Hao-HW3

Bruce Hao

September 15, 2016

```
#setwd('/Users/brucehao/Google Drive/CUNY/git/DATA606/Lab3')
library(IS606)
library(ggplot2)
```

3.2 (see normalPlot)

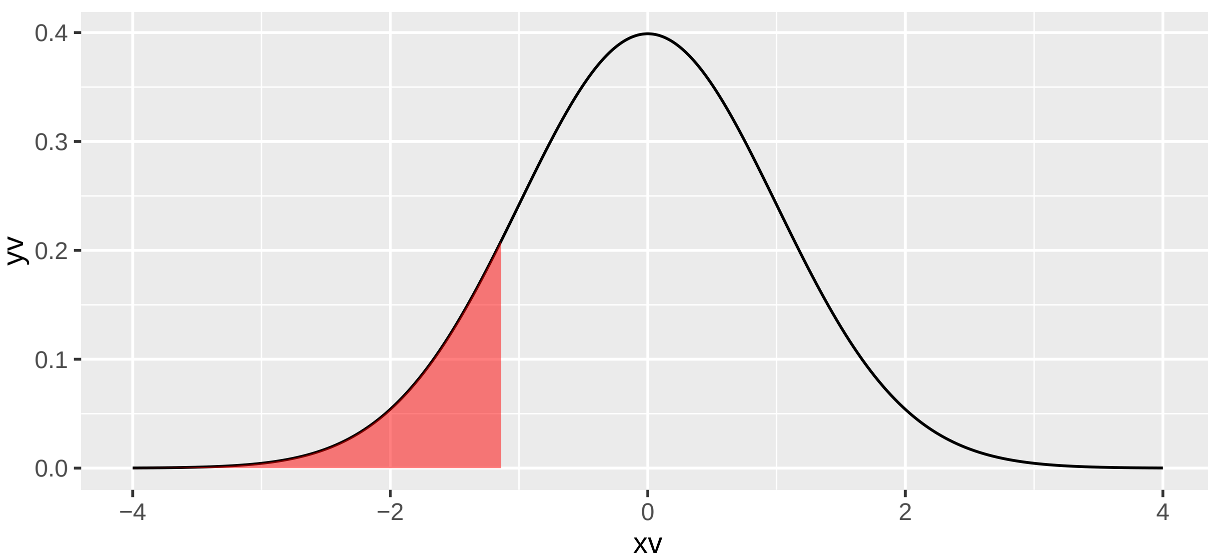
```
# setup objects for normal distribution plots
xv = seq(-4, 4, 0.01)
yv = dnorm(xv)
df = data.frame(xv, yv)
p = ggplot(df, aes(xv, yv)) + geom_line()
```

a) Z < -1.35

```
shade = df[xv < (-1.13),]
pnorm(-1.13)
```

```
## [1] 0.1292381
```

```
p + geom_ribbon(data = shade, aes(ymax=yv, ymin=0), fill = 'red', alpha = 0.5)
```

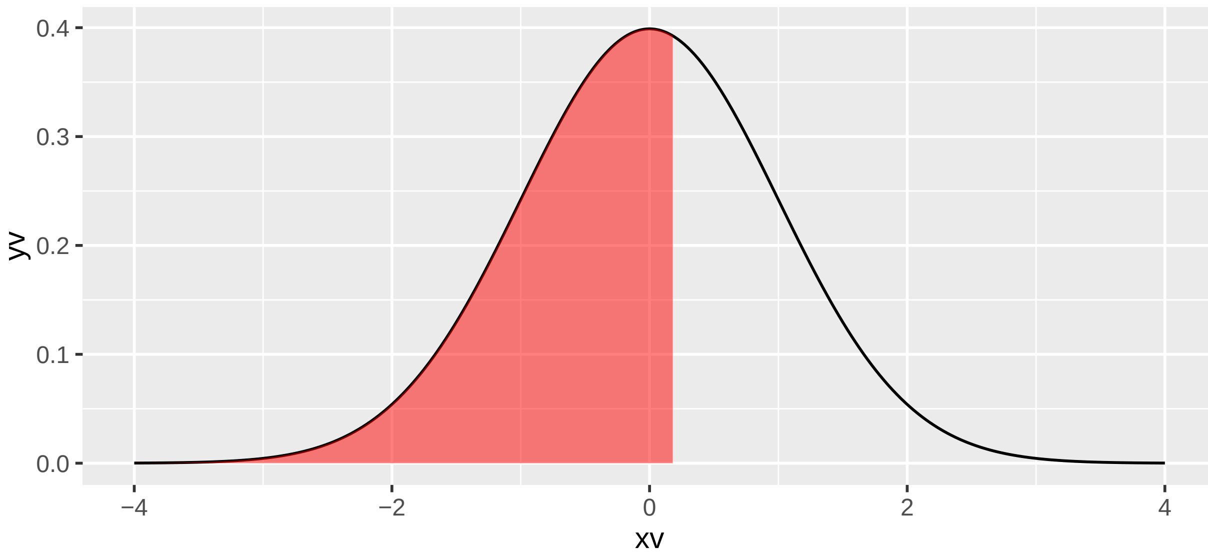


b) Z < 0.18

```r
shade = df[xv < (0.18),]
pnorm(0.18)
```

```
## [1] 0.5714237
```

```r
p + geom_ribbon(data = shade, aes(ymax=yv, ymin=0), fill = 'red', alpha = 0.5)
```

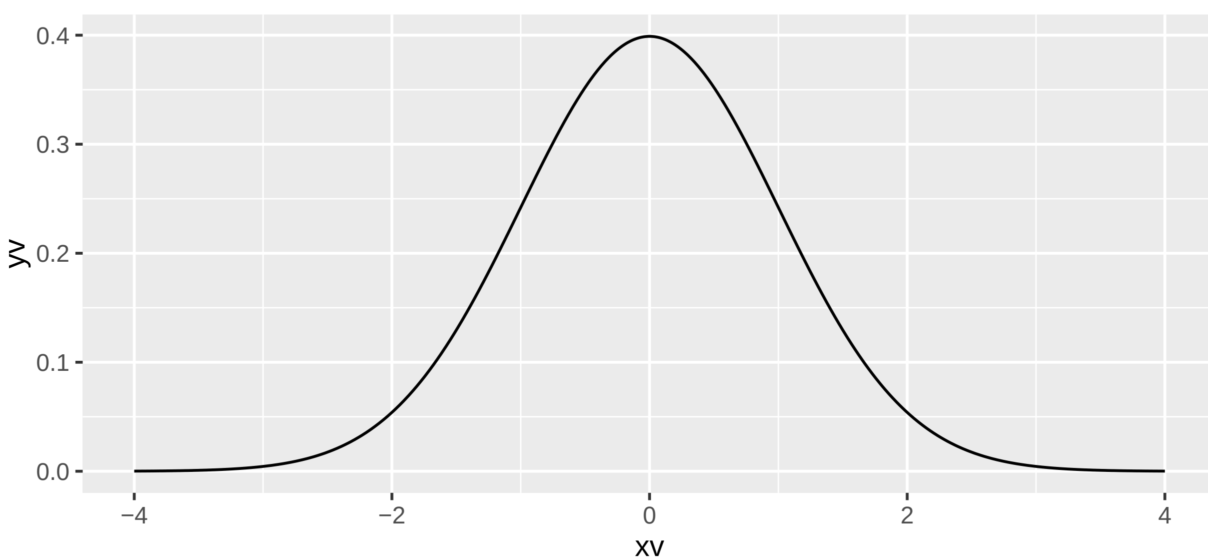


a) Z > 8

```r
shade = df[xv > 8,]
1 - pnorm(8)
```

```
## [1] 6.661338e-16
```

```r
p
```

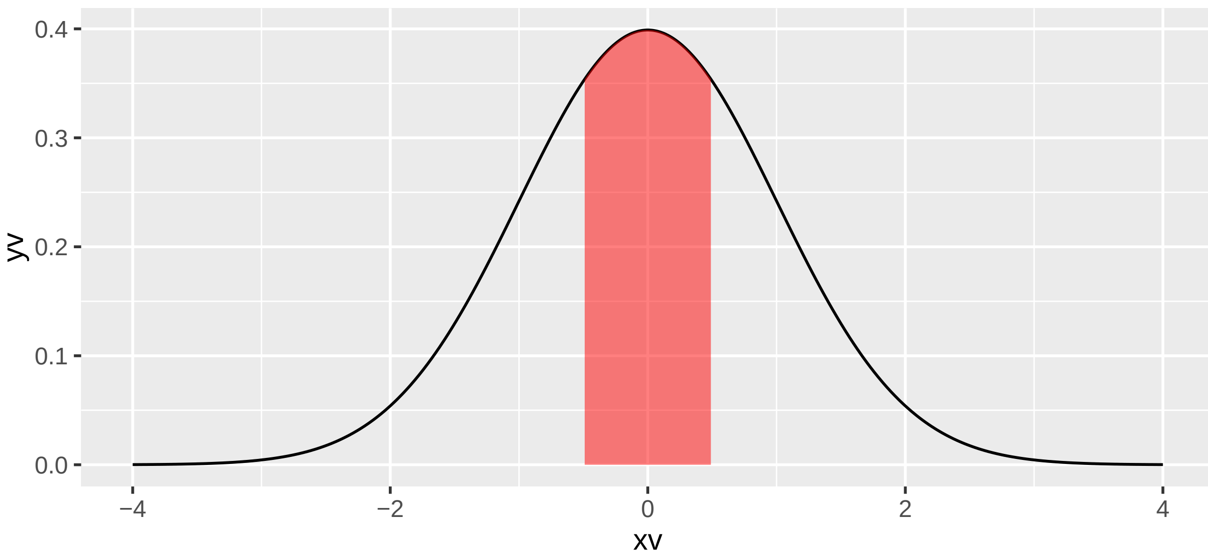


a) |Z| < 0.5

```
shade = df[xv > (-0.5) & xv < 0.5,]
pnorm(0.5) - pnorm(-0.5)
```

```
## [1] 0.3829249
```

```
p + geom_ribbon(data = shade, aes(ymax=yv, ymin=0), fill = 'red', alpha = 0.5)
```



3.4

a) N(mu=4313, sig=583); N(mu=5261, sig=807)

b) Leo's Z-score of +1.09 suggests that he was ~1 sd slower than average within his group, and Mary's Z-score of 0.31 suggests than she was ~0.3 sd slower than average within her group.

```
# Leo's Z-score
(4948 - 4313) / 583
```

```
## [1] 1.089194
```

```
# Mary's Z-score
(5513 - 5261) / 807
```

```
## [1] 0.3122677
```

c) Mary ranked better within her group because while she was slower than average, she was closer to it than Leo.

d) Leo finished faster than ~13.8% of his group.

```r
1 - pnorm(4948, 4313, 583)
```

```
## [1] 0.1380342
```

e) Mary finished faster than ~37.7% of her group.

```r
1 - pnorm(5513, 5261, 807)
```

```
## [1] 0.3774186
```

f) Potentially, yes because the probabilities at each Z-score would be different under different types of distributions.

3.18 (use qqnormsim from lab 3)

a). The results below suggest the heights are normally distributed.

```r
v = c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,65,67,67,69,73)
mean = mean(v)
sd = sd(v)
# percentage within 1 sd should approximate 68% if normally distributed
length(v[v > (mean - sd) & v < (mean + sd)]) / length(v)
```

```
## [1] 0.68
```

```r
# percentage within 2 sd should approximate 95% if normally distributed
length(v[v > (mean - sd*2) & v < (mean + sd*2)]) / length(v)
```

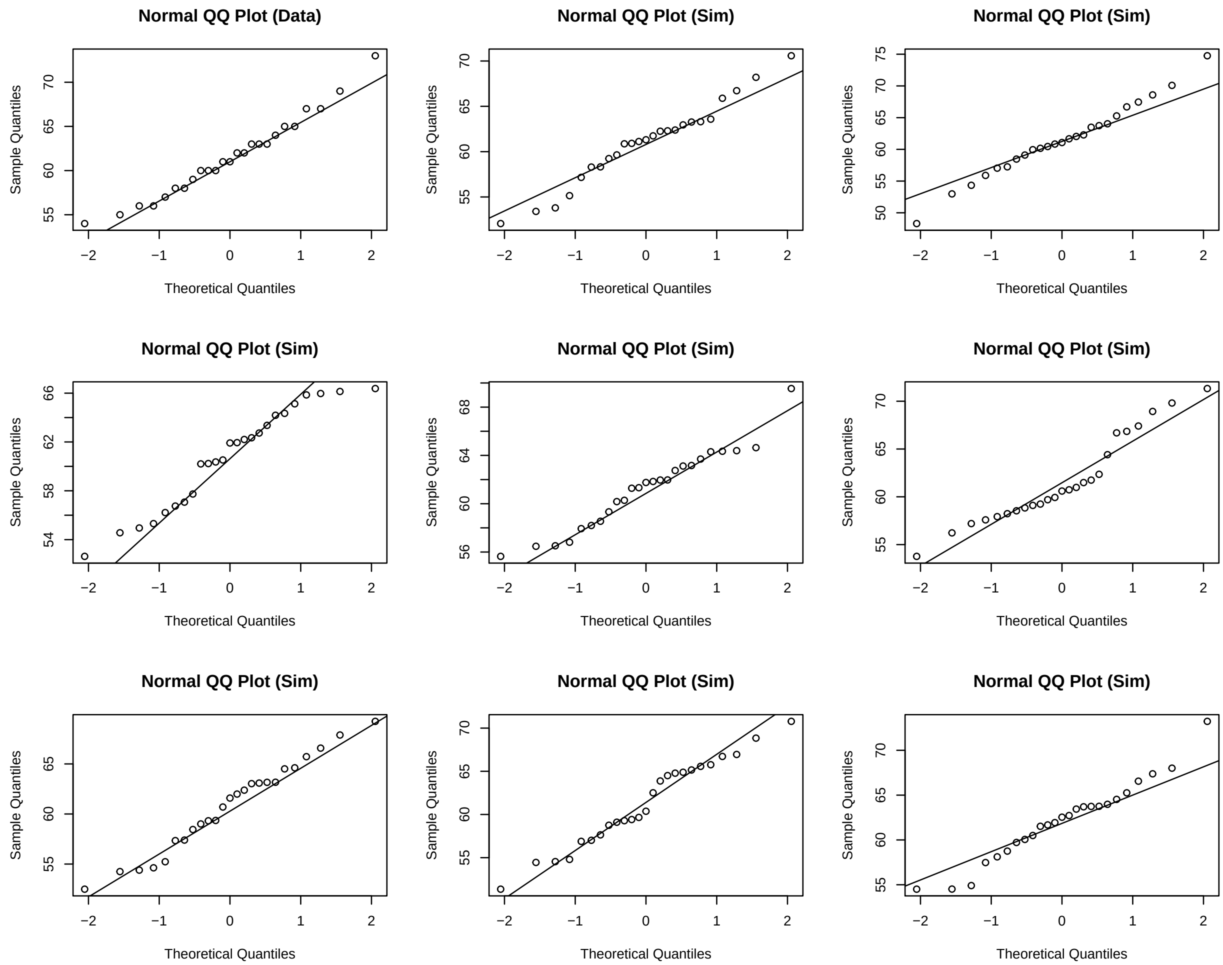
```
## [1] 0.96
```

```r
# percentage within 3 sd should approximate 99.7% if normally distributed
length(v[v > (mean - sd*3) & v < (mean + sd*3)]) / length(v)
```

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

b) The results do appear normally distributed as most of the points fall on or near the qq-plot line. Based on the simulation comparison, the actual data fit the line better than the simulated normal data.

```r
qqnormsim(v)
```



3.22

a) 1.7%

```
defect = 0.02
(1- defect)^9 * defect^1
```

```
## [1] 0.01667496
```

b) 13.3%

```
(1 - defect)^100
```

```
## [1] 0.1326196
```

c) mean wait time = 50 units; sd of wait time = 50 units

```
1 / defect
```

```
## [1] 50
```

```r
((1 - defect) / defect**2)**0.5
```

```
## [1] 49.49747
```

d) mean wait time = 20 units; sd of wait time = 20 units

```r
defect2 = 0.05
1 / defect2
```

```
## [1] 20
```

```r
((1 - defect2) / defect2**2)**0.5
```

```
## [1] 19.49359
```

e) As the defect rate increase, the mean and sd of the wait time decreases proportionally.

3.38

a) 38.2%

```r
boy = 0.51
choose(3, 2) * boy^2 * (1 - boy)
```

```
## [1] 0.382347
```

b) The 3 scenarios are shown in the columns below:

```r
combn(3, 2)
```

```
##      [,1] [,2] [,3]
## [1,]    1    1    2
## [2,]    2    3    3
```

```r
(boy * boy * (1 - boy)) + (boy * (1 - boy) * boy) + ((1- boy) * boy * boy)
```

```
## [1] 0.382347
```

c) The approach in part b would be more tedious because first all 56 combinations would need be determined, and then the 56 probabilities would need to be calculated and added up.

3.42

a) 3.9%

```r
prob = 0.15
n = 10
k = 3
choose(n-1, k-1) * prob^3 * (1-prob)^7
```

```
## [1] 0.03895012
```

b) 15% since each serve is independent.

c) Part b is asking about an isolated, independent event. Part a is asking about a specific scenario of multiple events.