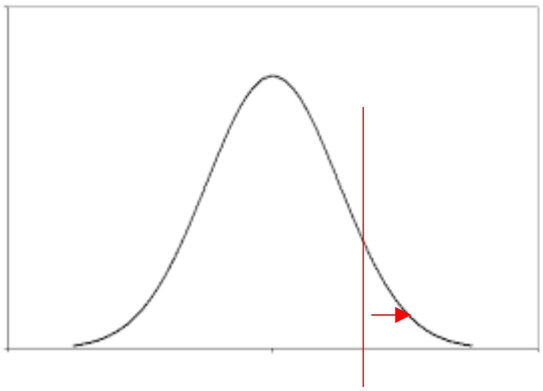
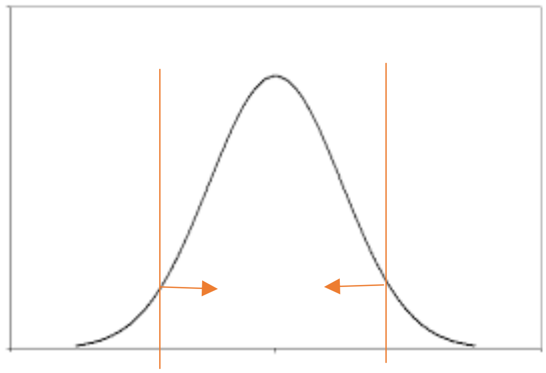
MSIS 5503 – Statistics for Data Science – Fall 2021 - Assignment 4





When asked to show probabilities on the normal distribution, draw vertical line(s) using the "Insert, Shapes" and then draw horizontal arrow(s) to show covered area(s). Example above shows P(X >x). The second example below shows probabilities for values between x1 and x2.

1. Births are approximately uniformly distributed between the 52 weeks of the year. They can be said to follow a uniform distribution from week 1 to week 53 (spread of 52 weeks).
   a. $X \sim U(1,53)$
   b. $f(X=x) = 1/(b-a) = 1/52$ for $1 \leq X \leq 53$; $x=0$ for $X \leq 1$ and $X \geq 53$
      $F(X=x) = (x-a)/(b-a) = (x-1)/52$ for $1 \leq X \leq 53$; $x=0$ for $X \leq 1$ and $X \geq 53$
      (Note: Either function acceptable as answer)
   c. $\mu = (53 + 1)/2 =$ **27**
   d. $\sigma = 52/\text{SQRT}(12) =$ **15.011**
   e. What is the probability that a person is born at the exact moment week 19 starts? That is, find $P(x = 19) =$ **0**
      (Here the answer MUST be 0; There is NO probability for an exact value of X in a continuous distribution)

      Do in R (Practice and check with calculator)
   f. $P(2 < x < 31) = (31-2)/(53-1) = 29/52 = 0.5577$
```
> print(punif(31, 1, 53) - punif(2, 1, 53))
[1] 0.5576923
```
   g. Find the probability that a person is born after week 40 = $P(X > 40) = (53-40)/(53-1) = 13/52=.25$
```
> print(1 - punif(40, 1, 53))
[1] 0.25
```
   h. $P(12 < x | x < 28) = (28-12)/(28-1) = 16/27 = .592$
```
> print((punif(28, 1, 53) - punif(12, 1, 53))/punif(28, 1, 53))
[1] 0.5925926
```
      or

```
> print(1 - punif(12, 1, 28))
[1] 0.5925926
```

i. Find the 70th percentile. (x-1) = .7*52 so x = 37.4

```
> print(qunif(0.70, 1, 53))
[1] 37.4
```

j. Find the minimum for the upper quarter. Upper quarter is 75th percentile (x-1) = .75*52 = 39, so x = week 40.

```
> print(qunif(0.75, 1, 53))
[1] 40
```

2. During the years 1998–2012, a total of 29 earthquakes of magnitude greater than 6.5 have occurred in Papua New Guinea. Assume that the time spent waiting between earthquakes is exponential.

Do in R (Practice and check with calculator)

a. What is the probability that the next earthquake occurs within the next three months?

X is the time interval (in months) between earthquakes in Papua New Guinea. There were 29 earthquakes (magnitude > 6.5) in 1998 – 2012 = 12* 15 months = 180 months. (I am also accepting 168 months or 14 years since the question was not clear. Answers in red below also acceptable).)

So, expected number of earthquakes per month = λ = 29/180.
$\lambda = 29/(12*(2012-1998)) = 0.1726$ per month

So X ~ Exponential (λ = 29/180).
X ~ Exponential (λ = 29/168).

The expected time interval between earthquakes (i.e., expected duration without earthquakes) = 180/29 months.
$P(X \leq 3) = F(3) = 1 - e\text{-}^{\lambda x} = 1 - e^{-29*3/180}$
By calculator = 0.3833
By R:

```
> print(pexp(3, 29/180))
[1] 0.3832758
```

(0.4042)

b. Given that six months has passed without an earthquake in Papua New Guinea, what is the probability that the next three months will be *free* of earthquakes?

Due to the memoryless property P(X > 6 months + 3 months| X > 6 months) = P(X > 3) = 1 – 0.3833 = 0.6167.

(0.5958)

By R:

```
> print(1 - pexp(3, 29/180 ))
[1] 0.6167242
```

c. What is the probability of zero earthquakes occurring in 2014?

We are looking for the probability of *number of earthquakes* in 12 months, which is Poisson. Now, $\lambda = 29/180$ per month, so for 12 months, $\lambda = 12*29/180$.

By calculator: $P(X = 0) = \lambda^x * e^{-\lambda}/x! = \lambda^0 * e^{-\lambda}/0! = e^{-\lambda} = e^{-12*29/180} = 0.1447$
By R:

```
> print(ppois(0, 12*29/180))
[1] 0.1446652
```

(This must be done using Poisson)
(0.126)

d. What is the probability that *at least* two earthquakes will occur in 2014?

By calculator: $1 - P(X=0) - P(X=1) = 1 - 0.1447 - (12*29/180)^1 * e^{-12*29/180} / 1! = 0.5756$
By R:

```
> print(1 - ppois(1, 12*29/180))
[1] 0.5756488
```

(This must be done using Poisson)
(0.613)

3. IQ is normally distributed with a mean of 100 and a standard deviation of 15. Suppose one individual is randomly chosen. Let X = IQ of an individual.

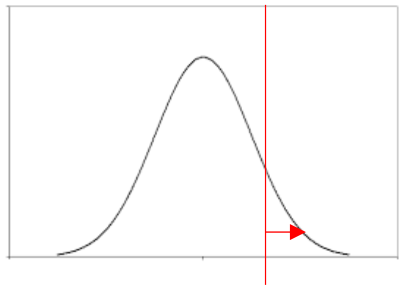
a. $X \sim N(100,15)$

<u>Do in R (Practice and Check with Standard Normal Tables)</u>

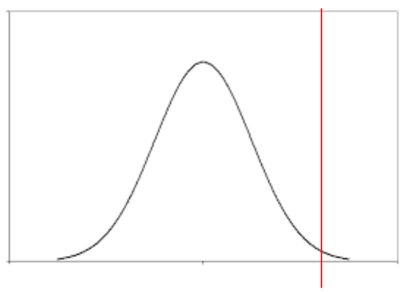
b. Write the probability statement in terms of both X and Z, show the area (for X) and find the probability that the person has an IQ greater than 120.

$P(X > 120) = P(Z > (120-100)/15) = P(Z > 1.33) = 0.0908$ (from tables)
From R:

```
> print(1 - pnorm(120, 100, 15))
[1] 0.09121122
```



c. MENSA is an organization whose members have the top 2% of all IQs. Find the minimum IQ needed to qualify for the MENSA organization. Sketch the graph by hand, and write the probability statement (in terms of both X and Z).

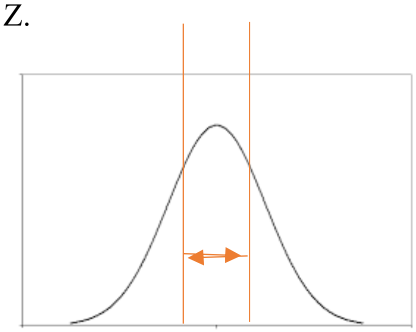


.
We need $P(X > x) = P(Z > z) = 0.02$.
That is: $P(Z \leq z) = 0.98$.

From tables: $P(Z \leq 2.06) = 0.98$.
So, $x = 100 + 2.06*15 = 130.9$

From R:
```
> print(qnorm(0.98, 100, 15))
[1] 130.8062
```

d. The middle 50% of IQs fall between what two values? Sketch the graph by and write the probability statement. Give the probability statement in terms of X and Z.



We need:
That is: $P(X \le x_2) - P(X \le x_1) = 0.50 = P(Z \le z_2) - P(Z \le z_1)$

Since this is symmetric, we first find z_2 such that $P(Z \le z_2) = 0.75$.
From tables: $z_2 = 0.675$.
Then we find $z_1$such that $P(Z \le z_1) = 0.25$.
From tables: $z_1 = - 0.675$.
So the two values are:
$x_1 = 100 -0.675*15 = 89.8765$
$x_2 = 100 +0.675*15 = 110.125$

From R:

```
> print(paste("Lower value ", qnorm(0.25, 100, 15), "Upper value ",qnorm(0.75, 100, 15)))
[1] "Lower value  89.8826537470588 Upper value  110.117346252941"
```

In terms of Z they fall between (89.883-100)/15= - 0.6745 and (110.117-100)/15 =0.6745 for $Z \sim N(0,1)$.

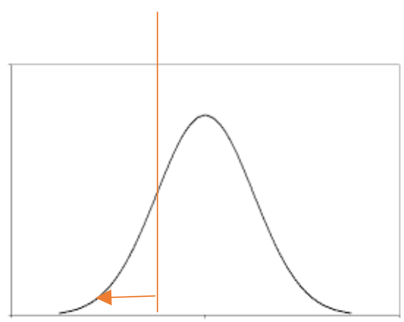
```
>   print(paste("25th percentile is  ",round(qnorm(0.25,0,1,lower.tail = TRUE,log = FALSE),4)))
[1] "25th percentile is   -0.6745"
>
>   print(paste("75th percentile is  ",round(qnorm(0.75,0,1,lower.tail = TRUE,log = FALSE),4)))
[1] "75th percentile is   0.6745"
```

4. Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet.

   a. If X = distance in feet for a fly ball, then X ~ N(250,50)

   <u>Do in R (Practice and Check with Standard Normal Tables)</u>

   b. If one fly ball is randomly chosen from this distribution, what is the probability that this ball traveled fewer than 220 feet? Write the probability statement in terms of both X and Z, show the area (for X).



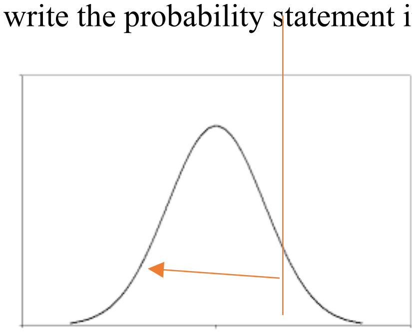
We need:

$P(X \leq 220) = P(Z \leq ((220 - 250)/50) = P(Z \leq -0.6)$
i.e., $Z \leq -0.6$.

From Tables:
$P(X \leq 220) = 0.27425$

From R:
```
> print(pnorm(220, 250, 50))
[1] 0.2742531
```

   c. Find the 80th percentile of the distribution of fly balls. Show the area (for X), and write the probability statement in terms of X and Z.



$P(X \leq x) = P(Z \leq z) = 0.8$
From tables: z = 0.845

```
>   print(paste("80th percentile is  ",round(qnorm(0.80,0,1,lower.tail = TRUE,log
 = FALSE),4)))
[1] "80th percentile is   0.8416"
```

x = 250 + 0.8416*50 = 292.08

From R:

```
> print (qnorm(0.8, 250, 50))
[1] 292.0811
```

5. A $1 scratch off lotto ticket will be a winner one out of 10 times. Out of a shipment of n = 200 lotto tickets, *using the Binomial, Poisson and Normal distributions* in each case, find the probability for the lotto tickets that there are:

   We can treat the underlying random variable as:
   Binomial(n=200, p=0.1)
   Poisson(np = 20)
   Normal(np = 20, √npq = √18)

   <u>**Do in R**</u>

   a. somewhere between 75 and 95 prizes.

   ```
   > print(pbinom(95, 200, 0.1) - pbinom(75, 200, 0.1))
   [1] 0
   > print(ppois(95, 20) - ppois(75, 20))
   [1] 0
   > print(pnorm(95, 20, sqrt(18)) - pnorm(75, 20, sqrt(18)))
   [1] 0
   ```

   b. somewhere between 15 and 25 prizes.

   ```
   > print(pbinom(25, 200, 0.1) - pbinom(15, 200, 0.1))
   [1] 0.7564673
   > print(ppois(25, 20) - ppois(15, 20))
   [1] 0.7313019
   > print(pnorm(25, 20, sqrt(18)) - pnorm(15, 20, sqrt(18)))
   [1] 0.7614072
   ```

   c. more than 50 prizes.

   ```
   > print(1 - pbinom(50, 200, 0.1))
   [1] 2.96305e-10
   > print(1 - ppois(50, 20))
   [1] 4.828738e-09
   > print(1 - pnorm(50, 20, sqrt(18)))
   [1] 7.687184e-13
   ```

   d. If a customer keeps buying tickets till she finds a winner, find the probability that her 10^{th} ticket will be a winner.

   This would be a Geometric distribution, with p = probability of "success" = 0.1.

   ```
   > print(dnbinom(9, 1, 0.1))
   [1] 0.03874205
   ```