DATA 606 Homework 3

Ilya Kats

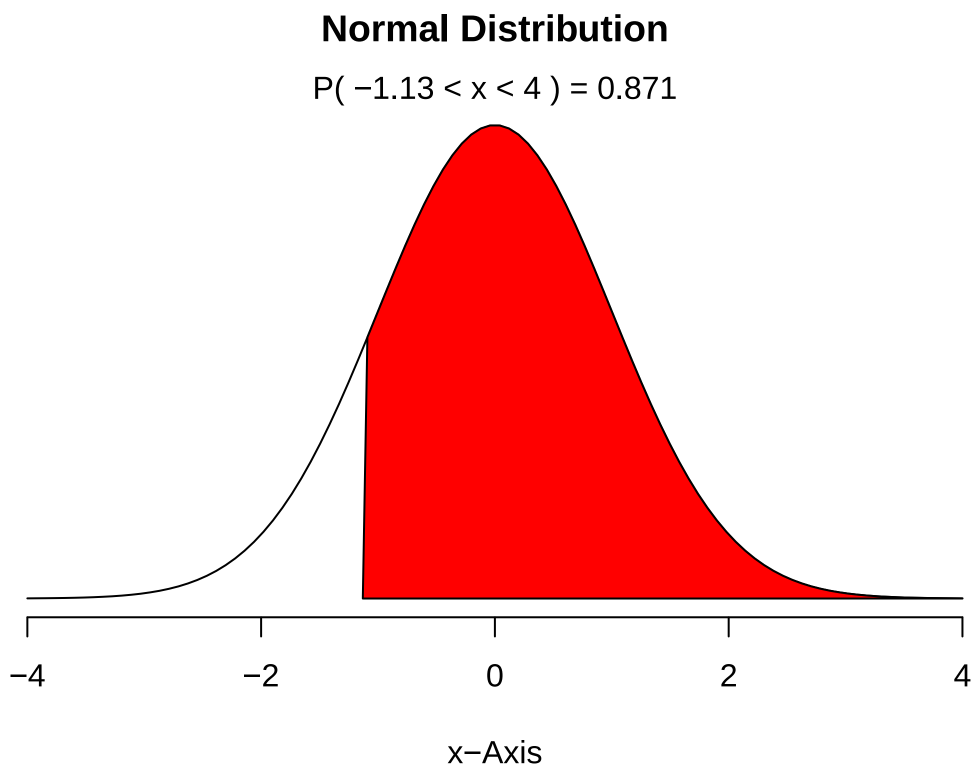
March 4, 2017

```r
library('DATA606')
```

3.2 Area under the curve, Part II.

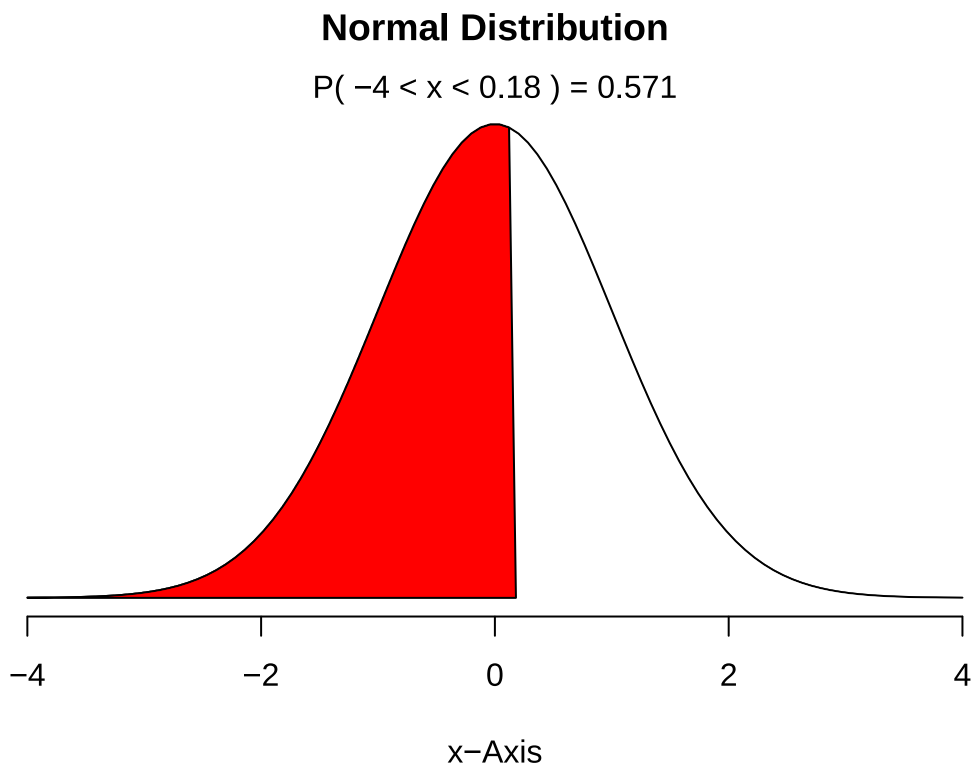
a. $Z > -1.13$: The area is $1 - 0.1292381 = 0.8707619$ or 87.08%.

```r
normalPlot(mean = 0, sd = 1, bounds=c(-1.13,4), tails = FALSE)
```



b. $Z < 0.18$: The area is 0.5714237 or 57.14%.

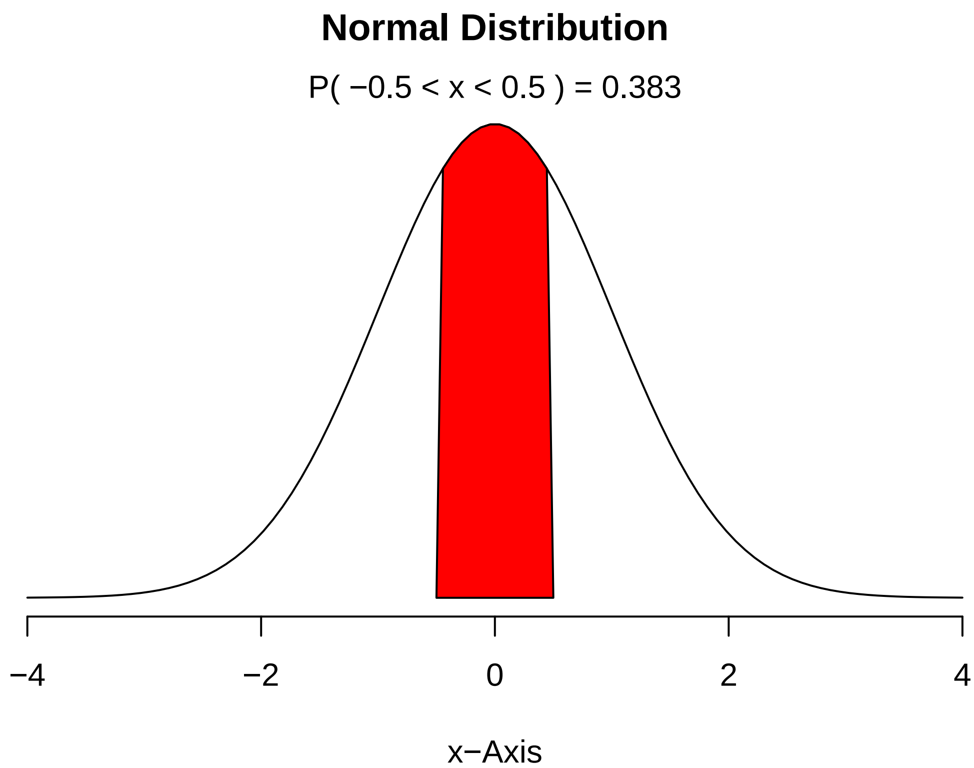
```r
normalPlot(mean = 0, sd = 1, bounds=c(-4,0.18), tails = FALSE)
```



c. $Z > 8$: The probability of a value in normal distribution being 8 standard deviations away from mean is well under 0.01%, so the area is almost nearly 0.

d. $|Z| < 0.5$: The area is $0.6914625 - 0.3085375 = 0.3829249$ or 38.3%.

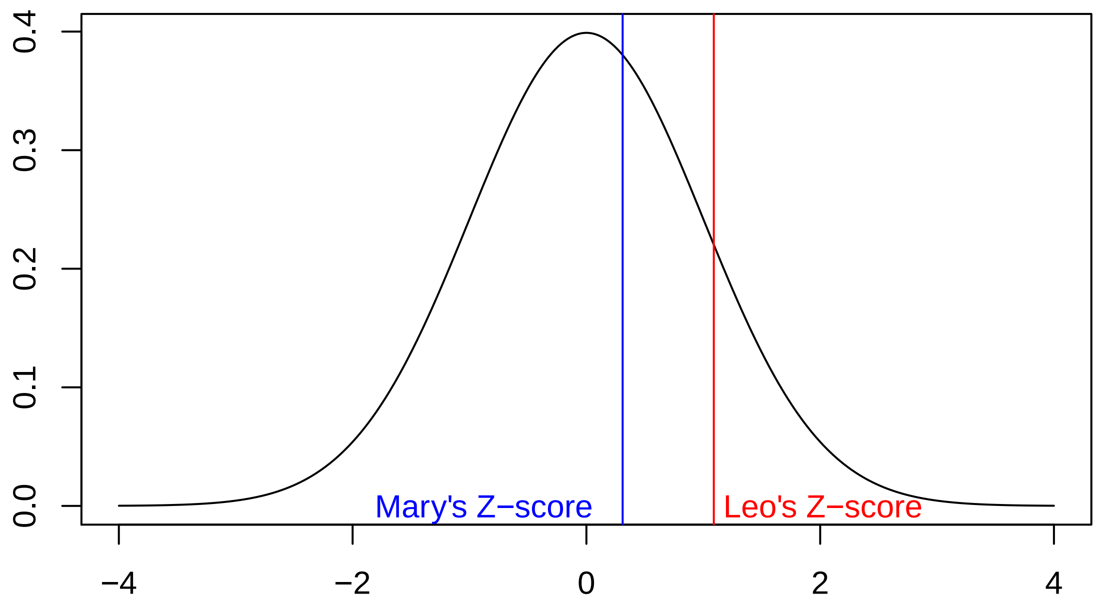
```
normalPlot(mean = 0, sd = 1, bounds=c(-0.5,0.5), tails = FALSE)
```



3.4 Triathlon times, Part I.

a. Men, Ages 30-34: $N(\mu = 4313, \sigma = 583)$, and Women, Ages 25-29: $N(\mu = 5261, \sigma = 807)$.

b. $Z_{Leo} = \frac{x-\mu}{\sigma} = \frac{4948-4313}{583} \approx 1.0892$ and $Z_{Mary} = \frac{x-\mu}{\sigma} = \frac{5513-5261}{807} \approx 0.3123$; Leo finished the race about 1.09 standad deviations above the mean, while Mary finished the race about 0.31 standard deviations above the mean.



c. Please note that because a better performance corresponds to a faster finish, lower Z-scores correspond to better performance. **Mary ranked better** in her group since her Z-score is better than Leo's.

d. Leo's Z-score corresponds to probability 0.8619672. Since higher Z-score corresponds to slower finish, Leo finished faster than $1 - 0.8619672 = 0.1380328$ or 13.8%.

e. Mary's Z-score corresponds to probability 0.6225937. Since higher Z-score corresponds to slower finish, Mary finished faster than $1 - 0.6225937 = 0.3774063$ or 37.74%.

f. If distributions are not nearly normal, then part (b) will remain the same since Z-scores can still be calculated. However, parts (d) and (e) rely on the normal model for calculations, so the results would change.

3.18 Heights of female college students.

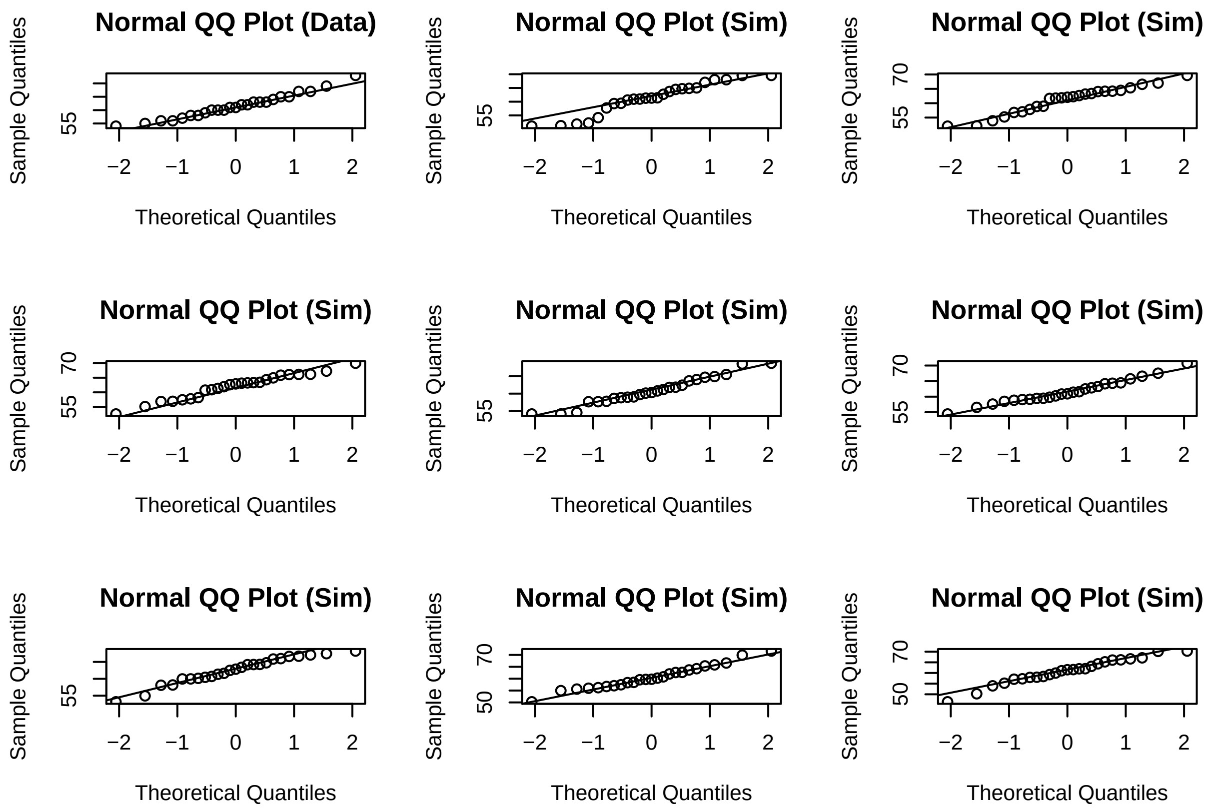
```
heights <- c(54, 55, 56, 56, 57, 58, 58, 59, 60, 60, 60, 61, 61,
             62, 62, 63, 63, 63, 64, 65, 65, 67, 67, 69, 73)
hgt_m <- mean(heights)
hgt_m
```

```
## [1] 61.52
```

```
hgt_sd <- sd(heights)
hgt_sd
```

```
## [1] 4.583667
```

```r
qqnormsim(heights)
```



Looking at the QQ plots, the plot for actual data mostly follows the line with a few outliers at the edges. It appears better than some QQ plots for simulated data with normal distribution. As such I think we can conclude that the heights data follows a normal distribution.

```r
# Values one standard deviation away from mean
pnorm(hgt_m + hgt_sd, mean = hgt_m, sd = hgt_sd) -
  pnorm(hgt_m - hgt_sd, mean = hgt_m, sd = hgt_sd)
```

```
## [1] 0.6826895
```

```r
# Values two standard deviation away from mean
pnorm(hgt_m + 2 * hgt_sd, mean = hgt_m, sd = hgt_sd) -
  pnorm(hgt_m - 2 * hgt_sd, mean = hgt_m, sd = hgt_sd)
```

```
## [1] 0.9544997
```

```r
# Values three standard deviation away from mean
pnorm(hgt_m + 3 * hgt_sd, mean = hgt_m, sd = hgt_sd) -
  pnorm(hgt_m - 3 * hgt_sd, mean = hgt_m, sd = hgt_sd)
```

```
## [1] 0.9973002
```

Using normal distribution probability, we can confirm that the heights **follow the 68-95-99.7% rule very closely**.

3.22 Defective rate.

$p = 0.02$

a. $P(10th\ transistor\ is\ the\ first\ with\ a\ defect) = (1-p)^{n-1}p = (1-0.02)^9 * 0.02 = 0.016675$

b. $P(no\ defects\ in\ a\ batch\ of\ 100) = (1-p)^{100} = 0.98^{100} = 0.1326196$

c. $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$ and $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{0.98}{0.0004}} = \sqrt{2450} = 49.4974747$

d. If $p = 0.05$, then $\mu = \frac{1}{0.05} = 20$ and $\sigma = \sqrt{\frac{0.95}{0.0025}} = 19.4935887$.

e. When probability of an event is higher, the event is more common, so the expected number of trials before it occurs and the standard deviation are lower.

3.38 Male children.

a. If $p = 0.51$, $n = 3$ and $k = 2$, then $P(two\ boys\ out\ of\ three\ kids) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \frac{3!}{2!} * 0.51^2 * 0.49 = 0.382347$.

b. Possible combinations include:

- boy, boy, girl
- boy, girl, boy
- girl, boy, boy

$P(two\ boys\ out\ of\ three\ kids)$

$= (P(boy) * P(boy) * P(girl)) + (P(boy) * P(girl) * P(boy)) + (P(girl) * P(boy) * P(boy))$

$= 3 * 0.51 * 0.51 * 0.49 = 0.382347$

c. If using approach b to calculate the probability that a couple with 8 kids has 3 boys, it is first necessary to determine the number of combinations of having 3 boys among 8 kids. The list will be significantly longer than with 3 kids (in fact, there are 56 combinations). With approach a, all that is necessary is just to plug in the numbers into the formula.

3.42 Serving in volleyball.

$p = 0.15$

a. This is a negative binomial distribution (serves are independent with each one being either a success or a failure, probability of success is the same for each serve and the last serve is a success) with $n = 10$ and $k = 3$.

$P(3rd\ success\ on\ the\ 10th\ try) = \binom{n-1}{k-1}p^k(1-p)^{n-k} = \frac{9!}{2!*7!} * 0.15^3 * 0.85^7 = 0.0389501$

b. Serves are independent events and previous outcomes have no effect on future events. The probability of the success on the 10th serve is 0.15.

c. Part a is looking for the probability of a specific combination of successes withing 10 serves. Although each serve is independent, we are considering all 10 serves in determining the probability of the desired pattern. Contrary to this part b is only concerned with one serve. Previous outcomes are irrelevant because events are independent.