pnorm and dnorm function

SMWadgymar, KGSmith

2022-08-16

In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo competed in the Men, Ages 30 - 34 group while Mary competed in the Women, Ages 25 - 29 group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they each did within their respective groups. Can you help them? Here is some information on the performance of their groups:

~The finishing times of the Men, Ages 30 - 34 group has a mean of 4313 seconds with a standard deviation of 583 seconds.

~The finishing times of the Women, Ages 25 - 29 group has a mean of 5261 seconds with a standard deviation of 807 seconds.

~The distributions of finishing times for both groups are approximately Normal.

Remember: a better performance corresponds to a faster finish.

1) What is the shorthand notation for these two normal distributions?

A: Men: N(4313, 583) Women: N(5261, 807)

2) What are the z-scores for Leo's and Mary's finishing times? What do these z-scores tell you?

```
(4948-4313)/586 #Leo, 1.089
```

```
## [1] 1.083618
```

```
(5513-5261)/807 #Mary, 0.312
```

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

A: The z-scores tell you the number of standard deviations away from the mean the observation is. It ives you a way to compare observations from different groups.

3) Did Leo or Mary rank better in their respective gropus?

A: With respect to their groups, Mary had a better time than Leo because her z-score was smaller.

4) What percent of the triathletes did Leo finish faster than in his group? That is, what percent of triathletes were slower than Leo? To answer this question, we will use the pnorm function, which will give you the cumulative probability below a specified value for a given normal distribution.

The arguments for pnorm are pnorm(q=, mean=, sd=) where q is the value of interest.

Use the pnorm function to estimate the percent of triathletes that finished slower than Leo.

```
1-pnorm(4948, mean=4313, sd=583)
```

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

5) What percent of the triathletes did Mary finish faster than in her group? In other words, what percent of triathletes were slower than Mary?

A:

```
1 - pnorm(5513, mean = 5261, sd = 807)
```

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

6) The qnorm function will report the cutoff value associated with the percentile for a given normal distribution. The arguments are:

qnorm(p=percentile, mean=, sd=, lower.tail=T)

Note that the lower.tail argument equals T by default.

Use qnorm() to calculate the finishing time that would distinguish the fastest 10% of men in the men, 31-35 triathlon. Calculate it two ways, once with lower.tail=T and once with lower.tail=F.

```
qnorm(p=0.1, mean=4313, sd=583, lower.tail=T)
```

```
## [1] 3565.855
```

```
qnorm(p=0.9, mean=4313, sd=583, lower.tail=F)
```

```
## [1] 3565.855
```

Use qnorm() to calculate the finishing time that would distinguish the slowest 25% of men in the men, 31-35 triathlon? Calculate it two ways, once with lower.tail=T and once with lower.tail=F.

```
qnorm(p=0.75, mean=4313, sd=583, lower.tail=T)
```

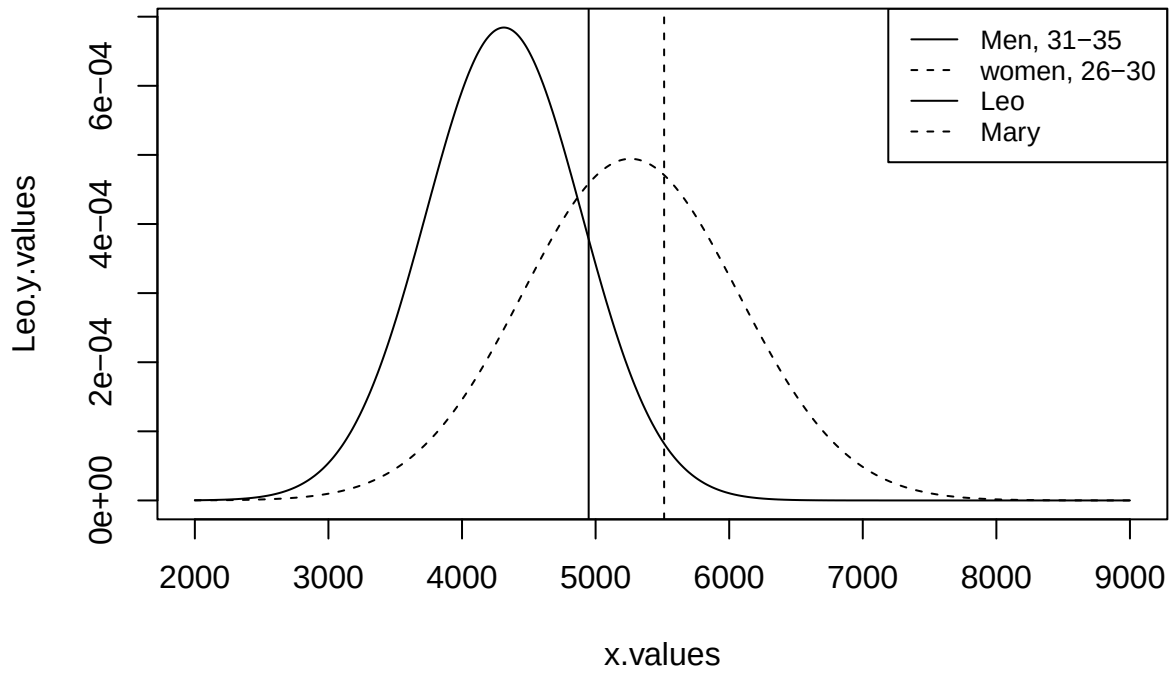
```
## [1] 4706.228
```

```
qnorm(p=0.25, mean=4313, sd=583, lower.tail=F)
```

```
## [1] 4706.228
```

7) Let's make a graph. Below each line of code is a # for you to annotate what each line is doing. Work through these lines one by one and annotate while you go.

```r
x.values=seq(2000, 9000, length=1000)
# Produces a vector with 1000 values in it called x.values that spans the range of possible times that

Leo.y.values=dnorm(x.values, mean=4313, sd=583)
# Produces exact probabilities of each value in the vector x.values given the normal distribution assoc

Mary.y.values=dnorm(x.values, mean=5261, sd=807)
# Produces exact probabilities of each value in the vector x.values given the normal distribution assoc

plot(x.values, Leo.y.values, type="l")
# Produces a probability distribution representing the men, 31-35 finishing times

#The lines() function, below, lets us add a new line to a plot. If we used plot(), it would produce a s

lines(x.values, Mary.y.values, type='l', lty=2)
# Adds a probability distribution representing the women, 26-30 finishing times

abline(v=c(4948, 5513), lty=c(1,2))
#Adds vertical lines to the plot that show where Leo's and Mary's finishing times are

legend('topright', legend=c("Men, 31-35", "women, 26-30", "Leo", "Mary"), lty=1:2, cex=0.8)
```




```r
#Adds a legend which distinguishes the probability distributions for the men, 31-55 and women, 26-30 fi
```

Q: The dnorm() function is similar to dbinom(). What do you think dnorm() is doing in the code above?

A: dnorm() will return the exact probability or denisty of a given value ofr a given normal distribution.