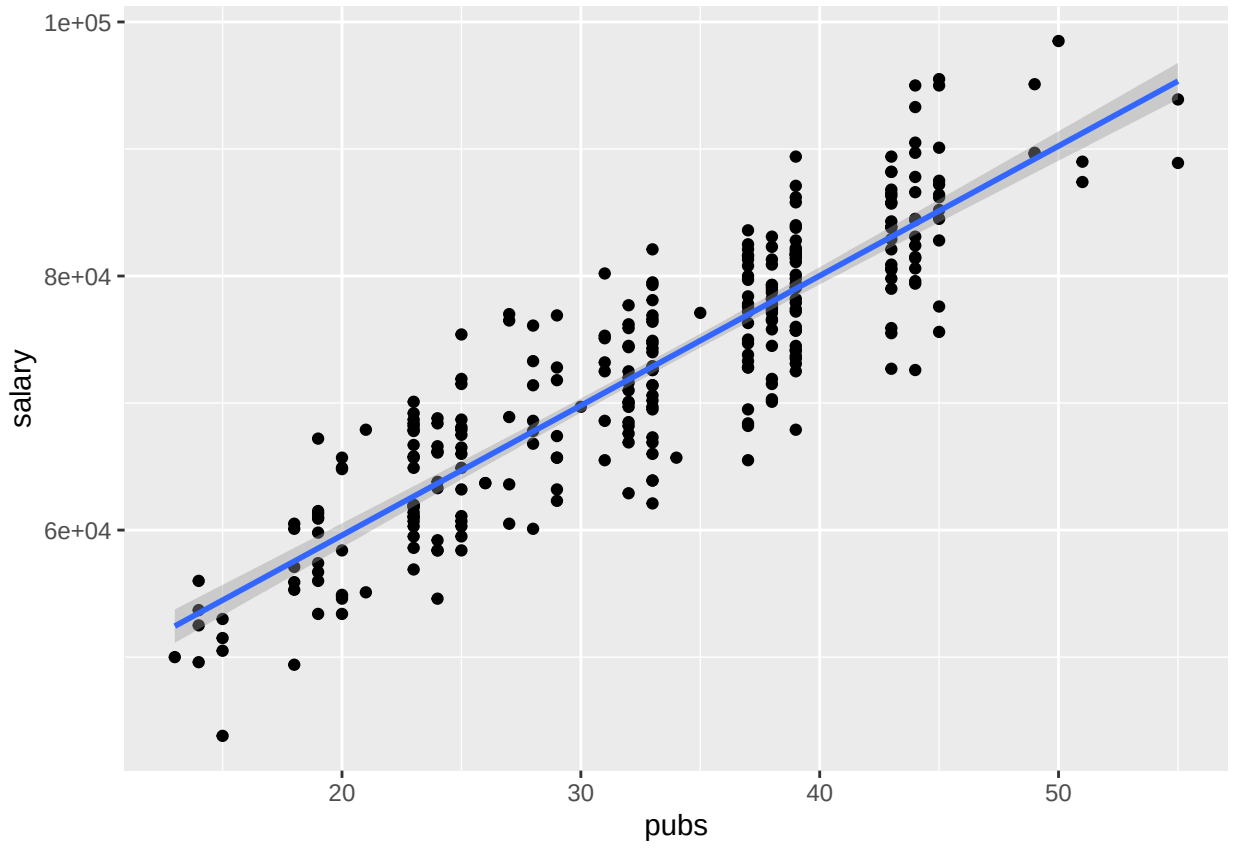



```
print(paste("These data appear to be relativley normal although there is one lower bound outlier"))
```

```
## [1] "These data appear to be relativley normal although there is one lower bound outlier"
```

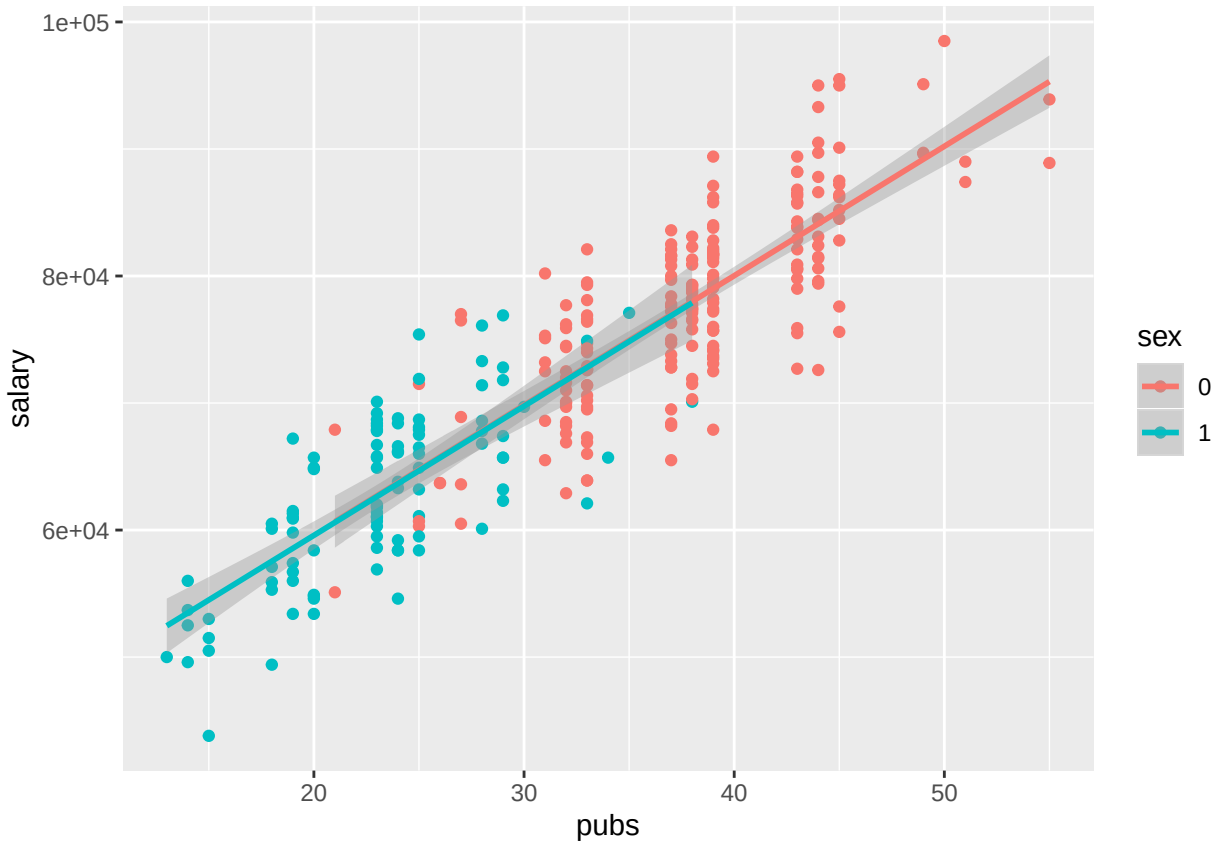
### Problem 1C

```
out.scat.one <- ggplot(data=all.dat, aes(x=pubs, y=salary)) +
  geom_point() +
  geom_smooth(method='lm')
print(out.scat.one)
```




### Problem 1D

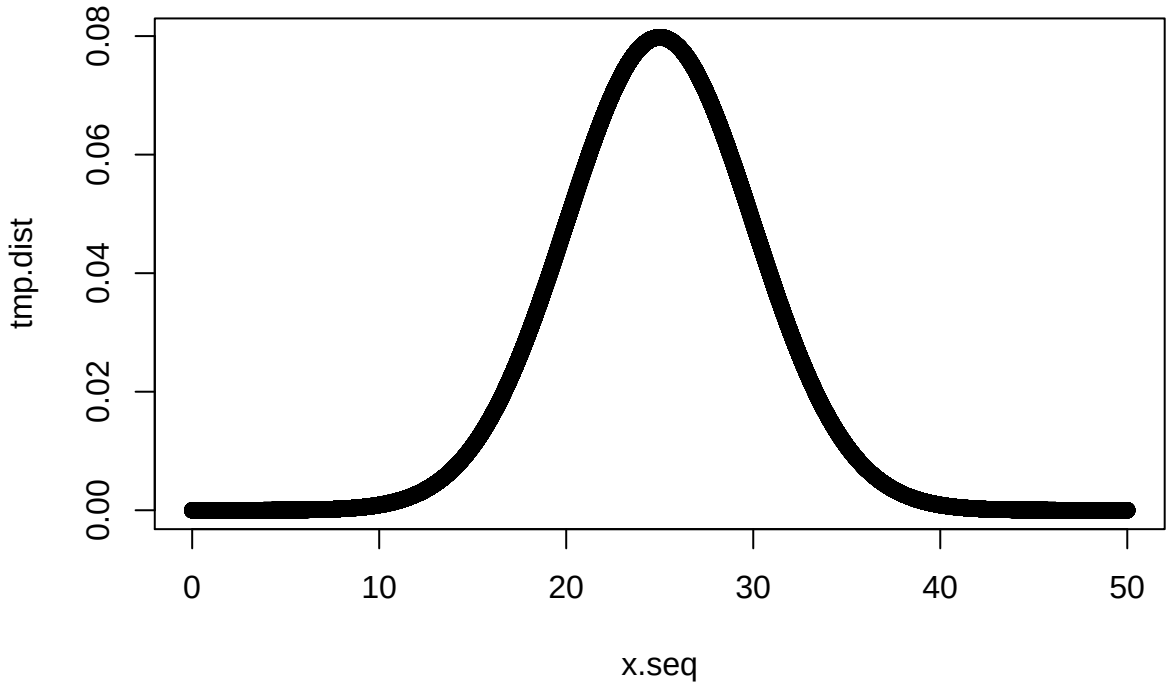
```
all.dat$sex <- factor(all.dat$sex)
out.scat.two <-  ggplot(data=all.dat, aes(x=pubs, y=salary, group=sex, color=sex)) +
  geom_point() +
  geom_smooth(method='lm')
print(out.scat.two)
```




## The following code will be used to answer question 2

### First create the theoretical PDF

```
x.seq <- seq(0, 50, by=.001)
tmp.dist <- dnorm(x=x.seq, mean=25, sd=5)
## Now plot this distribution
plot(x.seq, tmp.dist)
```

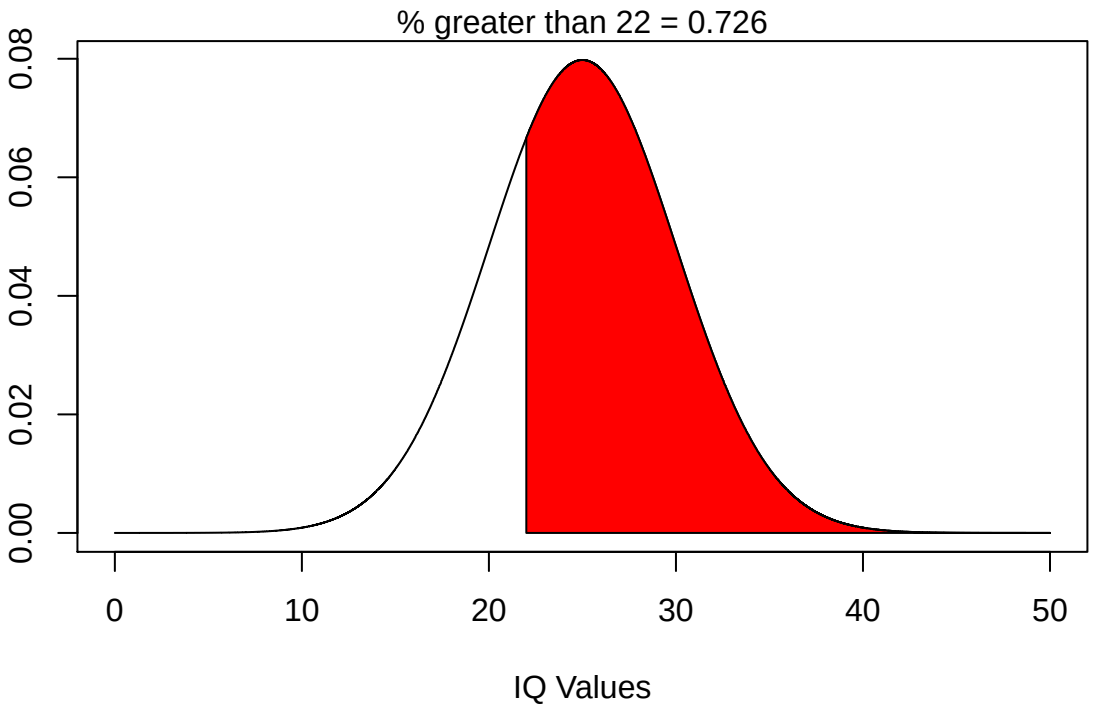



### Problem 2A

```
## First plot what we want
mean=25; sd=5
lb=22
x <- x.seq
hx <- dnorm(x,mean,sd)

plot(x, hx, type="n", xlab="IQ Values", ylab="",
  main="", axes=TRUE)

i <- x >= lb
lines(x, hx)
polygon(c(lb,x[i],50), c(0,hx[i],0), col="red")
area <- 1 - pnorm(lb, mean, sd)
result <- paste("% greater than 22 =", round(area, 3))
mtext(result,3)
```

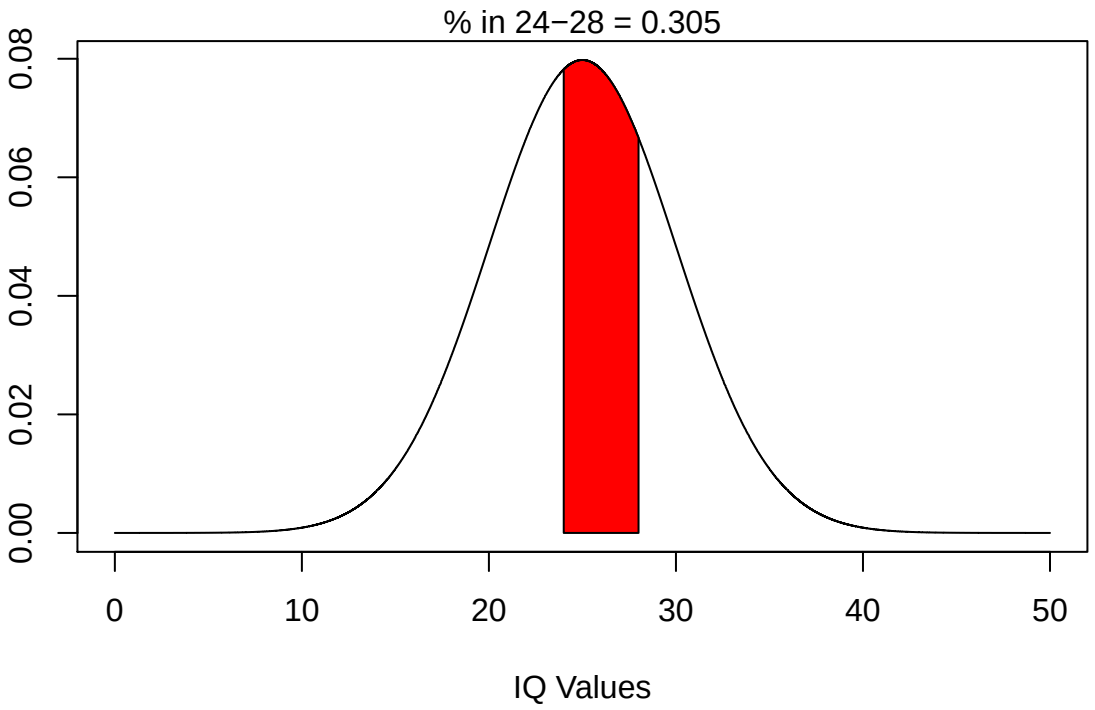



### Problem 2B

```
## First plot what we want
mean=25; sd=5
lb=24 ; ub=28
x <- x.seq
hx <- dnorm(x,mean,sd)

plot(x, hx, type="n", xlab="IQ Values", ylab="",
  main="", axes=TRUE)

i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
area <- pnorm(ub, mean, sd) - pnorm(lb, mean, sd)
result <- paste("% in 24-28 =", round(area, 3))
mtext(result,3)
```

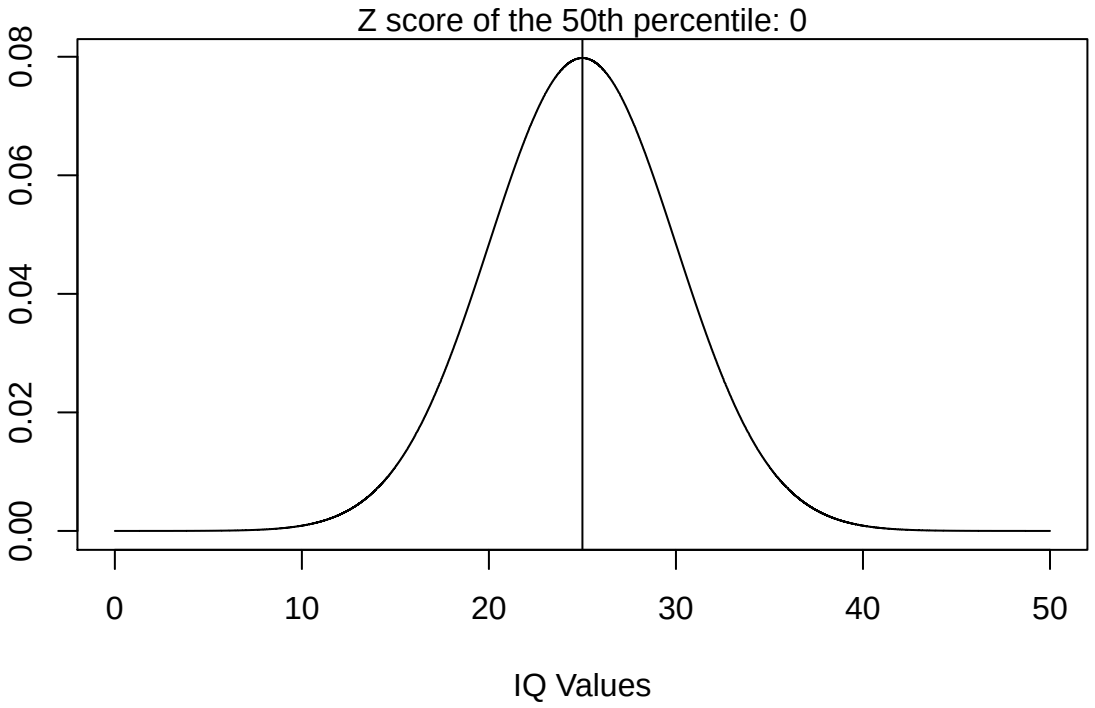



### Problem 2C

```
## First plot what we want
mean=25; sd=5
x <- x.seq
hx <- dnorm(x,mean,sd)

plot(x, hx, type="n", xlab="IQ Values", ylab="",
  main="", axes=TRUE)
lines(x, hx)
abline(v=mean)

## Now calculate the z score... of the fifty percentile.. which I know is 0
z_score_f <- qnorm(.5)
result <- paste("Z score of the 50th percentile:", z_score_f)
mtext(result,3)
```



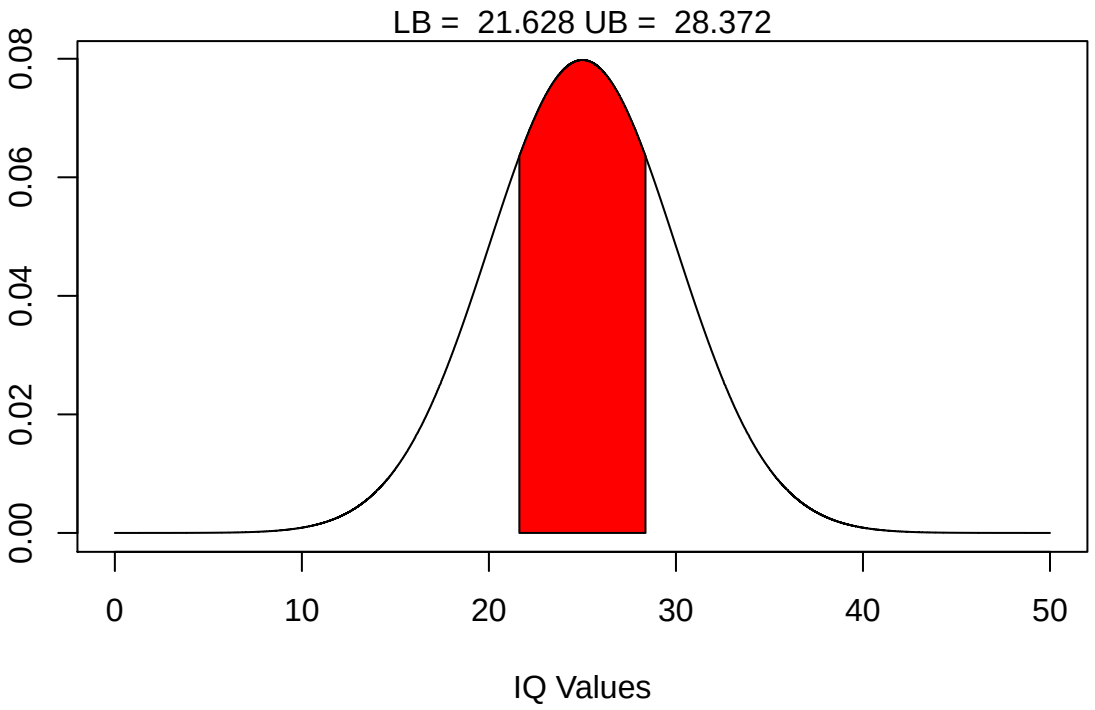

### Problem 2D

```
## First plot what we want
mean=25; sd=5
x <- x.seq
hx <- dnorm(x,mean,sd)

lb <- 25 + (qnorm(.25)*5)
ub <- 25 + (qnorm(.75)*5)

plot(x, hx, type="n", xlab="IQ Values", ylab="",
  main="", axes=TRUE)

i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
area <- pnorm(ub, mean, sd) - pnorm(lb, mean, sd)
result <- paste("LB = ", round(lb, 3), "UB = ", round(ub, 3))
mtext(result,3)
```




```
answer_p1 <- paste("The lower bound z score is: ", round(qnorm(.25),3), "The associated raw value is: "
answer_p2 <- paste("The upper bound z score is: ", round(qnorm(.75),3), "The associated raw value is: "
print(answer_p1)
```

```
## [1] "The lower bound z score is:  -0.674 The associated raw value is:  21.628"
```

```
print(answer_p2)
```

```
## [1] "The upper bound z score is:  0.674 The associated raw value is:  28.372"
```

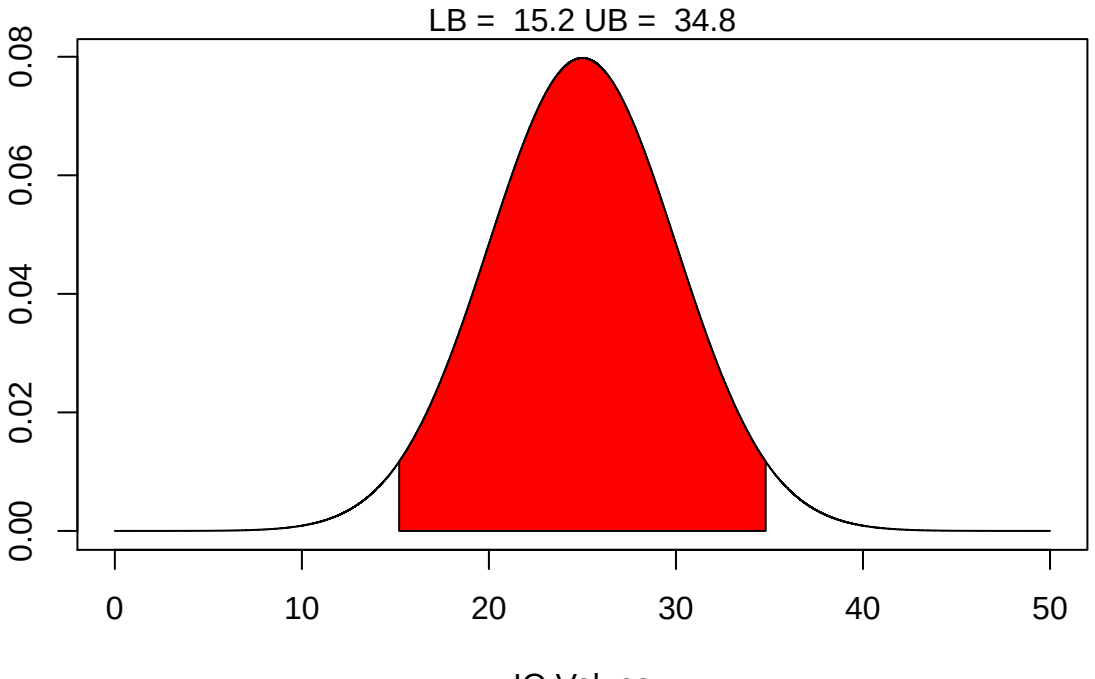
### Problem 2E

```
## First plot what we want
mean=25; sd=5
x <- x.seq
hx <- dnorm(x,mean,sd)

lb <- 25 + (qnorm(.025)*5)
ub <- 25 + (qnorm(.975)*5)

plot(x, hx, type="n", xlab="IQ Values", ylab="",
  main="", axes=TRUE)

i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")
area <- pnorm(ub, mean, sd) - pnorm(lb, mean, sd)
result <- paste("LB = ", round(lb, 3), "UB = ", round(ub, 3))
mtext(result,3)
```




```
answer_p1 <- paste("The lower bound z score is: ", round(qnorm(.025),3), "The associated raw value is:
answer_p2 <- paste("The upper bound z score is: ", round(qnorm(.975),3), "The associated raw value is:
print(answer_p1)
```

```
## [1] "The lower bound z score is:  -1.96 The associated raw value is:  15.2"
```

```
print(answer_p2)
```

```
## [1] "The upper bound z score is:  1.96 The associated raw value is:  34.8"
```