Lecture 9

Part 2 Probability and Distributions

Expected Value of a Uniform Distribution

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{a}^{b} x \times \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2} \right] |_a^b \\
&= \frac{b^2 - a^2}{2(b-a)} \\
&= \frac{(b+a)(b-a)}{2(b-a)} \\
&= \frac{a+b}{2}
\end{aligned}
$$

Variance

- To calculate variance, we use the shortcut
  - Mean $E(X) = \frac{a+b}{2}$
  - $E(X^2)$:

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{a^2 + ab + b^2}{3}$$

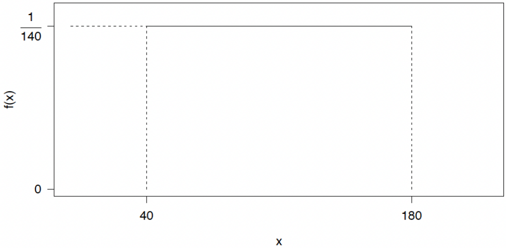
  - Apply the shortcut: $V(X) = E(X^2) - [E(X)]^2$
- We have

$$V(X) = \frac{(b-a)^2}{12}$$

Example of Uniform Distribution: Patient Waiting Times

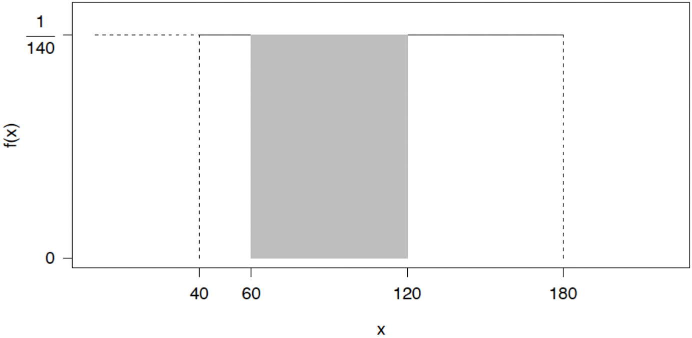
- The length of time patients wait to see a doctor is uniformly distributed between 40 mins and 3 hrs.
- Let X be the waiting time (in mins).

$$f(x) = \frac{1}{140}, \quad 40 \leq x \leq 180$$



Find the Probability of Waiting...

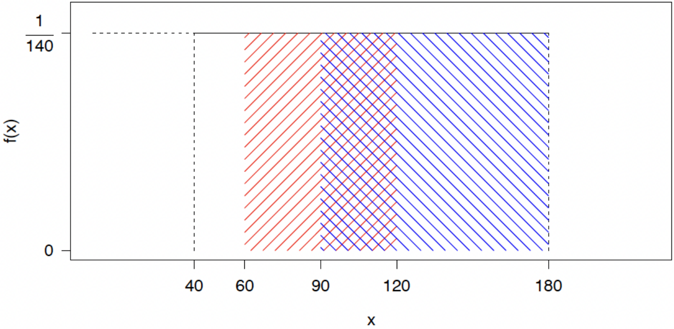
- Between 1 and 2 hrs.



$$P(60 < X < 120) = (120 - 60) \times \frac{1}{140} = \frac{3}{7}$$

Find the Probability of Waiting...

- Between 1 and 2 hrs, given you waited over 90 min.



$$P(\{60 < X < 120\}|\{X > 90\}) = \frac{P(90 < X < 120)}{P(X > 90)} = \frac{\frac{3}{14}}{\frac{9}{14}} = \frac{1}{3}$$

Calculate the...

- Mean and standard deviation of the waiting times.

$$E(X) = \frac{a+b}{2} = \frac{40+180}{2} = 110 \text{ min}$$

$$V(X) = \frac{(b-a)^2}{12} = \frac{(180-40)^2}{12} = 1633.3333 \text{ min}^2$$

$$SD(X) = \sqrt{V(X)} = \sqrt{1633.333} = 40.4145 \text{ min}$$

Parachutist

- A parachutist lands at a random point along the straight line between two markers, A and B.
- Suppose the point at which they land is uniformly distributed between A and B.
- Find the probability that their distance to A is more than three times their distance to B.

Probability Density Function

- Let X be the point at which the parachutist lands.
- Since we know $X \sim U(A, B)$, we know the PDF is

$$f(x) = \frac{1}{B - A}, \quad A \leq x \leq B$$

- We also know that X lies between A and B.

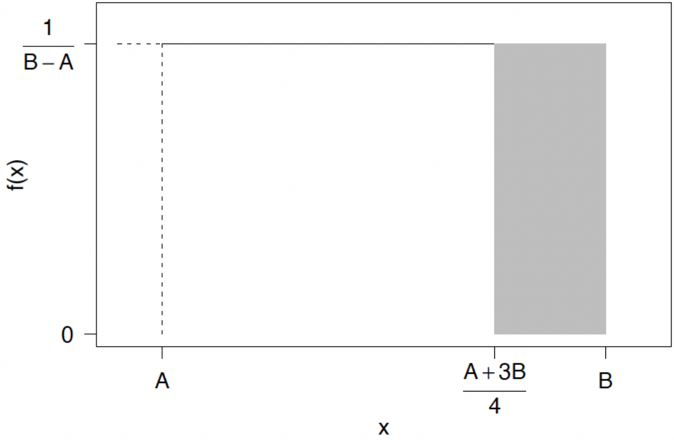
Distance to Markers

- Distance from X to A is $X - A$.
- Distance from X to B is $B - X$.
- We want:

$$\begin{aligned} P(X - A > 3(B - X)) &= P(4X > A + 3B) \\ &= P\left(X > \frac{A + 3B}{4}\right) \end{aligned}$$

- Let's sketch the PDF and shade the region of interest.

Region of Interest



$$P\left(X > \frac{A+3B}{4}\right) = (B - (\frac{A+3B}{4})) \times \frac{1}{B-A} = \frac{1}{4}$$

Normal Distribution

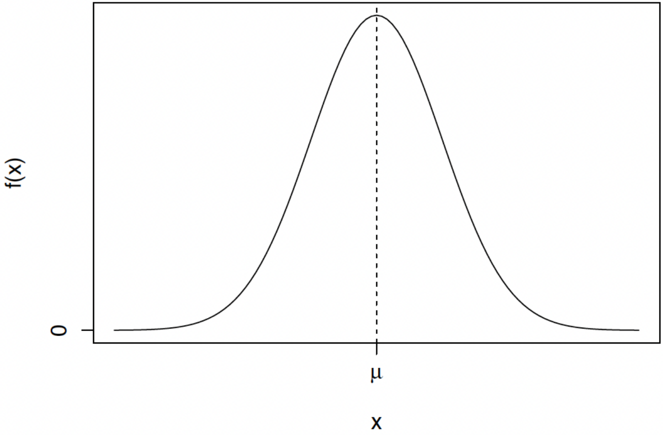
- A continuous random variable X is said to have a **normal distribution** with mean μ and variance σ^2 if its PDF is given by the following function:

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

- We use the notation $X \sim N(\mu, \sigma^2)$.
- There are two *parameters* that define the normal distribution, namely, μ and σ^2.

Probability Density Function

- Bell-shaped, symmetric about μ, reaches highest point at $x = \mu$, tends to zero as $x \to \pm\infty$.



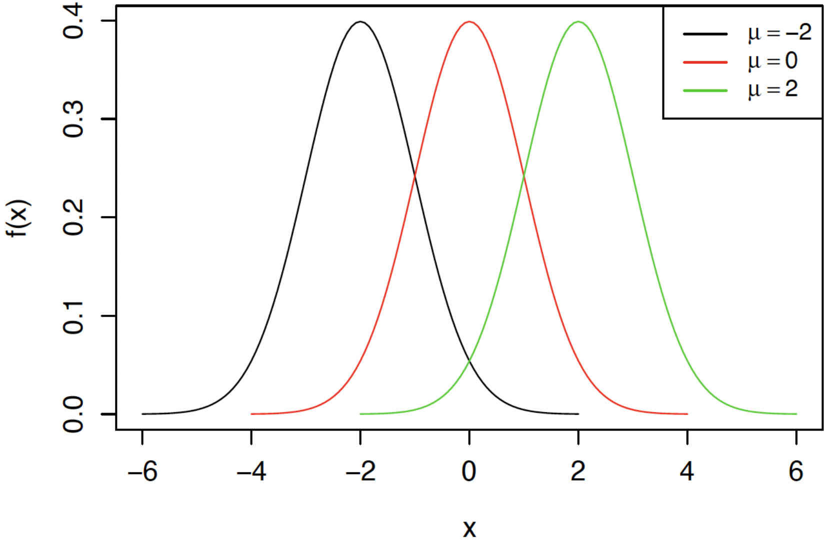
Properties

- Total area under the curve equals 1.
- If we apply the formulae for $E(X)$ and $V(X)$ and use some calculus, we can show that:

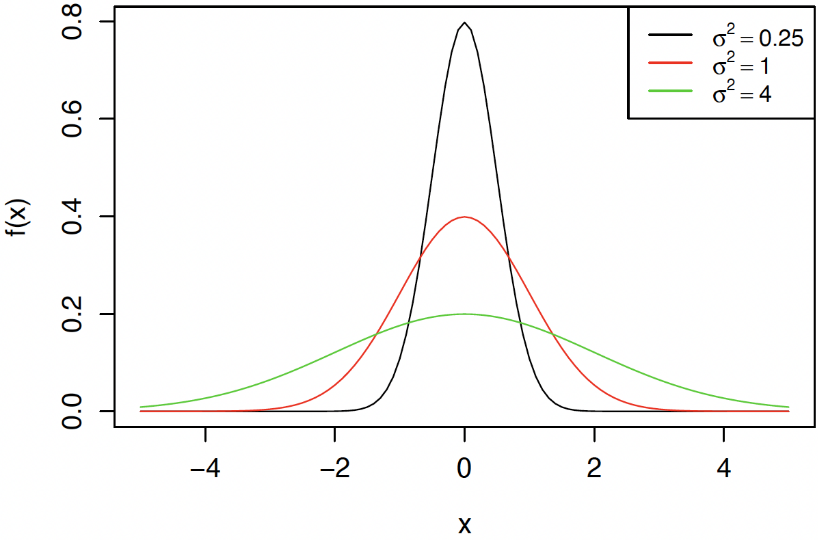
$$E(X) = \mu$$
$$V(X) = \sigma^2$$

- Changing μ (different means) will shift the PDF curve left and right along the x-axis.
- Changing σ^2 (different variances) will make the PDF curve become more peaked or more flattened.

Different Means, Same Variances

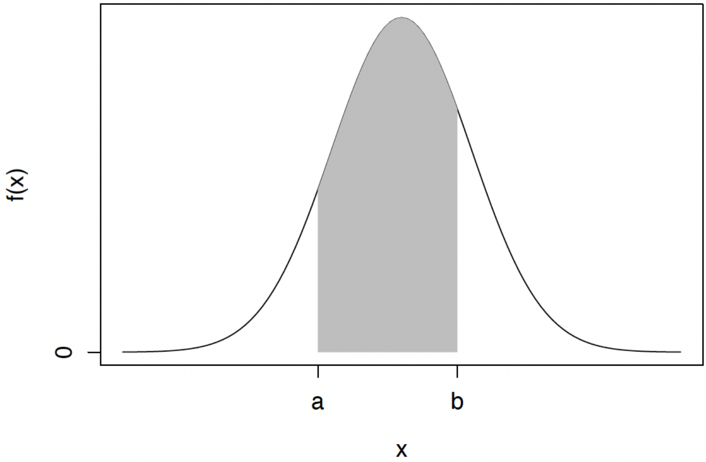


Different Means, Same Variances



Calculating Probabilities

- How can we find $P(a < X < b)$, where a and b can be either positive, negative or $\pm\infty$?



Calculating Probabilities

- We need to find the area under the PDF curve between a and b...
- That is, we need to perform the following integration:

$$
\begin{aligned}
P(a < X < b) &= \int_a^b f(x)dx \\
&= \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}
$$

- Not easy to do!

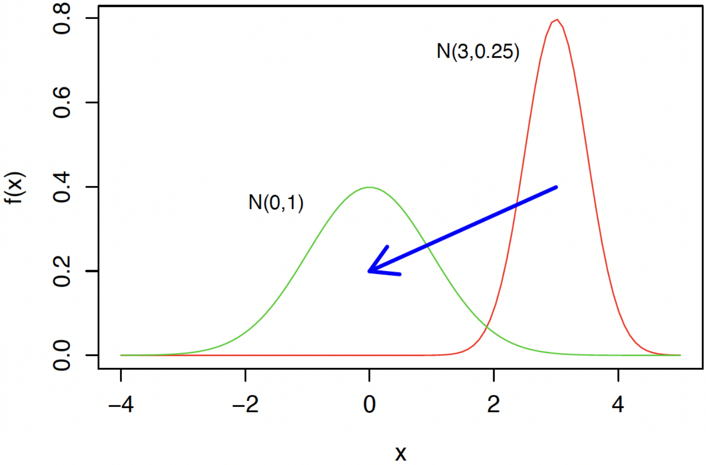
Statistical Tables

- Statistical tables are available that list $P(X < a)$ for various values of a for a normal distribution.
- However, there are infinite number of normal distributions out there!
- Specifically, there is a different one for every different pair of values of μ and σ^2.
- How many tables do we need???

Statistical Tables

- It turns out we only need one statistical table for one specific normal distribution called the **standard normal distribution**.
- The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma^2 = 1$, i.e., $N(0, 1)$.
- Any other normal distribution can be *standardized* to become a $N(0, 1)$ distribution.

Standardizing a Normal Distribution



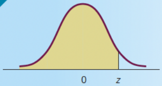
Standardizing a Normal Distribution

- If $X \sim N(\mu, \sigma^2)$, then the linear transformation

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

  standardizes X to be a $N(0, 1)$ random variable.
- The random variable Z can be interpreted as the number of standard deviations X is away from μ.
- The table of probabilities for a $Z \sim N(0, 1)$ distribution is called a z-table and it lists probabilities of the form $P(Z < z)$ for various values of z.

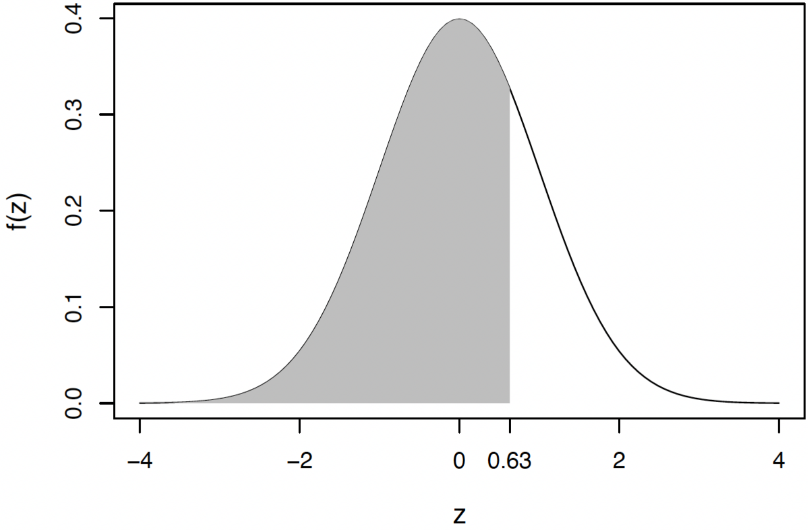
z-Table



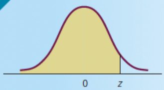
$P(-\infty < Z < z)$.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Find $P(Z < 0.63)$



Find $P(Z < 0.63)$

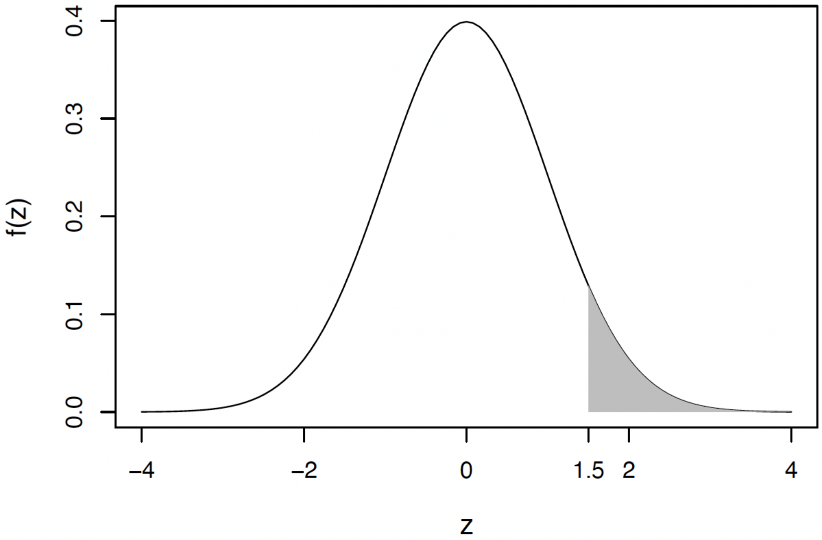


$P(-\infty < Z < z)$.

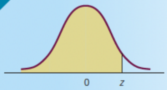
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

$$P(Z < 0.63) = 0.7357$$

Find $P(Z > 1.5)$



Find $P(Z > 1.5)$

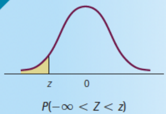


$P(-\infty < Z < z)$.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

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

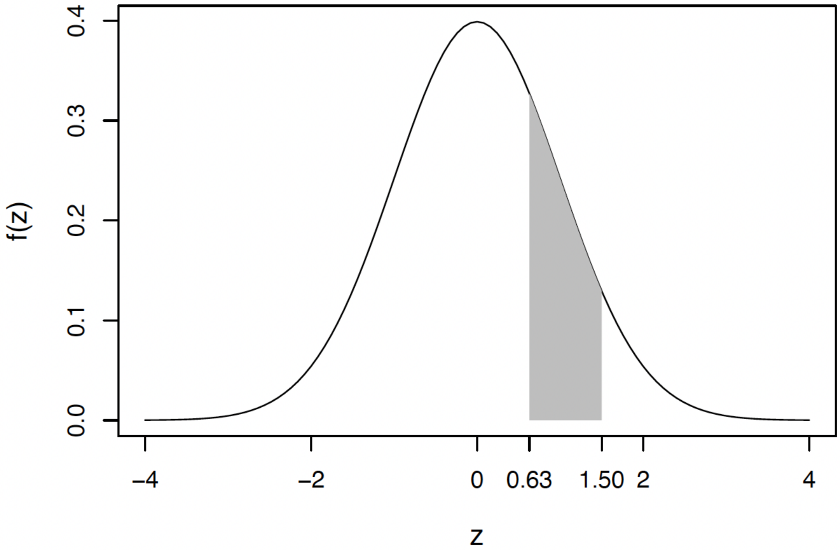
Find $P(Z > 1.5)$



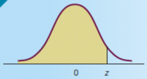
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

$$P(Z > 1.5) = P(Z < -1.5) = 0.0668$$

Find $P(0.63 < Z < 1.5)$



Find $P(0.63 < Z < 1.5)$



$P(-\infty < Z < z)$.

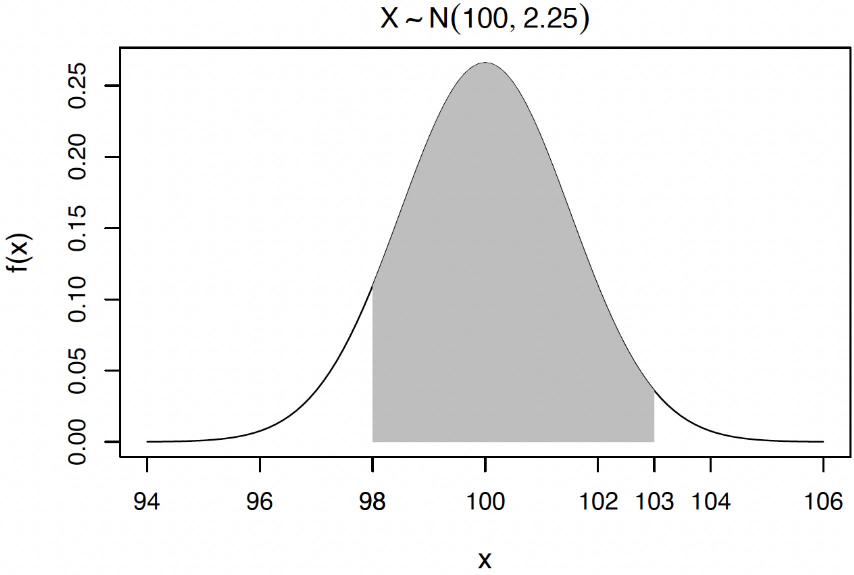
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

$$P(0.63 < Z < 1.5) = P(Z < 1.5) - P(Z < 0.63)$$
$$= 0.9332 = 0.7357 = 0.1975$$

Example 1

- The length of metallic strips produced by a machine are normally distributed with a mean of 100cm and a variance of 2.25cm^2.
- Only strips that are between 98cm and 103cm are acceptable.
- What proportion of strips are acceptable?
- Let X be the length of a metallic strip in cm.
- Then $X \sim N(\mu = 100, \sigma^2 = 2.25)$.

Example 1



X ~ N(100, 2.25)
f(x)
0.00
0.05
0.10
0.15
0.20
0.25
94
96
98
100
102
103
104
106
x

Example 1

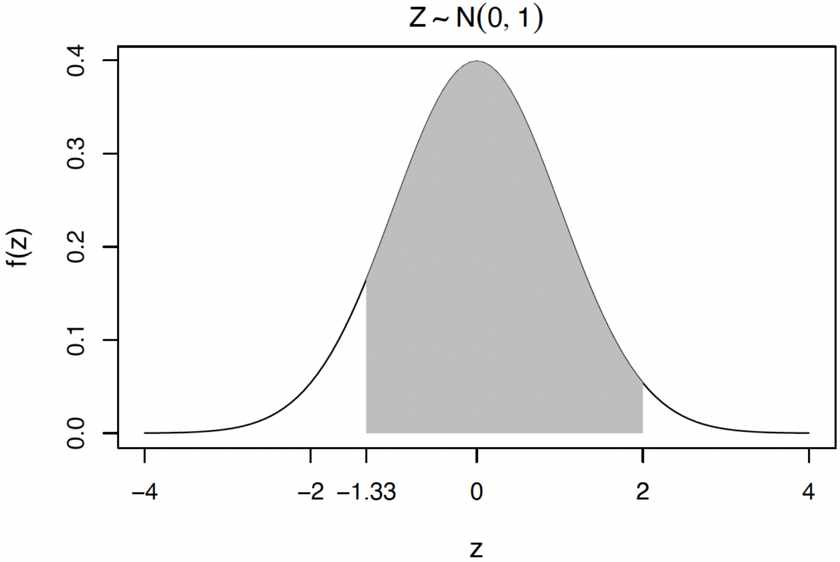
- Firstly, we must standardize the X-distribution so that we can find the values in the Z-distribution that correspond to $X = 98$ and $X = 103$.
- Standardize $X = 98$:

$$Z = \frac{X - \mu}{\sigma} = \frac{98 - 100}{\sqrt{2.25}} = -1.33$$

- Standardize $X = 103$:

$$Z = \frac{X - \mu}{\sigma} = \frac{103 - 100}{\sqrt{2.25}} = 2$$

Example 1: After Standardization



Example 1

- The area of the shaded region is equal to:

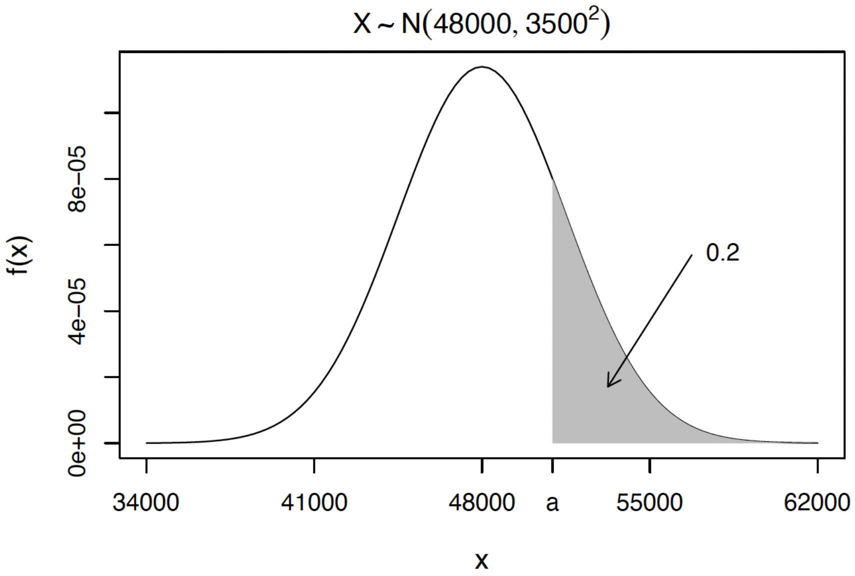
$$\begin{aligned} P(-1.33 < Z < 2) &= P(Z < 2) - P(Z < -1.33) \\ &= 0.9772 - 0.0918 \\ &= 0.8854 \end{aligned}$$

- So 88.54% of metallic strips produced are acceptable.

Example 2

- Salaries of workers in a factory are normally distributed with mean \$48,000 and standard deviation \$3,500.
- What is the minimum salary of the top 20% of workers?
- Let X be the salary of a worker.
- Then $X \sim N(\mu = 48000, \sigma^2 = 3500^2)$.
- We want to find a such that $P(X > a) = 0.2$.

Example 2

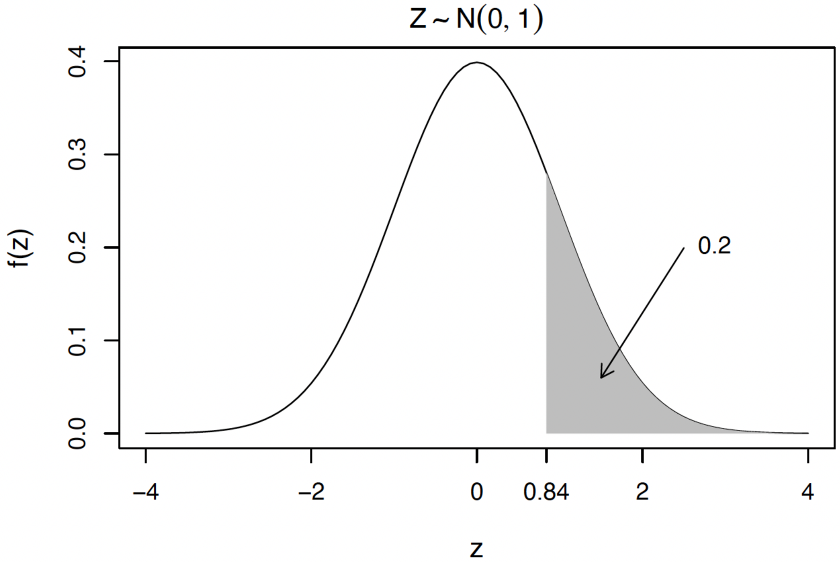


$X \sim N(48000, 3500^2)$
f(x)
0e+00
4e−05
8e−05
0.2
34000
41000
48000
a
55000
62000
x

Example 2

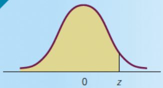
- This example is slightly different from the previous.
- This time, we essentially have to work *backwards*.
- That is, we need to:
  - Start with the probability (i.e., 0.2).
  - Find the Z-value that cuts off this probability in the upper tail of the standard normal distribution.
  - De-standardize this Z-value to arrive at the corresponding X-value (i.e., a), in dollars.

Example 2: After Standardization



Example 2

- $P(Z > z) = 0.2$ is the same as $P(Z < z) = 0.8$.



$P(-\infty < Z < z)$.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Example 2

- We now have to de-standardize $Z = 0.84$ to find the corresponding value in the X-distribution, which is $N(48000, 3500^2)$.

$$Z = \frac{X - \mu}{\sigma}$$
$$\Rightarrow 0.84 = \frac{X - 48000}{3500}$$

- Solving this for X we get $X = 50940$.
- Therefore, the minimum salary for the top 20% of workers is $50,940.

Example 3

- A soft drink machine can be regulated so that it pours an average of μml per cup.
- If the amount it pours is normally distributed with standard deviation 0.3ml, give the setting for μ so that 8ml cups will overflow only 1% of the time.
- Let X be the amount that the machine pours.
- Then $X \sim N(\mu, 0.3^2)$.
- We want to find μ which makes 8ml cups overflow only 1% of the time.

Example 3

- So we want:
$$P(X > 8) = 0.01$$
- From this probability, we can work backwards to find the corresponding Z-value.
- That is, find z such that $P(Z > z) = 0.01$, or in other words, $P(Z < z) = 0.99$.
- From the z-tables, we get $z = 2.33$.

Example 3

- Therefore, if we standardize $X = 8$, we should get $Z = 2.33$:

$$Z = \frac{X - \mu}{\sigma}$$
$$\Rightarrow 2.33 = \frac{8 - \mu}{0.3}$$

- Solving this for μ we get $\mu = 7.301$.
- So the setting for μ which results in 8ml cups overflowing only 1% of the time is $\mu = 7.301$.