0521 Slides

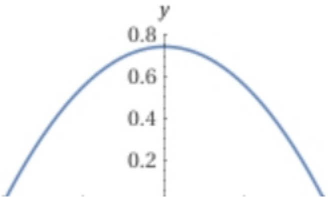
Cesai Li

May 2025

Recall: Suppose an experiment has a continuous numerical sample space. If we associate a veriable x to the experimental outcome, then x is said to be a continuous random variable, and there is a probability distribution $f(x)$ such that $f(i) \geq 0$ for any i in the sample space, and the area under the graph (curve) of $f(x)$ is exactly 1.

We've seen examples like $f(x) = 1$ on $0 \leq x \leq 1$ and

$$f(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{for } x \in [-1, 1], \\ 0 & \text{otherwise.} \end{cases}$$



Those are all **good probability density/probability distribution functions** in the sense that $f(i) \geq 0$ for all i and the underlying area is 1. However, real-world data can never be a perfect curve.

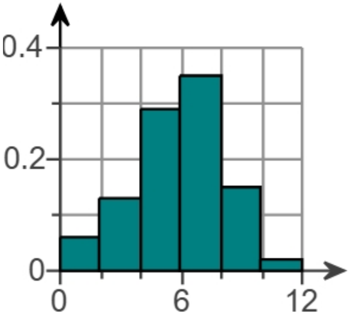
Approximation by a model $F(X)$

Suppose we perform an experiment and get a sample data set. We plot the sample points and notice the graph look *roughly* like a simple functions $F(x)$ (straight line, bell curve, etc.), we can probably make use of the nice properties of $F(x)$ to obtain useful information about our data set.

- We need to make sure that it is **reasonable** to approximate our data set by $F(x)$. There are methods with different strength, from looking at graphs or comparing means/median/standard deviation to performing careful numerical analysis.
- We obtain the [means/median/standard deviation/other info] from our nice simple function $F(x)$, and claim that the [means/median/standard deviation/other info] of our data set is approximately that of $F(x)$.

Population vs. Sample vs. Model

If we claim that the probability distribution of our random variable x **can be approximated** by a nice mathematical function $F(x)$, then we call $F(x)$ a **model**.



This is a histogram of a data set. By looking at its shape, one may claim that this data set can be approximated by a normal curve.

Real experiment: **Population** vs. **Sample** vs. **Model**

In a physics experiment, I want to measure properties of an acoustic box. A speaker is places inside the acoustic box to produce sound wave. A microphone is used to detect sound wave. The microphone is connected to an oscilloscope to give me Amplitude and Frequency data points.

I propose a theoretical **model**

$$Amp = \frac{F_o}{\sqrt{1 + (\frac{2\pi}{\gamma f})^2 \cdot (f^2 - f_o^2)^2}}$$

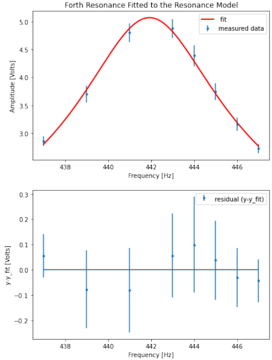
I want to check if this model describes the properties of the acoustic box. This will not be on quiz/exam. Just to help understand the population vs. sample vs. model distinction.

I perform my experiment and got a data set (sample). In this setting, my **population** is a continuous quantity 'Amplitude vs. frequency during an interval of time'. My **sample** data set is

	A	B	C	D	E	F	G	H	I	J	K	L
1	# frequency (f) in Hz	#u[f] in Hz	#Amplitude (A) in V	#u[A] in V	#A1 in V	#A2 in V	#A3 in V	#A4 in V	#A5 in V	#Amplitude (A) in V	#STDEV[A] in V	#u[A] in V
2	290	0.4	1.9340	0.0075	1.93	1.95	1.93	1.95	1.91	1.934	0.016733201	0.0074833
3	295	0.1	2.738	0.016	2.72	2.76	2.74	2.69	2.78	2.738	0.034928498	0.0156205
4	297	0.1	2.9900	0.0089	3.01	2.97	2.99	3.01	2.97	2.99	0.02	0.0089443
5	299	0.1	3.238	0.019	3.24	3.29	3.18	3.26	3.22	3.238	0.041472883	0.0185472
6	302	0.2	3.338	0.010	3.37	3.33	3.35	3.31	3.33	3.338	0.022803509	0.010198
7	304	0.1	3.2500	0.0095	3.25	3.27	3.24	3.27	3.22	3.25	0.021213203	0.0094868
8	306	0.1	2.9740	0.0075	2.99	2.97	2.95	2.99	2.97	2.974	0.016733201	0.0074833
9	308	0.1	2.6700	0.0089	2.69	2.65	2.67	2.65	2.69	2.67	0.02	0.0089443
10	310	0.1	2.432	0.010	2.42	2.44	2.46	2.4	2.44	2.432	0.022803509	0.010198
11	392	0.2	3.908	0.015	3.94	3.9	3.86	3.9	3.94	3.908	0.033466401	0.0149666
12	393	0.3	4.520	0.028	4.56	4.6	4.52	4.44	4.48	4.52	0.063245553	0.0282843
13	394	0.3	5.136	0.025	5.16	5.1	5.06	5.16	5.2	5.136	0.055497748	0.0248193
14	395	0.2	5.136	0.025	5.16	5.2	5.1	5.16	5.06	5.136	0.055497748	0.0248193
15	396	0.5	4.740	0.013	4.74	4.78	4.74	4.74	4.7	4.74	0.028284271	0.0126491
16	398	0.5	3.3940	0.0075	3.41	3.39	3.37	3.39	3.41	3.394	0.016733201	0.0074833
17	400	0.4	2.6940	0.0075	2.69	2.71	2.67	2.71	2.69	2.694	0.016733201	0.0074833
18	404	1.4	1.962	0.020	2.02	1.98	1.94	1.9	1.97	1.962	0.04494441	0.0200998
19	425	0.3	2.0120	0.0086	2.02	1.98	2.02	2.03	2.01	2.012	0.019235384	0.0086023
20	430	0.2	3.4640	0.0051	3.47	3.45	3.46	3.48	3.46	3.464	0.011401754	0.005099
21	440	0.1	5.152	0.015	5.2	5.16	5.16	5.12	5.12	5.152	0.033466401	0.0149666
22	441	0.1	5.468	0.015	5.46	5.5	5.5	5.46	5.42	5.468	0.033466401	0.0149666
23	442	0.3	5.692	0.055	5.58	5.64	5.9	5.7	5.64	5.692	0.123773988	0.0553534
24	443	0.2	5.424	0.080	5.28	5.7	5.5	5.28	5.36	5.424	0.178549713	0.0798499
25	450	0.3	1.700	0.026	1.65	1.68	1.65	1.78	1.74	1.7	0.057879185	0.0258844
26	460	1.5	0.3456	0.0027	0.344	0.348	0.352	0.336	0.348	0.3456	0.0060663	0.0027129
27	488	0.5	0.6872	0.0049	0.69	0.694	0.668	0.694	0.69	0.6872	0.010917875	0.0048826
28	490	0.6	0.9536	0.0052	0.952	0.956	0.944	0.972	0.944	0.9536	0.011523888	0.0051536
29	491	0.2	1.0980	0.0066	1.1	1.09	1.12	1.1	1.08	1.098	0.014832397	0.0066332
30	492	0.6	1.1880	0.0066	1.17	1.18	1.19	1.21	1.19	1.188	0.014832397	0.0066332
31	493	1.8	1.144	0.013	1.18	1.17	1.12	1.13	1.12	1.144	0.028809721	0.0128841
32	494	1.1	1.0100	0.0071	1.02	1.01	0.99	1.03	1	1.01	0.015811388	0.0070711
33	497	0.5	0.5176	0.0047	0.516	0.508	0.524	0.532	0.508	0.5176	0.010430724	0.0046648
34	500	1.1	0.260	0.026	0.257	0.264	0.207	0.218	0.355	0.2602	0.058358376	0.0260987
35	541	10	0.1950	0.0022	0.199	0.196	0.189	0.191	0.2	0.195	0.00484768	0.0021679
36	545	0.4	0.7824	0.0020	0.776	0.784	0.78	0.784	0.788	0.7824	0.004560702	0.0020396
37	550	0.6	0.9200	0.0036	0.912	0.928	0.92	0.928	0.912	0.92	0.008	0.0035777
38	553	1.4	0.9296	0.0069	0.936	0.912	0.92	0.952	0.928	0.9296	0.015388307	0.0068819
39	554	0.9	0.945	0.022	0.912	0.928	1.03	0.936	0.92	0.9452	0.048241061	0.0215741
40	555	1	0.8962	0.0045	0.88	0.896	0.904	0.896	0.905	0.8962	0.010009995	0.0044766
41	560	0.8	0.6760	0.0036	0.676	0.684	0.668	0.684	0.668	0.676	0.008	0.0035777
42	575	0.2	2.0080	0.0097	2.04	1.99	2.02	2	1.99	2.008	0.021679483	0.0096954
43	580	0.1	3.734	0.011	3.71	3.76	3.73	3.71	3.76	3.734	0.025099801	0.011225
44	583	0.1	5.272	0.015	5.28	5.24	5.32	5.28	5.24	5.272	0.033466401	0.0149666
45	584	0.1	5.488	0.079	5.44	5.36	5.4	5.44	5.8	5.488	0.177538728	0.0793977
46	585	0.1	5.168	0.015	5.12	5.16	5.2	5.16	5.2	5.168	0.033466401	0.0149666
47	586	0.1	4.696	0.027	4.72	4.68	4.72	4.6	4.76	4.696	0.060663004	0.0271293

Real experiment: **Population** vs. **Sample** vs. **Model**

I then fit my sample data points (after some data analysis) to my model.

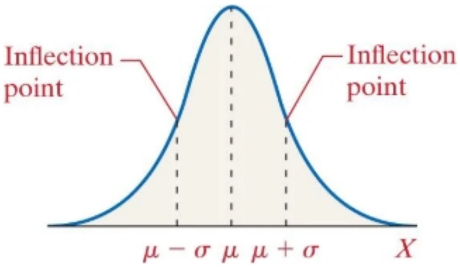


This seems to be a goot approximation. I may then claim that the acoustic box's amplitude vs. frequency behaves roughly like my model. Let P be a mathematical properties of my model. I can claim that the acoustic box's amplitude vs. frequency roughly has this property P.

Normal distribution / bell curve

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-1/2[(x-\mu)/\sigma]^2)$$

A (hypothetical) variable x is called a **normal variable** if its probability distribution is exactly this $f(x)$. If $\mu = 0$ and $\sigma = 1$ we say such a normal variable is a **standard normal variable** and $f(x)$ is a standard normal distribution. We often denote a standard normal variable by z.



Properties of a bell curve

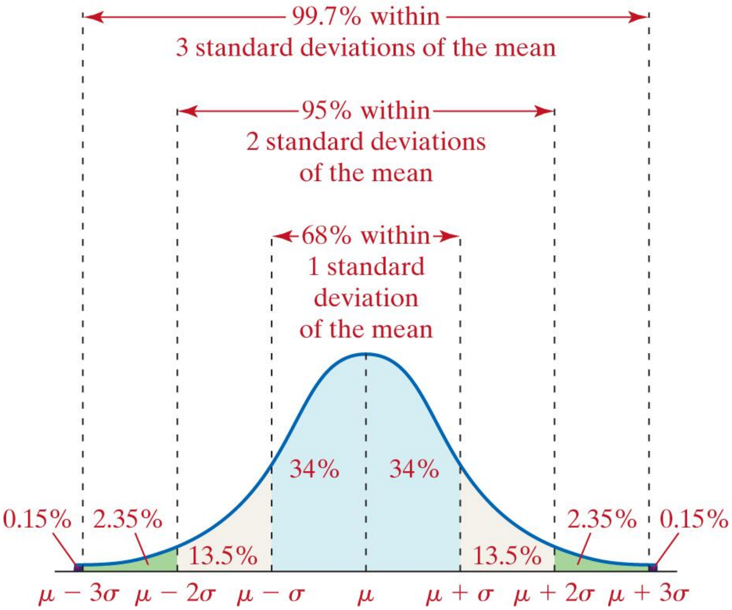
1. The bell curve is symmetric about its mean μ.
2. The area under the bell curve is 1.
3. The curve gets closer and closer to 0 as x increases/decreases away from μ, but never really gets to 0.

7. The Empirical Rule:

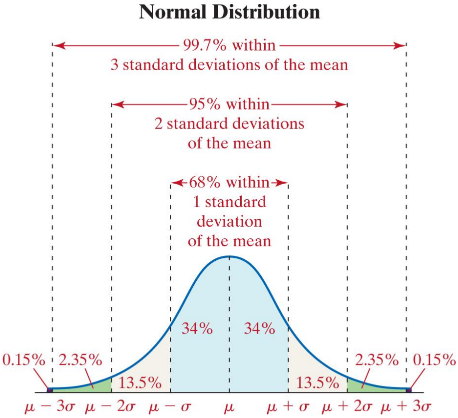
- Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$;
- approximately 95% of the area is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$;
- approximately 99.7% of the area is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$.

4.
5. $\mu \pm \sigma$ are **inflection points**.

Normal Distribution



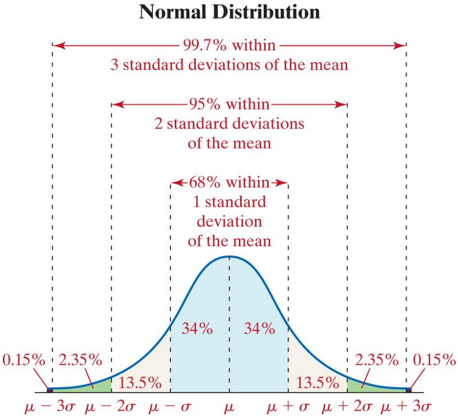
Example 1



Suppose a normal variable x is of $\mu = 1$ and $\sigma = 1$.

1. What is $P(x \geq 1)$?
2. What is $P(0 \leq x \leq 2)$?
3. What is $P(x \leq 1)$?

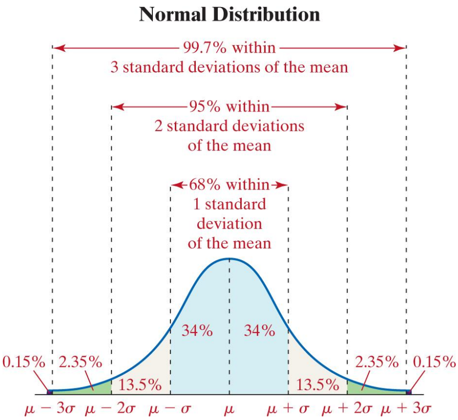
Example 2



Suppose a normal variable x is of $\mu = 1$ and $\sigma = 1$.

1. What is $P(x \geq 2)$?
2. What is $P(x \leq 0)$?

Example 3

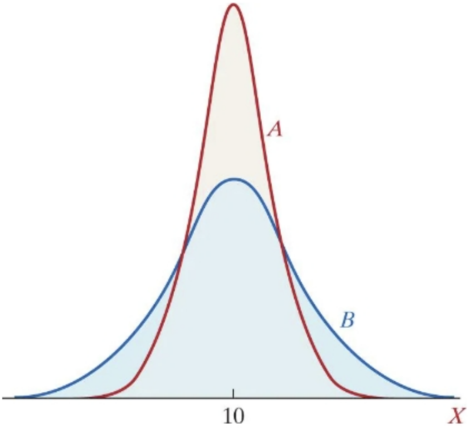


What is $P(x \leq \mu - \sigma)$ for the following two normal variable x?

1. A normal variable x of $\mu = 1$ and $\sigma = 1$
2. A normal variable x of $\mu = 0$ and $\sigma = 2.5$

Compare two normal distribution

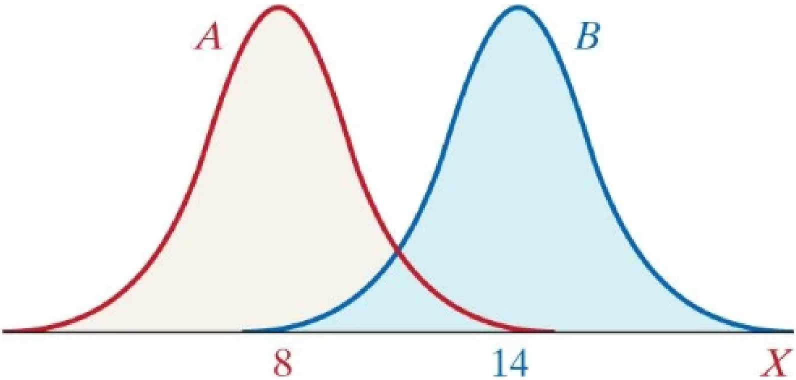
Compare two normal curves A and B.



What information can we get?

Compare two normal distribution

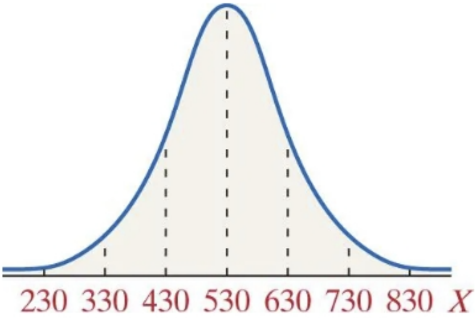
Compare two normal curves A and B.



What information can we get?

Example 4

Given a normal curve as below, what are its mean μ and standard deviation σ?



Poll Activity 7

The function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2}[(x-\mu)/\sigma]^2\right)$$

is called the Normal distribution. It is a sort of

1. Sample
2. Population
3. Experiment outcome
4. Variable
5. None of the above

This question will not be on quiz/exam.

Poll Activity 8

Suppose x is a normal random variable with mean μ and standard deviation σ. What is $P(x < \mu - \sigma)$?

1. 34%
2. 50%
3. 13.5%
4. 16%
5. None of the above
6. Depends on μ and σ

Standard normal variable

A normal variable is called a **standard normal variable** if it has a mean $\mu = 0$ and $\sigma = 1$. We often denote a standard normal variable by z or Z.

Let's read the Standard Normal Distribution Table. Given a z-value a, this table tells $P(z \leq a)$.

Standard Normal Distribution Table

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.927[illegible]
1.5	[illegible]	[illegible]	[illegible]	[illegible]	0.9382	0.9394	[illegible]

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.927
1.5	[illegible]	[illegible]	[illegible]	[illegible]	0.9382	0.9394	[illegible]

1. What is $P(z \leq -1.33)$?
2. What is $P(z \leq 1.33)$?
3. What is $P(z \geq -1.33)$?
4. What is $P(z \geq 1.33)$?

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.927[illegible]
1.5	[illegible]	[illegible]	[illegible]	[illegible]	0.9382	0.9394	[illegible]

What is $P(1.03 \leq z \leq 1.33)$?

Change of Variable

Change of variable $x \to z$ is a key concept

Given a non-standard normal variable x with a mean μ and a standard deviation σ, we may convert it to a normal variable z by the formula

$$z = \frac{x - \mu}{\sigma}$$

Lemma.

$$P(\frac{x-\mu}{\sigma} \leq a) = P(x \leq a\sigma + \mu)$$

$$P(\frac{x-\mu}{\sigma} \geq a) = P(x \geq a\sigma + \mu)$$

for any number a.

Proposition

Suppose x is a normal variable with mean μ and standard deviation σ. Let $z = \dfrac{x-\mu}{\sigma}$. Then:

$$P(x \leq a) = P(z \leq \frac{a-\mu}{\sigma})$$

$$P(x \geq a) = P(z \geq \frac{a-\mu}{\sigma})$$

for any number a.

Example 5

Proposition

Suppose x is a normal variable with mean μ and standard deviation σ. Let $z = \frac{x-\mu}{\sigma}$. Then:

$$P(x \leq a) = P(z \leq \frac{a-\mu}{\sigma})$$

$$P(x \geq a) = P(z \geq \frac{a-\mu}{\sigma})$$

for any number a.

Let x be a normal variable with $\mu = 150$ and $\sigma = 10$. Use the change of variable method and the Standard Normal Distribution Table to find out $P(x \geq 160)$.

Example 6

Proposition

Suppose x is a normal variable with mean μ and standard deviation σ. Let $z = \dfrac{x-\mu}{\sigma}$. Then:

$$P(x \leq a) = P(z \leq \frac{a-\mu}{\sigma})$$

$$P(x \geq a) = P(z \geq \frac{a-\mu}{\sigma})$$

for any number a.

Let x be a normal variable with $\mu = 150$ and $\sigma = 10$. Use the change of variable method and the Standard Normal Distribution Table to find out $P(155 \leq x \leq 160)$.

Key questions to ask:

1. If we know a data set is (approximately) a normal distribution, how can we extract information we need from the normality? (Be careful with **Population** vs. **Sample** vs. **Model**)
2. Given a data set, how can we tell if it can be approximated by a normal distribution?

Example. The GRE is a test required for admission to many US graduate schools. Suppose students' scores on some GRE test can be approximated by a normal distribution with mean 150 and standard deviation 10. What proportion of the students scored between 155 and 160? **Answer: Approximately 14.98%**

Assessing normality

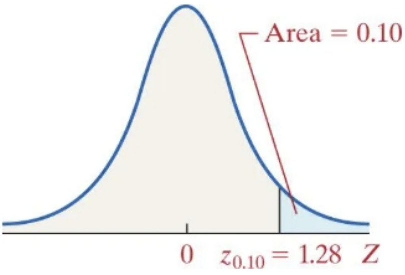
Given a variable x, how do we know if x is approximately a normal variable? (I.e. if the probability density function $f(x)$ can be approximated by a normal distribution.)

1. Method 1. Plot $f(x)$ and compare its graph to that of a normal distribution.
2. Method 2. Compare $\overline{x} \pm s$, $\overline{x} \pm 2s$, $\overline{x} \pm 3s$ to 68%, 95%, 99.7%, resp.
3. Methods using technology or more advanced math tools (not covered in this course).

Be careful: In the textbook the term 'z-score' has two different meanings. Here we do not call z_α a 'z-score'.

Definition of z_α

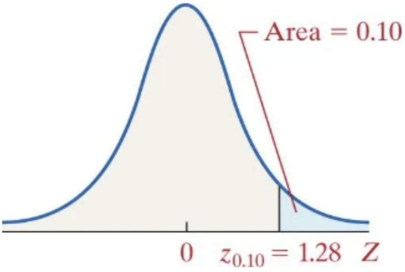
Given a number $0 \le \alpha \le 1$, z_α is defined by $P(z \ge z_\alpha = \alpha)$.



Example

Definition of z_α

Given a number $0 \leq \alpha \leq 1$, z_α is defined by $P(z \geq z_\alpha = \alpha)$.



Use the Standard Normal Distribution Table to find out $z_{0.3}$.

Property of z_α

Look at the diagram, the symmetry implies the following result.

Lemma

For any $0 \leq \alpha \leq 1$ we have $-z_\alpha = z_{1-\alpha}$.

Important concepts: Population Parameter vs. Sample Statistic.

	Population Parameter	**Sample Statistic**
Mean	μ	$\overline{x}$
Median	η	M
Variance	σ^2	s^2
Standard Deviation	σ	s