

DATA 606 - Homework 3

Joshua Sturm

09/11/2017

Contents

3.2 Area under the curve, Part II	1
3.4 Triathlon times, Part I	5
3.18 Heights of female college students.	6
3.22 Defective Rate	7
3.38 Male Children	8
3.42 Serving in Volleyball	8

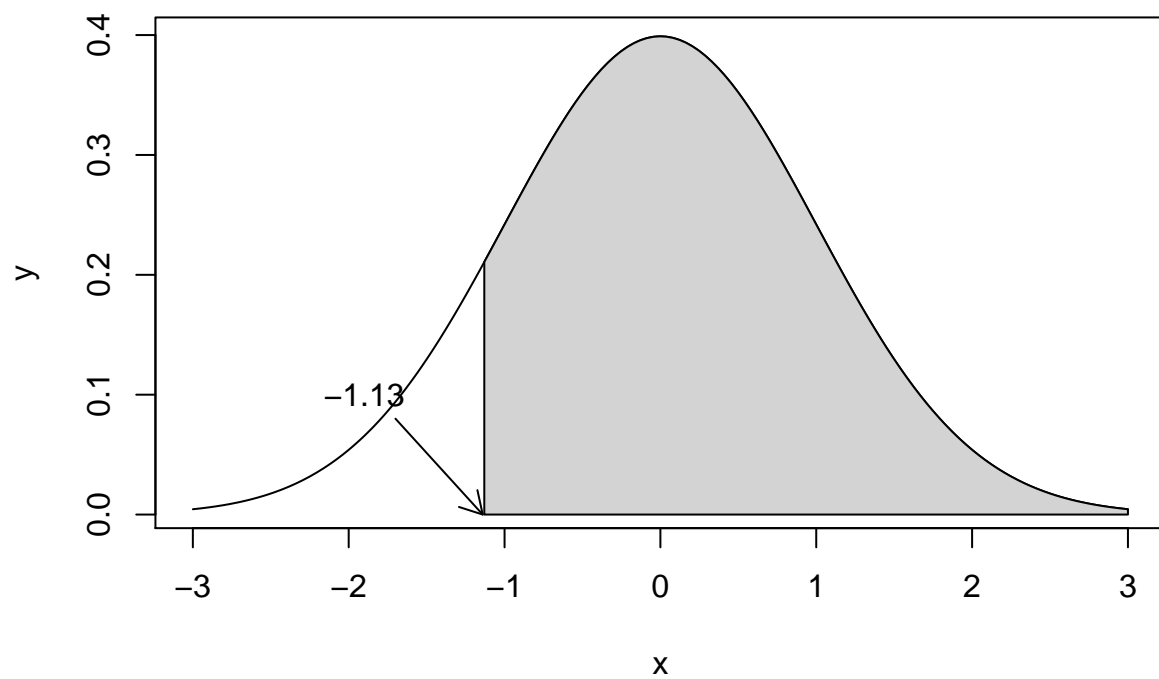
3.2 Area under the curve, Part II.

What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

(a) $Z > -1.13$ (b) $Z < 0.18$ (c) $Z > 8$ (d) $|Z| < 0.5$.

(a) $Z = \frac{x-\mu}{\sigma} \rightarrow -1.13 = \frac{x-0}{1} \rightarrow x = -1.13$.

```
x=seq(-3,3,length=500)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(-1.13,3,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(-1.13,x,3),c(0,y,0),col="lightgrey")
arrows(-1.7,0.08,-1.14,0.0,length=.13)
text(-1.9,0.1,"-1.13")
```

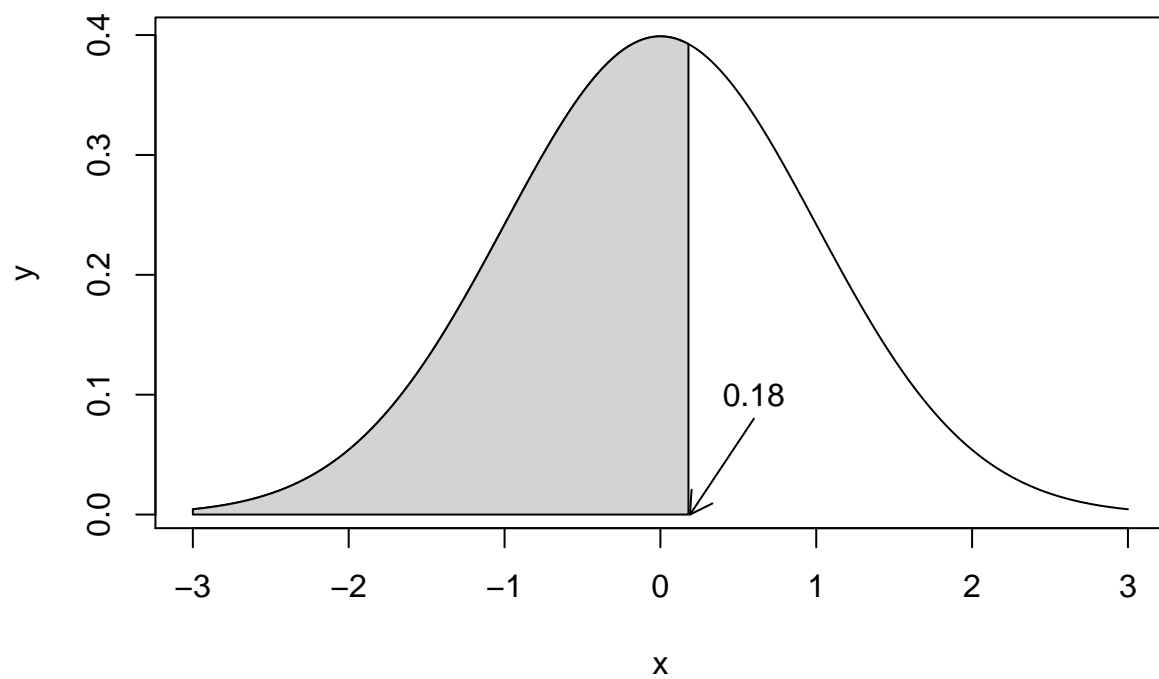


```
1-pnorm(-1.13)
```

```
## [1] 0.8707619
```

(b)

```
x=seq(-3,3,length=500)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(-3,0.18,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(-3,x,0.18),c(0,y,0),col="lightgrey")
arrows(0.6,0.08,0.19,0.0,length=.13)
text(0.6,0.1,"0.18")
```

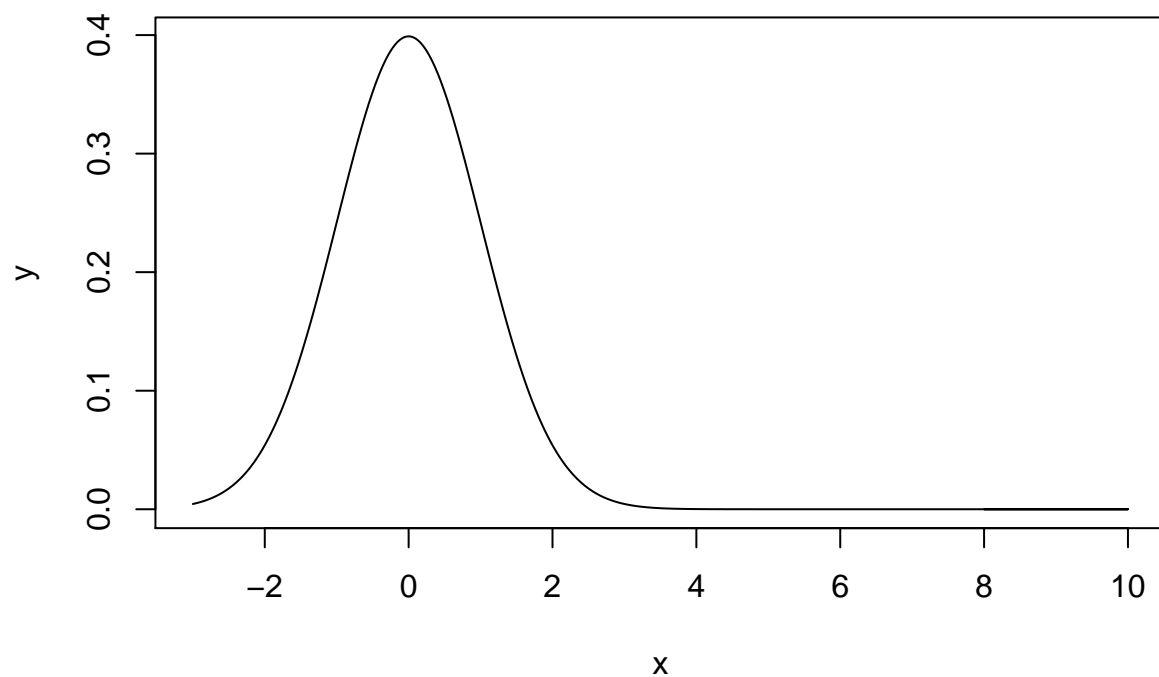


```
pnorm(0.18)
```

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

(c)

```
x=seq(-3,10,length=500)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(8,10,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(8,x,10),c(0,y,0),col="lightgrey")
```



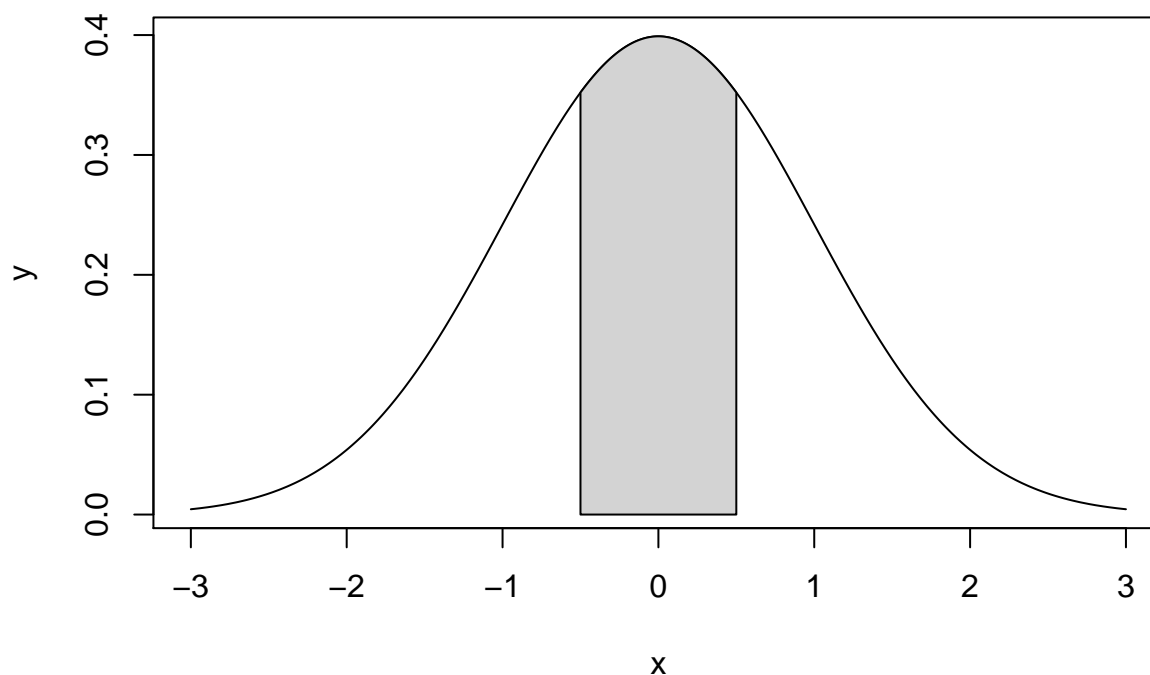
```
1 - pnorm(8)
```

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

This is an extreme example, so it barely shows on the graph.

(d)

```
x=seq(-3,3,length=500)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(-0.5,0.5,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(-0.5,x,0.5),c(0,y,0),col="lightgrey")
```



```
pnorm(0.5)
```

```
## [1] 0.6914625
```

```
pnorm(-0.5)
```

```
## [1] 0.3085375
```

3.4 Triathlon times, Part I

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 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.

- Write down the short-hand for these two normal distributions.
- What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?

- (c) Did Leo or Mary rank better in their respective groups? Explain your reasoning.
- (d) What percent of the triathletes did Leo finish faster than in his group?
- (e) What percent of the triathletes did Mary finish faster than in her group?
- (f) If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

Solutions:

(a) For the men's group: $N(\mu = 4313, \sigma = 583)$. For the women's group: $N(\mu = 5261, \sigma = 807)$.

(b) $Z = \frac{x - \mu}{\sigma}$.

$Z_L = \frac{4948 - 4313}{583} = 1.0892$. $Z_M = \frac{5513 - 5261}{807} = 0.3123$.

Leo scored 1.09 standard deviations above the mean, while Mary scored 0.31 standard deviations above the mean, each in their respective groups.

(c) Since Mary's z-score is smaller, it means she finished closer to the mean of her group, so she did better in her respective group.

(d) $P(Z_L > 1.0892)$

```
1 - pnorm(1.0892)
```

```
## [1] 0.1380328
```

~0.86 finished faster than leo, which means Leo finished faster than $1 - 0.86 = 0.13$.

(e) $P(Z_M > 0.3123)$

```
1 - pnorm(0.3123)
```

```
## [1] 0.3774063
```

Using the same method as (d), we find that Mary was faster than 0.37 of her group.

- (f) The Z-scores would remain the same, since we can calculate them in non-normal distributions. The rest of the answers would change, though, because we can't calculate probabilities and percentiles using Z-scores for non-normal distributions.

3.18 Heights of female college students.

- (a) The mean height is 61.52 inches with a standard deviation of 4.58 inches. Use this information to determine if the heights approximately follow the 68-95-99.7% Rule.
- (b) Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.

Solutions:

(a)

The 68-95-99.7% Rule states that for normal distributions, 68% of the data fall within 1 standard deviation. To verify, let's input the data:

```
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)
mean <- 61.52
sd <- 4.58
```

```
# for 68%:
x1 <- mean - sd
x2 <- mean + sd
v1 <- pnorm(x2,mean,sd) - pnorm(x1,mean,sd)
v1
```

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

```
# for 95%:
y1 <- mean - 2*sd
y2 <- mean + 2*sd
v2 <- pnorm(y2,mean,sd) - pnorm(y1,mean,sd)
v2
```

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

```
# for 99.7%:
z1 <- mean - 3*sd
z2 <- mean + 3*sd
z3 <- pnorm(z2,mean,sd) - pnorm(z1,mean,sd)
z3
```

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

We can see that the heights adhere to the 68-95-99.7 rule almost perfectly.

- (b) The histogram does closely fit the normal curve, and the normal probability plot is almost perfect.

3.22 Defective Rate

A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others.

- What is the probability that the 10th transistor produced is the first with a defect?
- What is the probability that the machine produces no defective transistors in a batch of 100?
- On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?
- Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation?
- Based on your answers to parts (c) and (d), how does increasing the probability of an event affect the mean and standard deviation of the wait time until success?

Solutions:

- (a) Since we're trying to find the *first* instance that satisfies our condition, we're dealing with a geometric distribution. The formula for Geometric distribution is: $(1 - p)^{n-1} \cdot p$.

Let $p = 0.02$ be the probability that the transistor is defective. Then, $(1 - p) = 0.98$.

$$(0.98)^9 \cdot 0.02 = 0.0167.$$

- (b) Since p is defined as the transistor being defective, we want the first 100 transistors to be working, i.e. $(1 - p)^{100}$.

$$(1 - p)^{100} = (0.98)^{100} = 0.1326.$$

- (c) For Geometric distribution, the mean (expected value) is $\frac{1}{p}$, and the standard deviation is $\sqrt{\frac{1-p}{p^2}}$.

$$\mu = \frac{1}{0.02} = 50 \quad \sigma = \sqrt{\frac{0.98}{(0.02)^2}} = 49.4975.$$

So we can expect to produce 50 working transistors before making a defective one, with a standard deviation of 49.4975.

- (d) Using the same formulas as in (c), we get:

$$\mu = \frac{1}{0.05} = 20, \quad \sigma = \sqrt{\frac{0.95}{(0.05)^2}} = 19.4936.$$

For this machine, we can expect to produce 20 working transistors before making a defective one, with a standard deviation of 19.4936.

- (e) If we increase the probability of producing a defective transistor, we can expect to produce fewer working ones before making our first defective one. Similarly, our standard deviation will shrink, as well.

3.38 Male Children

While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

- (a) Use the binomial model to calculate the probability that two of them will be boys.
- (b) Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes.

Confirm that your answers from parts (a) and (b) match.

- (c) If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a).

Solutions:

- (a) The general formula for the binomial distribution is: $\binom{n}{k}p^k(1-p)^{n-k}$. $\binom{3}{2}(0.51)^2 \cdot (0.49)^1 = 0.382347$.
- (b) There are three possible scenarios, with their corresponding probabilities:

- boy, boy, girl $0.51 \times 0.51 \times 0.49$
- boy, girl, boy $0.51 \times 0.49 \times 0.51$
- girl, boy, boy $0.49 \times 0.51 \times 0.51$

Since they are disjoint events, we can sum the probabilities:

```
x1 <- 0.51*0.51*0.49
x2 <- 0.51*0.49*0.51
x3 <- 0.49*0.51*0.51
s <- sum(x1,x2,x3)
s
```

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

- (c) If we would use the method from (b), we'd have to map out each possible permutation ($\binom{8}{3} = 56$), which is a lot of extra, unnecessary work.

3.42 Serving in Volleyball

A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other.

- (a) What is the probability that on the 10th try she will make her 3rd successful serve?
- (b) Suppose she has made two successful serves in nine attempts. What is the probability that her 10th serve will be successful?
- (c) Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be different. Can you explain the reason for this discrepancy?

Solutions:

- (a) We're dealing with a negative binomial distribution. The general formula is: $\binom{n-1}{k-1} \cdot p^k \cdot (1-p)^{n-k}$. $\binom{9}{2}(0.15)^3(0.85)^7 = 0.039$.

- (b) Since these events are independent, the success in previous attempts has no effect on subsequent serves. Therefore, the probability of success is still 0.15.

- (c) For (a), we're looking for the probability of the kth success in the nth trial, so we have to calculate the negative binomial. In (b), we're not concerned with previous successes, just the probability of success in a single scenario, so it's just p .