# Sampling Distribution of Means

# Sampling Distribution of the Means

- The sampling distribution of means is what you get if you consider all possible samples of size *n* taken from the same population and form a distribution of their means.
- Each randomly selected sample is an independent observation.

# Central Limit Theorem

- http://onlinestatbook.com/2/sampling_distributions/clt_demo.html
- As sample size goes large and number of buckets are high, the means will follow a normal distribution with same mean (μ) and $\frac{1}{n}$ of variance ($\sigma^2$).

# Expectation and Variance for $\bar{X}$

$$E(\bar{X}) = \mu$$

**Mean of all sample means of size *n* is the mean of the population.**

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Standard deviation of $\bar{X}$ tells how far away from the population mean the sample mean is likely to be and is called the **Standard Error of the Mean**, and is given by

$$Standard\ Error\ of\ the\ Mean = \frac{\sigma}{\sqrt{n}}$$

If X ~ N($\mu$, $\sigma^2$), then $\bar{X}$ ~ N($\mu$, $\sigma^2$/n)

# When an Attribute is Not Normal

- Let us assume it is a sample from infinite data
- So, if we take many such samples of large sample size (>30 as a thumb rule), the mean values, $\bar{x}$, will be hovering close to the population mean, $\mu$, with a standard deviation, $s = \frac{\sigma}{\sqrt{n}}$, where $\sigma$ is the population standard deviation and $n$ is the sample size.

# Using the Central Limit Theorem

Let us say the mean number of Gems per packet is 10, and the variance is 1. If you take a sample of 30 packets, what is the probability that the sample mean is 8.5 Gems per packet or fewer?


# Using the Central Limit Theorem

We know that $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, $\mu$ = 10, $\sigma^2$ = 1 and n = 30.

We need the value of P($\bar{X}$ < 8.5) when $\bar{X}$ ~ N(10, 0.0333).

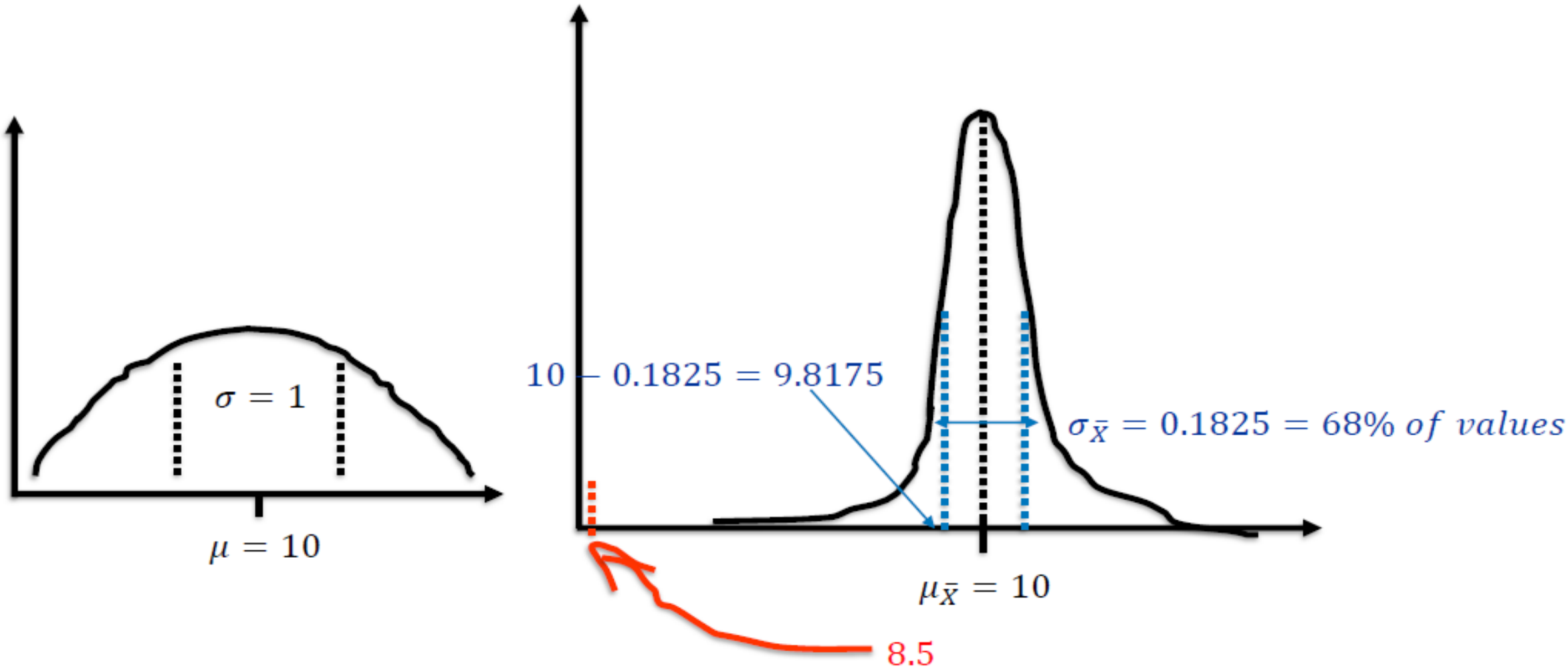
$$z = \frac{8.5 - 10}{\sqrt{0.0333}} = -8.22$$

P(Z < z) = P(Z < -8.22)

This doesn't exist in probability tables.  What does it mean?

# Using the Central Limit Theorem

How do we visualize it?

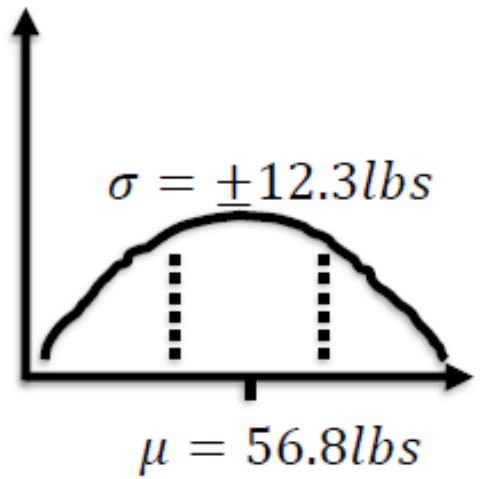


# Using the Central Limit Theorem

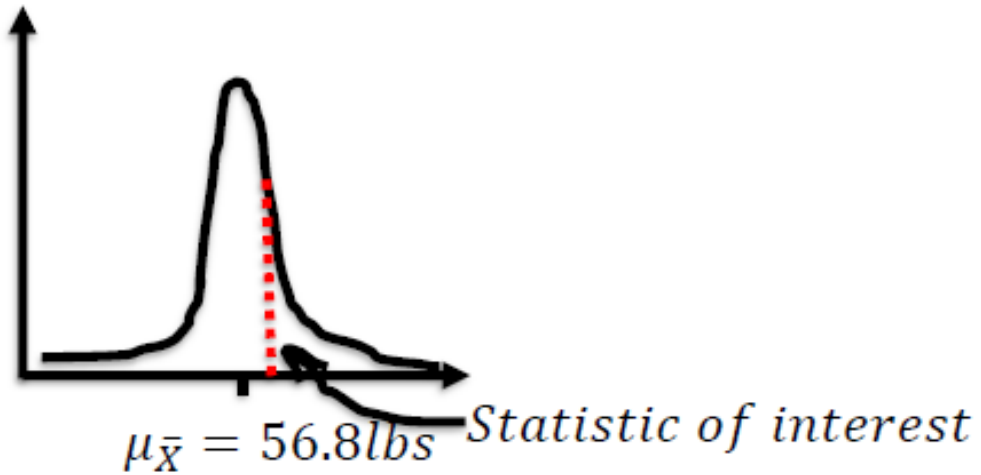
The Aluminum Association of America reports that the average American household uses 56.8 lbs of aluminium in a year. A random sample of 51 households is monitored for one year to determine aluminium usage. If the population standard deviation of annual usage is 12.3 lbs, what is the probability that the sample mean will be $> 60\ lbs$?

# Sampling Distribution

*Population distribution*

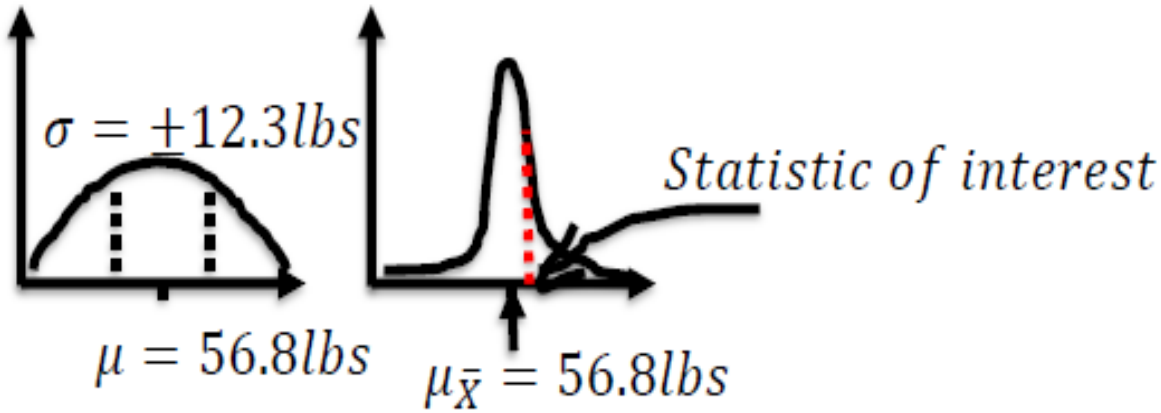



*Sampling distribution of sample mean when* $n = 51$



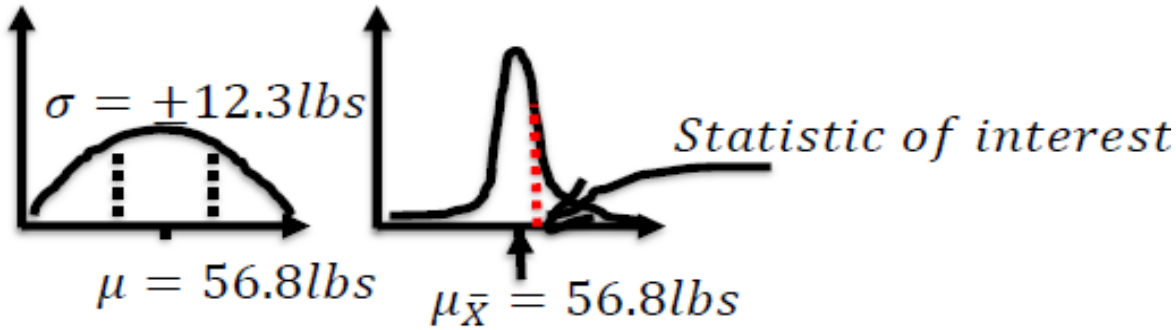

- Step 1: List all known parameters and values
- Step 2: Calculate others, or estimate if cannot be calculated
- Step 3: Find probabilities using tables, Excel or R

# Sampling Distribution




- Step 1: List all known parameters and values
  - Population mean, $\mu = 56.8\ lbs$
  - Population standard deviation, $\sigma = 12.3\ lbs$
  - Sample size, $n = 51$
  - Sample mean, $\bar{x} > 60\ lbs$
  - Mean of sample means, $\mu_{\bar{X}} = \mu = 56.8\ lbs$

# Sampling Distribution



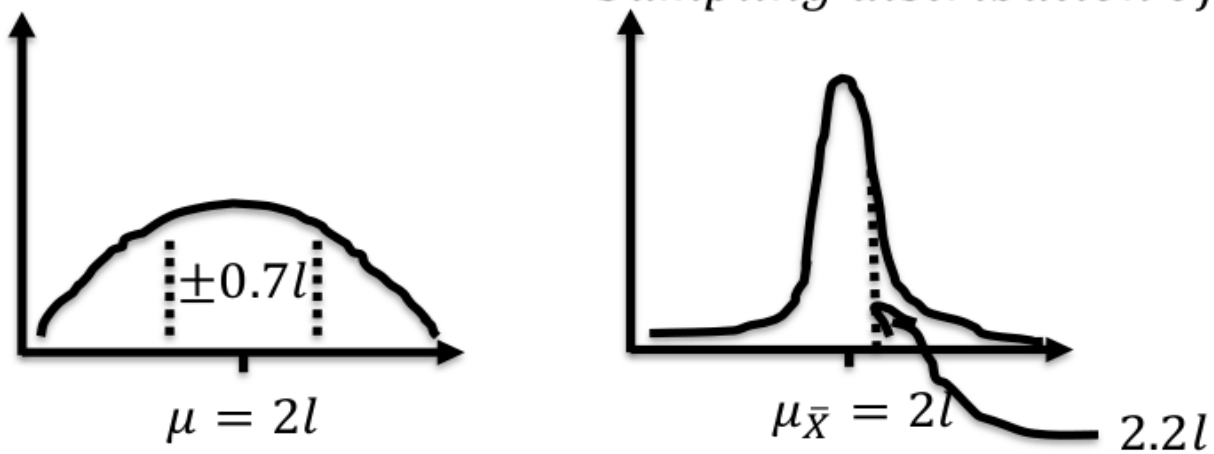

- Step 2: Calculate others or estimate, if cannot be calculated
  - Standard deviation of sample means, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12.3}{\sqrt{51}} = 1.72$
  - $\therefore z = \frac{60-56.8}{1.72} = 1.86$
- Step 3: Find probabilities using tables, Excel or R
  - Excel: 1-NORM.S.DIST(z,TRUE) = 0.0316
  - Please calculate these for:
    - $> 58\ lbs$
    - $> 56\ lbs < 57\ lbs$
    - $< 50\ lbs$

# Sampling Distribution

The average male drinks $2l$ of water when active outdoors with a standard deviation of $0.7l$. You are planning a trip for 50 men and bring $110l$ of water. What is the probability that you will run out of water?

$\mu = 2, \sigma = 0.7$

$P(run\ out) \Rightarrow P(use > 110l) \Rightarrow P(average\ water\ use\ per\ male > 2.2l)$




$\mu_{\bar{X}} = \mu = 2l, {\sigma_{\bar{X}}}^2 = \frac{\sigma^2}{n} = \frac{0.49}{50}$

$\Rightarrow \sigma_{\bar{X}} = 0.099$

$z = \frac{2.2 - 2}{0.099} = 2.02$

$P(\bar{X} < 2.02) = 0.9783$

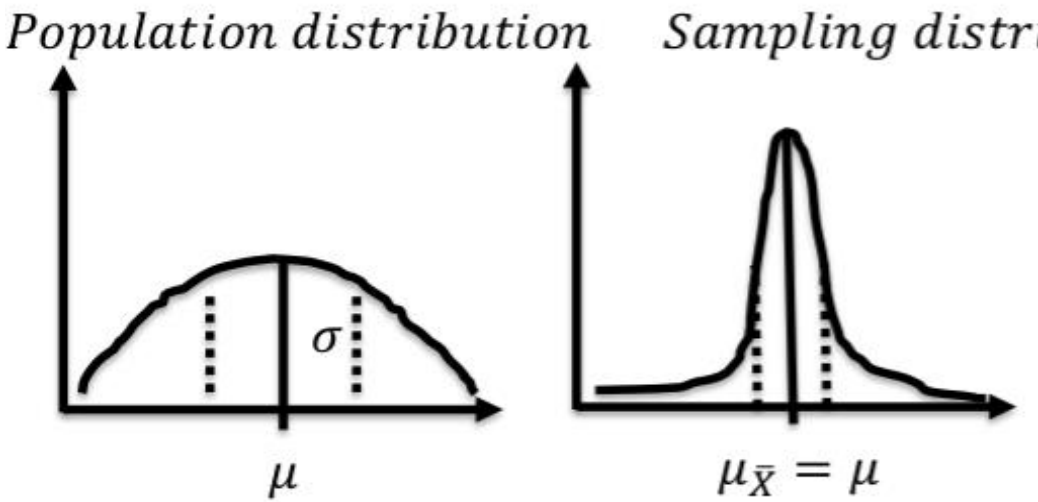
The probability of running out is

$1\text{-}0.9783 = 0.0217$ or $2.17\%$

# Using the Central Limit Theorem

You sample 36 apples from your farm's harvest of 200,000 apples. The mean weight of the sample is 112g with a 40g sample standard deviation. What is the probability that the mean weight of all 200,000 apples is between 100 and 124g?

# Sampling Distribution




$$\sigma_{\bar{X}}{}^2 = \frac{\sigma^2}{n} = \frac{\sigma^2}{36} \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{6}$$

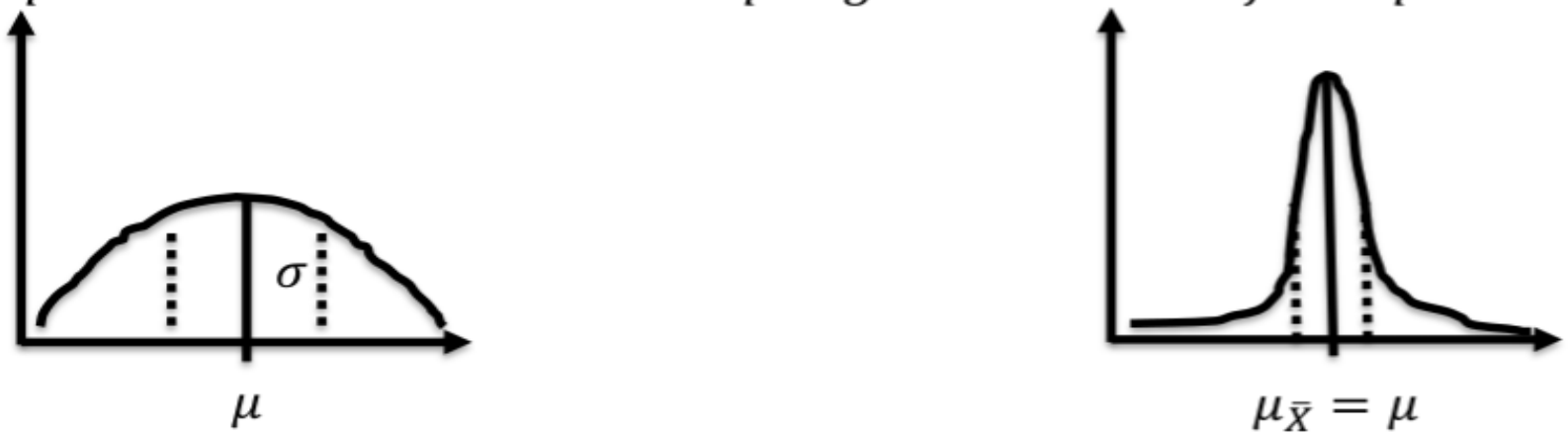
What are we trying to find out?

We need to know if population mean, $\mu$, is within $\pm$12g of the sample mean, $\bar{X}$.

This is the same as saying that we need to know if sample mean, $\bar{X}$, is within $\pm$12g of the population mean, $\mu$. Since $\mu = \mu_{\bar{X}}$, we can now use the sampling distribution of the means.

# Sampling Distribution

$Population\ distribution$ $Sampling\ distribution\ of\ sample\ mean\ when\ n = 36$




We need to find out how many standard deviations away from $\mu_{\bar{X}}$ is 12g. But, we don't know $\sigma_{\bar{X}}$ because we don't know $\sigma$. We use the sample standard deviation, $s$ (40g), as the best estimate of population standard deviation. $\sigma \approx s = \pm 40g.\ \therefore \sigma_{\bar{X}} = \frac{\sigma}{6} = \frac{40}{6} = 6.67.$ So 12g is 12/6.67 = 1.8 standard deviations.

The z-table gives the probability as 0.9641 but that is the entire region below +1.8 z.

Find the region between -1.8 and +1.8 z.

0.9282. How would you get this answer if you did not have the negative z table?

# Normal Distribution

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 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 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 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 |
| 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 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |

(0.9641-0.5000)*2
= 0.4641*2
= 0.9282

# Activity – R

According to National Center for Health Statistics of the US, the distribution of serum cholesterol levels for 20-74 year old males has a mean of 211mg/dl with a standard deviation of 46mg/dl.

- What is the probability that the serum cholesterol level of a male is >230mg/dl?
- What is the probability that the average serum cholesterol level of a random sample of 25 males will be >230mg/dl?

Answer: 34.0%, 1.9%