STA05 Discussion 3 Extra Exercise

Shuangjie Zhang

October 2019

1 Exercise 1

A normal data with Mean = 50, SD =10

1. How many percentage of total population would lie between [40, 60]?
2. How many percentage of total population would lie between [30, 60]?
3. How many percentage of total population would lie between [60, 70]?
4. Find 90th percentile for this data.
5. Find a symmetric region which covers 75% total population.

Solution:

1. How many percentage of total population would lie between [40, 60]?

Standardized the two interval boundaries.

$\frac{40-50}{10} = -1; \frac{60-50}{10} = 1;$

After normalizing the two interval boundaries, this question is same to asking us what is area between [-1, 1] under the **standard normal distribution**.

From the table ,we read that when value $z = 1$, the area is 68.27%. So 68.27% total population would lie between [40, 60].

2. How many percentage of total population would lie between [30, 60]?
Standardized the two interval boundaries.
$\frac{30-50}{10} = -2; \frac{60-50}{10} = 1;$

After normalizing the two interval boundaries, this question is same to asking us what is area between [-2, 1] under the **standard normal distribution**. Notice that this is not a symmetric interval.

From the table ,we read that when value $z = 1$, the area is 68.27%. It means that the area between [-1, 1] is 68.27%. When value $z = 2$, the area is 95.45%. It means that the area between [-2, 2] is 95.45%.

Since our normal density plot is symmetric, then we can get that [-1, 0] and [0, 1] will have the same percentage $\frac{68.27\%}{2} = 34.135\%$. And [-2, 0] and [0, 2] shares the same percentage $\frac{95.45\%}{2} = 47.725\%$. So the interval [-2,1] can be decomposed into [-2, 0] and [0,1]. So the sum of area of [-2, 0] and [0,1] are [-2, 1] will be $\frac{68.27\%}{2} + \frac{95.45\%}{2} = 34.135\% + 47.725\% = 81.86\%$. (More easily to get the idea of symmetric after you have drawn the picture below.)

So 81.86 % of total population would lie between [30, 60].

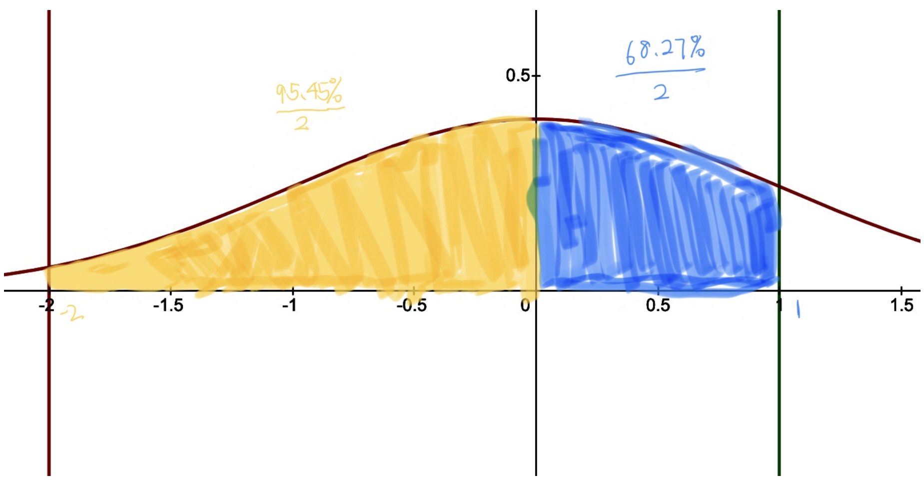


Figure 1: QUESTION 2

3. How many percentage of total population would lie between [60, 70]?
Standardized the two interval boundaries.
$\frac{60-50}{10} = 1; \frac{70-50}{10} = 2;$

After normalizing the two interval boundaries, this question is same to asking us what is area between [1, 2] under the **standard normal distribution**.

From the table ,we read that when value $z = 1$, the area is 68.27%. It means that the area between [-1, 1] is 68.27%. When value $z = 2$, the area is 95.45%. It means that the area between [-2, 2] is 95.45%.

Since our normal density plot is symmetric, then we can get that [0, 1] will have the percentage $\frac{68.27\%}{2} = 34.135\%$. And [0, 2] has the percentage $\frac{95.45\%}{2} = 47.725\%$. So the interval [1, 2] can be decomposed into [0, 2] minus [0,1]. So the overall are between [-2, 1] will be $\frac{95.45\%}{2} - \frac{68.27\%}{2} = 47.725\% - 34.135\% = 13.59\%$. (More easily to get the idea of symmetric after you have drawn the picture below.)

So 13.59 % of total population would lie between [60, 70].

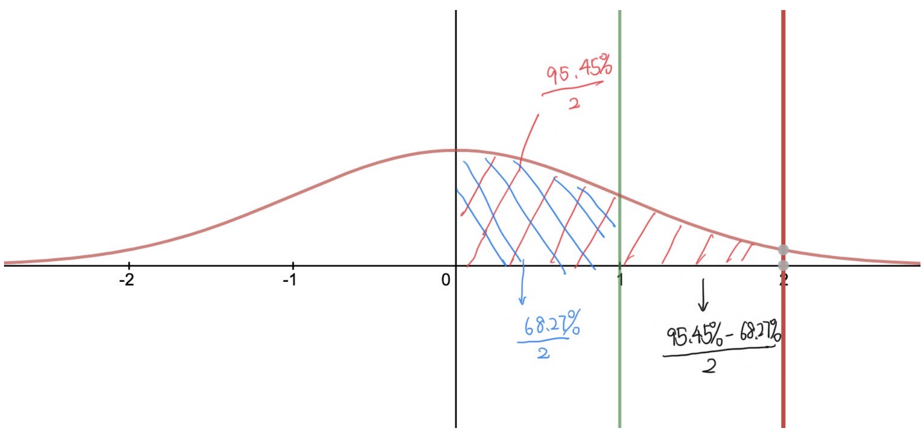


Figure 2: QUESTION 3

4. Find 90th percentile for this data.

90th percentile is a number which means that 90% of total population is below this number. That is, 10% of total population is above this number. Because of symmetry, the area between [-90th percentile, 90th percentile] will be 80%. It means from the table we need to find a z which has area about 80%. Here we have $z = 1.30$ with middle area 80.64% which is close to 80%.

Is 1.30 our final answer?

No. Here 1.30 means 1.30SD above the mean. So we need to convert to the original data scale. The final answer should be mean$+1.30\times$ SD$=50+1.30\times 10 = 63$.

5. Find a symmetric region which covers 75% total population.

This question is more like a inverse question of question 1. In question 1, we are asked to compute the area under a symmetric interval. But in this question we want 75% area. Then we need to compute two boundaries for the interval.

First, form the table we look at the area column to find a number which is close to 75. Luckily, we have got when z=1.15, the area is 74.99% which is most close to 75%. 1.15 here mean the two boundaries number is 1.15 SD away from the mean. So the interval will be [mean-1.15× SD, mean+1.15× SD]= [38.5, 61.5].

2 Exercise 2

A normal data with 10% percentile -2 and 90% percentile 3.2

1. What is the mean and the sd for this data?
2. Find a symmetric 90% interval
3. What is the height in density plot for number 3.6?

Solution:

1. Since symmetry, 10% percentile and 90% percentile means the middle area is 80%. So from the table, we know that the two interval numbers is mean-1.30×SD and mean+×SD.

Then we have two equations: mean-1.30×SD = -2, and mean+×SD=3.2.

After solving the equations, we have mean =0.6, and SD=2.

Another way to think of this question. 90% percentile 3.2 means that 90% of total population is less than 3.2. That is, 10% of total population is bigger than 3.2. Since symmetry, we will have 80% in the middle. From the table, we find the number 1.30. It means 3.2 is 1.30SD above the mean. One equation is derived: Mean+1.30SD=3.2. The other equation is got from the same way.

2. Find a symmetric 90% interval

In this question we want 90% area. Then we need to compute two boundaries for the interval.

First, form the table we look at the area column to find a number which is close to 90. Luckily, we have got when z=1.65, the area is 90.11% which is most close to 90%. 1.65 here mean the two boundaries number is 1.65 SD away from the mean. So the interval will be [mean-1.65× SD, mean+1.65× SD]= [0.6-1.65*2, 0.6+1.65*2] = [-2.7, 3.9].

3. What is the height in density plot for number 3.6?

Standardized 3.6. $\frac{3.6-0.6}{2} = 1.5$. From the table we know that when $z = 1.5$, the height is 12.95.

3 Correlation

Calculate the correlation between (1, 2, 3, 4, 5) and (5, 4, 3, 2, 1). And come up with a possible scenario where that interesting correlation occurs.

Solution:
Here x,y is not standardized.
x = (1, 2, 3, 4, 5); mean(x) = 3: SD(x) = $\sqrt{2}$
y = (5, 4, 3, 2, 1); mean(y) = 3: SD(y) = $\sqrt{2}$
xy = (5,8,9,8,5); mean(xy) = 7

$$r = \frac{mean(xy) - mean(x)mean(y)}{SD(x) \times SD(y)}$$

$$r = \frac{7 - 3 \times 3}{\sqrt{2} \times \sqrt{2}} = -1$$

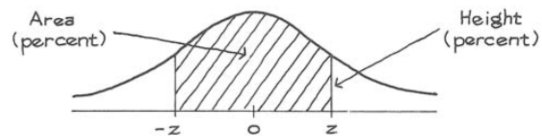
Another way to calculate the correlation is that:

$$r = mean(x_{standard} \times y_{standard})$$

First you need to standardize x and y, then calculate the mean of standardized xy. It returns the same result -1.

-1 means negatively correlated. And points are close to the line. It mean when x increase 1, y will decrease 1. From our data, we see that from 1 to 2 we increase x by 1, then in y it decreases from 5 to 4 b y 1.

Tables



A NORMAL TABLE

z	*Height*	*Area*	*z*	*Height*	*Area*	*z*	*Height*	*Area*
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	9.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991