ASTANA IT UNIVERSITY

Spring 2024

Probability and Statistics

Lecture 6 – The Normal and Chi-square distributions.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

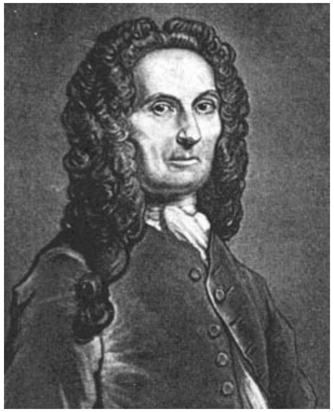
Lecture overview

1. The normal distribution
2. Finding the null and alternative hypothesis to compare the observed and theoretical distributions
3. The chi squared statistic, χ^2
4. The degrees of freedom for given sample of observations
5. Testing the null hypothesis using χ^2
6. Testing the binomial and Poisson distributions as models

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

The Normal Distribution

$$X \sim \mathrm{N}(\mu, \sigma^2)$$



Abraham de Moivre (1667-1754)

- Discovered in 1733 by de Moivre as an approximation to the Binomial distribution when the number of trails is large
- Derived in 1809 by Gauss

Karl F. Gauss (1777-1855)

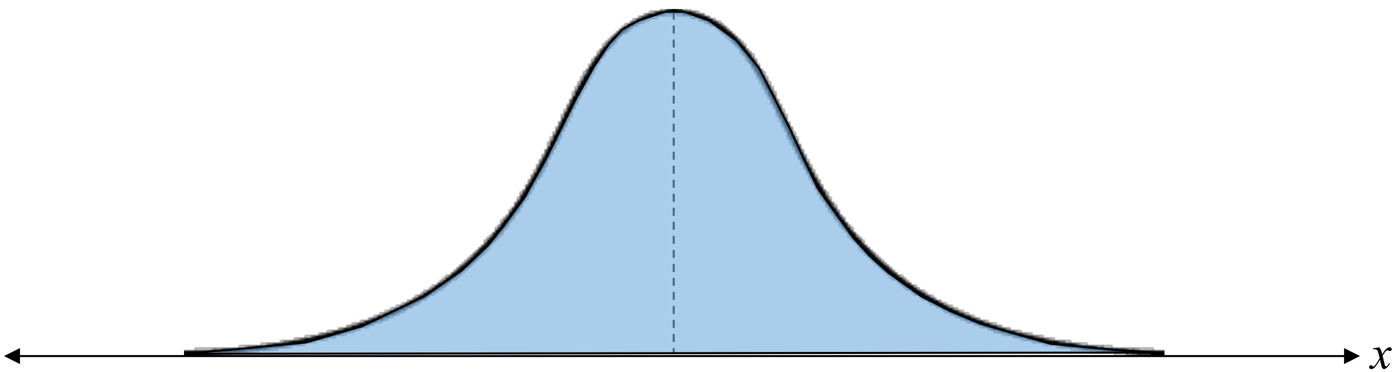
Normal Distribution is used to **model**

- Salaries of working people
- Marks of students
- Time spent on travelling
- Heights of people
- Errors in calculations
- Blood Pressure

Properties of Normal Distributions

1. The normal distribution is **bell-shaped** and is **symmetric** about the **mean**.

2. The **mean**, **median**, and **mode** are **equal**.

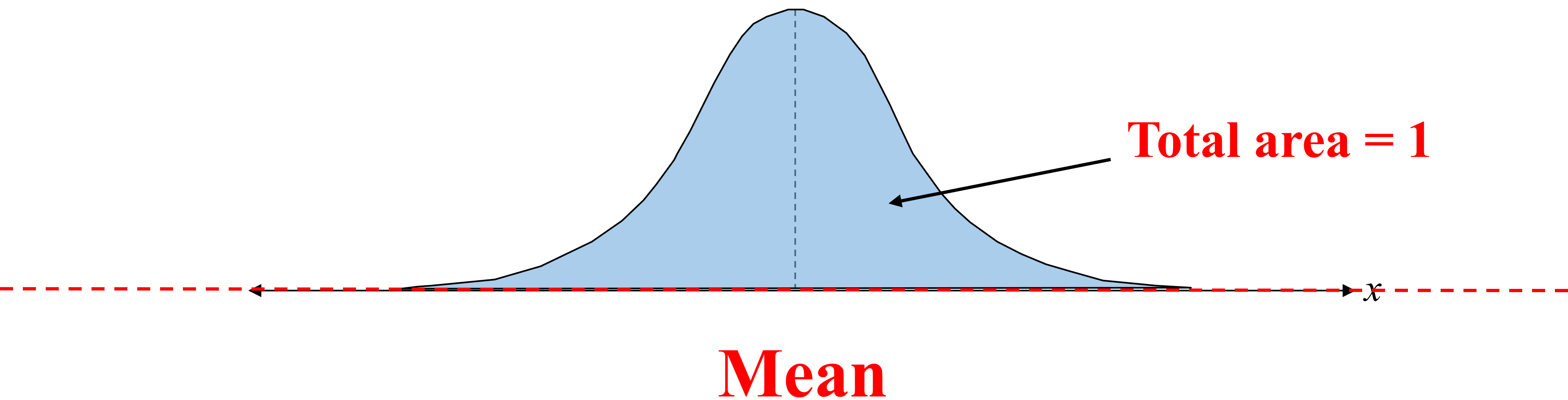


Median = Mean = Mode

Properties of Normal Distributions

3. The total area under the normal curve is equal to 1.

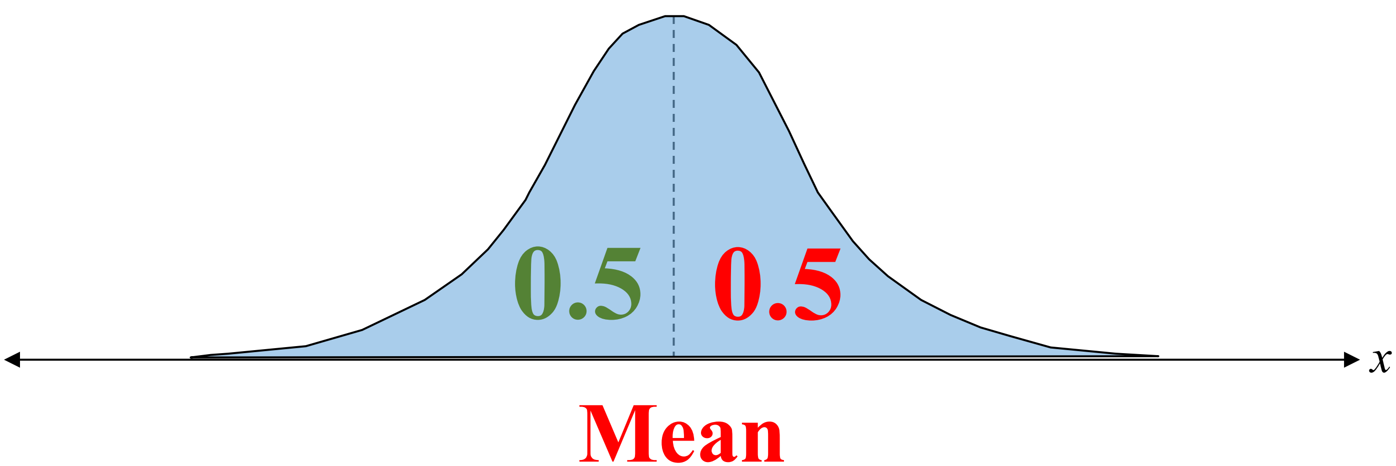
4. The normal curve approaches, but never touches, the *x*-axis as it extends farther and farther away from the mean.



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Properties of Normal Distributions

5. 50% of area (probability) on the right from the mean (origin) and 50% of the area (probability) on the left



The Normal Distribution

- A continuous random variable is said to be normally distributed with mean μ and variance σ^2 if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} e^{-\frac{1}{2}\left(\frac{\mathbf{X}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)^2}$$

Where e = the mathematical constant approximated by 2.71828
π = the mathematical constant approximated by 3.14159
μ = the population mean
σ = the population standard deviation
X = any value of the continuous variable

The Normal Distribution

It is **cumbersome work** to integrate pdf of Normal distribution, every time we need to find probability values

$$f(x) = \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} e^{-\frac{1}{2}\left(\frac{\mathbf{X}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)^2}$$

There is a more preferable option known as

Standard Normal table

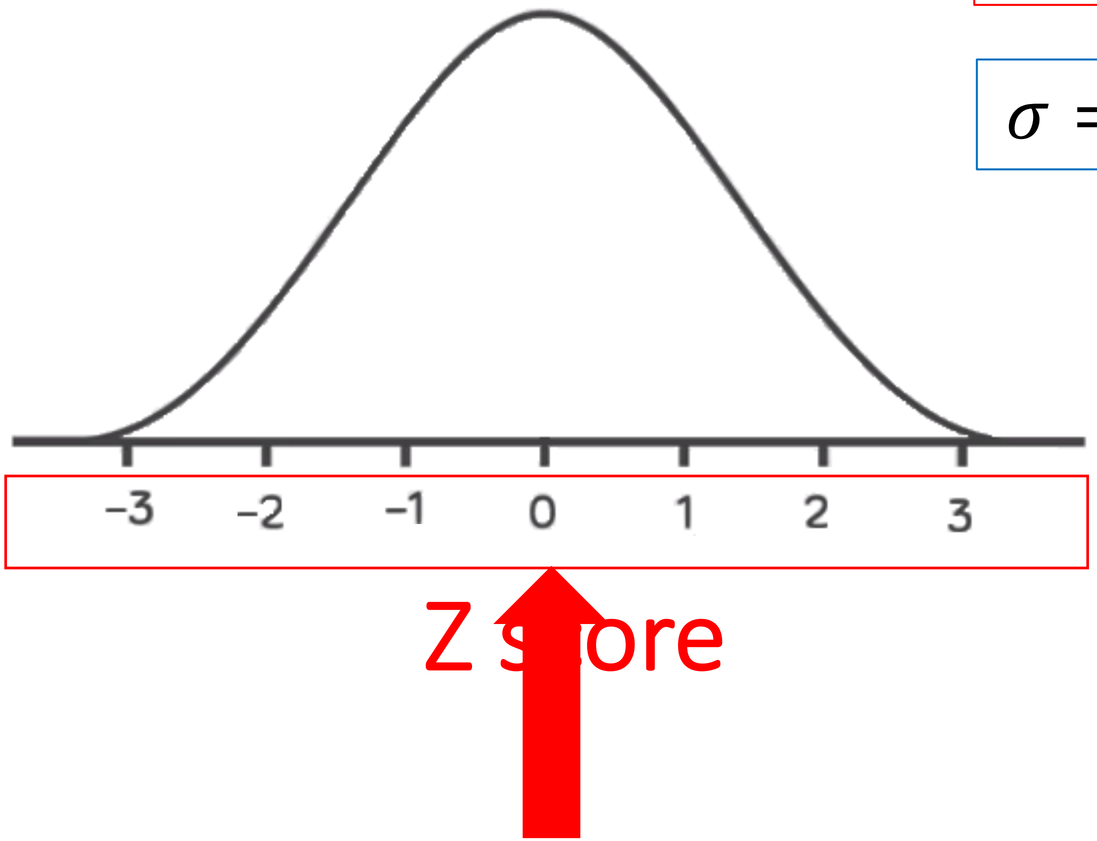
Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Standard Normal Distribution

$Z \sim N(0, 1^2)$

$\mu = 0$

$\sigma = 1$



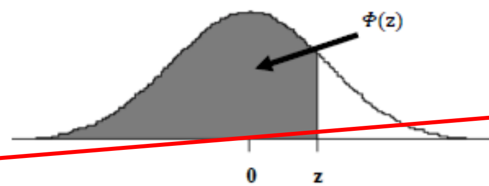
Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

And we can find that AREA using table that is called "Z score table", also known as Standard Normal table

THE NORMAL DISTRIBUTION

If Z has a normal distribution with mean 0 and variance 1 then for each value of z, the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \leq z)$.

For negative values of z use $\Phi(-z) = 1 - \Phi(z)$



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
											ADD								
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

In this table the top row and first column correspond to z values

All values inside the table corresponds to areas

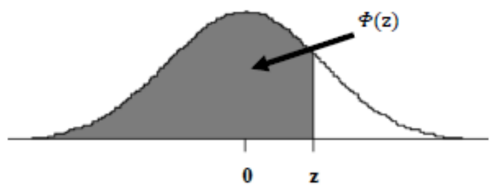
Example

Use Standard Normal table to find probability.
Find $P(Z<0.51)=?$

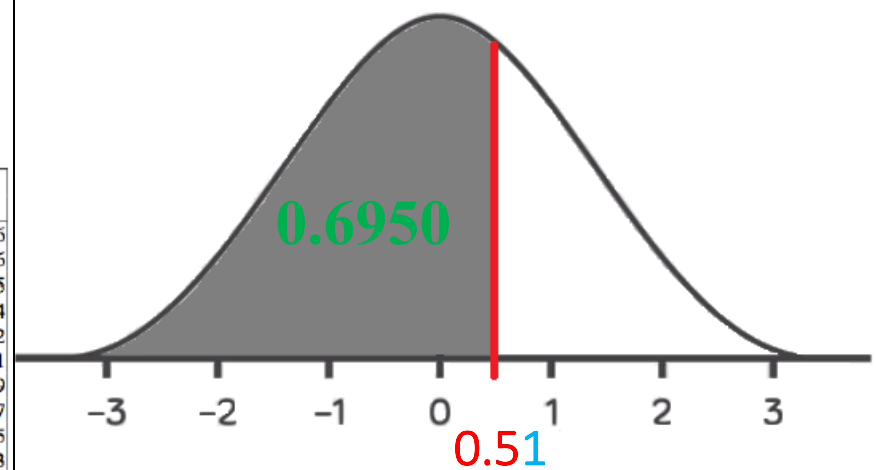
THE NORMAL DISTRIBUTION

If Z has a normal distribution with mean 0 and variance 1 then for each value of z, the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \le z)$.

For negative values of z use $\Phi(-z) = 1-\Phi(z)$



z	0	1	2	3	4	5	6	7	8	9	ADD 1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0



$P(Z<0.51) = 0.6950$

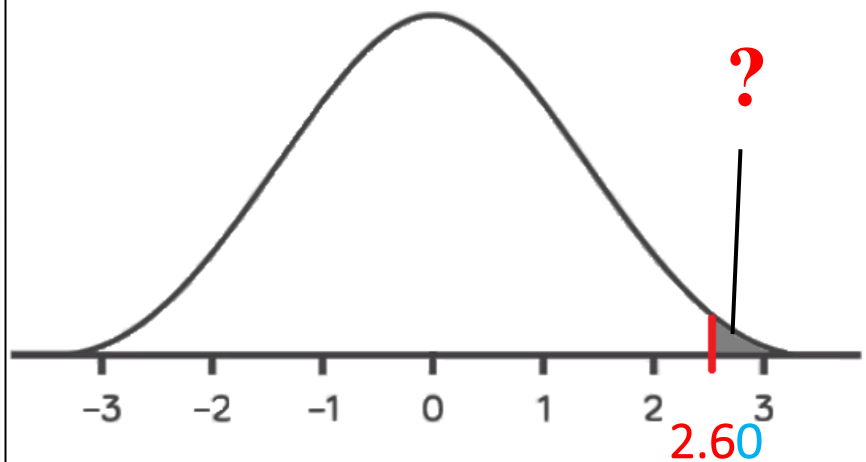
Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Exercise

Use Standard Normal table to find probability.

Find:

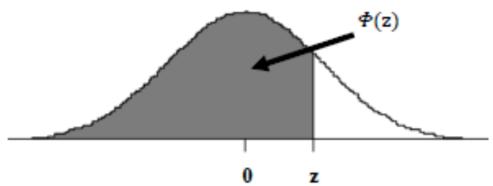
P(Z>**2.60**)=?



THE NORMAL DISTRIBUTION

If Z has a normal distribution with mean 0 and variance 1 then for each value of z, the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \le z)$.

For negative values of z use $\Phi(-z) = 1-\Phi(z)$



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Solution

Use Standard Normal table to find probability.

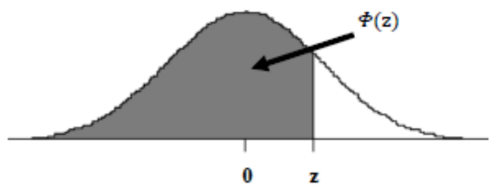
Find:

P(Z>**2.60**)=?

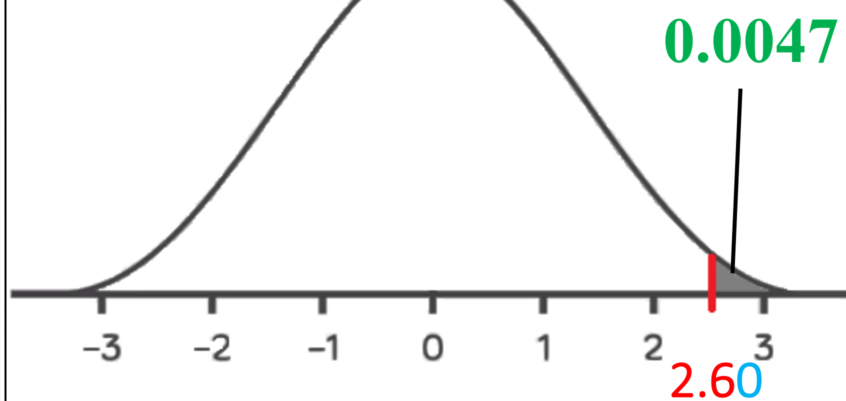
THE NORMAL DISTRIBUTION

If Z has a normal distribution with mean 0 and variance 1 then for each value of z, the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \le z)$.

For negative values of z use $\Phi(-z) = 1 - \Phi(z)$



z	0	1	2	3	4	5	6	7	8	9	ADD 1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0



P(Z>**2.60**) = 1- P(Z<**2.60**) =

= 1- 0.9953 =

= 0.0047

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Use table to find z given a probability

In some of the problems you won't be given value of z, but instead probability. And you will be asked to find value of z.

Whenever possible this table should be used to find z given a value for $p = P(Z > z)$.

If $P(Z < a)$ is greater than 0.5, then a will be > 0.
If $P(Z < a)$ is less than 0.5, then a is less than 0.

If $P(Z > a)$ is less than 0.5, then a will be > 0.
If $P(Z > a)$ is more than 0.5, then a will be < 0.

Example

Find the value of the constant $\boldsymbol{a}$ such that

a. $P(Z < \boldsymbol{a}) = 0.7611$

b. $P(Z > \boldsymbol{a}) = 0.0287$

c. $P(Z < \boldsymbol{a}) = 0.0170$

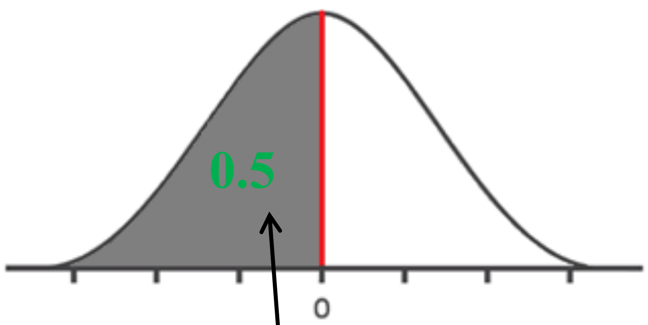
Example

Find the value of the constant ***a*** such that

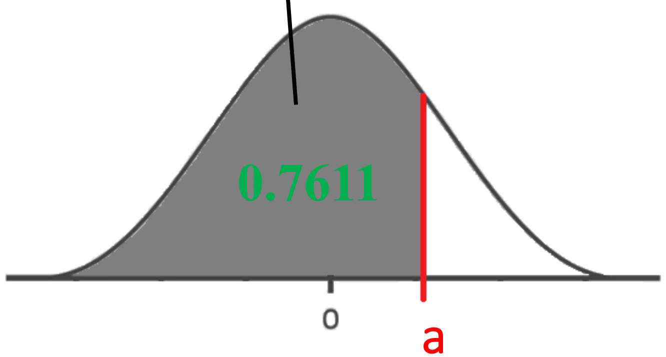
a. $P(Z < a) = \mathbf{0.7611}$

Solutions:

Since $P(Z < \mathbf{0}) = \mathbf{0.5}$



Thus, ***a*** must be definitely be **greater** than **0** (to the **right** from **0**)



Because $\mathbf{0.7611 > 0.5}$

By using Standard normal table

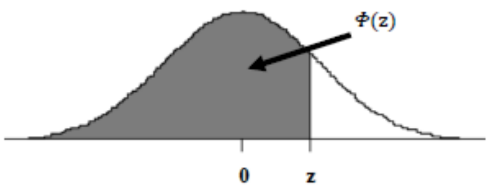
Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

a. $P(Z < a) = 0.7611$

THE NORMAL DISTRIBUTION

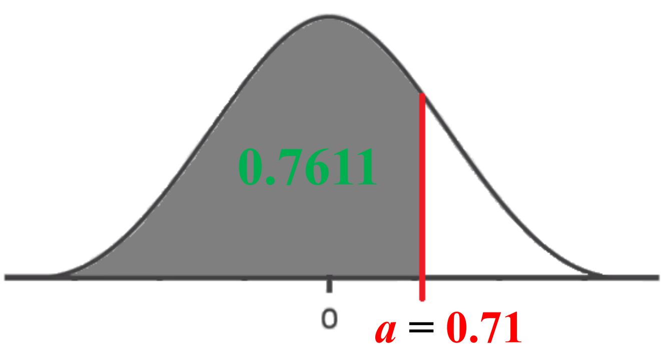
If Z has a normal distribution with mean 0 and variance 1 then for each value of z the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \leq z)$.

For negative values of z use $\Phi(-z) = 1-\Phi(z)$



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
											ADD								
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Thus, from
Standard normal table
$a = 0.71$

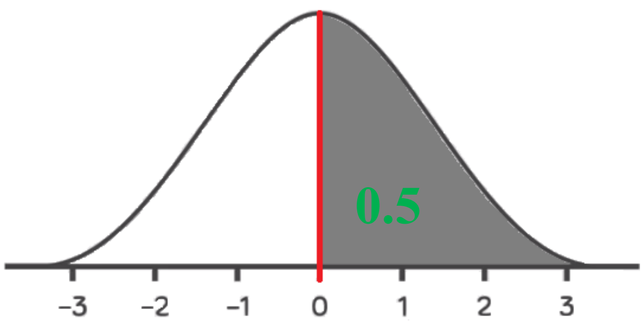


Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

Find the value of the constant $\boldsymbol{a}$ such that

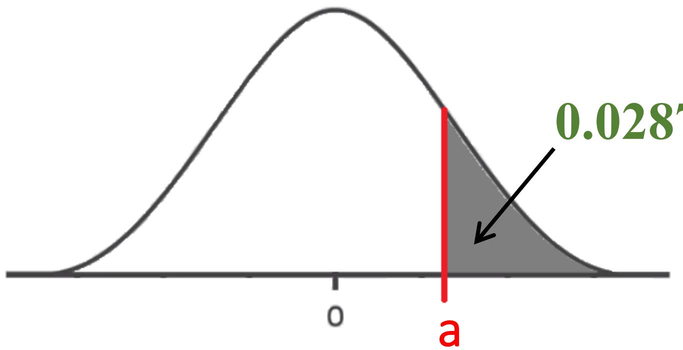
b. $P(Z > a) = 0.0287$



Solutions:

Since $P(Z > 0) = \mathbf{0.5}$

Thus, $\boldsymbol{a}$ must be definitely be **greater** than **0** (to the **right** from **0**)



Because $\mathbf{0.0287 < 0.5}$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

b. $P(Z > a) = \mathbf{0.0287}$

$P(Z < a) = 1 - \mathbf{0.0287} = \mathbf{0.9713}$

THE NORMAL DISTRIBUTION

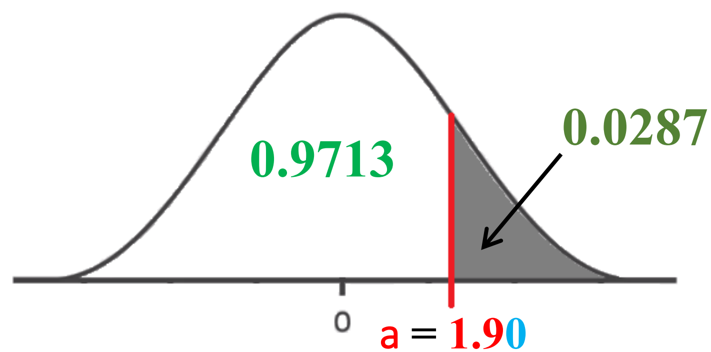
If Z has a normal distribution with mean 0 and variance 1 then for each value of z, the table gives the value of $\Phi(z)$, where $\Phi(z) = P(Z \le z)$.

For negative values of z use $\Phi(-z) = 1-\Phi(z)$

z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Thus, from Standard normal table

$a = 1.90$



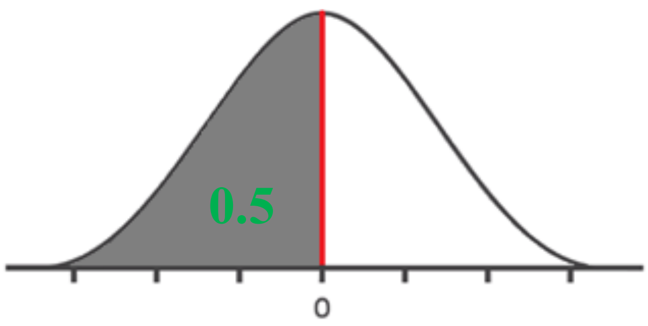
Example

Find the value of the constant a such that

c. $P(Z < a) = 0.0170$

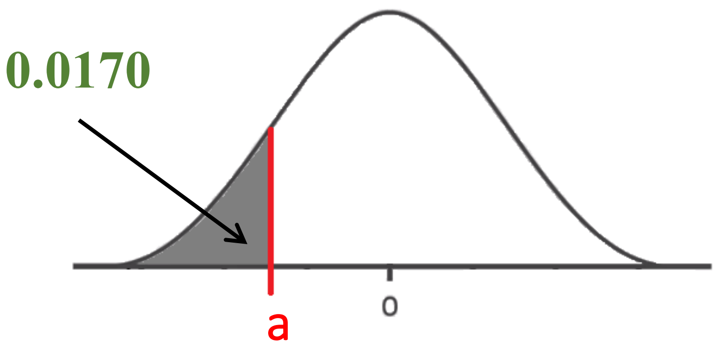
Solutions:

Since $P(Z < 0) = 0.5$



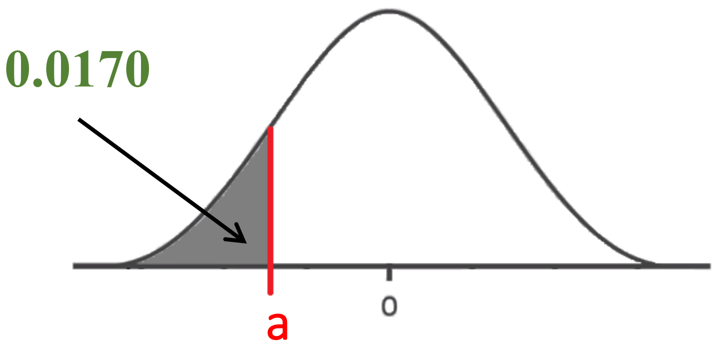
Thus, a must be definitely **smaller** than **0** (to the **left** from **0**)

Because **$0.0170 < 0.5$**

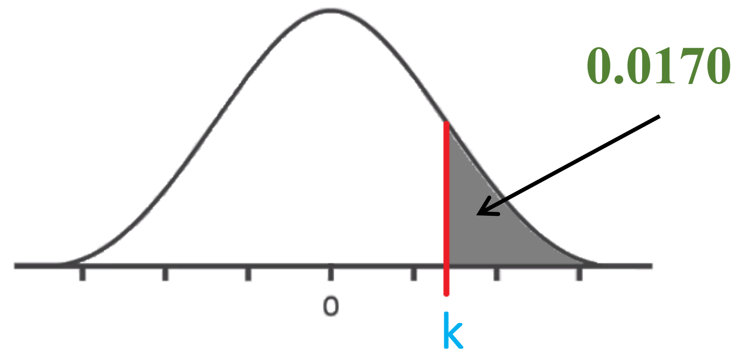


$P(Z < a) = 0.0170$

$P(Z > k) = 0.0170$

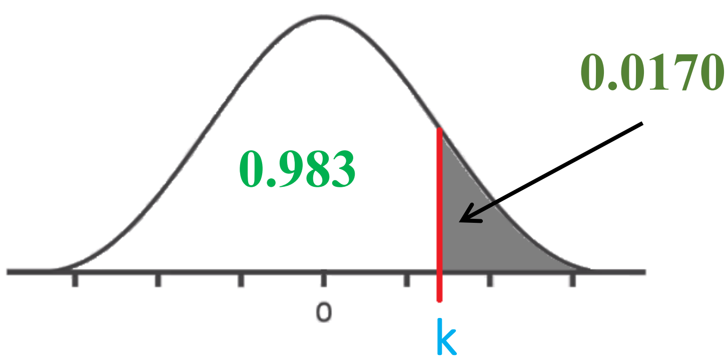


Let $k = -a$



$P(Z > k) = 0.0170$

$P(Z < k) = 1 - 0.0170 = 0.983$

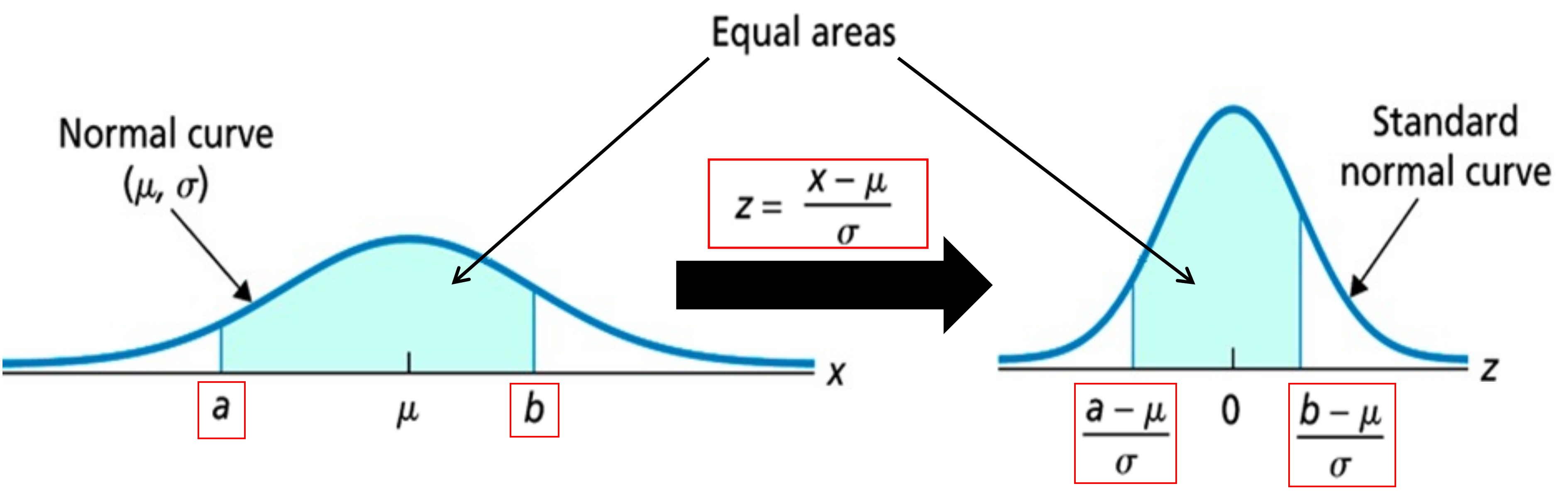


$k = 2.12 \rightarrow a = -2.12$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Use standard normal table to find probabilities of any normal distribution

Standard Normal distribution allow us to find probabilities of **Normal distribution** by using technique known as "**Standardization**" (converts x-value into a z-score).



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

$$\mu = 175cm$$

$$\sigma = 10cm$$

Find probability of a specific student been taller than 190cm?

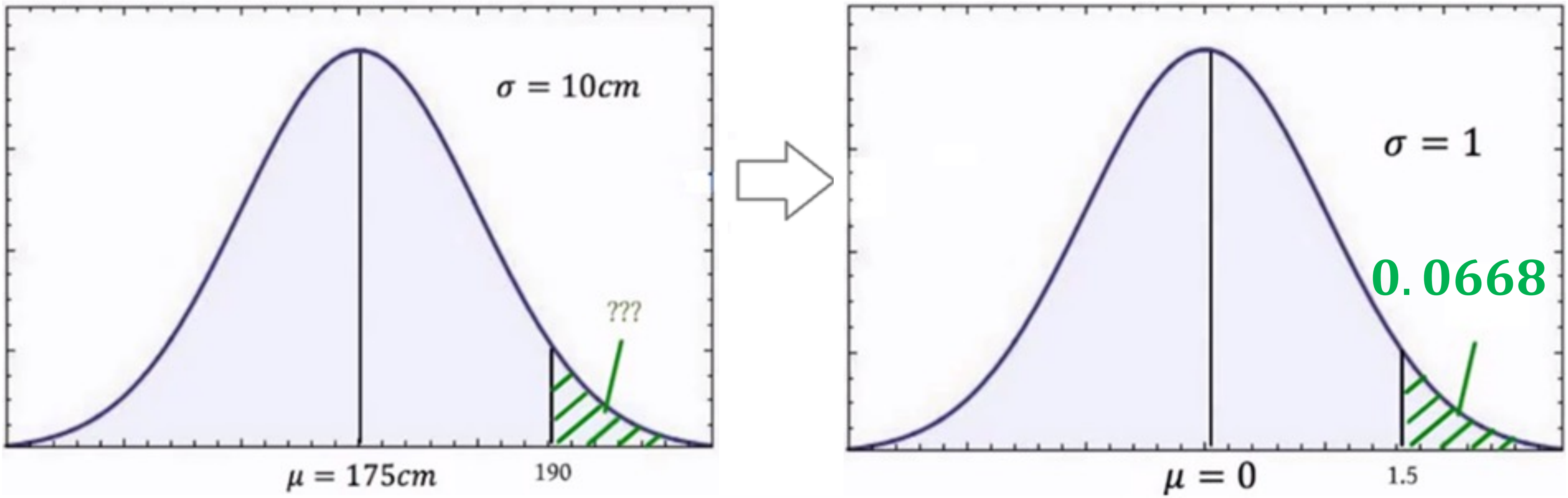
Note: The arrangement of the students on the picture above does not represent histogram.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

Solutions:

$$\boxed{z = \frac{X - \mu}{\sigma}}$$



$$P(H > 190) = P\left(Z > \frac{190 - 175}{10}\right) = P(Z > 1.5) =$$

$$= 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

The random variable $X \sim N(50, 4^2)$.

Find $P(X \leq \mathbf{45}) =$?

Solutions:

$$P(X \leq \mathbf{45}) = P\left(Z \leq \frac{45-50}{4}\right) =$$

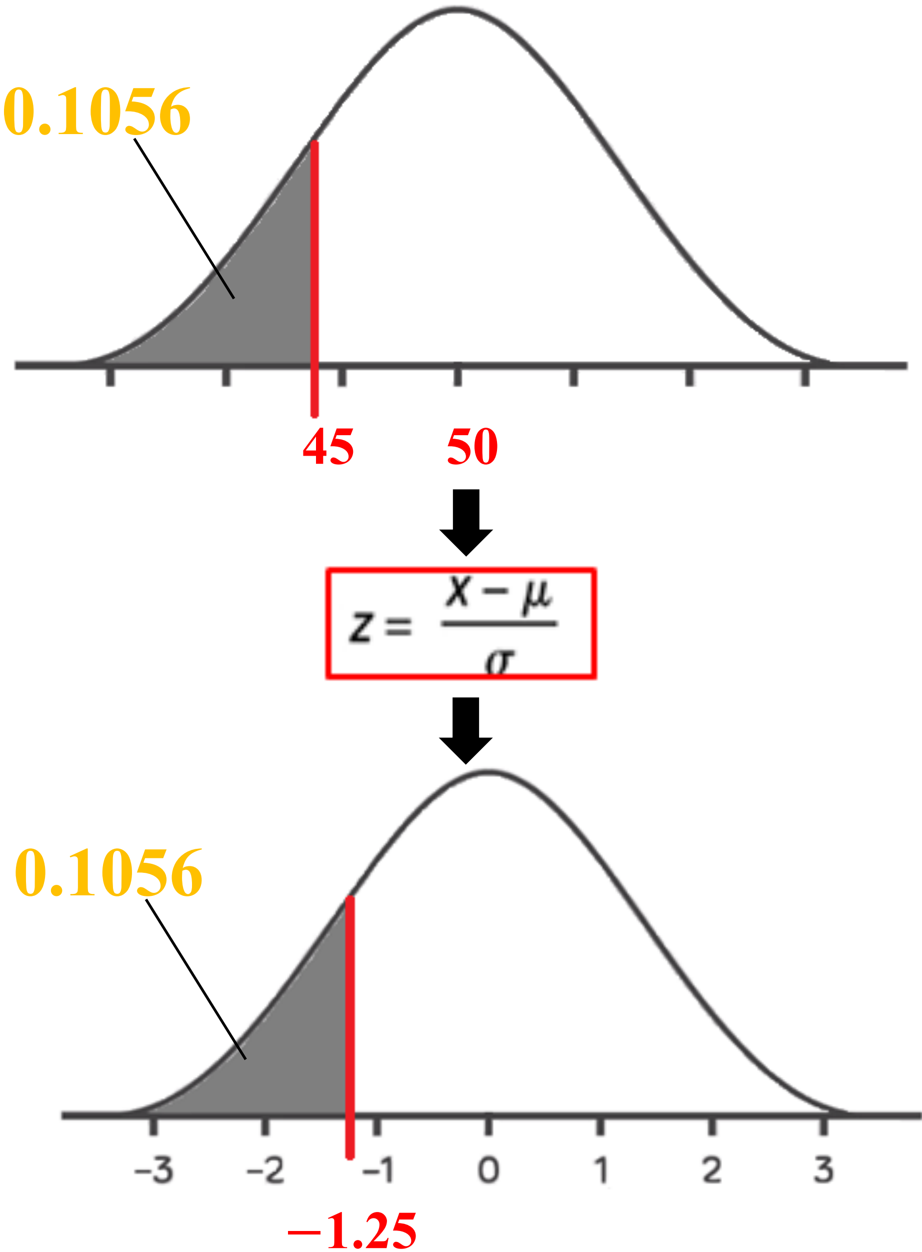
$$= P(Z \leq -1.25) =$$

$$= P(Z \geq 1.25) =$$

$$= 1 - P(Z \leq 1.25) =$$

$$= 1 - 0.8944 =$$

$$= \mathbf{0.1056}$$



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Find the fourth unknown given any three in the equation $P(Z>(k-\mu)/\sigma)=\alpha$

- If $X \sim N(\mu, \sigma^2)$ and $P(X > a) = \alpha$, where α is a probability, you write this statement as

$$P\left(Z > \frac{a-\mu}{\sigma}\right) = \alpha.$$

- Sometimes neither μ nor σ is given, in which case you will have to solve simultaneous equations.

Exercise

The random variable $X \sim N(50, \sigma^2)$.

Given that $P(X<46) = 0.2119$, find the value of σ.

Solutions:

$$P(X < 46) = 0.2119$$

$$P\left(Z < \frac{46 - 50}{\sigma}\right) = 0.2119$$

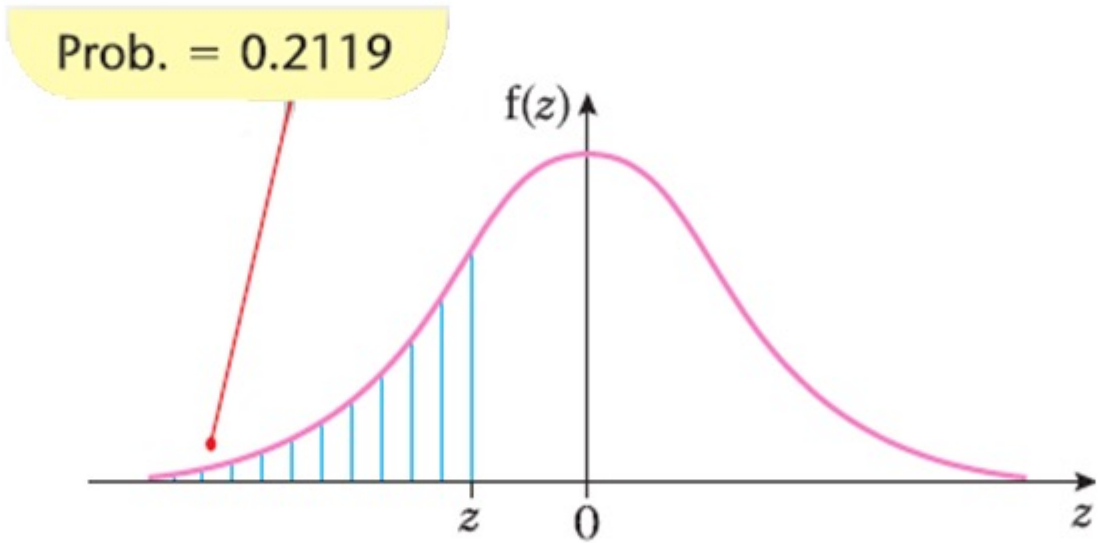
$$1 - 0.2119 = 0.7881$$

$$P(Z < 0.80) = 0.7881$$

$$\frac{46 - 50}{\sigma} = -0.80$$

$$\frac{-4}{-0.80} = \sigma$$

$$\sigma = 5$$



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example 10

The heights of a large group of women are normally distributed with a mean of 165 cm and a standard deviation of 3.5 cm.

Steven is looking for a woman whose height is between 168 cm and 174 cm for a part in his next film.

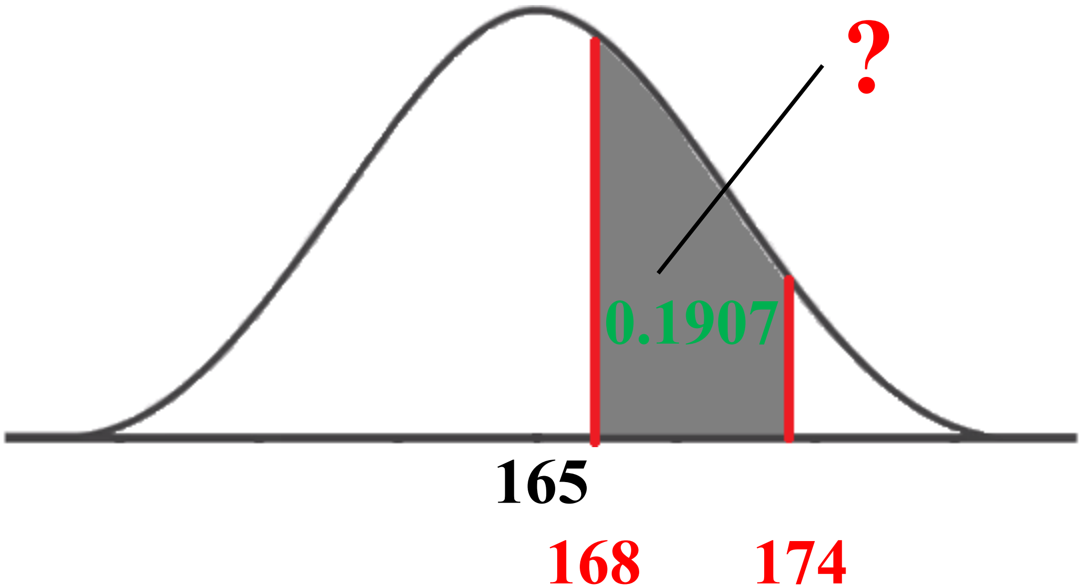
Find the proportion of women from this group who meet Steven's criteria. Let H = the height of a woman from this group.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

Solutions:

$H \sim N(165, 3.5^2)$



$P(168<H<174) =$

$$= P\left(\frac{168-165}{3.5}<Z<\frac{174-165}{3.5}\right) =$$

$$= P(0.857<Z<2.571) =$$

$$= P(Z<2.571) - P(Z<0.857) =$$

$$= 0.9949 - 0.8042 = \mathbf{0.1907}$$

Exercise

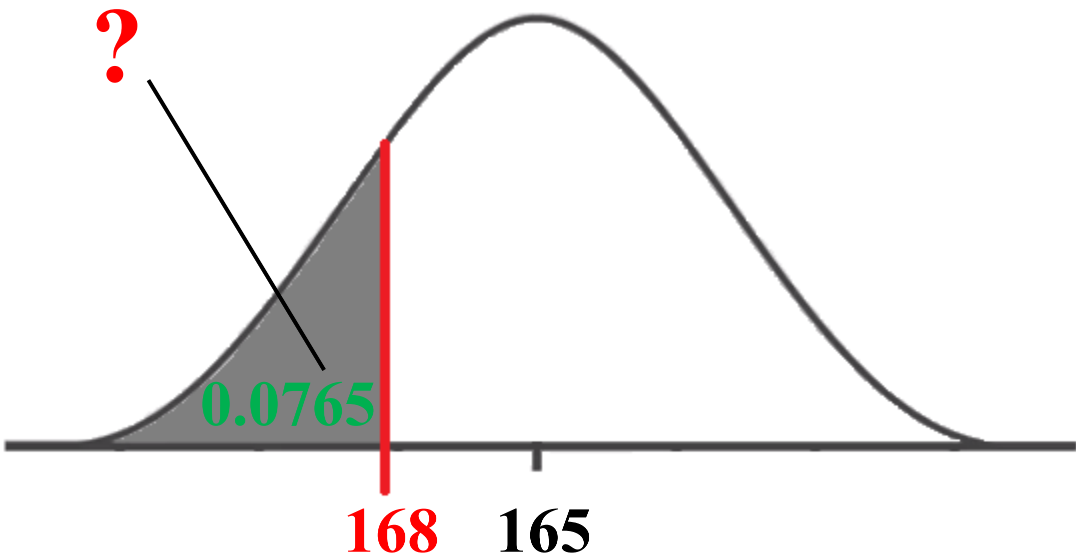
The heights of a large group of women are normally distributed with a mean of 165 cm and a standard deviation of 3.5 cm.

A woman is selected at random from this group.

Find the probability that she is shorter than 160 cm.

Solutions:

$H \sim N(165, 3.5^2)$



$P(H<160) =$

$= P\left(Z< \frac{160-165}{3.5}\right) =$

$= P(Z< -1.429) =$

$= P(Z>1.429) =$

$= 1 - P(Z<1.429)=$

$= 1- 0.9235 =$ **0.0765**

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Summary of key points

1 The random variable X that has a normal distribution with mean μ and standard deviation σ is represented by

$$X \sim \mathrm{N}(\mu, \sigma^2)$$

where σ^2 is the variance of the normal distribution.

2 If $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Z \sim \mathrm{N}(0, 1^2)$ then

$$Z = \frac{X - \mu}{\sigma}$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Example

Let's take a die and throw it 120 times.

The observed results can be summarised in a frequency distribution table.

Number, n	1	2	3	4	5	6
Observed frequency	23	15	25	18	21	18

If the die is unbiased what would you expect the distribution to be ?

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

If the die is unbiased you would expect the numbers 1 to 6 to appear the same number of times.

Number, n	1	2	3	4	5	6
Expected frequency	20	20	20	20	20	20

The observed and the expected frequencies are not identical but this is no surprise even if the die is unbiased.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

H_0 : there is no difference between the observed and the theoretical distributions.

H_1 : there is a difference between the observed and theoretical distributions.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Calculate the chi squared statistic, χ^2

Number, n	1	2	3	4	5	6
Observed frequency	23	15	25	18	21	18
Expected frequency	20	20	20	20	20	20

We want to come up with a number which measures how different the observed values are from the theoretical values. If this number is 'small' we will accept that the the observed data is a sample from the theoretical distribution. The number we calculate is χ^2

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Number on die, n	1	2	3	4	5	6
Observed frequency	23	15	25	18	21	18
Expected frequency	20	20	20	20	20	20
$O_i - E_i$	3	-5	5	-2	1	-2
$(O_i - E_i)^2$	9	25	25	4	1	4

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = \frac{9}{20} + \frac{25}{20} + \frac{25}{20} + \frac{4}{20} + \frac{1}{20} + \frac{4}{20}$$

$$= \frac{68}{20} = 3.4$$

Let' look at an alternative calculation

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{n} \frac{O_i^2 + E_i^2 - 2O_iE_i}{E_i}$$

$$= \sum_{i=1}^{n} \frac{O_i^2}{E_i} + \sum_{i=1}^{n} E_i - 2\sum_{i=1}^{n} O_i$$

$$= \sum_{i=1}^{n} \frac{O_i^2}{E_i} + N - 2N = \sum_{i=1}^{n} \frac{O_i^2}{E_i} - N$$

In our example n was 6 (six numbers on the die) and N the total number of observations, 120

$$\sum O = \sum E = N$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Number on die, n	1	2	3	4	5	6
Observed frequency	23	15	25	18	21	18
Expected frequency	20	20	20	20	20	20

$$\chi^2 = \sum_{i=1}^{6} \frac{O_i^2}{E_i} - N$$

$$= \frac{23^2}{20} + \frac{15^2}{20} + \frac{25^2}{20} + \frac{18^2}{20} + \frac{21^2}{20} + \frac{18^2}{20} \quad - \quad 120$$

$$= 3.4$$

Which calculation do you prefer ?

The second one involves less arithmetic ?

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

We calculated $\frac{(O_1-E_1)^2}{E_1}$ $\frac{(O_2-E_2)^2}{E_2}$ $\frac{(O_3-E_3)^2}{E_3}$ $\frac{(O_4-E_4)^2}{E_4}$ $\frac{(O_5-E_5)^2}{E_5}$ and $\frac{(O_6-E_6)^2}{E_6}$ but do we actually have six variables ?

No ! Because $\sum O = \sum E = N$=120

If we know any of the five O's or E's we can determine the sixth. We don't have 6 degrees of freedom we have 5. Because we fixed N we loose one degree of freedom.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$$\sum_{i=1}^{n} E_i = N$$

Is called a constraint. This constraint applies to all the examples where we compare an observed distribution with a theoretical distribution.

In some examples the expected frequencies can only be calculated by estimating the parameters of the theoretical distribution using the mean of the observed data. When a parameter is estimated it is an additional constraint.

Number of restrictions or constraints = k

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Number paired O_i and E_i in final table = n

If some of the E_i values are very small, less than 5, they will have to be combined with adjacent cells and so the n reduces accordingly.

Degrees of freedom = ν = n - k

In our example n =6 and k=1 so the degrees of freedom = ν = n - k = 5

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

In our example we have $\chi^2 = 3.4$ and degrees of freedom = 5

The larger this value the more likely the observations don't follow the probability distribution we are using for comparison.

This chi squared statistic can only take positive values. If the expected and observed frequencies are identical its value is zero.

Let's find a critical value at the 5% level,

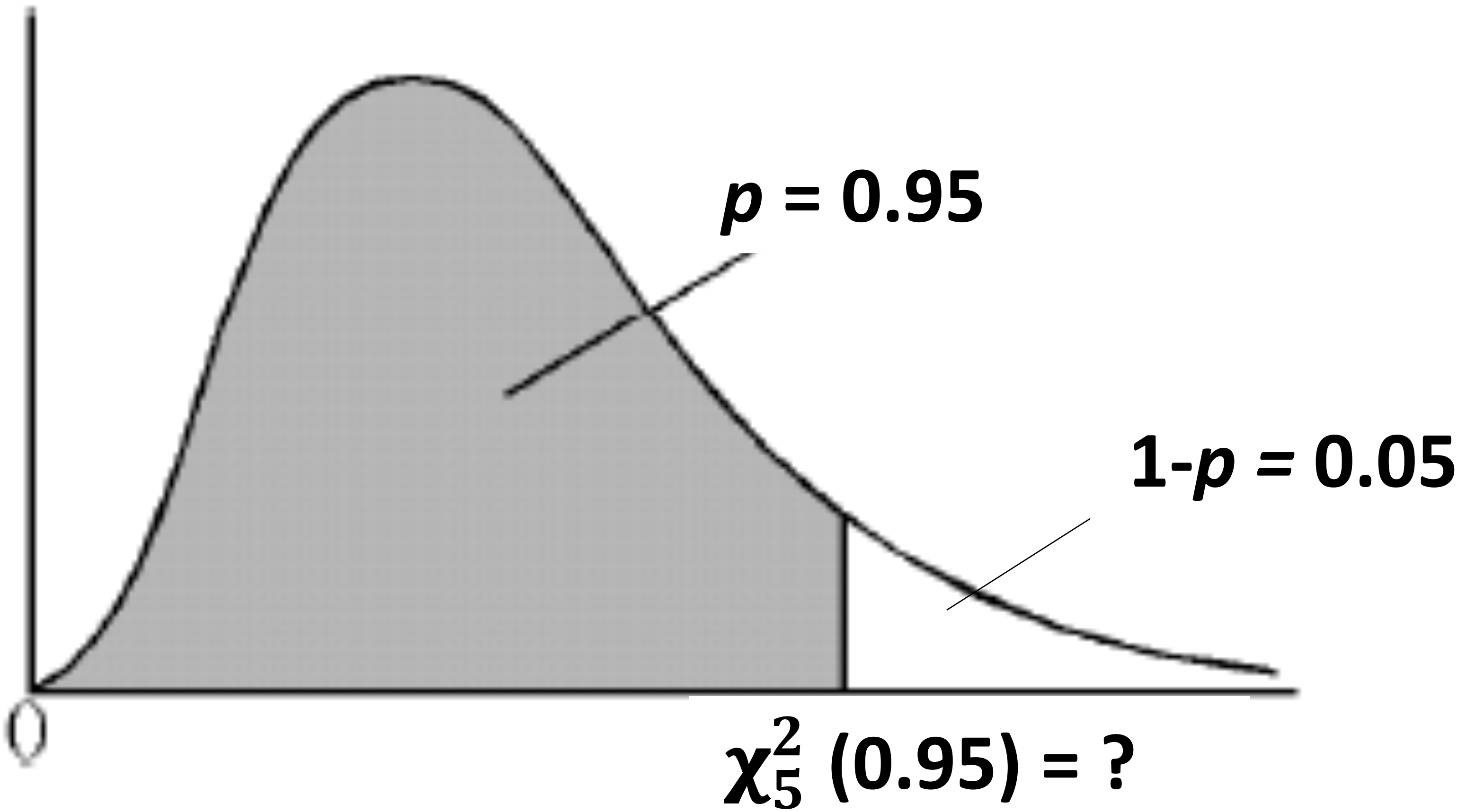
$\chi^2_5(0.95) = ?$

The subscript 5 reminds that the degrees of freedom are 5. Use the tables: p=0.95 and $\nu = 5$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Testing the null hypothesis using χ^2 tables

χ_5^2 (0.95) = critical value = ?



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Here are the χ^2 tables

Probability (area to the left of the critical value)

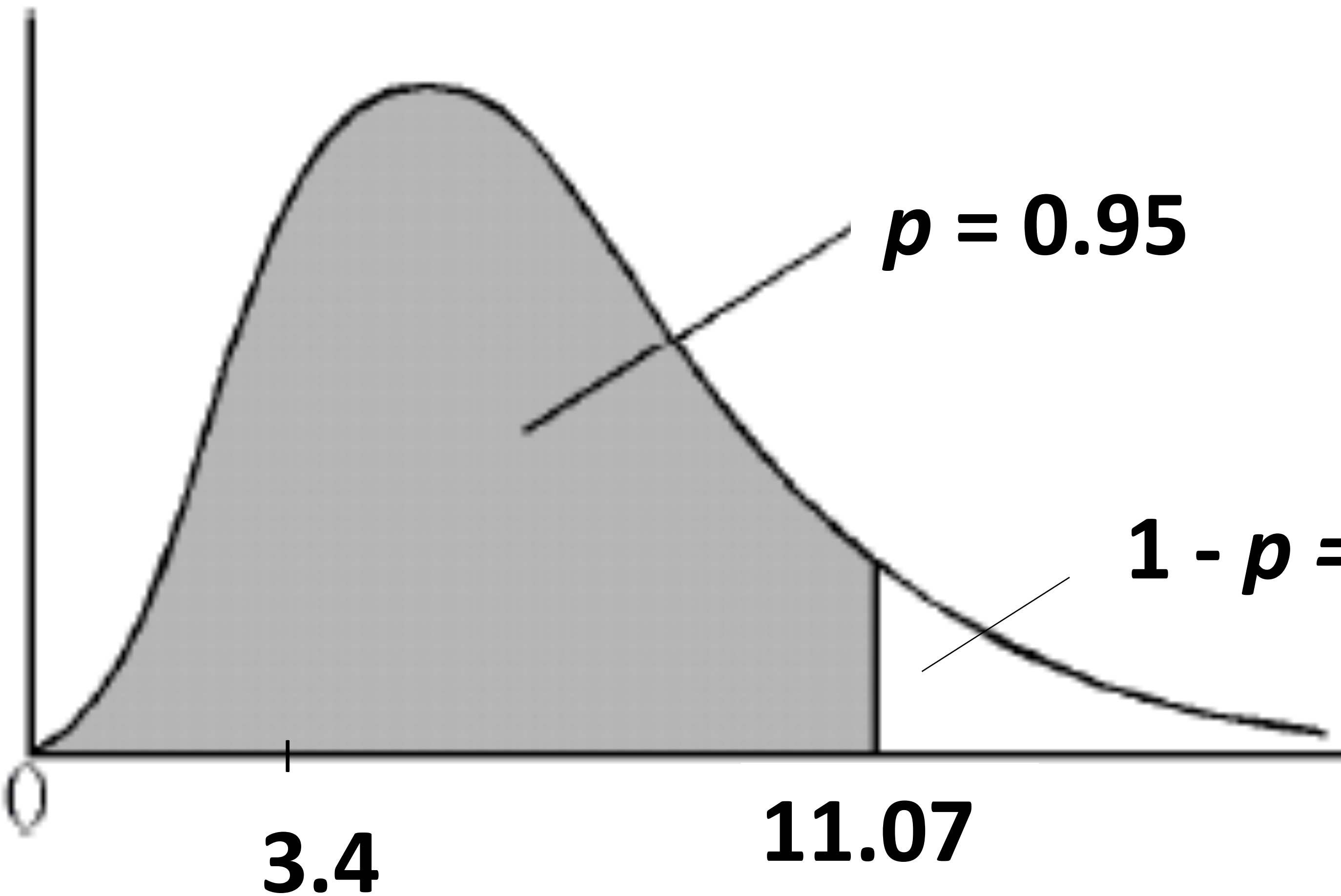
p	0.01	0.025	0.05	0.9	0.95	0.975	0.99	0.995	0.999
$\nu = 1$	$0.0^3 1571$	$0.0^3 9821$	$0.0^2 3932$	2.706	3.841	5.024	6.635	7.879	10.83
2	0.02010	0.05064	0.1026	4.605	5.991	7.378	9.210	10.60	13.82
3	0.1148	0.2158	0.3518	6.251	7.815	9.348	11.34	12.84	16.27
4	0.2971	0.4844	0.7107	7.779	9.488	11.14	13.28	14.86	18.47
5	0.5543	0.8312	1.145	9.236	11.07	12.83	15.09	16.75	20.51
6	0.8721	1.237	1.635	10.64	12.59	14.45	16.81	18.55	22.46
7	1.239	1.690	2.167	12.02	14.07	16.01	18.48	20.28	24.32
8	1.647	2.180	2.733	13.36	15.51	17.53	20.09	21.95	26.12
9	2.088	2.700	3.325	14.68	16.92	19.02	21.67	23.59	27.88

degrees of freedom

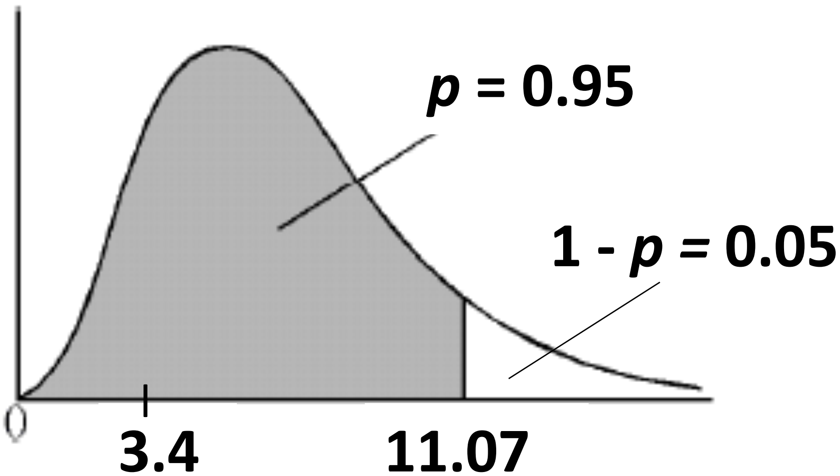
The critical value is 11.07

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$$\chi^2_5\ (0.95) = 11.07$$



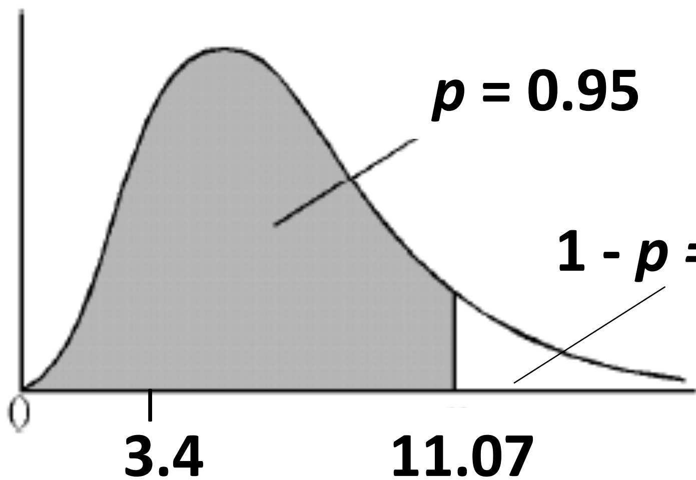
$$\chi^2_5\ (0.95) = 11.07$$



3.4 < critical value therefore accept H_0 ,
the observed and theoretical distributions are the same.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$\chi^2_5\,(0.95) = 11.07$



If the observed distribution has the same distribution as the theoretical distribution. Then 5% of the samples with 120 observations will have a χ^2 statistic that exceeds 11.07 .

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

The binomial distribution as a model

Here is a sampling distribution

x	0	1	2	3	4	5	6	Total
Observed frequencies	12	16	8	3	1	0	0	40

Is this a binomial distribution ?

Answer this question using a 5% goodness of fit test.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2	3	4	5	6	Total
Observed frequencies	**12**	**16**	**8**	**3**	**1**	**0**	**0**	**40**

If *X~B(n,p)* then $P(X=x)=\binom{n}{x}p^x\,(1-p)^{n-x}$

But what is n and what is p ?

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2	3	4	5	6	Total
Observed frequencies	12	16	8	3	1	0	0	40

The table implies that *X* takes the values 0, 1, 2, 3, 4, 5 and 6 therefore n=6

$$\mu = \frac{\sum xO}{\sum O} = \frac{0+16+16+9+4+0+0}{40} = \frac{45}{40} = 1.125$$

$$\mu = np \quad \therefore \quad p = \frac{\mu}{n} = \frac{1.125}{6} = 0.1875$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

We can now find our expected frequencies if

$$\boldsymbol{X \sim B(6, 0.1875)}$$

$$\boldsymbol{P(X=0) = \binom{6}{0} (0.1875)^0 \, (0.8125)^6 = 0.2877}$$

$$\boldsymbol{\therefore E_0 = 40 \times 0.2877 = 11.51}$$

Note that the mean of our expected frequencies is $\boldsymbol{1.125 = 6 \times 0.1875}$ This is another constraint !

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$X \sim B(6, 0.1875)$ **and the total number of observations was 40**

$$E_0 = 40 \times P(X = 0) = 40 \binom{6}{0} (0.1875)^0 \, (0.8125)^6 = 11.51$$

$$E_1 = 40 \times P(X = 1) = ?$$

$$E_2 = 40 \times P(X = 2) = ?$$

$$E_3 = 40 \times P(X = 3) = ?$$

$$E_4 = 40 \times P(X = 4) = ?$$

$$E_5 = 40 \times P(X = 5) = ?$$

$$E_6 = 40 \times P(X = 6) = ?$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

$$E_0 = 40 \times P(X = 0) = 40 \binom{6}{0} (0.1875)^0 \ (0.8125)^6 = 11.51$$

$$E_1 = 40 \times P(X = 1) = 40 \binom{6}{1} (0.1875)^1 \ (0.8125)^5 = 15.93$$

$$E_2 = 40 \times P(X = 2) = 40 \binom{6}{2} (0.1875)^2 \ (0.8125)^4 = 9.19$$

$$E_3 = 40 \times P(X = 3) = 40 \binom{6}{3} (0.1875)^3 \ (0.8125)^3 = 2.83$$

$$E_4 = 40 \times P(X = 4) = 40 \binom{6}{4} (0.1875)^4 \ (0.8125)^2 = 0.49$$

$$E_5 = 40 \times P(X = 5) = 40 \binom{6}{5} (0.1875^5 \ (0.8125)^1 = 0.50$$

$$E_6 = 40 \times P(X = 6) = 40 \binom{6}{6} (0.1875)^6 \ (0.8125)^0 = 0.00$$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2	3	4	5	6	Total
Observed frequencies	12	16	8	3	1	0	0	40
Expected frequencies	11.51	15.93	9.19	2.83	0.49	0.05	0.00	40

To use the chi squared test the smallest Expected frequency has to be 5 or more, hence we need to combine the last 5 values.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2 or more	Total
Observed frequencies	12	16	12	40
Expected frequencies	11.51	15.93	12.56	40

Now we have our expected and observed frequencies and the expected frequencies all exceed 5 so we can carry out the goodness of fit test.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

STEP 1 State Null and Alternative hypotheses

STEP 2

Calculate $\chi^2 = \sum_{i=1}^{n} \frac{O_i^2}{E_i} - N \;\; or \;\; \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

STEP 3 Degrees of freedom = ν = n - k

STEP 4 Use tables to find Critical value

STEP 5 Accept or reject H_0

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

STEP 1 H_0 : The given distribution is binomial

H_1 : The given distribution is NOT binomial

STEP 2

$$\chi^2 = \sum_{i=1}^{3} \frac{O_i^2}{E_i} - N = \frac{12^2}{11.51} + \frac{16^2}{15.93} + \frac{12^2}{12.56} - 40 = 0.05$$

STEP 3

n = 3 *We have 3 pairs of values.*

Restrictions = 2 $\sum O = \sum E = 40$ *and*

Degrees of freedom = 1 *mean= 1.125*

STEP 4 Critical value = $\chi_1^2\,(0.95) = 3.84$

STEP 5 Accept H_0

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

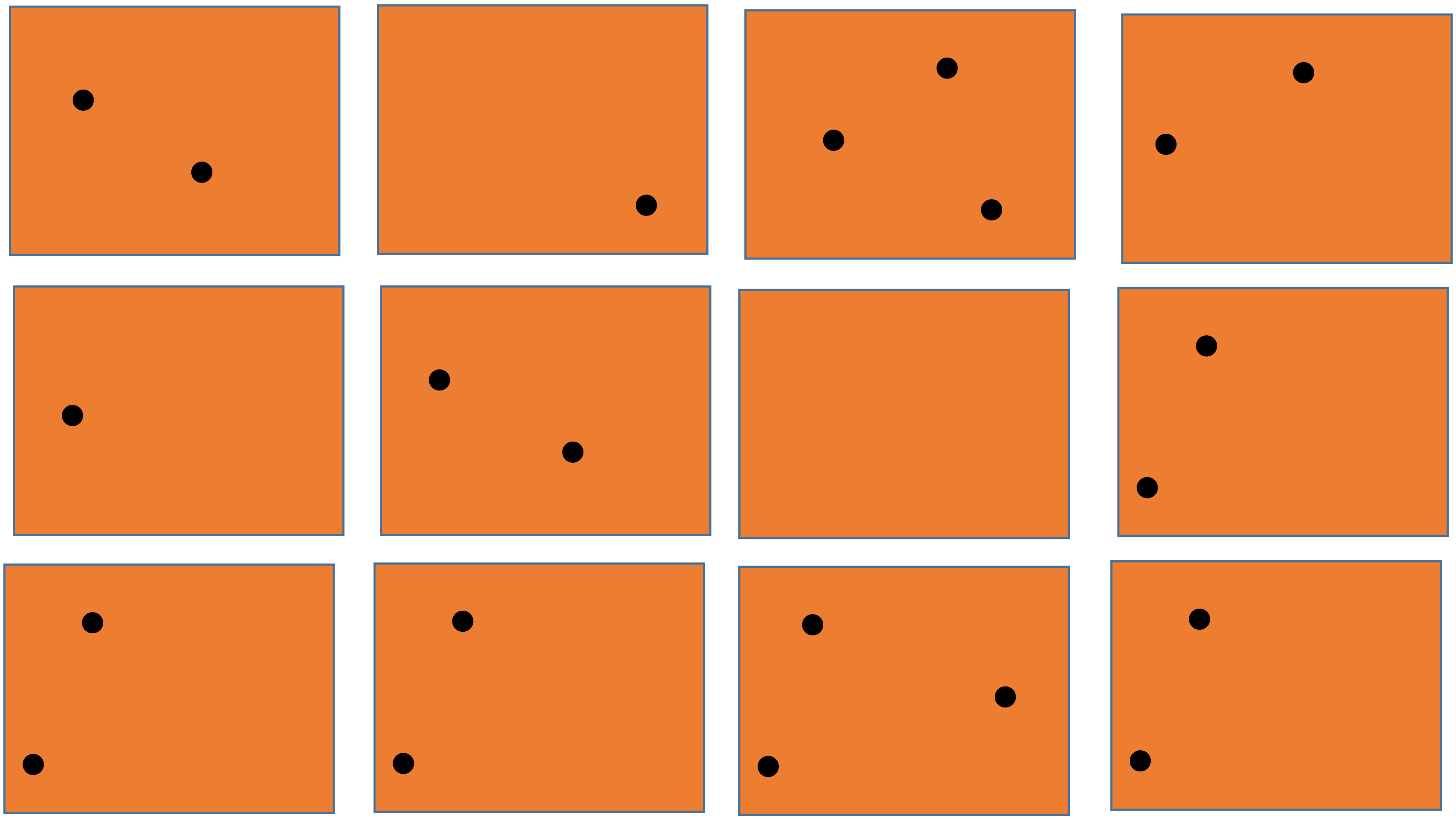
The Poisson distribution as a model

Do you remember the example we used to introduce the Poisson distribution?

"Consider a machine producing square metal sheets. The number of flaws on each of 100 sheets is noted ."

Let's carry out a 5% goodness of fit test on a similar set of data.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz



Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

Number of flaws	Frequency
0	8
1	31
2	30
3	16
4	10
5	4
6	1

Is this a Poisson distribution ?

The mean number of flaws per sheet is $205/100 = 2.05$

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2	3	4	5	6+
O	8	31	30	16	10	4	1
E	12.87	26.39	27.05	18.48	9.47	3.88	1.86

The expected frequencies have been calculated using

$$100P(X = x) = e^{-2.05}\frac{2.05^x}{x!}$$

Note that $\sum E = \sum O = 100$

and we had to use the observed data to find λ

We have TWO constraints

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

The Poisson expected frequencies can be calculated very quickly.

$$100P(X = 0) = e^{-2.05} = 12.87$$

multiply this number by 2.05 to give 26.39

multiply this number by 2.05 and divide by 2 to give 27.05

multiply this number by 2.05 and divide by 3 to give 18.48

multiply this number by 2.05 and divide by 4 to give 9.47

multiply this number by 2.05 and divide by 5 to give 3.88

....

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

x	0	1	2	3	4	5	6+
O	8	31	30	16	10	4	1
E	12.87	26.39	27.05	18.48	9.47	3.88	1.86

We need to combine cells because some expected frequencies are less than 5

x	0	1	2	3	4	5+
O	8	31	30	16	10	5
E	12.87	26.39	27.05	18.48	9.47	5.74

Now we are ready to do our goodness of fit test!

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

STEP 1 H_0 : The given distribution is Poisson

H_1 : The given distribution is NOT Poisson

STEP 2

$$\chi^2 = \sum_{i=1}^{6} \frac{O_i^2}{E_i} - N = \frac{8^2}{12.87} + \frac{31^2}{26.39} + \frac{30^2}{27.05} + \frac{16^2}{18.48} + \frac{10^2}{9.47} + \frac{5^2}{5.74} - 100 = 3.43$$

STEP 3 n = 6 *We have 6 pairs of values.*

Restrictions = 2 $\sum O = \sum E = 100$ *and*

Degrees of freedom = 4 *mean= 2.05*

STEP 4 Critical value = $\chi_4^2(0.95) = 9.488$

STEP 5 Accept H_0

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

IMPORTANT SUMMARY EXAMPLE

4 coins were tossed 200 times with the following results:

Heads	0	1	2	3	4	TOTAL
Frequency	9	42	73	61	15	200

Decide whether the coins are biased using a χ^2 goodness of fit 5% test.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

IMPORTANT SUMMARY EXAMPLE

Heads	0	1	2	3	4	TOTAL
Frequency	9	42	73	61	15	200

We need to calculate expected frequencies and their total must be 200, first constraint.

Do we need to find the mean of the given data to find the expected frequencies?

If Yes then we have a second constraint.
If No then we don't have the second constraint.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

IMPORTANT SUMMARY EXAMPLE

Heads	0	1	2	3	4	TOTAL
Frequency	9	42	73	61	15	200

To decide whether the coins are biased we need to use B(4, 0.5).

We don't need to find p from the given data. Hence we only have the one constraint.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

IMPORTANT SUMMARY EXAMPLE

Heads	0	1	2	3	4	TOTAL
O	9	42	73	61	15	200
E	12.5	50	75	50	12.5	200

No need to combine any cells . All the expected frequencies exceed 5.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

STEP 1 H_0 : The given distribution is B (4, 0.5)

H_1 : The given distribution is NOT B (4,0.5)

STEP 2 $\chi^2 = 5.23$

STEP 3 n = 5

Restrictions = 1

Degrees of freedom = 4

We have 5 pairs of values.

$\sum O = \sum E = 200$

STEP 4 Critical value = $\chi_4^2\,(0.95) = 9.49$

STEP 5 Accept H_0

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz

SUMMARY: degrees of freedom for Binomial and Poisson fit

MODEL	Number of cells (after combining expected frequencies less than 5)	Estimated parameters	Degrees of freedom
Binomial	n	None (n and p given)	n-1
Binomial	n	1 (p estimated using mean of observed frequencies)	n-2
Poisson	n	None (λ given)	n-1
Poisson	n	1 (λ estimated from mean of expected frequencies)	n-2

References:

1. Palin A., Park A., Whiteley C., (2012), A-level mathematics for Edexcel Statistics 1, CGP, UK.
2. Attwood, G., Clegg, A., Dyer, G. and Dyer, J (2008), Edexcel AS and A-Level Modular Mathematics series S2, Pearson, Harlow, UK.
3. Lecture notes, Statistics and Math for Life Sciences courses, NUFYP, Nazarbayev University.

Adina Amanbekkyzy, a.amanbekkyzy@astanait.edu.kz