統計学

講義資料その6
正規分布

兵庫県立大学 社会情報科学部

山本 岳洋

t.yamamoto@sis.u-hyogo.ac.jp

2021年度前期・火曜3限　神戸商科キャンパス 全学共通科目

この資料の内容

- 教科書「確率統計」
  - 3章2節
- 参考書 小波先生「統計学入門」
  - 5章
- 参考書 東大出版会「統計学入門」
  - 5章・6章

この資料の内容

- 正規分布
- 標準正規分布
- 標準正規分布と正規分布の関係
- 正規分布上の確率計算の例題

この資料の確認ポイント

- **正規分布の性質が分かる**
  - 山の形のような分布
  - 期待値 μ で左右対称の形をしている
- **標準正規分布$N(0,1)$と正規分布$N(\mu,\sigma^2)$の関係が分かる**
  - p.54 の関係式は暗記してもよいくらい重要
- **標準正規分布$N(0,1)$上の確率を用いて正規分布$N(\mu,\sigma^2)$上の確率が計算できる**
- **Excel（Googleスプレッドシート）を用いて正規分布$N(\mu,\sigma^2)$上の確率が計算できる**

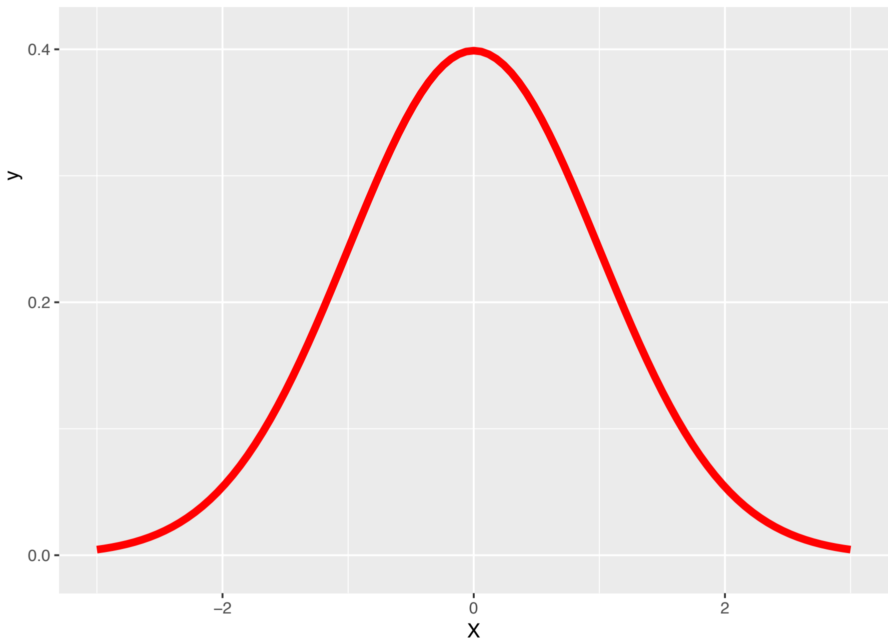
ここからの内容

- いよいよ正規分布が登場します
- 正規分布の確率密度関数の式に惑わされず，その性質を理解してください
  - 積分を展開することはありませんので，ビビらないように
- この資料の演習ができれば，本講義後半の区間推定や仮説検定もほとんど理解できたようなものです

正規分布

正規分布

- **世の中の自然現象や社会現象の多くに当てはまる分布**
  - *正規分布なしに近代の統計理論はありえない*
    - 参考書: 東大出版会「統計学入門」 p.120 より引用



正規分布が分かると･･･

- **大・中都市圏に住む14歳男子の50m走のタイムは平均7.5秒，　標準偏差0.6秒であった.**
  - 参考: 平成29年政府「体力・運動能力調査」より

- **タイムが正規分布に従うと仮定すると，**
  - 6.3秒から8.7秒の間に入る生徒の割合は？
  - 中間95%層（上位2.5%から97.5%）は何秒から何秒？
  - 6.9秒で走る学生は上位何%に入っているか？
  - 上位2.5%に入るためには何秒以上で走る必要があるか？
- **このように，平均と標準偏差だけからいろんな確率が計算できるようになる**

正規分布の確率密度関数

- 期待値 μ , 分散 σ^2 の正規分布の確率密度関数は以下で与えられる

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$f(x)$の定義域はすべての実数値 $(-\infty < x < +\infty)$

e は自然対数の底（ネイピア数） $e = 2.718 \ldots$

正規分布の確率密度関数（別表記）

- 期待値 μ , 分散 σ^2 の正規分布の確率密度関数は以下で与えられる

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

$\exp(x) = e^x$ （この方が印刷上見やすいので $\exp(x)$ の記法の方がよく用いられる）

正規分布の期待値・分散

- **前ページの確率密度関数に従う確率変数は期待値 $E(X) = \mu$, 分散 $V(X) = \sigma^2$ となることが知られている**
  - 中間レポートで発展課題として出す予定
  - ガウス積分とよばれる公式を使うと解けます

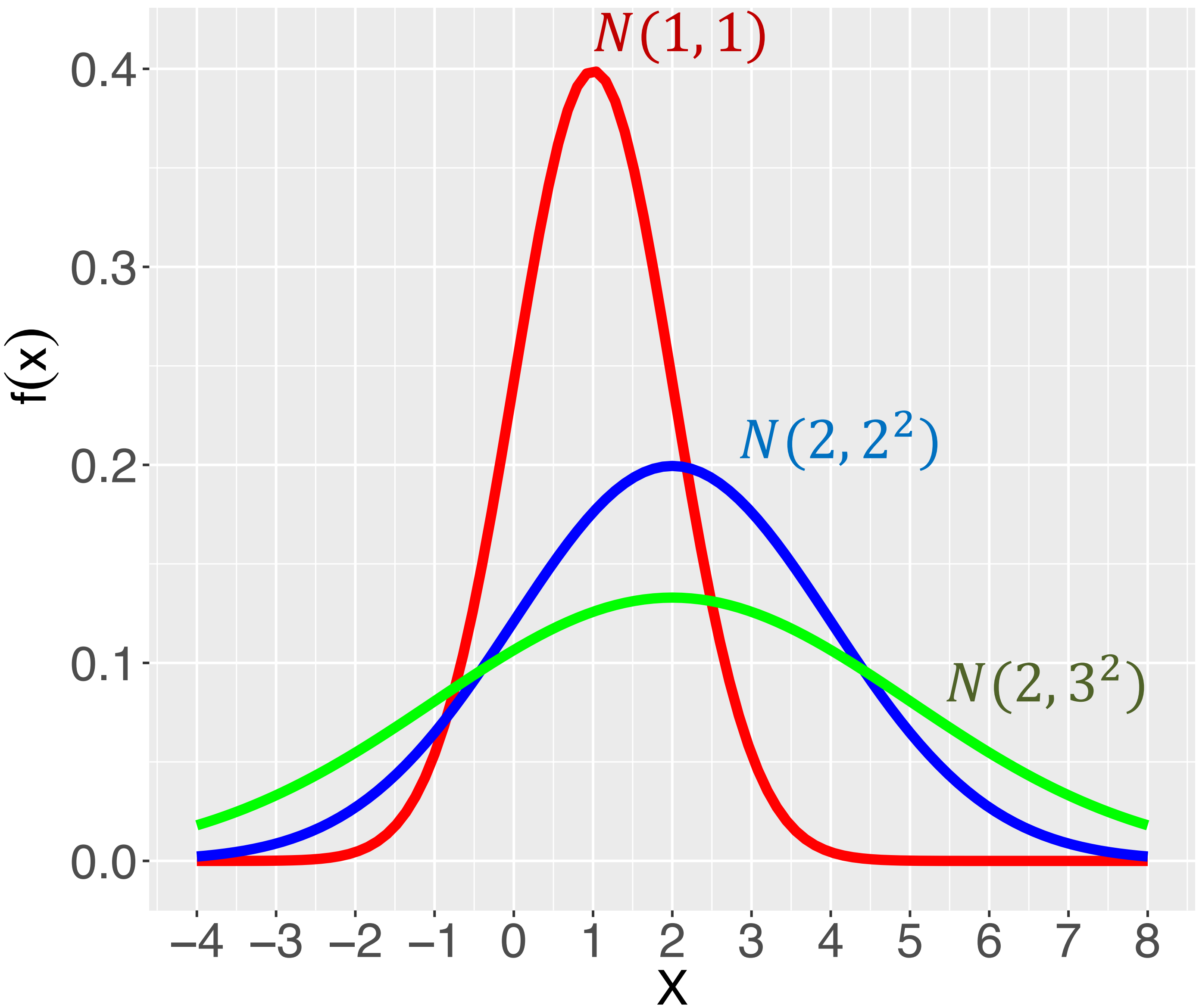
- $E(X) = \int_{-\infty}^{+\infty} xf(x)dx \ = \cdots = \ \mu$
- $V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx \ = \cdots = \sigma^2$

正規分布 $N(\mu, \sigma^2)$ に従う

- **確率変数 X が 期待値 μ , 分散 σ^2 の正規分布に従うとき，X は正規分布 $N(\mu, \sigma^2)$ に従うとあらわす**

- 例: *50m走のタイムが 正規分布 $N(7.5, 0.6^2)$ に従うと仮定するとき⋯,* と表現したりする

正規分布の具体的な形



- **分布の形はなめらか**
- **最頻値**（モード）**は**
  $$X = \mu$$
- **σ^2が大きいほど分布はなだらか**
  - あるいは，σ^2が小さいほど分布は尖っている

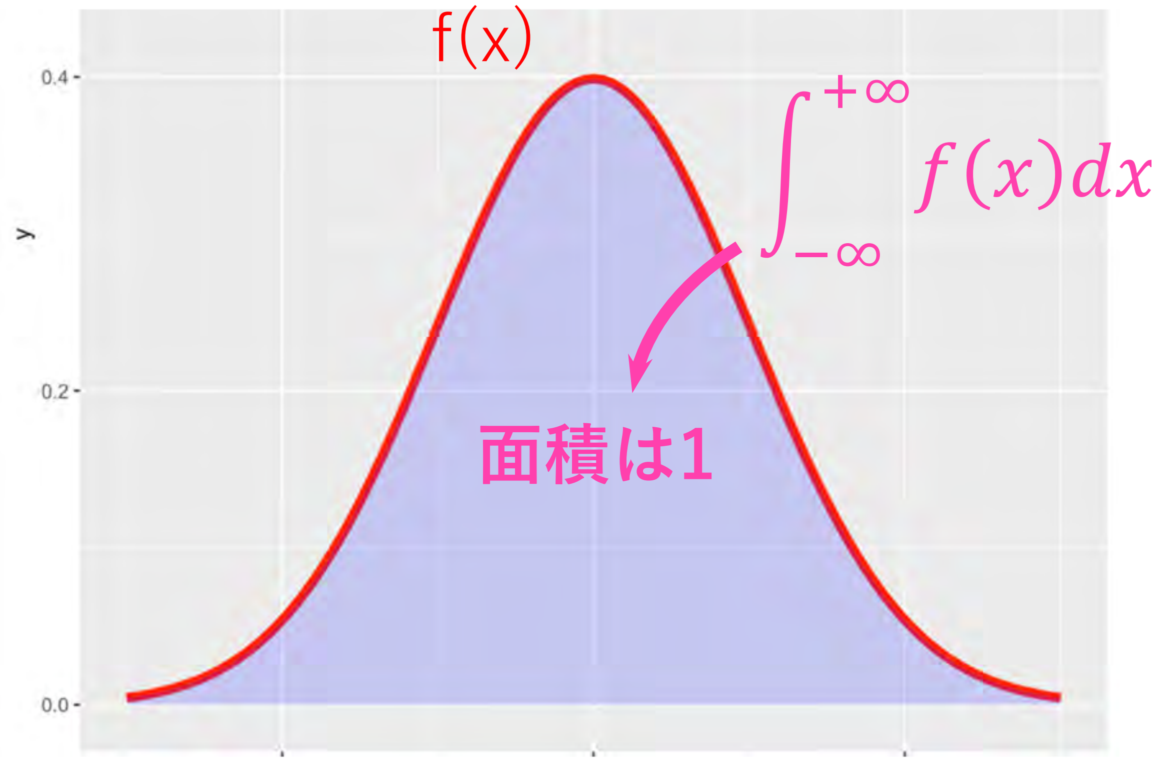
正規分布の性質

- Xはすべての実数値$(-\infty \leq x \leq +\infty)$をとりうる

- 確率密度関数を積分したらちゃんと1になる
  - $P(-\infty \leq x \leq +\infty)$

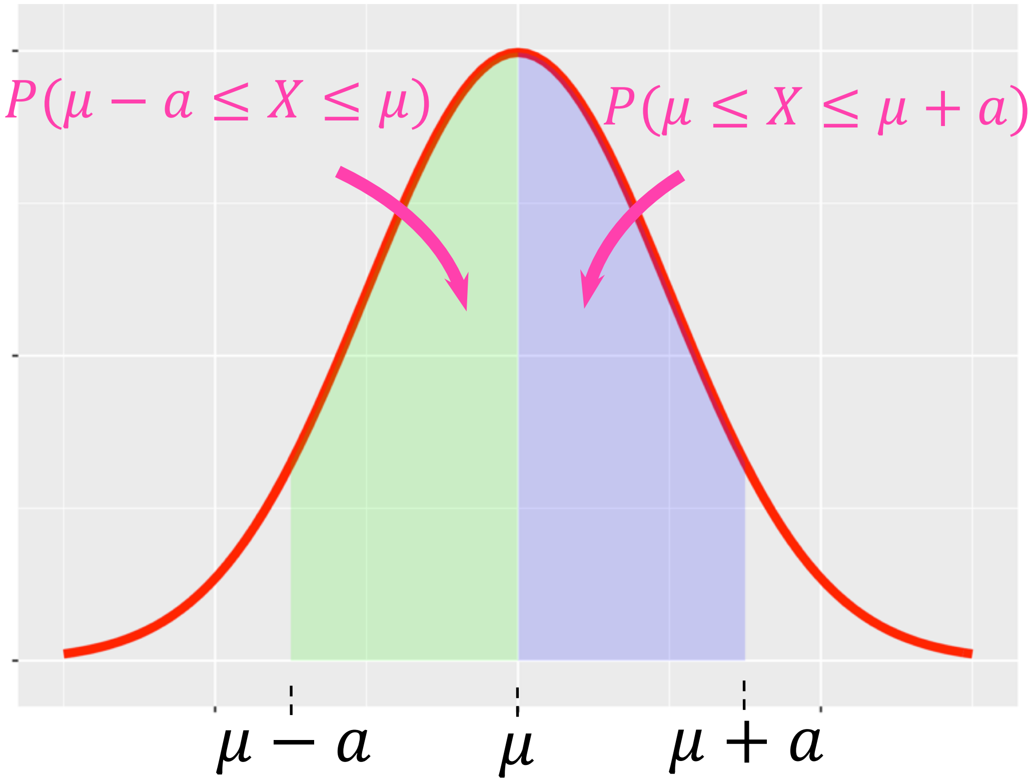
$$= \int_{-\infty}^{+\infty} f(x)dx$$

$$= 1$$



正規分布の対称性 (重要)

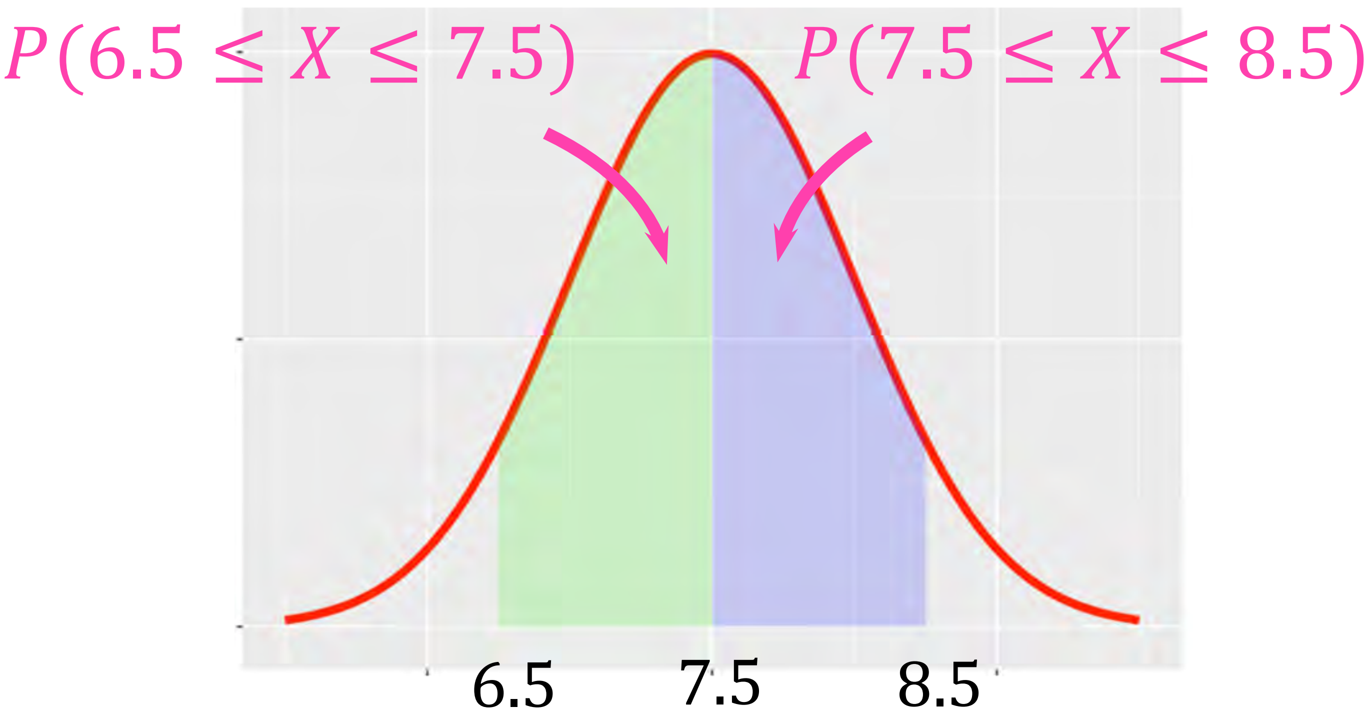
- **期待値 μ に関して左右対称**
  - いま，$a > 0$とすると
    - $P(\mu - a \leq X \leq \mu) = P(\mu \leq X \leq \mu + a)$
    - $P(X \leq \mu - a) = P(\mu + a \leq X)$



正規分布の対称性（具体例）

50m走のタイムが 正規分布 $\mathrm{N}(7.5, 0.6^2)$ **に従うと仮定するとき**

- 6.5秒以上7.5秒以下で走る人の割合と,
  7.5秒以上8.5秒以下で走る人の割合は等しい
- 6.5秒以下で走る人の割合と,
  8.5秒以上で走る人の割合も等しい



参考: 正規分布のさまざまな呼び方

- 正規分布
- Normal Distribution
- ガウス分布
- Gaussian Distribution

すべて正規分布の呼び方

– この講義では正規分布と呼びます

標準正規分布

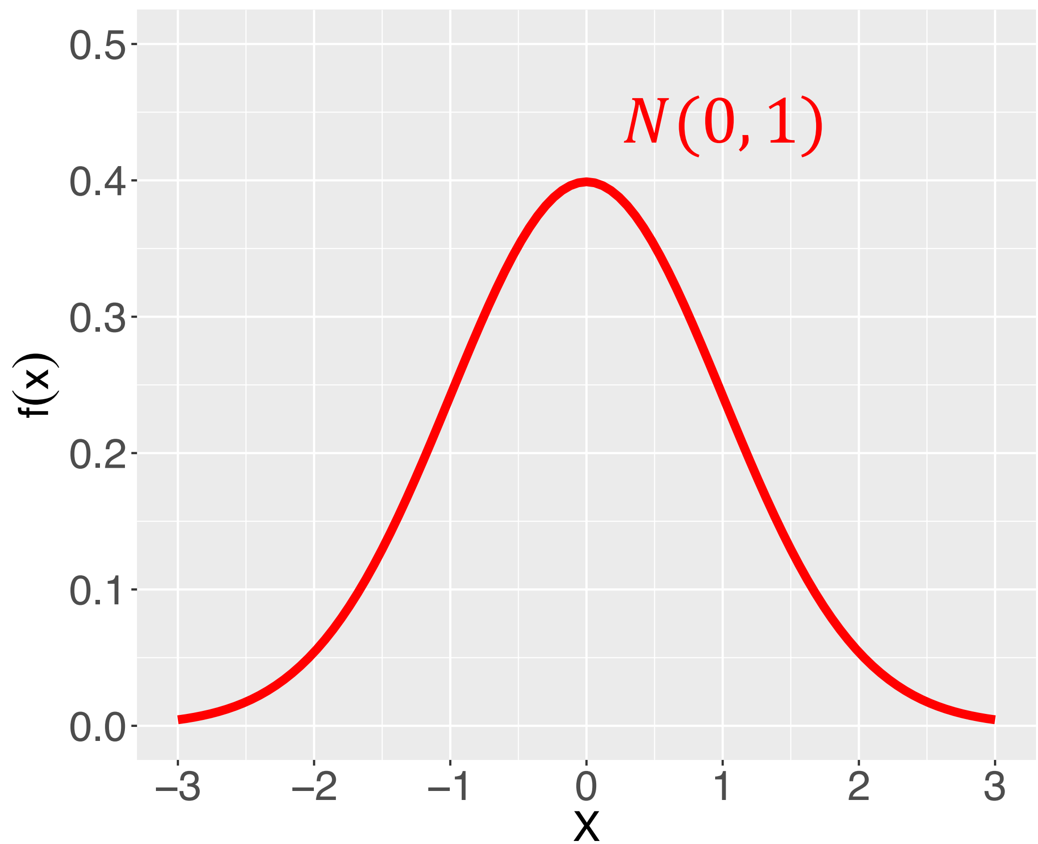
標準正規分布

- **期待値 0，標準偏差1 の正規分布を**
  標準正規分布 といい，$N(0,1)$ とあらわす
  - 記述統計の標準化と関連

- **確率密度関数**

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

p.9の$f(x)$に$\mu = 0, \sigma^2 = 1$ を代入した式になっている



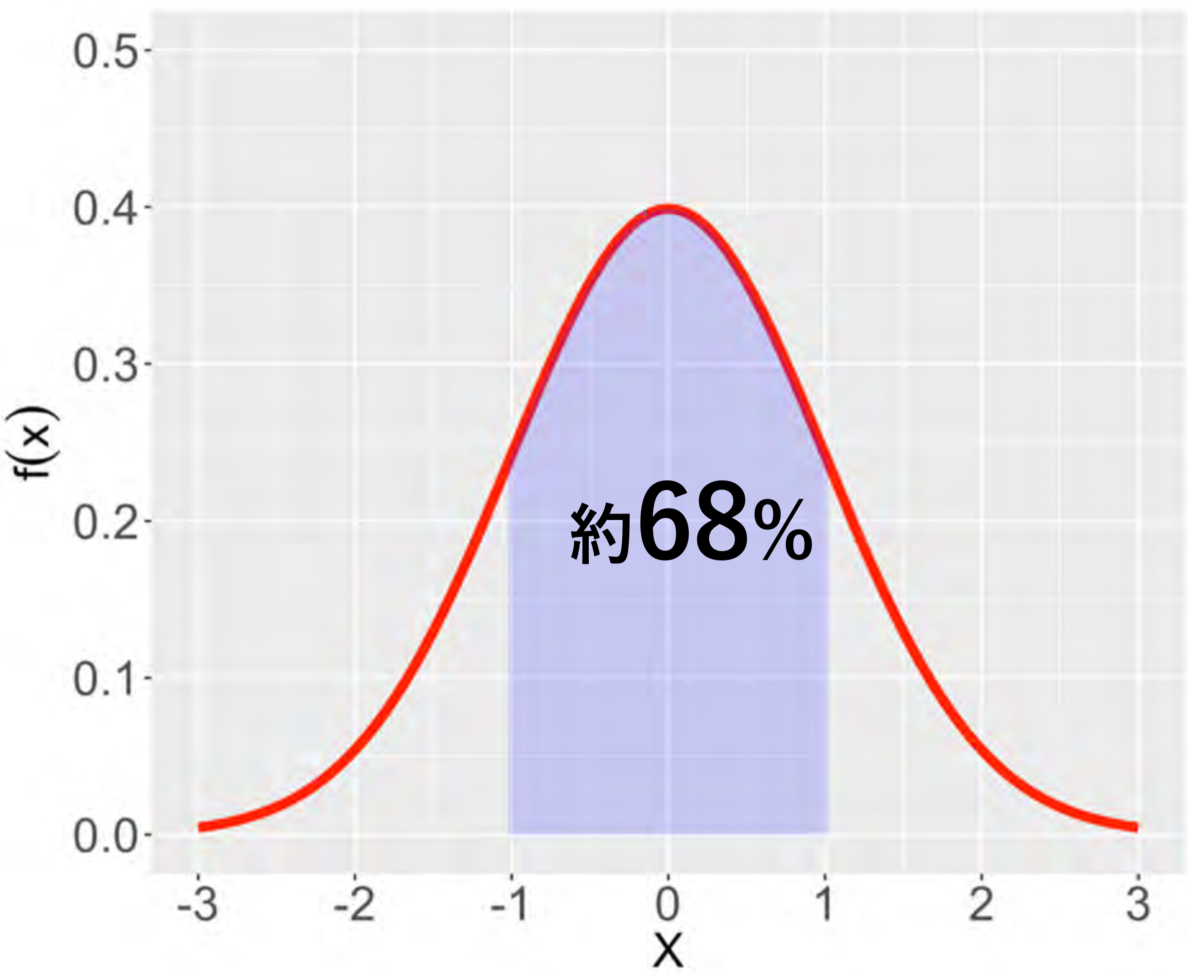
標準正規分布の役割

- 後述するスライドでみるように,
  標準正規分布 $N(0,1)$ における確率さえ
  計算できれば，任意の正規分布$N(\mu,\sigma^2)$に
  おける確率が計算できる

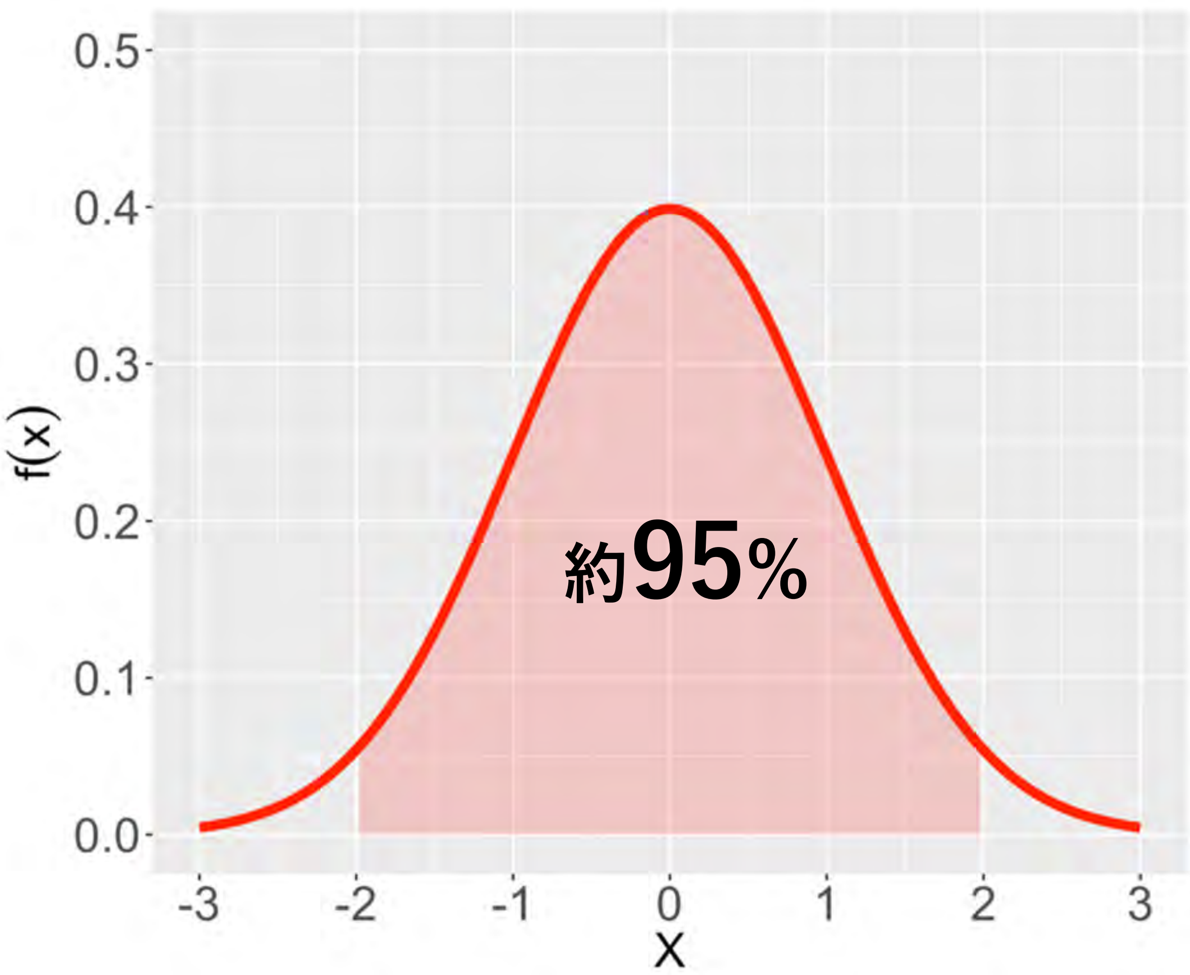
標準正規分布の有名な確率
（覚えても損はないので覚えてください）

- **確率変数 Z が標準正規分布$N(0,1)$に従うとき，以下の確率がよく知られている**
  - 補足: Z は標準正規分布に従う確率変数を表すときによく用いられる記号
  - $P(-1 \leq Z \leq 1) =$ 約0.68
  - $P(-2 \leq Z \leq 2) =$ 約0.9545
  - $P(-3 \leq Z \leq 3) =$ 約0.9973

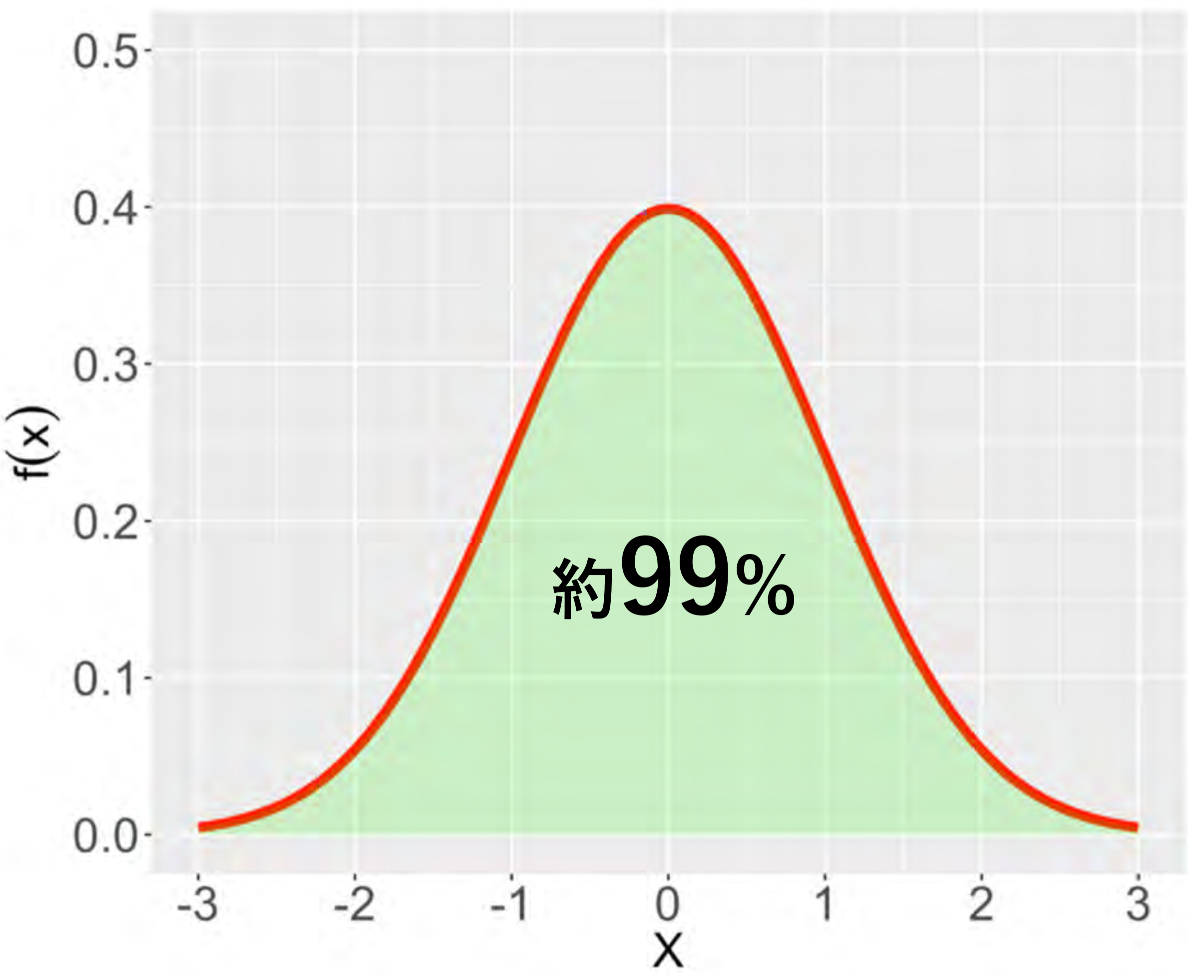
標準正規分布の有名な確率



標準正規分布の有名な確率



標準正規分布の有名な確率

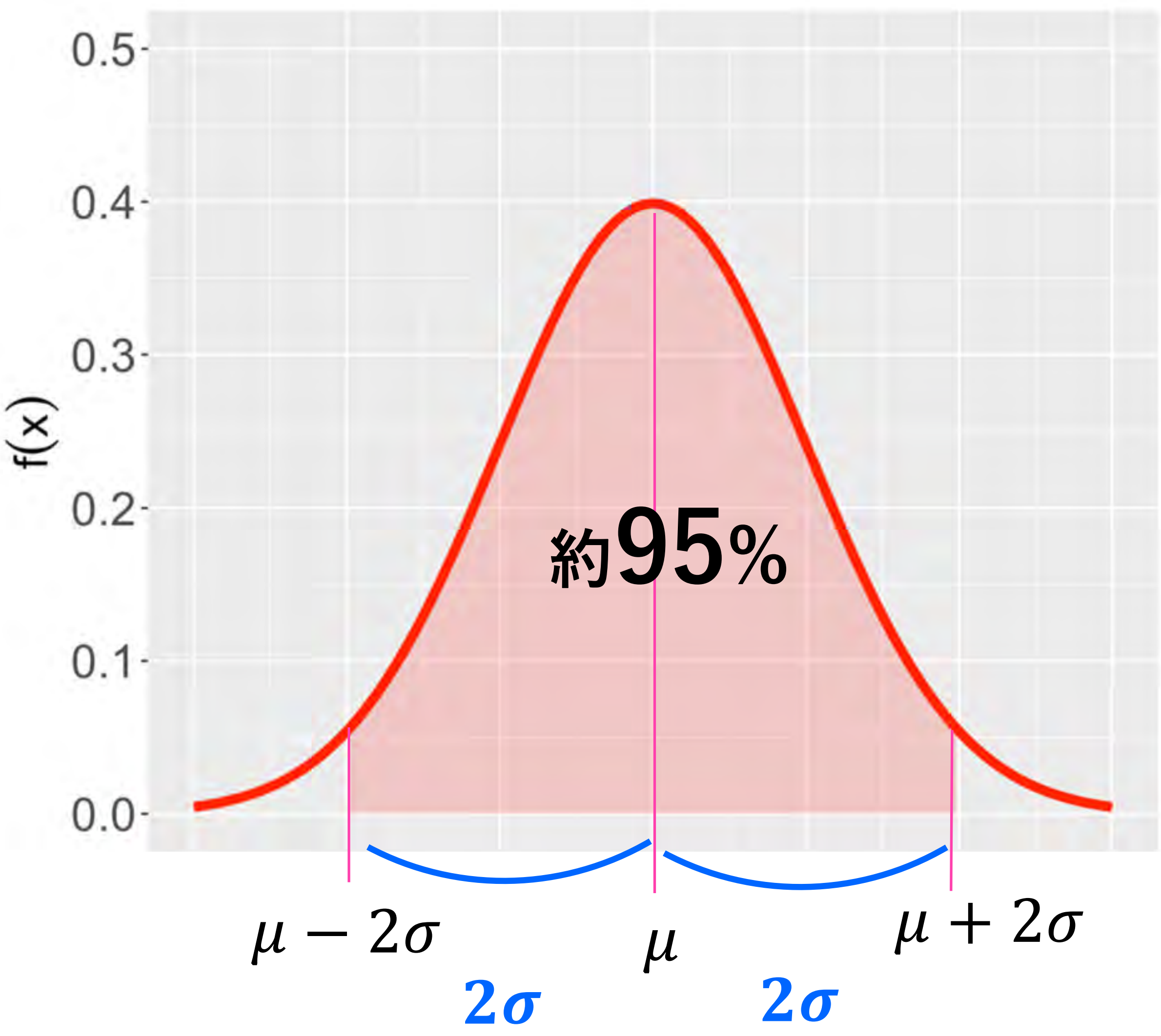


正規分布の有名な確率（これも覚えておいて損はない）

- いまみた標準正規分布の有名な確率と，後述する正規分布と標準正規分布の対応関係を使うと，以下のことが分かる
- 確率変数 X が$N(\mu,\sigma^2)$ に従うとき
  - $P(\mu-\sigma \le X \le \mu+\sigma)=$ 約68%
  - $P(\mu-2\sigma \le X \le \mu+2\sigma)=$ 約95%
  - $P(\mu-3\sigma \le X \le \mu+3\sigma)=$ 約99%

正規分布の有名な確率

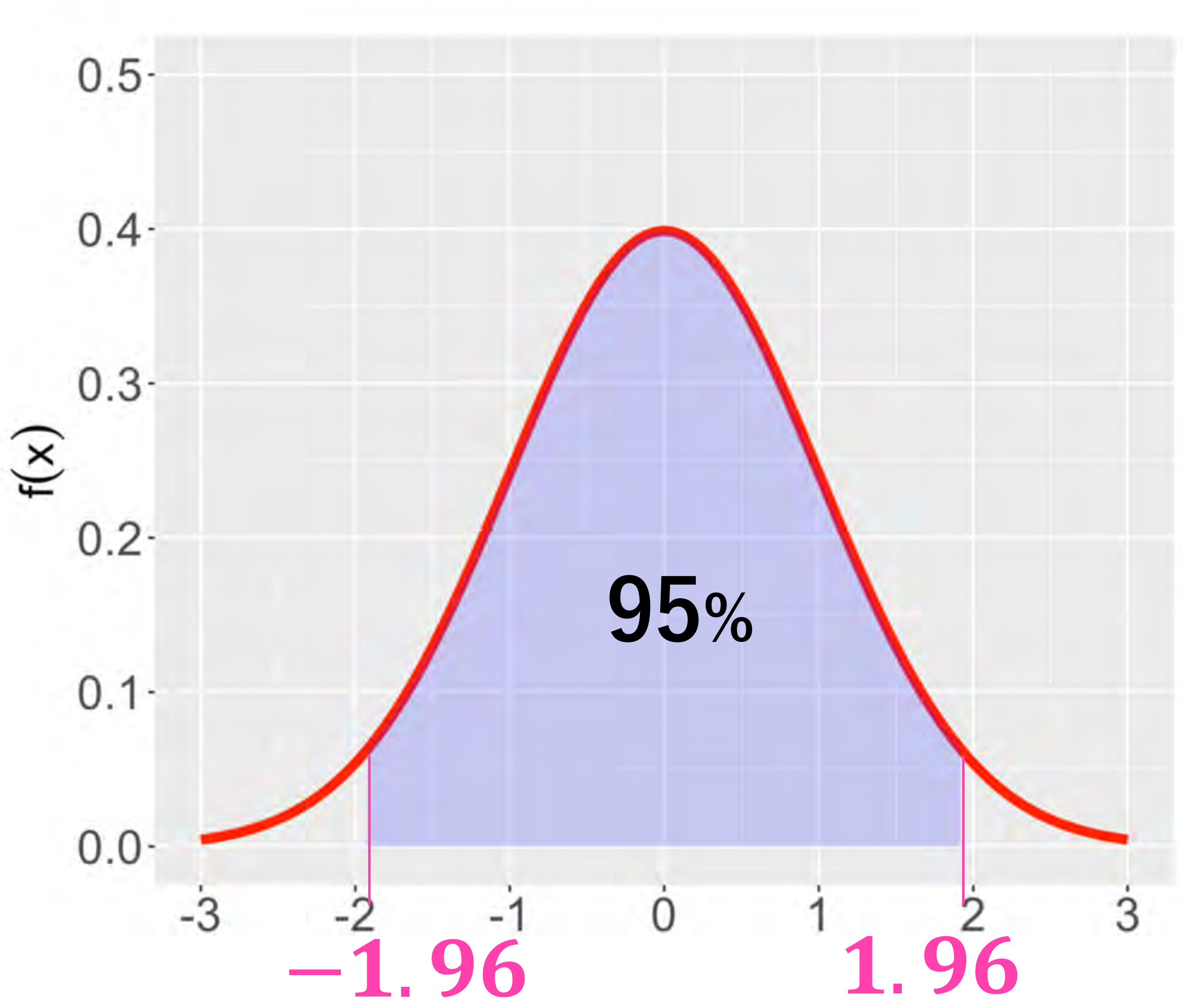
$\mu - 2\sigma$から$\mu + 2\sigma$の間に約95%くらいのデータが入っている



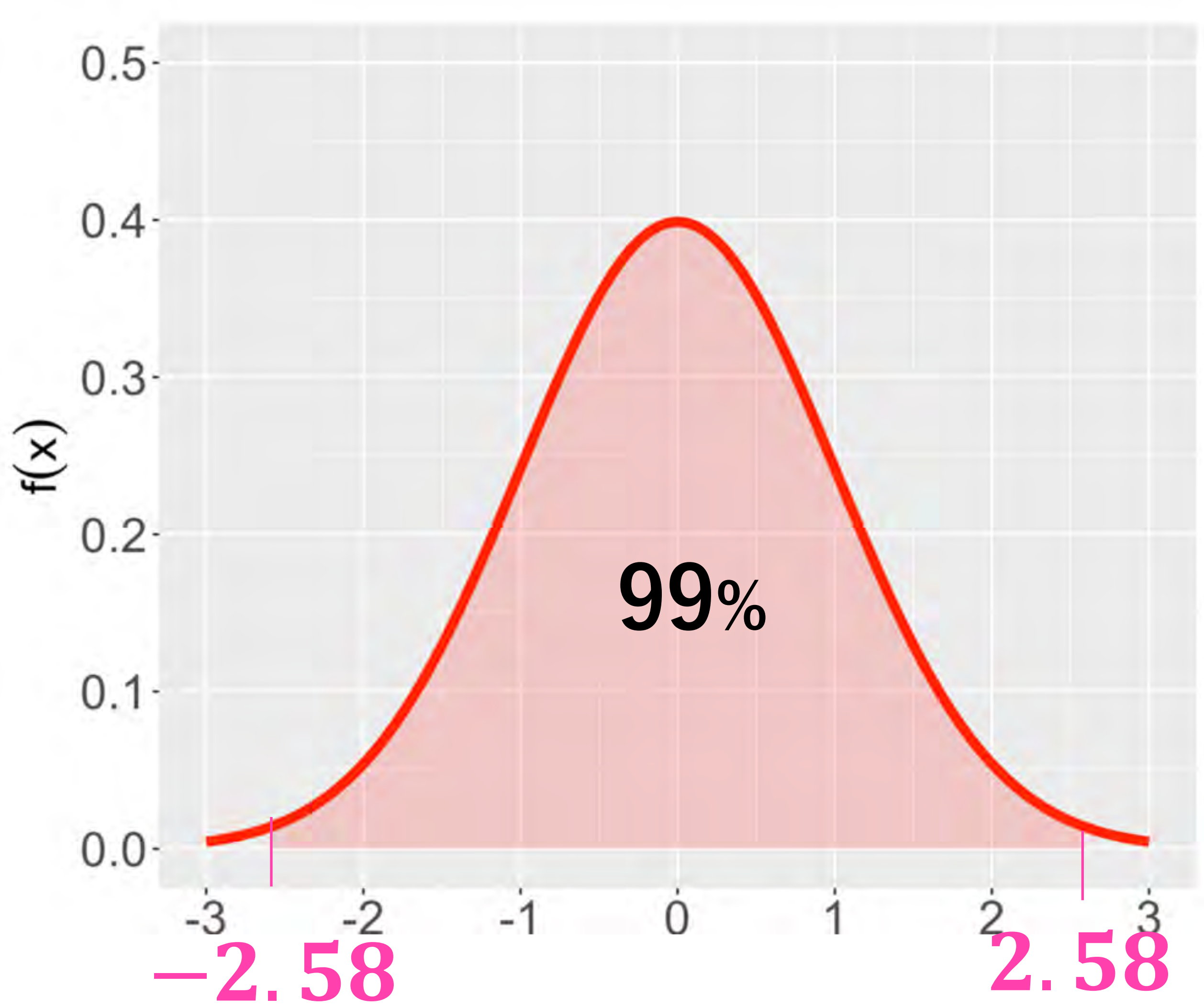
標準正規分布の有名な確率その2
（覚えても損はないので覚えてください）

- Zが標準正規分布に従うとき，以下の確率がよく知られている
  - 本講義後半はこちらの方をよく使います
  - こちらの方が，より95%と99%に正確
  - $P(-1.96 \leq Z \leq 1.96) = 0.9500$
  - $P(-2.58 \leq Z \leq 2.58) = 0.9901$

標準正規分布の有名な確率その2



標準正規分布の有名な確率その2



標準正規分布上の確率の求め方

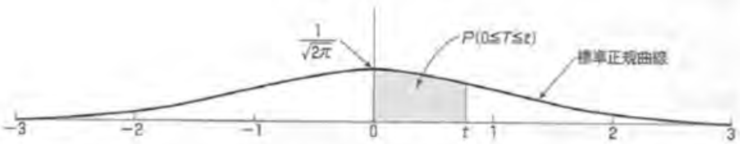
- **標準正規分布上の確率をどのように求めれば良いのか？**
  - 標準正規分布表を使う

標準正規分布表

- 統計の教科書にはほぼ必ず巻末に標準正規分布の表がある

引用: 岡本和夫，確率統計，
実教出版(2012） p.127

正規分布表



t	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998

標準正規分布表の見方

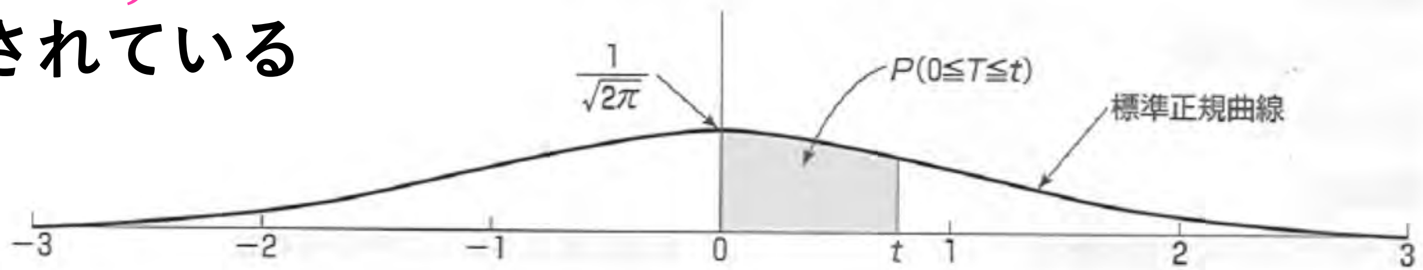
1. 教科書「確率・統計」における標準正規分布表の使い方

2. 小波先生参考書「統計学入門」における標準正規分布表の使い方

標準正規分布表の見方
(教科書「確率・統計」の説明)

$P(0 \leq Z \leq z)$の確率
が記載されている

正規分布表



t	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177

引用: 岡本和夫，確率統計，実教出版(2012） p.127

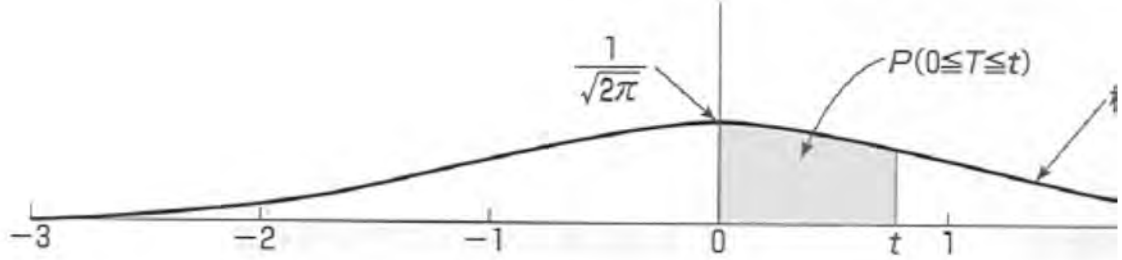
標準正規分布表の見方
（教科書の説明）

左の行の数字（0.6）に
右の列の数字（4）を
繋げた 0.64 が図中のtに該当

0.64のセルの数値（0.2389）
が $P(0 \leq Z \leq 0.64)$ を表す

$$P(0 \leq Z \leq 0.64) = 0.2389$$

正規分布表



t	0	1	2	3	4	5	6	7
				0120	0.0160	0.0199	0.0239	0.02
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.067
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.106
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.144
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.180
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.215
0.6	.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.248
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.279
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.307
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.357
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.379
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.414
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.429
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.441

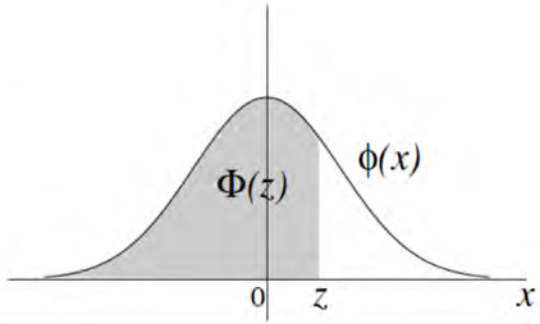
標準正規分布表の見方
（小波先生参考書の説明）

教科書「確率・統計」とは異なり，**$P(-\infty \leq Z \leq z)$の確率**が掲載されている

B.2　正規分布表

正規分布表には積分範囲が異なるものがあるので注意すること．

$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$



	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50398	.50797	.51196	.51595	.51993	.52392	.52790	.53188	.53585
0.1	.53982	.54379	.54775	.55171	.55567	.55961	.56355	.56749	.57142	.57534
0.2	.57925	.58316	.58706	.59095	.59483	.59870	.60256	.60641	.61026	.61409
0.3	.61791	.62171	.62551	.62930	.63307	.63683	.64057	.64430	.64802	.65173
0.4	.65542	.65909	.66275	.66640	.67003	.67364	.67724	.68082	.68438	.68793
0.5	.69146	.69497	.69846	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72574	.72906	.73237	.73565	.73891	.74215	.74537	.74857	.75174	.75490
0.7	.75803	.76114	.76423	.76730	.77035	.77337	.77637	.77935	.78230	.78523
0.8	.78814	.79102	.79389	.79673	.79954	.80233	.80510	.80784	.81057	.81326
0.9	.81593	.81858	.82121	.82381	.82639	.82894	.83147	.83397	.83645	.83891

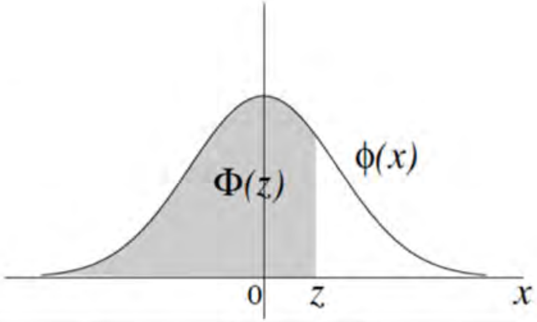
引用: 小波秀雄，統計学入門(2020)　付録 B.2

標準正規分布表の見方
（小波先生参考書の説明）

B.2 正規分布表

正規分布表には積分範囲が異なるものがあるので注意すること．

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$



$$P(Z \le 0.64) = 0.73891$$

	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50398	.50797	.51196	.51595	.51993	.52392	.52790	.53188	.53585
0.1	[illegible]	[illegible]	[illegible]	[illegible]	.55567	.55961	.56355	.56749	.57142	.57534
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.59483	.59870	.60256	.60641	.61026	.61409
0.3	[illegible]	[illegible]	[illegible]	[illegible]	.63307	.63683	.64057	.64430	.64802	.65173
0.4	.65542	.65909	.66275	.66640	.67003	.67364	.67724	.68082	.68438	.68793
0.5	.69146	.69497	.69846	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72574	.72906	.73237	.73565	.73891	.74215	.74537	.74857	.75174	.75490
0.7	.75803	.76114	.76423	.76730	.77035	.77337	.77637	.77935	.78230	.78523
0.8	.78814	.79102	.79389	.79673	.79954	.80233	.80510	.80784	.81057	.81326
0.9	.81593	.81858	.82121	.82381	.82639	.82894	.83147	.83397	.83645	.83891

引用: 小波秀雄，統計学入門(2020） 付録 B.2

標準正規分布表の見方
（小波先生参考書の説明）

$$P(Z \leq 0.64) = 0.73891$$

いま，正規分布は期待値 μ で左右対称なので,
標準正規分布上では

$$P(Z \leq 0) = 0.5$$

$$P(0 \leq Z \leq 0.64) = P(Z \leq 0.64) - P(Z \leq 0)$$

$$= 0.73891 - 0.5$$

$$= 0.23891$$

教科書「確率統計」
の結果と一致

ミニ演習

- **教科書もしくは参考書の標準正規分布表を用いて以下を計算せよ**

1. $P(Z \leq 2)$
2. $P(Z \leq z) = 0.05$ となる z の値

- 考え方
  - **正規分布の図を書いてみる（大事！）**

ミニ演習1 (教科書を使用する場合)

1. $P(Z \leq 2)$の値
   - 教科書は $P(0 \leq Z \leq 2)$しか分からないので,

$$P(Z \leq 2) = P(Z \leq 0) + P(Z \leq 2)$$
$$= 0.5 + P(0 \leq Z \leq 2)$$

$z = 2.00$ のところを確認すると, 0.4772なので,
$P(0 \leq Z \leq 2) = 0.4772$ がわかる

$$P(Z \leq 2) = 0.5 + 0.4772$$
$$= 0.9772$$

ミニ演習1 （小波先生参考書を使用する場合）

- こちらの場合は直接求まる

- 標準正規分布表でz=2.00 のところを確認すると，　$P(Z \leq 2) = 0.97724$

ミニ演習2
教科書の場合

2. $P(Z \leq z) = 0.05$ となる z の値

- 対称性より，$P(-z \leq Z) = 0.05$
- $-z = t$とおくと，$P(t \leq Z)$ =0.05 となるtを求めたい
- $P(0 \leq Z) = 0.5$なので,
  - $P(0 \leq Z) = P(0 \leq Z \leq t) + P(t \leq Z) = 0.5$
  - $P(0 \leq Z \leq t) = 0.45$
- つまり，$P(0 \leq Z \leq t) = 0.45$ となる t を
  分布表からもとめて，マイナスをつければ良い
- すると， $t = 1.64$ (1.65でもよい)
  だと分かるので $P(Z \leq z) = 0.95$
  となるようなzは z = -1.64

ミニ演習2

小波先生参考書の場合

2. $P(Z \leq z) = 0.05$ となる z の値

- $P(Z \leq t) = 0.95$となるようなtを求めて，マイナスをつければよい
- 分布表より，t = 1.64 （1.65でもよい）のとき，$P(Z \leq 1.64) = 0.95$ であることが分かる
- 従って，$P(Z \leq z) = 0.95$ となるようなzは
z = -1.64

正規分布の確率計算のポイント

- 自分で正規分布の図を書き，どの部分の確率をいま求めようとしているのかを**図示すること**が重要

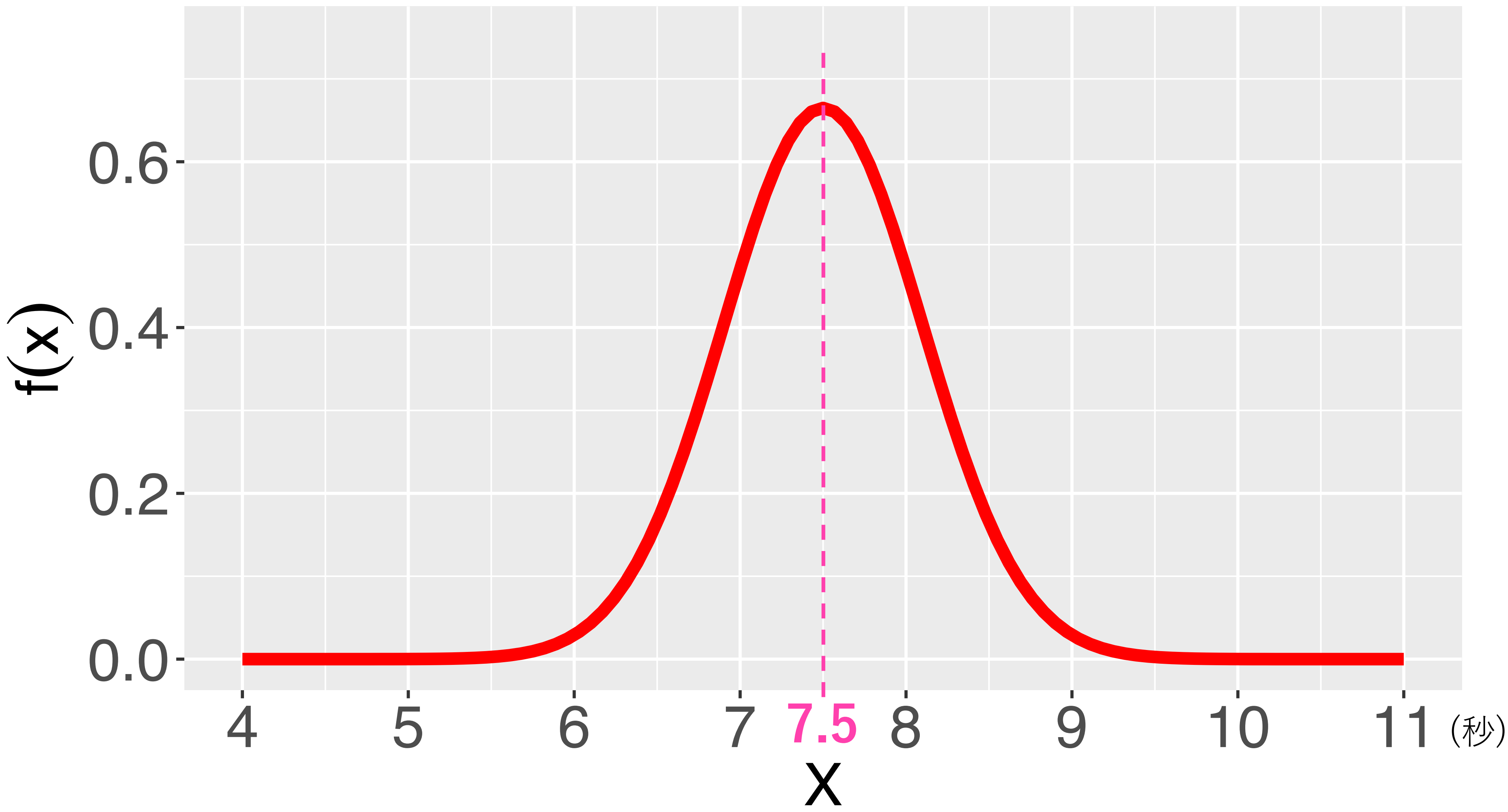
標準正規分布と正規分布の関係

例題

- **大・中都市圏に住む14歳男子の50m走のタイムは平均7.5秒，標準偏差0.6秒であった．タイムが正規分布 $N(7.5, 0.6^2)$ に従うと仮定するとき，以下を求めよ**

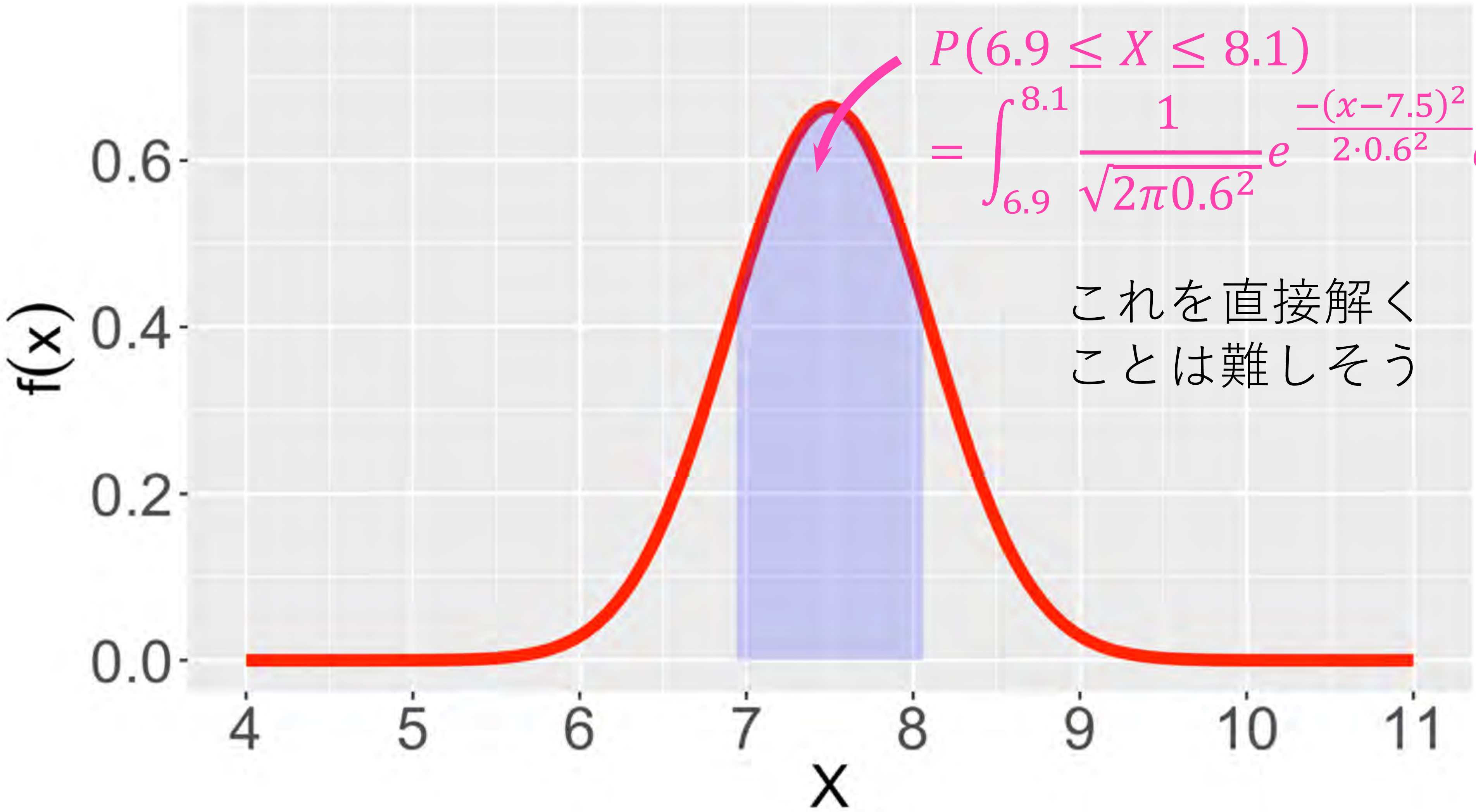
  – 6.9秒から8.1秒の間に入る生徒の割合

参考:正規分布 $N(7.5, 0.6^2)$ の形



6.9秒から8.1秒の間に入る生徒の割合

- つまり, $P(6.9 \leq X \leq 8.1)$ を求めたい



$$P(6.9 \leq X \leq 8.1) = \int_{6.9}^{8.1} \frac{1}{\sqrt{2\pi 0.6^2}} e^{\frac{-(x-7.5)^2}{2\cdot 0.6^2}} dx$$

これを直接解く
ことは難しそう

標準正規分布と正規分布の関係

- 標準正規分布 $N(0,1)$ 上での確率さえ計算できれば，任意の正規分布$N(\mu,\sigma^2)$上の確率が計算できることが知られている

正規分布の変数変換

- **正規分布について，以下の性質が知られている**
  - 証明はこの講義の範囲を超えるので略
    興味がある人は東大出版会の参考書5-6章参照

確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとき
新しい確率変数 $X' = aX + b$ も
正規分布 $N(a\mu + b, a^2\sigma^2)$ に従う

すべての確率分布（つまり，正規分布以外の分布）に
ついても成り立つ性質というわけではないので注意すること

正規分布の標準化

- 前ページの性質より，確率変数 X が $N(\mu, \sigma^2)$に従うとき，以下の変換を行った確率変数 Z は $N(0,1)$ にしたがう

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

- これを**標準化**という
  - 参考: 記述統計の標準化と考え方は同じ

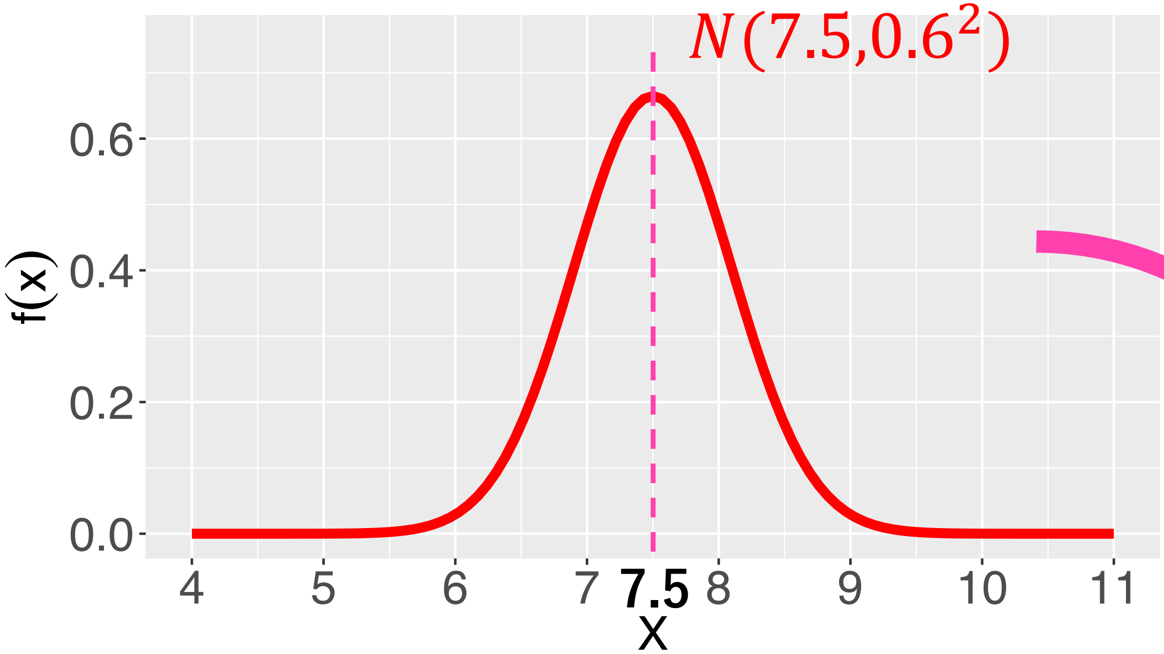
正規分布の標準化（具体的に）

- $N(7.5, 0.6^2)$に従うとき，以下の変換を行った確率変数 Z は $N(0,1)$ にしたがう

$$Z = \frac{X - 7.5}{0.6}$$

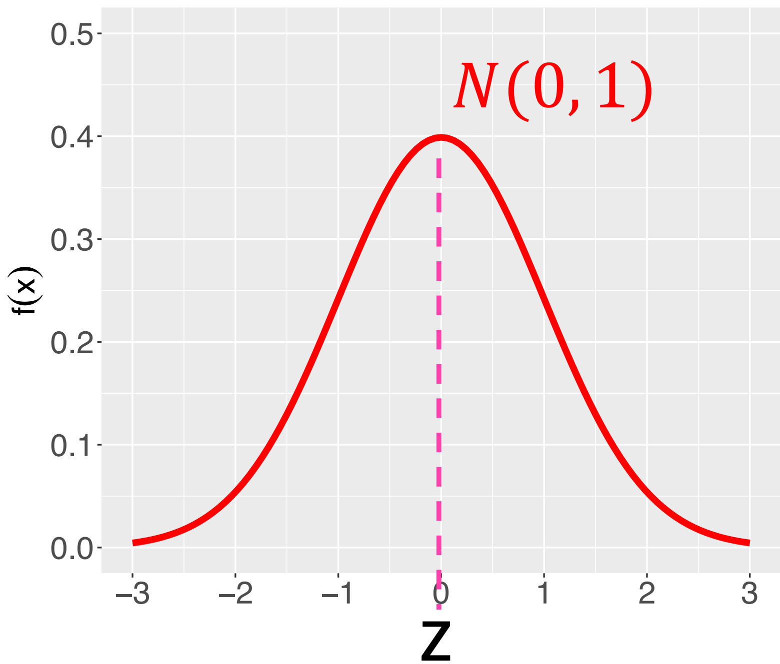
- x=7.5秒のとき，z=0
- x=8.1秒 のとき，z=1
- x=8.7秒のとき，z=2

正規分布の標準化のイメージ



標準化

$$Z = \frac{X - 7.5}{0.6}$$



標準化すると何が嬉しいか？

- 標準正規分布 $N(0,1)$ の確率さえ分かれば，一般的な正規分布$N(\mu,\sigma^2)$の確率もわかる

$$Z = \frac{X-\mu}{\sigma} \quad \text{より} \quad X = \sigma Z + \mu$$

- これらの式に従って，標準正規分布上での確率や正規分布上での確率を変換してやればよい

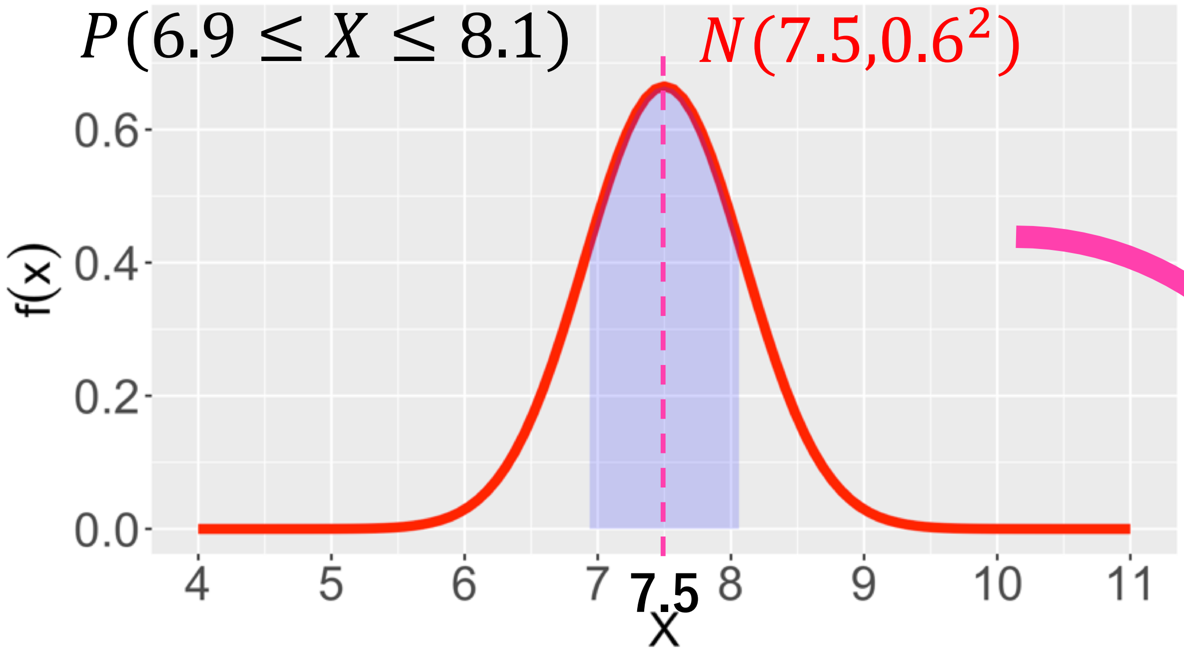
標準正規分布と正規分布の対応（重要）

- 具体的には，Xが$N(\mu, \sigma^2)$に従い，Zが$N(0, 1)$に従うとき，以下の2つが成り立つ
- 大変重要なので必ず理解し覚えてください

$$P(a \leq X \leq b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$$

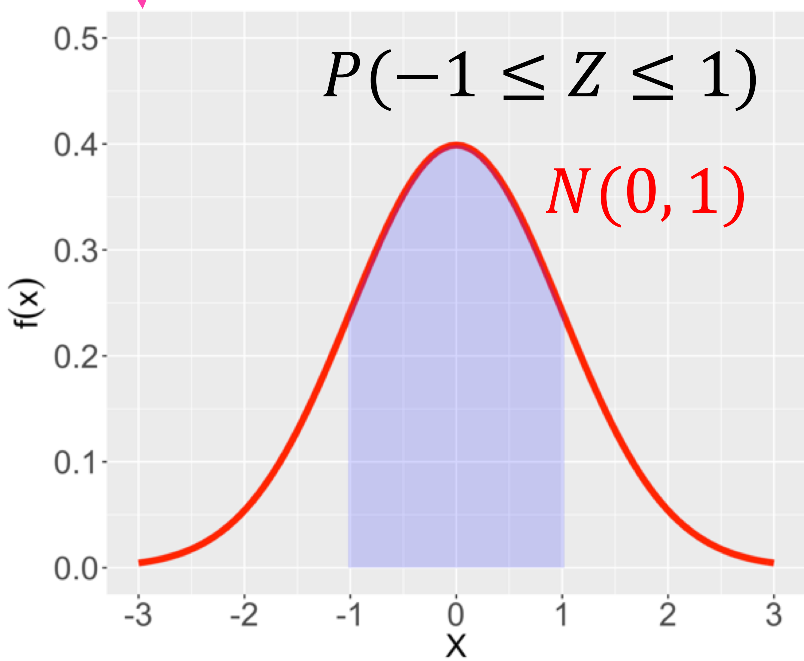
$$P(a \leq Z \leq b) = P(\sigma a + \mu \leq X \leq \sigma b + \mu)$$

例題をこの関係式を使って解くと



$$\frac{6.9 - 7.5}{0.6} = -1$$

$$\frac{8.1 - 7.5}{0.6} = 1$$



つまり，$P(6.9 \leq X \leq 8.1)$ の代わりに $P(-1 \leq Z \leq 1)$を計算すれば良い.

標準正規分布表から計算すると，**約68%**であることが分かる

まとめ

- 標準正規分布上の確率から正規分布上の確率を計算することができる.
- X が $N(\mu, \sigma^2)$ に従い, Z が $N(0,1)$ に従うとき, 以下の2つが成り立つ

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

$$P(a \leq Z \leq b) = P(\sigma a + \mu \leq X \leq \sigma b + \mu)$$

正規分布の確率計算の例題

正規分布の確率計算の方法

1. **標準正規分布 $N(0,1)$ における確率をもとに計算**
   - 標準正規分布表を用いる

2. **計算機（ExcelやRなど）を用いて $N(\mu,\sigma^2)$ における各種確率を直接計算**

- **両者ともできるようになることが大事**
  - 前者は理論の理解に必要

例題その1

- **大・中都市圏に住む14歳男子の50m走のタイムは平均7.5秒，標準偏差0.6秒であった．タイムが正規分布 $N(7.5, 0.6^2)$ に従うと仮定するとき，以下を計算してみよう**

1. タイムが8.7秒以下の生徒の割合
2. 上位5%のタイムは何秒以下か？

これを標準正規分布表を用いて計算せよ

例題その1-1

1. タイムが 8.7秒以下の生徒の割合

求めたい確率は $P(X \leq 8.7)$

- $Z = \frac{8.7-7.5}{0.6} = 2$より，
- $P(X \leq 8.7) = P(Z \leq 2)$
  - $P(Z \leq 2)$ を分布表を用いて計算すればよい

$P(Z \leq 2)$=0.9772 だと分かるので，

$P(X \leq 8.7)$=0.9772　約**97.7%**

例題その1-2.

2. 上位5%のタイムは何秒以下か？

- 求めたいものは $P(X \leq x) = 0.05$ となるようなx
- $P(Z \leq z) = 0.05$となるような z を分布表から計算すると $z = -1.64$ (-1.65でもOK)
- x=$\sigma z + \mu$より$x = 0.6z + 7.5$してやればよい
- x=$0.6 \cdot (-1.64) + 7.5 = 6.516$

従って，上位5%のタイムは x ＝ 約**6.52** 秒

例題その2

- **大・中都市圏に住む14歳男子の50m走のタイムは平均7.5秒，標準偏差0.6秒であった．タイムが正規分布** $N(7.5, 0.6^2)$ **に従うと仮定するとき，以下を計算してみよう**

1. 6.9秒から8.1秒の間に入る生徒の割合
2. 中間95%層（上位2.5%から上位97.5%までの生徒）のタイムは何秒以上何秒以下か？

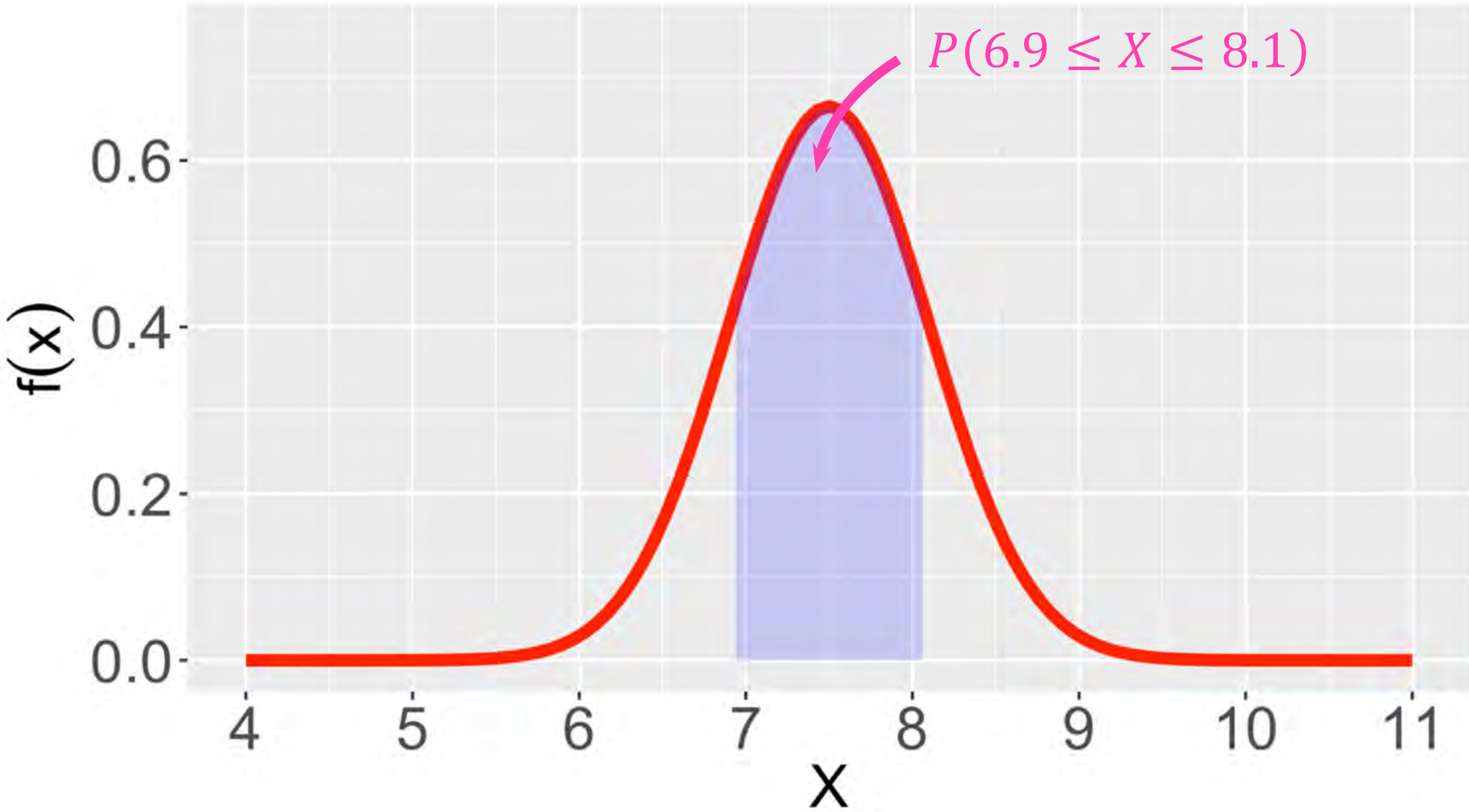
正規分布の確率計算の方法

1. 標準正規分布 $N(0,1)$ における確率をもとに計算
   - 標準正規分布表を用いる

2. 計算機（ExcelやRなど）を用いて $N(7.5, 0.6^2)$ における各種確率を計算
   - 今度はこちら側で解いてみる（もちろん，標準正規分布表からも求められます）

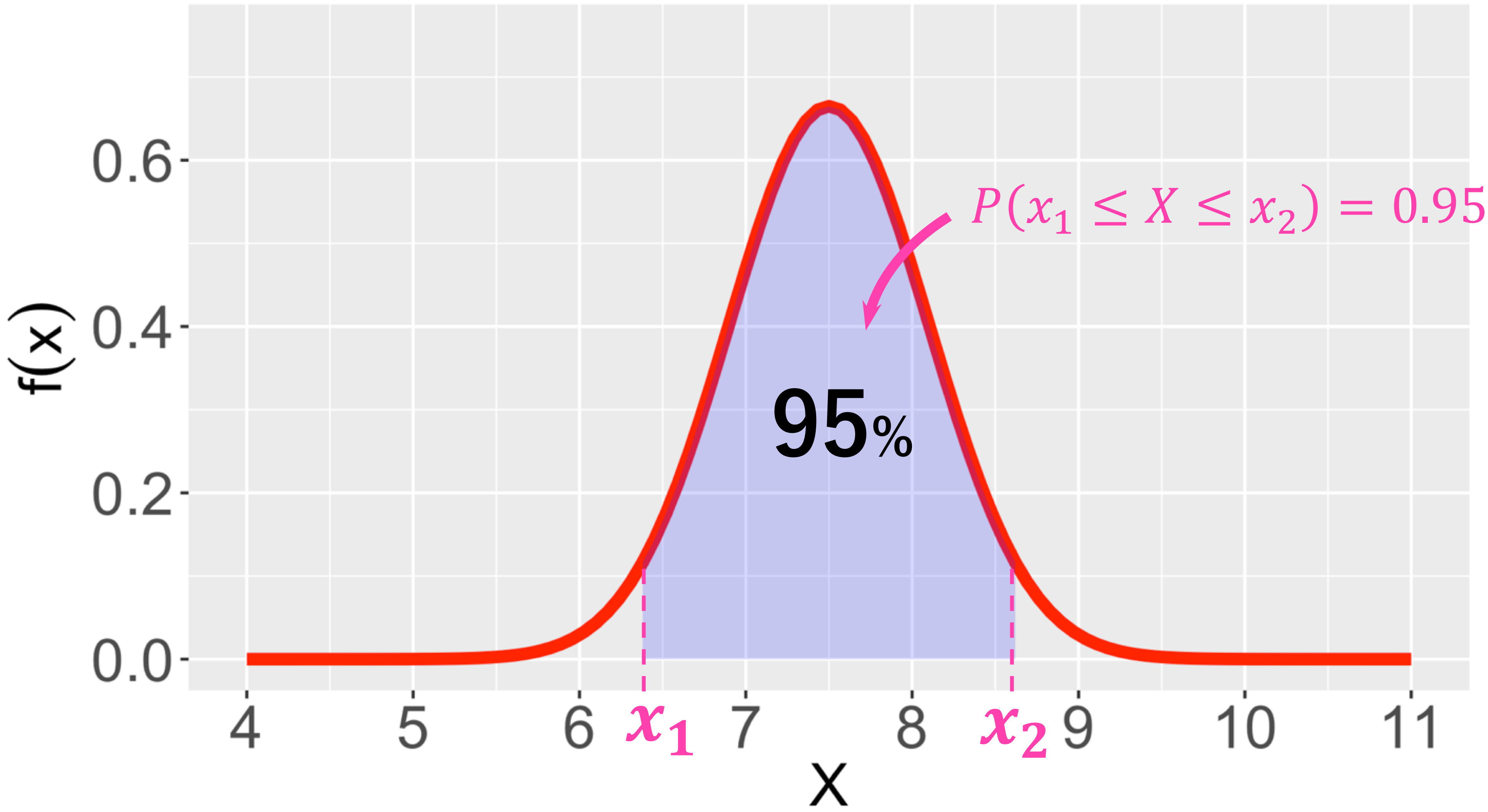
6.9秒から8.1秒の間に入る生徒の割合

- **つまり,** $P(6.9 \leq X \leq 8.1)$ **を求めたい**



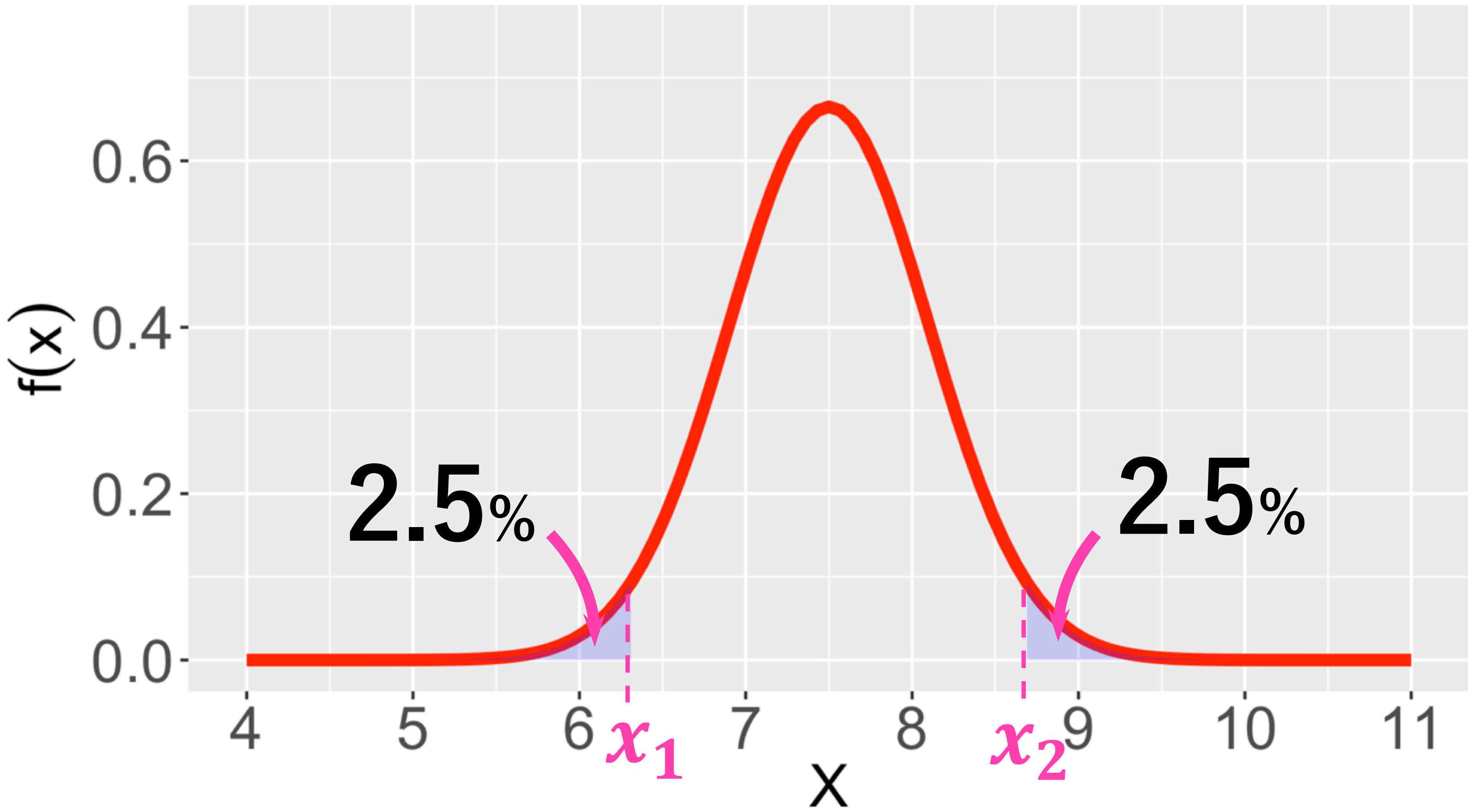
中間95%層のタイムとは？

- つまり, $P(x_1 \leq X \leq x_2) = \mathbf{0.95}$となる x_1, x_2 を求めたい



中間95%層のタイムとは？

- 言い換えれば，以下を満たす$\boldsymbol{x_1}, \boldsymbol{x_2}$を求める
  - $P(X \leq \boldsymbol{x_1}) = \mathbf{0.025}$ （タイム上位2.5%の値）
  - $P(X \leq \boldsymbol{x_2}) = \mathbf{0.975}$ （タイム下位2.5%の値）



Excelを用いた正規分布の計算その1

=NORM.DIST(x, μ, σ, TRUE)

- **上記関数は X が$N(\mu, \sigma^2)$ に従うときの $P(X \leq x)$を計算してくれる関数**
  - 分散ではなく標準偏差を入力する点に注意
  - なお，TRUEの代わりにFALSEと入力すると，$f(x)$ の値が計算される

=NORM.DIST(6.9, 7.5, 0.6, TRUE)

- **上記を正しく入力すれば，** 0.1586.. **と表示されるはず**
  - $P(X \leq 6.9) = 0.1586$ という意味

演習: Excelを用いて 例題その2をといてみましょう

- **6.9秒から8.1秒の間に入る生徒の割合**
  - $P(6.9 \leq X \leq 8.1)$

- **いろいろ方法はあるが，簡単なのは**
  - $P(6.9 \leq X \leq 8.1) = P(X \leq 8.1) - P(X \leq 6.9)$
    なので，この2つの確率を求めて引けばよい
  - 注: 連続型の確率変数については,
    等号の有無は気にしなくて良い
    （特定の値をとる確率は限りなく0に近いので）

Excelを用いて計算してみよう

- 正しく計算できれば， 0.6826… と表示されるはず

- つまり， 6.9秒から8.1秒の間に入る生徒の割合は約68%であることがわかる

Excelを用いた正規分布の計算その2

=NORM.INV(p, μ, σ)

- **上記関数はXが$N(\mu, \sigma^2)$に従うとき$P(X \leq x) = p$となるxを計算してくれる関数**
  - 混乱しやすいので注意

=NORM.INV(0.025, 7.5, 0.6)

- **上記を正しく入力すれば，6.324.. と表示されるはず**
  - $P(X \leq 6.324..) = 0.025$ という意味

演習: Excelを用いて 例題その2をといてみましょう

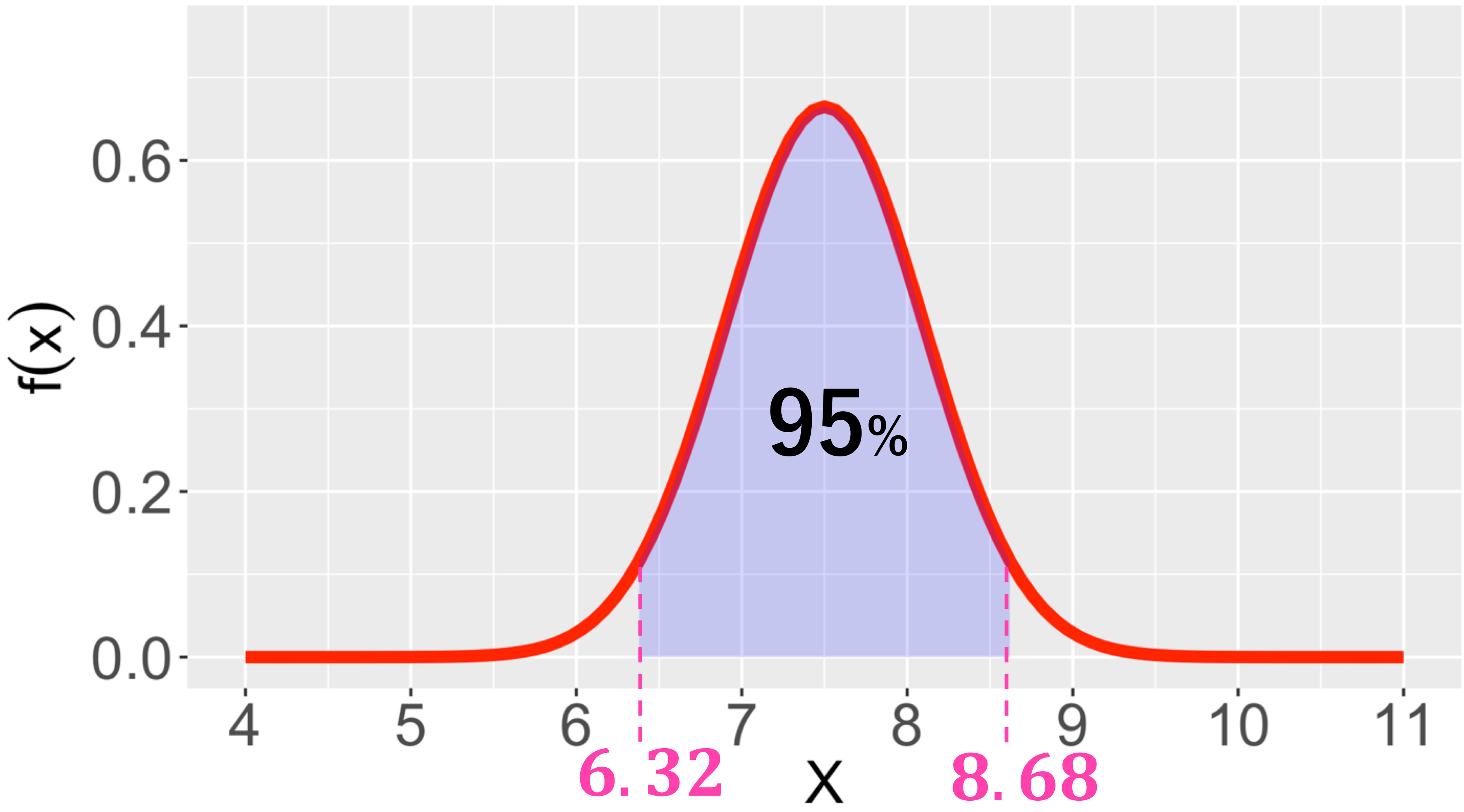
- **中間95%層のタイム，すなわち**
  - $P(X \leq x_1) = 0.025$
  - $P(X \leq x_2) = 0.975$
  - を満たす$\boldsymbol{x_1}, \boldsymbol{x_2}$をExcelを使って求めればよい

演習: Excelを用いて例題をといてみましょう

- 両者を正しく求められることができれば,
  - x_1=6.324..
  - x_2=8.675..

- つまり，中間95%層のタイムは
  6.32秒以上 8.68秒以下 の
  区間に入っている

中間95%層のタイムを図示するとこんな感じ



まとめ

- **正規分布の確率計算の2つの方法**
  - 標準正規分布表を用いる
  - ExcelやGoogleスプレッドシートで直接計算
  - **どちらも重要**なので，どちらもできるようになってください

- **もし筆記試験をする場合は，Excelを使用するわけはいかないので，標準正規分布表の重要な確率を明示します**

参考: 偏差値

- 偏差値Tは平均50，標準偏差10となるように変数変換したものであった．Tが$N(50, 10^2)$に従うと仮定したとき，$P(30 \leq T \leq 70)$ の確率を求めよ

- $Z = \frac{T-50}{10}$ なので，この確率を標準正規分布で表せば，$P(-2 \leq Z \leq 2) = 0.95$
  - つまり，偏差値30から70に約95%の生徒が入る

参考: 二項分布と正規分布の関係

- **二項分布**$B(n,p)$**は**n**が十分多いとき，正規分布**$N(np, np(1-p))$**に近づくことが知られている**
  - 参考: $B(n,p)$の期待値: np，分散: $np(1-p)$

- **つまり，n が十分大きければ二項分布を計算する代わりに正規分布で近似してしまってよい**

参考: 正規分布の関数の意味

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

- **いったいこの式はどこから出てきた？**
  - 以下のページが分かりやすい（山本の感想）
- **To-kei.net 「正規分布の密度関数を意味的に理解する」**
  - https://to-kei.net/distribution/normal-distribution/density-function-derivation/

この資料のまとめ

- **正規分布**
  - 標準正規分布との対応関係が重要

- **正規分布の確率計算**
  - Excel，標準正規分布表どちらもできるように
  - ここさえできれば，実は本講義後半の
    区間推定と仮説検定もほぼ理解できている