References:

z-value percentiles : https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/bs704_probability10.html

z-score table : http://www.z-table.com/

CHAPTER 5

1. (a) What is the probability of rolling a pair of dice and obtaining a total score of 9 or more? (b) What is the probability of rolling a pair of dice and obtaining a total score of 7?

(a)
Event : Rolling a pair dice
Total number of outcomes : 6*6 = 36 outcomes
Cases where total score is 9 or more : [(6,3), (5,4), (6,4), (6,5), (6,6), (5,5), (5,6), (4,5), (4,6), (3,6)] = 10 cases

Solution : Probability of rolling a pair of dice and obtaining a total score of 9 or more = 10/36 = **0.278**

(b)
Cases where total score is equal to 7 : [(6,1), (5,2), (4,3), (3,4), (2,5), (1,6)] = 6 cases

Solution : Probability of rolling a pair of dice and obtaining a total score 7 = 6/36 = **0.167**

3. A card is drawn at random from a deck. (a) What is the probability that it is an ace or a king? (b) What is the probability that it is either a red card or a black card?

Event : Drawing a card at random from a deck
Total number of outcomes : 52 outcomes

(a)
Cases where selected card is an ace : 4
Cases where selected card is a king: 4

Probability that the selected card is an ace or a king = P(card = ace) + P(card = king) = 4 Cases where selected card is an ace : 4/52 + 4/52 = 8/52 = 0.154

Solution : 0.154

(b)
Cases where selected card is red : 26
Cases where selected card is black : 26
Probability that the selected card is red or white (technically any card in the deck) = P(color = red) + P(color = black) = 26/52 + 26/52 = 52/52 = 1

Solution : 1

5. A fair coin is flipped 9 times. What is the probability of getting exactly 6 heads?

Event : Flipping a fair coin
Number of trials N : 9
Probability of getting a head P(H) = π = ½

Probability of getting exactly 6 heads = P(x=6)

Where, $P(x)=\frac{N!}{x!(N-x)!}\pi^x(1-\pi)^{N-x}$

P(6) = (9! $(1/2)^6(1/2)^3$) / (6! 3!)
= 7*8*9*$(0.5)^9$ / 6
=7*4*3* 0.00195
=0.1641

Solution : The probability of getting exactly 6 heads in 9 flips is 0.1641

7. You flip a coin three times. (a) What is the probability of getting heads on only one of your flips? (b) What is the probability of getting heads on at least one flip?

(a)
Event : Flipping a coin
Number of trials N : 3

Probability of getting a head P(H) = π = ½

Probability of getting heads on only one of your flips = P(x=1)

$P(1) = (3!\ (0.5)^1(0.5)^2\ )\ /\ 1!\ 2!$
=3*0.125 = 0.375

Solution : Probability of getting heads on only one of your flips = 0.375

(b)
Possible outcomes = (TTT, TTH, THH, HHH, HTT, HHT, HTH, THT)
Probability of getting heads on at least one flip = 7/8 = 0.875

Solution = Probability of getting heads on at least one flip = 0.875

9. A jar contains 10 blue marbles, 5 red marbles, 4 green marbles, and 1 yellow marble. Two marbles are chosen (without replacement). (a) What is the probability that one will be green and the other red? (b) What is the probability that one will be blue and the other yellow?

(a)
Event : Choosing 2 marbles from a jar of 20 without replacement
Number of trials N : 2

Probability of first getting a green marble P(green) = 4/20
Probability of then getting a red marble P(red) = 5/19 (cus no replacement)
The probability that one will be green and the other red = P(red) * P(green)
= 4/20 x 5/19 = 1/19

Probability of first getting a red marble P(red) = 5/20
Probability of then getting a green marble P(green) = 4/19 (cus no replacement)
The probability that one will be green and the other red in this case = P(green) * P(red) = 5/20 x 4/19 = 1/19

Solution : Total probability = 2/19 = **0.1053**

(b)
Probability of first getting a blue marble P(blue) = 10/20
Probability of then getting a yellow marble P(yellow) = 1/19 (cus no replacement)

The probability that one will be blue and the other yellow = P(blue) * P(yellow) = 10/20 x 1/19 = 1/38

Probability of first getting a yellow marble P(yellow) = 1/20
Probability of then getting a blue marble P(blue) = 10/19 (cus no replacement)
The probability that one will be yellow and the other blue in this case = P(yellow) * P(blue) = 1/20 x 10/19 = 1/38

Solution : Total probability = 2/38 = **0.0526**

11. You win a game if you roll a die and get a 2 or a 5. You play this game 60 times.

(a) What is the probability that you win between 5 and 10 times (inclusive)?
(b) What is the probability that you will win the game at least 15 times?
(c) What is the probability that you will win the game at least 40 times?
(d) What is the most likely number of wins.
(e) What is the probability of obtaining the number of wins in d?

(a)
Event : Rolling a die
Total number of outcomes N : 60 outcomes
Probability of getting a 2 or a 5 = π = 2/6 = 1/3
Probability of getting a win between 5 and 10 = P(5<=x<=10) =

$$P(x)=\frac{N!}{x!(N-x)!}\pi^x(1-\pi)^{N-x}$$

= P(x<=10) – P(x<5)

```
from scipy.stats import binom

print(binom.cdf(k=10, n=60, p=0.33) - binom.cdf(k=4, n=60, p=0.33))
```

```
0.003705783586844578
```

Solution : Probability that you win between 5 and 10 times (inclusive) = **0.003706**

(b)
Probability of winning at least 15 times = P(x>=15)
= 1 - P(x<15)

```python
from scipy.stats import binom

print(1 - binom.cdf(k=14, n=60, p=0.33))
```

```
0.9301317858346959
```

Solution : Probability of winning at least 15 times = 0.9301318

(c)
Probability of winning at least 40 times = P(x>=40)
= 1 - P(x<40)

```python
from scipy.stats import binom

print(1 - binom.cdf(k=39, n=60, p=0.33))
```

```
1.0019766605307012e-07
```

Solution : Probability of winning at least 40 times = 1.001977e-07

(d)
The most likely number of wins = Expectation of x = E(x) = N. π
Solution : E(x) = 60/3 = 20

(e)
Probability of obtaining 20 wins = P(x=20)

```python
from scipy.stats import binom

print(binom.pmf(k=20, n=60, p=0.33))
```

```
0.10856155863956825
```

Solution : Probability of obtaining 20 wins = 0.1085616

13. An unfair coin has a probability of coming up heads of 0.65. The coin is flipped 50 times. What is the probability it will come up heads 25 or fewer times? (Give answer to at least 3 decimal places).

Event : Flipping an unfair coin
Total number of trials n : 50 times
Number of successes X : 25
Probability of getting a head : p = 0.65
Probability of getting a head 25 or fewer times = (x<=25)
Using normal distribution to binomial distribution,
np = 50 * 0.65 = 32.5 >= 5
n(1-p) = 50 * (1-0.65) = 50 * 0.35 = 17.5 >= 5

Thus the sample size is large enough to use the normal approximation
We know that for binomial distribution $x<=25$ the normal distribution is $x<25.5$

Thus we need to compute $P(x<25.5)$

μ $= n*p = 50*0.65 = 32.5$

σ $= \sqrt{n*p*(1-p)} = \sqrt{50*0.65*0.35} = \sqrt{11.375} = 3.3727$

$z = (x - \mu) / \sigma = (25.5 - 32.5) / 3.3727 = -2.08$

Solution : $P(x<=25) = P(z<-2.08) = 0.0188$ (by looking it up in the z score table)

15. True/False: You are more likely to get a pattern of HTHHHTHTTH than HHHHHHHHTT when you flip a coin 10 times.

Total number of outcomes N : $2^{10} = 1024$ outcomes

Probability of getting a pattern HTHHHTHTTH = 1/1024
Similarly, the Probability of getting a pattern HHHHHHHHTT = 1/1024

Since both the probabilities are the same,
Solution : False, the chances of getting a pattern of HTHHHTHTTH are as likely as getting the pattern HHHHHHHHTT

CHAPTER 7

2. (a) What are the mean and standard deviation of the standard normal distribution? (b) What would be the mean and standard deviation of a distribution created by multiplying the standard normal distribution by 8 and then adding 75?

(a)
Mean of standard normal distribution = **0**
S.D of the standard normal distribution = **1**
(b)
New mean after multiplying the standard normal distribution by 8 and then adding 75 = $b\mu+A = (0*8) + 75 =$ **75**
New standard deviation = $b\sigma = 8*1 =$ **8**

4. (a) What proportion of a normal distribution is within one standard deviation of the mean? (b) What proportion is more than 2.0 standard deviations from the mean? (c) What proportion is between 1.25 and 2.1 standard deviations above the mean?

(a)
$P(Z<1) - P(Z<-1)$
$= 0.8413 - 0.1587 = 0.6826 = 68.26\%$
Solution : 68% of values lie within one S.D of the mean

(b)
$1 - (P(Z<2) - P(Z<-2))$
$= 1 - (0.97725 - 0.02275) = 1 - 0.9545 = 0.0455 = 4.55\%$
Solution : 5% of values lie outside 2 S.Ds from the mean

(c)
$P(Z<2.1) - P(Z<1.25)$
$=0.98214 - 0.89435 = 0.08779 = 8.78\%$
Solution : 9% of values lie outside between 1.25 and 2.1 S.Ds above the mean

6. Assume a normal distribution with a mean of 70 and a standard deviation of 12. What limits would include the middle 65% of the cases?

Given :

Mean: 70
Standard deviation: 12

Limits that include 65% of the cases :

$P(a<X<b) = 0.65$
where a and b are the lower and upper limit, respectively.

$P(-c<Z<c)=0.65$
$c = (1 - 0.65)/2 = 0.175$

looking it up in the z score table we have,
$P(-0.9346<Z<0.9346)=0.65$

We know that,
$Z = (X-\mu) / \sigma$
$X = Z \sigma + \mu$

For the lower limit, where X = a
$a = \mu - (0.9346 * \sigma)$
$= 70 - (0.9346 * 12)$
$= 58.785$

For the upper limit, we set X = b
$b = \mu + (0.9346 * \sigma)$
$= 70 + (0.9346 * 12)$
$= 81.215$

Therefore, $P(58.785<X<81.215) = 0.65$.

Solution : The upper and lower limits are (58.785, 81.215)

8. Assume the speed of vehicles along a stretch of I-10 has an approximately normal distribution with a mean of 71 mph and a standard deviation of 8 mph.

a. The current speed limit is 65 mph. What is the proportion of vehicles less than or equal to the speed limit?
b. What proportion of the vehicles would be going less than 50 mph?
c. A new speed limit will be initiated such that approximately 10% of vehicles will be over the speed limit. What is the new speed limit based on this criterion?
d. In what way do you think the actual distribution of speeds differs from a normal distribution?

Given :

Mean: 71
Standard deviation: 8

(a)
$P(X<=65)$:
$z = (X-\mu) / \sigma$
$z = (65 - 71) / 8 = -0.75$

Solution : $P(X<=65) = P(x<=-0.75) = 0.2266$ (from the z-score table)

(b)
$P(X<50)$:
$z = (50 - 71) / 8 = -2.625$

Solution : P(X<50) = **P(x<-2.63) = 0.0043** (from the z-score table)

(c)
If X is the new speed limit,
P(Z>X) = 1 – P(Z<=X) = 0.1
P(Z<=X) = 1 – 0.1 = 0.9

Z score of 90th percentile = 1.282
z = (X–μ) / σ
and hence X = z σ + μ
X = (1.282 * 8) + 71 = 81.256

Solution : The new speed limit = 81.256 mph

(d)
Solution :
The actual distribution of speeds differs from the normal distribution in that
i. It is not symmetric. The speeds are not symmetrically distributed about the central value. Two speeds that are an equal amount greater or lower than the mean speed, may
not have the same frequency as is required for a normal distribution.
ii. Secondly, the speeds further away from the average speed do not drop in frequency as rapidly as in normal distribution hence resulting in a fatter tail.
iii. There is a limitation that speed will always be positive.

10. You want to use the normal distribution to approximate the binomial distribution. Explain what you need to do to find the probability of obtaining exactly 7 heads out of 12 flips.

Probability of getting a head = p = ½
Total number of trials = n = 12
To find : P(X = 7)

Since the normal distribution is continuous, we need to use continuity correction
P(X = 7) = P(6.5<X<7.5)

We know that z = (X–μ) / σ
μ = np = 12/2 = 6
σ = √n*p*(1-p) = √12*0.5*0.5 = √3

P(6.5<X<7.5) = P((6.5 – 6)/ √3 < Z < (7.5 – 6)/ √3)

$= P(0.2887 < Z < 0.8660)$
$= P(Z < 0.87) - P(Z < 0.29)$
$= 0.8078 - 0.6141 = 0.1937$

Solution : Probability of obtaining exactly 7 heads out of 12 flips = 0.1937

12. Use the normal distribution to approximate the binomial distribution and find the probability of getting 15 to 18 heads out of 25 flips. Compare this to what you get when you calculate the probability using the binomial distribution. Write your answers out to four decimal places.

Event : Flipping a coin 25 times
Probability of getting a head : ½

To find : $P(15>=X>=18)$
In normal approximation, $P(14.5 > X > 18.5) = P(x<18.5) - P(x<14.5)$

We know that $z = (X-\mu) / \sigma$
$\mu = np = 25/2 = 12.5$
$\sigma = \sqrt{n*p*(1-p)} = \sqrt{25*0.5*0.5} = \sqrt{6.25} = 2.5$

$P(x<18.5) - P(x<14.5) = P (Z < (18.5-12.5)/2.5) - P (Z < (14.5-12.5)/2.5)$
$= P (Z < 2.4) - P (Z < 0.8)$
$= 0.9918 - 0.7881 =$ **0.2037**

The normal approximation of the binomial result = 0.2037

The binomial distribution is $p(x) = c(n,x) * p^x * q^{(n-x)}$
For x = 15 to 18,

$p(15) = (25!/15!*10!) * .5^{15} * .5^{10} = .0974166393$
$p(16) = (25!/16!*9!) * .5^{16} * .5^{9} = .0608853996$
$p(17) = (25!/17!*8!) * .5^{17} * .5^{8} = .0322334468$
$p(18) = (25!/18!*7!) * .5^{18} * .5^{7} = .0143259764$

$p(15) + p(16) + p(17) + p(18) = 0.2048614621$

The binomial distribution result = 0.2049

Solution : The binomial distribution result 0.2049 is slightly larger than the normal approximation of the binomial result 0.2037 but the results are quite comparable.

22. The following question uses data from the Angry Moods (AM) case study. For this problem, use the Anger Expression (AE) scores. (a) Compute the mean and standard deviation. (b) Then, compute what the 25th, 50th and 75th percentiles would be if the distribution were normal. (c) Compare the estimates to the actual 25th, 50th, and 75th percentiles.

(a)

```
# import libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy
import scipy

# Importing the statistics module
import statistics

# read by default 1st sheet of an excel file
angry_moods = pd.read_excel('angry_moods.xls')

Mean = statistics.mean(angry_moods['Anger_Expression'])
SD = statistics.stdev(angry_moods['Anger_Expression'])

print("The mean of the Anger Expression scores is :", Mean)
print("The standard deviation of the Anger Expression scores is :", SD)
```

```
The mean of the Anger Expression scores is : 37
The standard deviation of the Anger Expression scores is : 12.94142648553549
```

(b)

If the distribution is normal, $z = (X-\mu) / \sigma$, we know the following z values (reference mentioned at the top of the document)

Z value for 25th percentile = **-0.675**

Z value for 50th percentile = **0**

Z value for 75th percentile = **0.675**

(c)

Actual estimates of the 25th, 50th, and 75th percentiles :

```python
# import libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy
import scipy

# Importing the statistics module
import statistics

# read by default 1st sheet of an excel file
angry_moods = pd.read_excel('angry_moods.xls')

Mean = statistics.mean(angry_moods['Anger_Expression'])
SD = statistics.stdev(angry_moods['Anger_Expression'])

print("The mean of the Anger Expression scores is :", Mean)
print("The standard deviation of the Anger Expression scores is :", SD)
```

```
The mean of the Anger Expression scores is : 37
The standard deviation of the Anger Expression scores is : 12.94142648553549
```

Solution : The difference in the percentile values indicates that the distribution is not normal.

CHAPTER 9

1. A population has a mean of 50 and a standard deviation of 6. (a) What are the mean and standard deviation of the sampling distribution of the mean for N = 16? (b) What are the mean and standard deviation of the sampling distribution of the mean for N = 20?

(a)
The mean of the sampling distribution of the mean is the mean of the population from which the scores were sampled
Mean = 50
Standard deviation of the sampling distribution of the mean is S.D of the population divided by the square root of the sample size = $\sigma/\sqrt{n} = 6/\sqrt{16} = 6/4 =$ **1.5**

(b)
Mean of the sample = Mean of the population = **50**
Standard deviation of the sampling distribution of the mean = $6/\sqrt{20} =$ **1.3416**

3. What term refers to the standard deviation of the sampling distribution?

The **standard error** refers to the standard deviation of the sampling distribution of the mean.

5. A questionnaire is developed to assess women's and men's attitudes toward using animals in research. One question asks whether animal research is wrong and is answered on a 7-point scale. Assume that in the population, the mean for women is 5, the mean for men is 4, and the standard deviation for both groups is 1.5. Assume the scores are normally distributed. If 12 women and 12 men are selected randomly, what is the probability that the mean of the women will be more than 1.5 points higher than the mean of the men?

The distribution of the differences between means is the sampling distribution of the difference between means.

First compute the difference between the means = 5-4 = 1
The standard error difference is then the square root of the sum of each variance over each sample size.
$= \sqrt{(1.5)^2/12 + (1.5)^2/12} = 1.5 * \sqrt{1/6} = 0.6124$
$z = (X-\mu) / \sigma$
Compute Z when x = 1.5,
$Z = (1.5 - 1) / 0.6124$
Z = 0.82 which has a score of 0.7939
$P(x<1.5) = 0.7939$

$P(x>=1.5) = 1 - P(x<1.5) = 1 - 0.7939 = 0.2061$

Solution : 20.61% probability that the mean of the women will be more than 1.5 points higher than the mean of the men

7. If numerous samples of N = 15 are taken from a uniform distribution and a relative frequency distribution of the means is drawn, what would be the shape of the frequency distribution?

Solution : The Central Limit Theorem states that the relative frequency distribution of the means approaches a **normal distribution.**

9. What is the shape of the sampling distribution of r? In what way does the shape depend on the size of the population correlation?
Solution : The shape of the sampling distribution is **skewed** and the distribution is negatively skewed if the correlation is positive and the distribution in positively skewed if the correlation is negative. The shape

depends on the size of the population correlation in such a way that, **greater the value of population correlation ρ, the more pronounced the skew.**

11. A variable is normally distributed with a mean of 120 and a standard deviation of 5. Four scores are randomly sampled. What is the probability that the mean of the four scores is above 127?

Mean = 120
S.D = 5
Standard error = $5/\sqrt{4} = 5/2$

$z = (X-\mu) / \sigma$
compute z for X = 127
$= (127 - 120) / (5/2)$
$= 14/5 = 2.8$

$P(x>127) = 1- P(x<127)$
$1- P(Z < 2.8) = 1 - 0.9974 = 0.0026$

Solution : The probability that the mean of the four scores is above 127 is **0.26%**

13. The mean GPA for students in School A is 3.0; the mean GPA for students in School B is 2.8. The standard deviation in both schools is 0.25. The GPAs of both schools are normally distributed. If 9 students are randomly sampled from each school, what is the probability that:

(a) the sample mean for School A will exceed that of School B by 0.5 or more?
(b) the sample mean for School B will be greater than the sample mean for School A?

(a)
Mean of population A = 3
Mean of population B = 2.8
S.D of population A = 0.25
S.D of population B = 0.25
Sample size of A = 9
Sample size of B = 9

Mean difference = 3 – 2.8 = 0.2
Standard error of means = $\sqrt{(0.25^2/9 + 0.25^2/9)} = 0.1179$

$z = (X–\mu) / \sigma$
$z = (0.5 – 0.2)/0.1179$

The sample mean for School A will exceed that of School B by 0.5 or more = P (X>=0.5) = P(Z >= 2.55) = 1 – P(Z<2.55) = 1 – 0.9946 = 0.0054

Solution : The sample mean for School A will exceed that of School B by 0.5 or more = 0.54%

(b)
Probability that the sample mean for School B will be greater than the sample mean for School A = $P(B_s > A_s)$
$= P((B_s - A_s) > 0)$
= P(X>0)
Mean of Sample B – Mean of Sample A = 2.8 – 3 = -0.2
$z = (X–\mu) / \sigma$
$z = (0 – (-0.2))/0.1179 = 1.7$

P(X>0) = P(Z>1.7) = 1 – P(Z<=1.7)
= 1 – 0.9554 = 0.0446

Solution : Probability that the sample mean for School B will be greater than the sample mean for School A = 4.46%

15. When solving problems where you need the sampling distribution of r, what is the reason for converting from r to z'?

Solution : We convert r to z' in order to make it normally distributed with a known standard error.

17. True/false: The standard error of the mean is smaller when N = 20 than when N = 10.

Standard error of the mean = The standard error of the mean is the square root of the sum of each variance over each sample size.
Assuming var1 = 1and var2 = 1 with different sample sizes N1 = 20 and N2 = 10,
Standard error1 = √(1/20 + (1/20) = √1/10 = 0.31622
Standard error 2 = √(1/10) + (1/10) = √1/5 = 0.4472
Solution : True, the standard error of the mean is smaller when N = 20 then when N = 10

19. True/false: You choose 20 students from the population and calculate the mean of their test scores. You repeat this process 100 times and plot the distribution of the means. In this case, the sample size is 100.

Solution : False, even though the process is repeated 100 times, the sample size remains the same (i.e., 20 students)

21. True/false: The median has a sampling distribution.

Solution : True, all statistics have sampling distributions

23. In the Angry moods case study, (a) How many men were sampled? (b) How many women were sampled?

```
# import libraries
import pandas as pd
import numpy as np


# read by default 1st sheet of an excel file
angry_moods = pd.read_excel('angry_moods.xls')

angry_moods["Gender_Desc"] = np.where(angry_moods["Gender"].isin([1]), 'Male', 'Female')
angry_moods.groupby(['Gender_Desc'])['Gender_Desc'].count()
```

```
Gender_Desc
Female    48
Male      30
Name: Gender_Desc, dtype: int64
```

Solution : The number of males being sampled = 30. The number of females being sampled = 48

24. What is the mean difference between men and women on the Anger-Out scores?

```
# import libraries
import pandas as pd
import numpy as np


# read by default 1st sheet of an excel file
angry_moods = pd.read_excel('angry_moods.xls')

angry_moods["Gender_Desc"] = np.where(angry_moods["Gender"].isin([1]), 'Male', 'Female')
angry_moods.groupby(['Gender_Desc'])['Anger-Out'].mean()
```

```
Gender_Desc
Female    15.770833
Male      16.566667
Name: Anger-Out, dtype: float64
```

```
print("The mean difference between men and women is : ", 16.566667-15.770833)
```

```
The mean difference between men and women is :  0.7958339999999993
```

Solution : The mean difference between men and women on the Anger-Out scores = 0.7958

25. Suppose in the population, the Anger-Out score for men is two points higher than it is for women. The population variances for men and women are both 20. Assume the Anger-Out scores for both genders are normally distributed. Given this information about the population parameters:

(a) What is the mean of the sampling distribution of the difference between means?

The mean of the sampling distribution of the difference between means is nothing but the difference between sampling means.
Given : The Anger-Out score for men is two points higher than it is for women.
Hence the difference between means = 2

Solution : 2

(b) What is the standard error of the difference between means?

Standard error of the mean = The standard error of the mean is the square root of the sum of each variance over each sample size.

$= \sqrt{(20/30) + (20/48)} = 1.0408$

Solution : The standard error of the difference between means = 1.0408

(c) What is the probability that you would have gotten this mean difference (see #24) or less in your sample?

From #24, we know that mean difference = 0.7958
Mean of the sampling distribution = 2
Standard error of the mean = 1.0408
$z = (X - \mu) / \sigma$
$z = (0.7958 - 2) / 1.0408 = -1.157$
$P(X <= 0.7958) = P(Z <= -1.157) = 0.1230$

Solution : The probability = 12.3%