STAT 234 Lecture 8
Normal Distributions
Section 4.3

Yibi Huang
Department of Statistics
University of Chicago

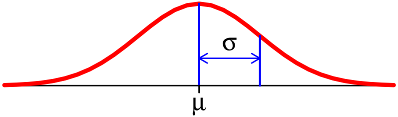
Normal Distributions

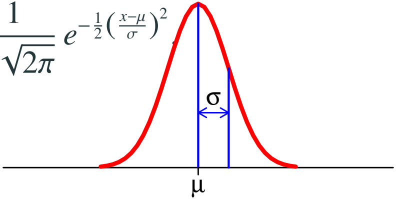
A random variable X is said to have a normal distribution (aka. Gaussian distributions) with a *mean* μ, and an *SD* σ denoted as

$$X \sim N(\mu, \sigma^2) \quad \text{in MMSA}$$
$$X \sim N(\mu, \sigma) \quad \text{in OpenIntro Statistics}$$

if its pdf is

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





The density curve is *bell-shaped* and *symmetric* about its mean μ.
To avoid the confusion, we use $N(\mu, \sigma^2)$ throughout the lectures and assignments

Expected Value and Variance of Normal Distributions

For the pdf of $X \sim N(\mu, \sigma^2)$

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

One can show that

- $\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1$ (not a trivial calculation, but can be found in most calculus textbooks)
- $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x; \mu, \sigma) dx = \mu$
- Var$(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x; \mu, \sigma) dx = \sigma^2$

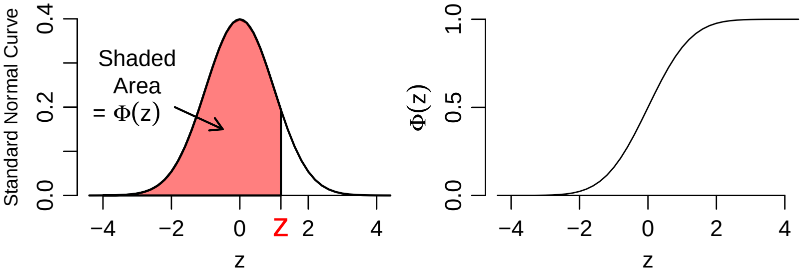
Standard Normal Distribution

A normal distribution with $\mu = 0$, and $\sigma = 1$ is called the *standard normal distribution*, denoted as $N(0, 1)$.

CDF of the Standard Normal Distribution

The cdf of the **standard normal distribution** $N(0, 1)$ is

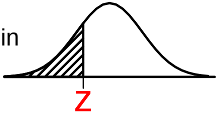
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$



- The cdf $\Phi(z)$ has no close-form formula
- The normal probability table (on p.792-793 in MMSA) gives the values of the cdf $\Phi(z)$ for different z's

The *normal probability table* (on p.792-793 in MMSA) gives

$\Phi(z) = P(Z \le z)$ = shaded area in



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

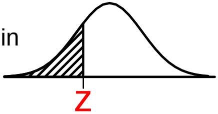
E.g., for $z = -0.83$, look at the row -0.8 and the column 0.03.

$\Phi(-0.83) = P(Z < -0.83) =$

= ________

The *normal probability table* (on p.792-793 in MMSA) gives

$\Phi(z) = P(Z \leq z)$ = shaded area in



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

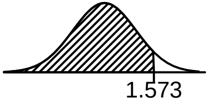
E.g., for $z = -0.83$, look at the row -0.8 and the column 0.03.

$\Phi(-0.83) = P(Z < -0.83) =$

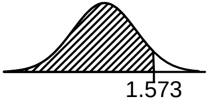
= 0.2033

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\Phi(1.573) = P(Z < 1.573) \quad =$



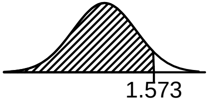
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\Phi(1.573) = P(Z < 1.573) =$ 

$=$ between $P(Z < 1.57)$ and $P(Z < 1.58)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\Phi(1.573) = P(Z < 1.573)$ =



= between $P(Z < 1.57)$ and $P(Z < 1.58)$

= between 0.9418 and 0.9429

Any value between 0.9418 and 0.9429 are acceptable in all HWs and exams.

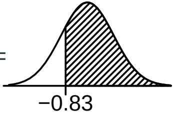
Find Normal Probabilities in R

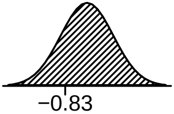
The R command `pnorm(z)` can find $P(Z \leq z)$ for $Z \sim N(0, 1)$

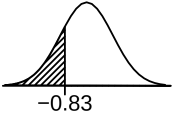
```
pnorm(-0.83)
[1] 0.2032694
pnorm(1.573)
[1] 0.9421406
```

Finding Upper Tail Probabilities

For $Z \sim N(0, 1)$

$P(Z > -0.83) =$ 

$=$ 

$-$ 

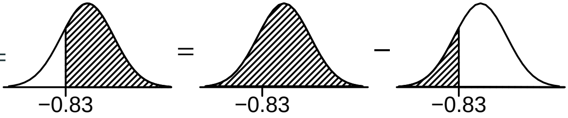
$$= 1 - \Phi(0.83)$$

$$= 1 -$$

Finding Upper Tail Probabilities

For $Z \sim N(0, 1)$

$P(Z > -0.83) =$



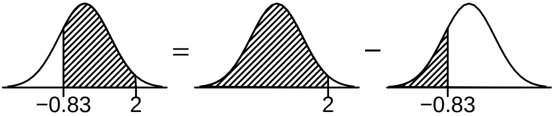
$= 1 - \Phi(0.83)$

$= 1 - 0.2033 = 0.7967$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148

```r
1 - pnorm(-0.83)
[1] 0.7967306
# another way to find upper tail area
pnorm(-0.83, lower.tail=FALSE)
[1] 0.7967306
```

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857

$P(-0.83 < Z < 2) =$ 

$$= P(Z < 2) - P(Z < -0.83)$$

$$= \Phi(2) - \Phi(-0.83)$$

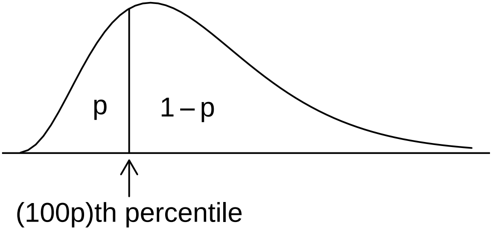
$$= 0.9772 - 0.2033 = 0.7739$$

```
pnorm(2) - pnorm(-0.83)
[1] 0.7739805
```

Percentile

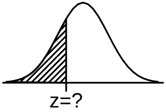
For $0 < p < 1$, the $(\mathbf{100p})$**th percentile** of a continuous random variable X is a value x such that

$$p = P(X \leq x).$$



Example: Percentiles of the Standard Normal (1)

Find the 25th percentile of the standard normal, i.e., the z such that

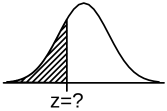
$\Phi(z) = P(Z < z) =$ shaded area in  $= 0.25.$

We need to search in the body of the normal table for the values closest to 0.25, which are 0.2514 and 0.2483, and find the z corresponding to those values, which are -0.67 and -0.68.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Example: Percentiles of the Standard Normal (1)

Find the 25th percentile of the standard normal, i.e., the z such that

$\Phi(z) = P(Z < z) =$ shaded area in 

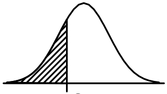
$= 0.25.$

We need to search in the body of the normal table for the values closest to 0.25, which are 0.2514 and 0.2483, and find the z corresponding to those values, which are -0.67 and -0.68.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Example: Percentiles of the Standard Normal (1)

Find the 25th percentile of the standard normal, i.e., the z such that

$\Phi(z) = P(Z < z) =$ shaded area in  $= 0.25.$

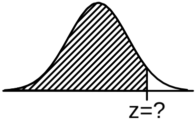
We need to search in the body of the normal table for the values closest to 0.25, which are 0.2514 and 0.2483, and find the z corresponding to those values, which are -0.67 and -0.68.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The 25th percentile of $N(0, 1)$ is between -0.67 and -0.68.

Example: Percentiles of the Standard Normal (2)

The 95th percentile of the standard normal is the z such that

$$\Phi(z) = P(Z < z) =$$  $$= 0.95.$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

The values closest to 0.95 in the body of the normal table are 0.9495 and 0.9505, with corresponding z-values are 1.64 and 1.65.

The 95th percentile is thus between 1.64 and 1.65.

Finding Standard Normal Percentiles in R

The R command `qnorm(p)` finds the **(100p)th percentile** for the Standard normal distribution, i.e., the z such that

$$P(Z < z) = p \quad \text{for } Z \sim N(0, 1).$$

E.g., the 25th percentile of $N(0, 1)$ is

```
qnorm(0.25)
[1] -0.6744898
```

E.g., the 95th percentile of $N(0, 1)$ is

```
qnorm(1-0.05)
[1] 1.644854
qnorm(0.05, lower.tail=F)  # alternative way
[1] 1.644854
```

Scaling Property of Normal Distributions

- If $X \sim N(\mu, \sigma^2)$, its *standardized z-score* $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
- Conversely, if $Z \sim N(0, 1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.
- This is because all normal distributions have the same *shape*; differ only in:
  - position given by the mean μ.
  - scale given by the standard deviation σ.

Remark. Not all distributions have scaling properties,
ex., $(X - a)/b$ is not Binomial even if X is Binomial

CDF of $N(\mu, \sigma^2)$

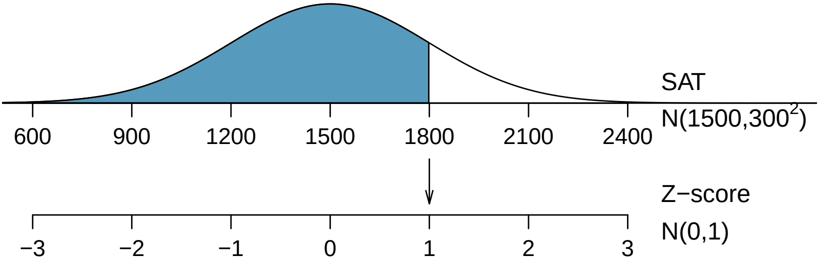
If $X \sim N(\mu, \sigma^2)$, then by definition, the CDF of X is

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

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

Example: SAT (1)

The distribution of SAT scores was about $N(\mu = 1500, \sigma^2 = 300^2)$. What percent of students scored below 1800 on the SAT?



Sol. Let X = SAT scores of a randomly selected test taker. We know $X \sim N(\mu = 1500, \sigma^2 = 300^2)$.

$$P(X \leq 1800) = P\left(\frac{X - 1500}{300} \leq \frac{1800 - 1500}{300}\right) = P(Z \leq 1)$$

where $Z = \frac{X-1500}{300} \sim N(0, 1)$.

From the table below, we can see that $P(Z < 1) = 0.8413$. So about 84% of students score below 1800 on the SAT.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

In R:

```
pnorm((1800-1500)/300)
[1] 0.8413447
pnorm(1800, mean = 1500, sd = 300)
[1] 0.8413447
```

Example: SAT (2)

What proportion of SAT takers scored between 1650 and 1800?

$$P(1650 \leq X \leq 1800) = P\left(\frac{1650-1500}{300} \leq \frac{X-1500}{300} \leq \frac{1800-1500}{300}\right)$$
$$= P(0.5 \leq Z \leq 1)$$
$$= \Phi(1) - \Phi(0.5)$$
$$= 0.8413 - 0.6915 = 0.1498.$$

as the normal table gives $\Phi(1) = 0.8413$ and $\Phi(0.5) = 0.6915$.

```
pnorm(1)-pnorm(0.5)
[1] 0.1498823
pnorm(1800, m = 1500, s = 300)-pnorm(1650, m = 1500, s = 300)
[1] 0.1498823
```

Example: What is the SAT score at the 90% percentile?

Want a score x such that $P(X < x) = 0.9$ for $X \sim N(\mu = 1500, \sigma = 300)$.

$$P(X < x) = P\left(\frac{X-1500}{300} < \frac{x-1500}{300}\right) = P\left(Z < \frac{x-1500}{300}\right)$$

$$= \Phi(z) = 0.9, \text{ where } z = \frac{x-1500}{300}.$$

What is the z such that $\Phi(z) = 0.9$?

z	0.05	0.06	0.07	0.08	0.09
1.1	.8749	.8770	.8790	.8810	.8830
1.2	.8944	.8962	.8980	.8997	.9015
1.3	.9115	.9131	.9147	.9162	.9177

So $z = \frac{x - 1500}{300} \approx 1.28$

$\Rightarrow \quad x = 1500 + 300 \times 1.28 = 1884$

The 90th percentile of the SAT score is 1884.

```r
qnorm(0.9, mean=1500, sd = 300)
[1] 1884.465
```

Example: What is the SAT score at the 90% percentile?

Want a score x such that $P(X < x) = 0.9$ for $X \sim N(\mu = 1500, \sigma = 300)$.

$$P(X < x) = P\left(\frac{X-1500}{300} < \frac{x-1500}{300}\right) = P\left(Z < \frac{x-1500}{300}\right)$$

$$= \Phi(z) = 0.9, \text{ where } z = \frac{x-1500}{300}.$$

What is the z such that $\Phi(z) = 0.9$? $z \approx 1.28$

z	0.05	0.06	0.07	0.08	0.09
1.1	.8749	.8770	.8790	.8810	.8830
1.2	.8944	.8962	.8980	.8997	.9015
1.3	.9115	.9131	.9147	.9162	.9177

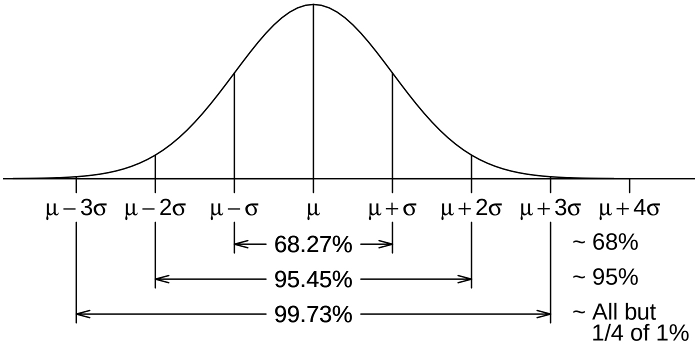
So $z = \dfrac{x-1500}{300} \approx 1.28$

$\Rightarrow \quad x = 1500 + 300 \times 1.28 = 1884$

The 90th percentile of the SAT score is 1884.

```
qnorm(0.9, mean=1500, sd = 300)
[1] 1884.465
```

68-95-99.7% Rule for the Normal Distributions



In terms of the standard normal CDF $\Phi(z)$:

$$P(-1 < Z < 1) = \Phi(1) - \Phi(-1) \approx 0.6827$$

$$P(-2 < Z < 2) = \Phi(2) - \Phi(-2) \approx 0.9545$$

$$P(-3 < Z < 3) = \Phi(3) - \Phi(-3) \approx 0.9973$$