

Data Analysis and Statistical Methods

Statistics 651

<http://www.stat.tamu.edu/~suhasini/teaching.html>

Lecture 9 (MWF) Calculations for the normal distribution

Suhasini Subba Rao

Evaluating probabilities for a normal distributions

- Suppose that we are studying the height of women which is normally distributed with mean 64.5 and standard deviation 2.5. How to we evaluate the percentile for a women of height 71 inches?
- One may think for every mean μ and standard deviation σ we require a new set of tables. This is not possible.
- In fact we only need the standard normal tables to evaluate any normal probability. You could also use statistical software such as JMP or free online Apps, to calculate the probabilities for us.

The z-transform and its link to chance

- Suppose the distribution of female heights has mean 64.5 inches and standard deviation 2.5 inches.
- A female is 71 inches tall. You ask yourself, is she exceptionally tall. Well she is 6.5 inches taller than the average. The standard deviation is 2.5, hence she is

$$z = \frac{71 - 64.5}{2.5} = 2.6$$

standard deviations from the mean.

- Now this information tells us she is far from the majority of data, but how far is far?

The z-transform and the normal

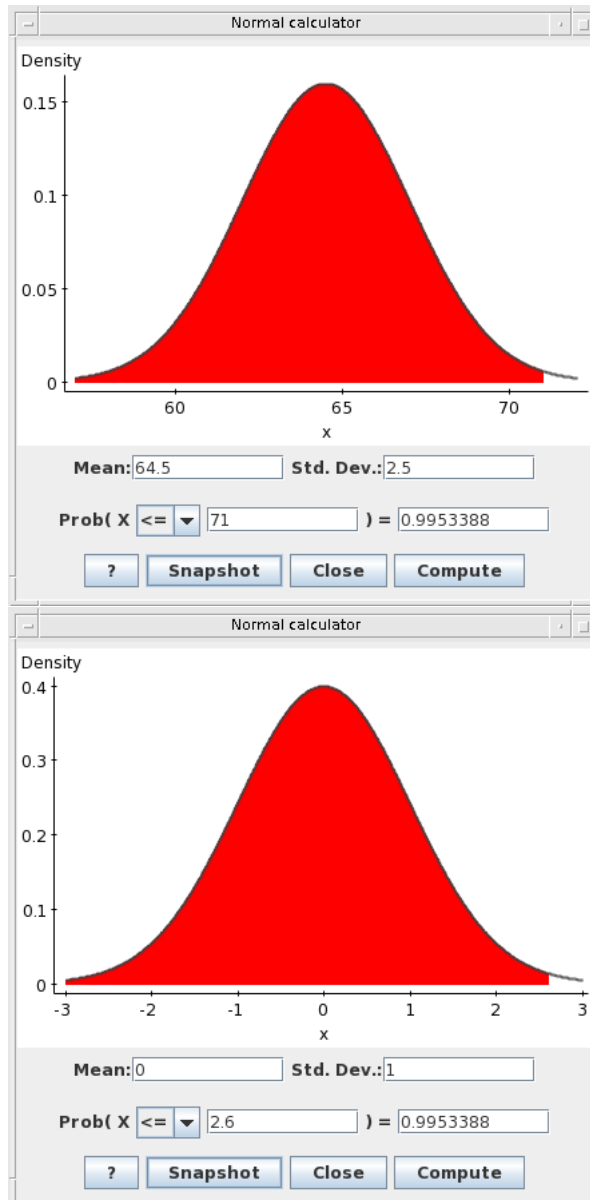
- We require her percentile. But this can only be calculated with additional knowledge of the distribution of heights.
- Using the assumption that heights are normally distributed and the z-transform $z = 2.6$ we can calculate the percentile.
- The z-transform

$$Z = \frac{X - \mu}{\sigma} = \frac{\text{observation} - \text{mean}}{\text{standard deviation/standard error}}$$

is essentially a way of converting normal data to the reference (standard normal) distribution.

- Looking up 2.6 in the tables give the probability 99.53%. This tells us she is in the 99.53% percentile. She is very tall!
- Remember When calculating a z-transform we are converting the normal data to a “reference” distribution, which is the standard normal (mean zero and standard deviation one). In the plots below we give both the original normal distribution and the standard normal. The red areas are the probabilities, you should observe that they are the same for both distributions.
- Important When calculating a z-transform don't exchange the observation with the mean.

Lecture 9 (MWF) The normal distribution calculations



The top plot gives the distribution of heights, together with her height. The lower plot gives the standard normal together with her corresponding z-score. Observe the areas (which corresponds to the percentiles) are the same.

In General

Summarising the previous calculations:

- Suppose a random variable has a normal distribution with mean μ and variance σ^2 we write this succinctly as $X \sim N(\mu, \sigma^2)$.
- We transform the random variable using the z-transform

$$Z = \frac{X - \mu}{\sigma} = \frac{\text{observation} - \text{mean}}{\text{standard deviation/standard error}}$$

- Then this new random variable $Z \sim N(0, 1)$, is a standard normal distribution with mean zero and variance one $Z \sim N(0, 1)$ (it is standard normal).

- The algebra is

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

- Since a , μ and σ are known, $\frac{a - \mu}{\sigma}$ is just a number. Recall $Z \sim N(0, 1)$, hence to evaluate $P(X < a)$ we need only to look it up in the standard normal tables.

Examples 1

Suppose $X \sim N(\mu = 4, \sigma = \sqrt{5} \approx 2.23)$ (mean 4 and standard deviation $\sqrt{5} = 2.23$)

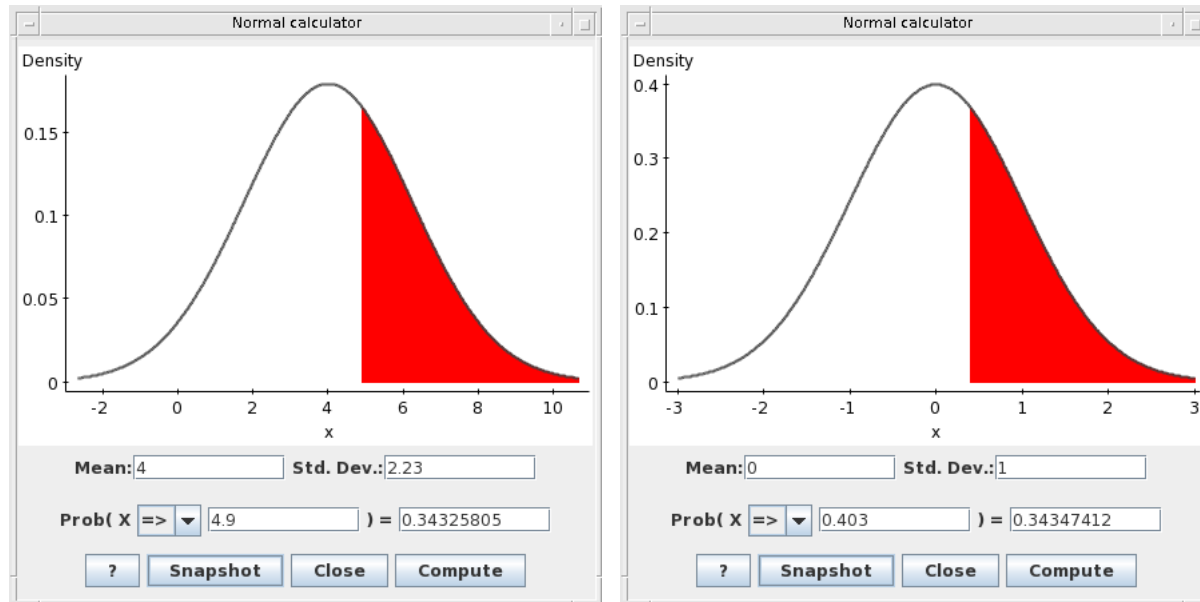
- Calculate (i) $P(X > 4.9)$, (ii) $P(X \leq 6)$, (iii) $P(X < 3.2)$ and (iv) $P(X \leq 2.3)$.
- Calculate (i) $P(3.2 < X \leq 6)$, (ii) $P(2.3 < X < 4.9)$ and (iii) $P(2.3 < X < 4.9 \text{ or } 3.2 < X \leq 6)$.

For each probability draw the picture.

- Hand written solutions with explanations and plots are given in

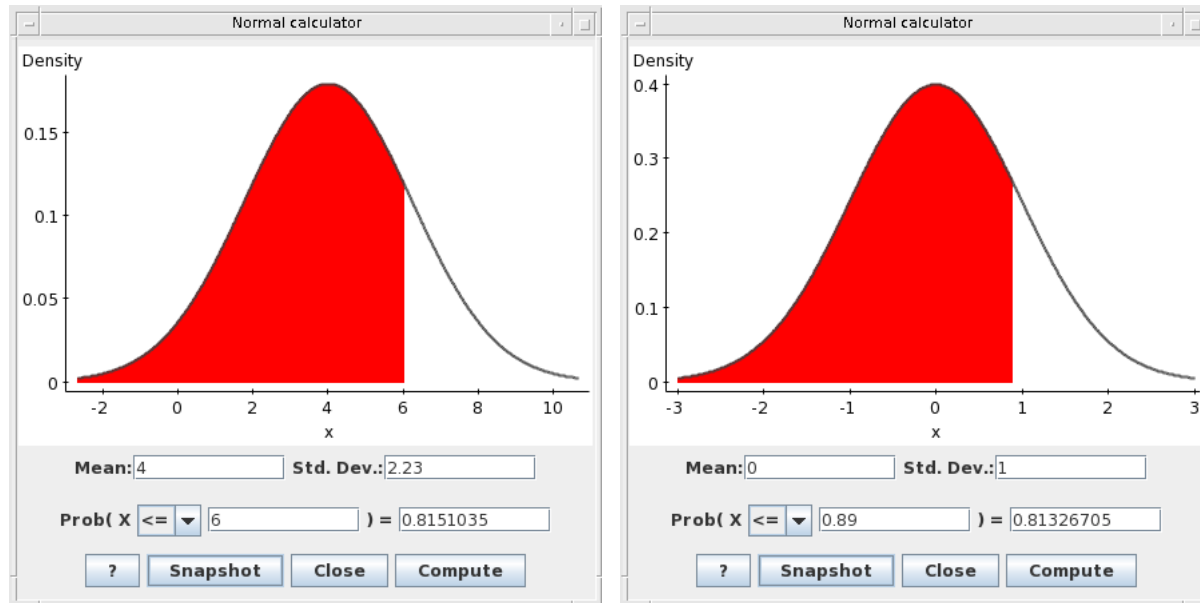
http://www.stat.tamu.edu/suhasini/teaching651/examples_normal_probailities_lecture9.pdf.

Solution 1(i)



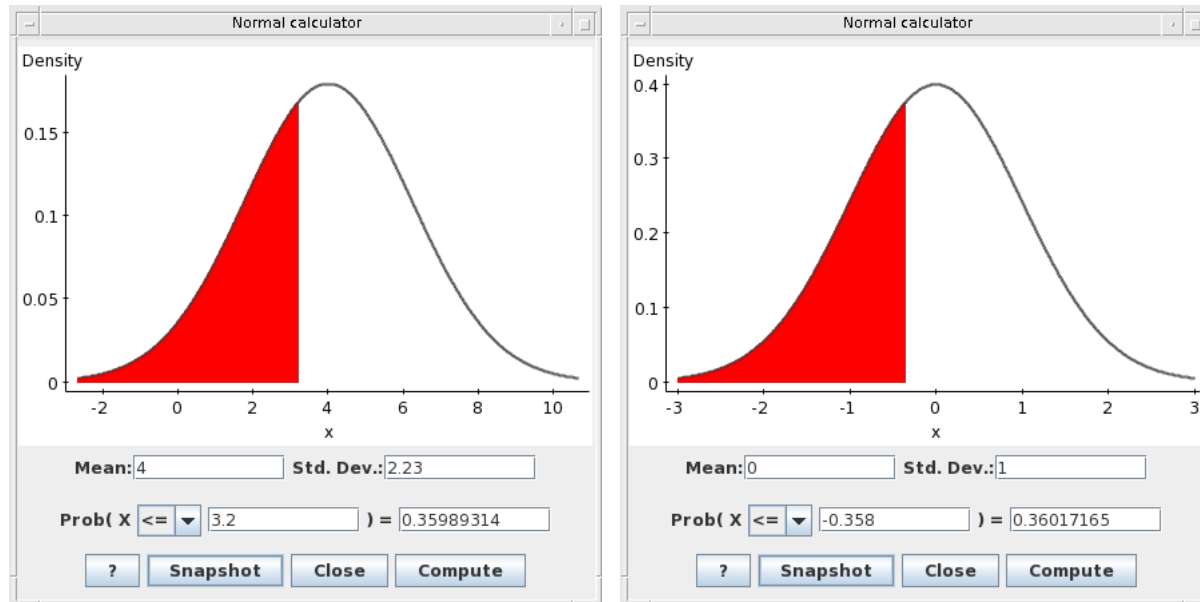
$$P(X > 4.9) = P(Z > 0.4) = 0.34.$$

Solution 1(ii)



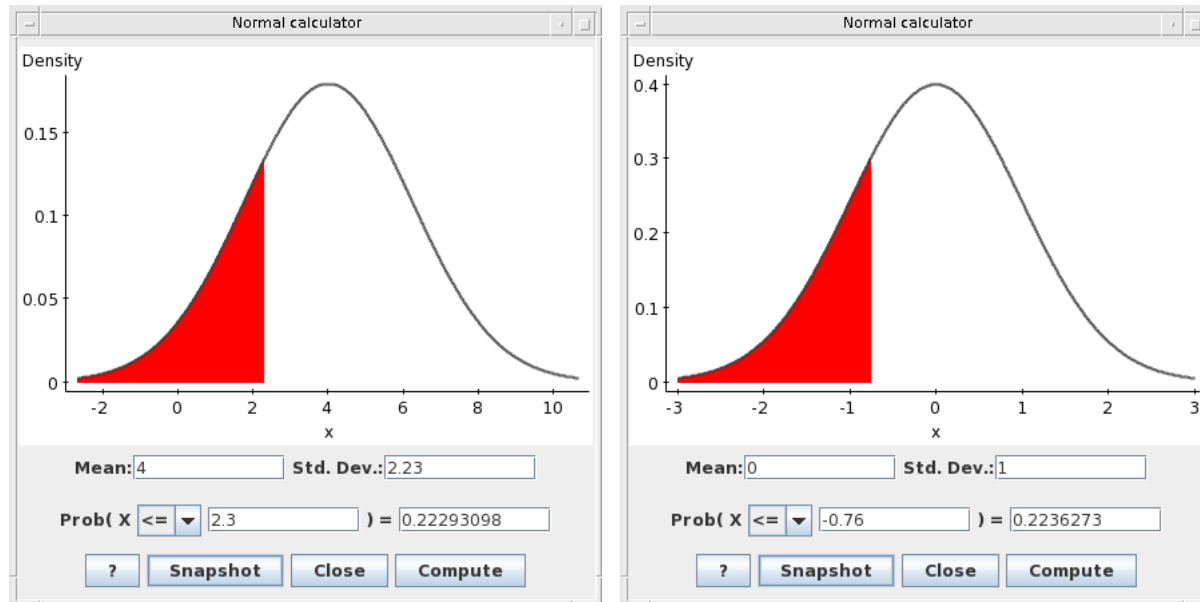
$$P(X \leq 6) = P(Z \leq 0.89) = 0.81$$

Solution 1(iii)



$$P(X < 3.2) = P(Z < -0.358) = 0.360$$

Solution 1(iv)



$$P(X \leq 2.3) = P(Z < -0.76) = 0.223$$

Examples 2

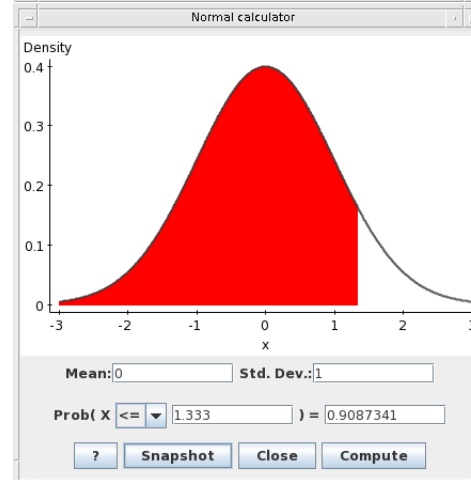
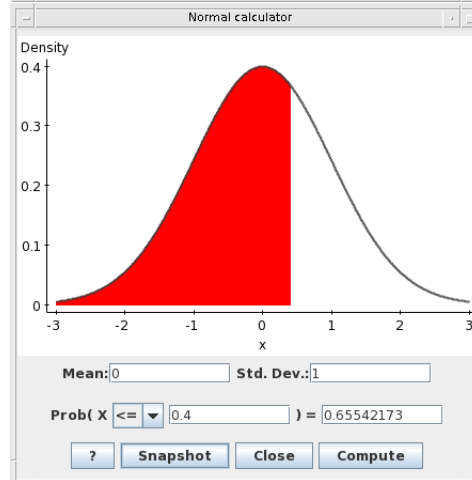
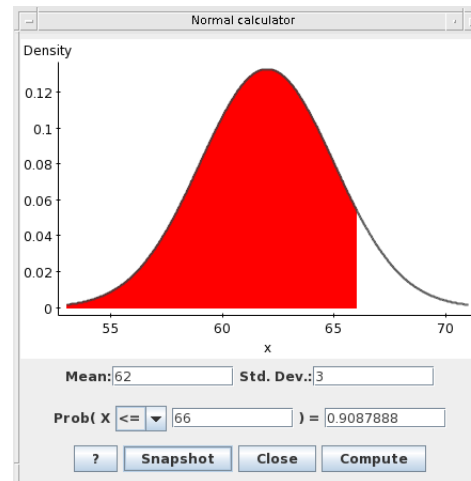
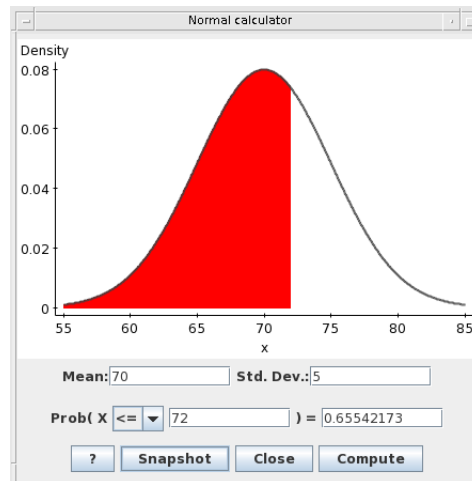
Suppose that the height of asian women follow (roughly) a normal distribution $N(\mu = 62, \sigma = 3)$ and the height of asian men also follow a normal distribution $N(\mu = 70, \sigma = 5)$. A sister and brother (both asian) are 66 inches and 72 inches respectively.

- Calculate the percentile of the brother's height.
- Calculate the percentile of the sister's height.
- Taking into account gender, who is taller the brother or the sister?

Solution 2

- Percentile of brother's height is $P(Y \leq 72) = P(Z \leq \frac{72-70}{5}) = P(Z \leq \frac{2}{5}) = 0.4$. The brother's height is on the 66th percentile.
- Percentile of sister's height $P(X \leq 66) = P(Z \leq \frac{66-62}{3}) = P(Z \leq \frac{4}{3}) = 0.91$. The sister's height is on the 91th percentile.
- Since the sister's percentile is larger than the brother's percentile, taking account of gender, the sister is taller.
- The plots corresponding to the probabilities (for both the normal distribution and the standard normal) are given below. Observe that the areas are the same.

Lecture 9 (MWF) The normal distribution calculations



Examples 3

Suppose $X \sim N(-2, 3)$ (mean is -2 and standard deviation is 3).

- Calculate (i) $P(X > -6)$, (ii) $P(X \leq 0)$, (iii) $P(X < 1.2)$ and (iv) $P(X < 2)$.
- Calculate (i) $P(-6 < X \leq 1.2)$, (ii) $P(2 < X < 1.2)$.
- For each answer make a plot and indicate the probability of interest.
- A quick summary of the solutions is given below.

Solutions 3

If $X \sim N(-2, \sigma = 3)$, then you will need evaluate the z-score $Z = \frac{X+2}{3}$ to calculate the probabilities.

To each of the solutions below add your own picture.

$$(a)(i) \quad P(X > -6) = 1 - P(X \leq -6) = 1 - P\left(\frac{X+2}{3} \leq \frac{-6+2}{3}\right) = 1 - P\left(Z \leq \frac{-4}{3}\right) = 1 - 0.0918 = 0.9082.$$

$$(ii) \quad P(X < 0) = P\left(\frac{X+2}{3} \leq \frac{2}{3}\right) = 0.7454$$

$$(iii) \quad P(X < 1.2) = P\left(\frac{X+2}{3} \leq \frac{1.2+2}{3}\right) = 0.8554$$

$$(iv) \quad P(X < 2) = P\left(\frac{X+2}{3} \leq \frac{2+2}{3}\right) = 0.9082.$$

$$(b)(i) \quad P(-6 < X \leq 1.2) = P(X < 1.2) - P(X \leq -6) = P\left(\frac{X+2}{3} \leq \frac{1.2+2}{3}\right) - P\left(\frac{X+2}{3} \leq \frac{-6+2}{3}\right) = 0.8554 - 0.0918.$$

(ii) This is a trick question! The event $(2 < X < 1.2)$ can never arise (think about it). So $P(2 < X < 1.2) = 0$.

Example 4

- Suppose $X \sim N(10, \sigma = 4)$ (mean is 10 and standard deviation is 4). Evaluate $P(8 \leq X \leq 15)$.

Solution 4

- First normalise: subtract mean and divide by standard deviation:

$$\begin{aligned} P(8 \leq X \leq 15) &= P\left(\frac{8 - 10}{4} \leq \frac{X - 10}{4} \leq \frac{15 - 10}{4}\right) \\ &= P\left(-\frac{1}{2} \leq Z \leq \frac{5}{4}\right) = P(Z \leq 1.25) - P(Z \leq -0.5). \end{aligned}$$

- Look up $P(Z \leq 1.25)$ and $P(Z \leq -0.5)$ in tables: $P(Z \leq 1.25) = 0.8944$ and $P(Z \leq -0.5) = 0.3085$.
- Therefore $P(8 \leq X \leq 15) = P(-\frac{1}{2} \leq Z \leq \frac{5}{4}) = P(Z \leq 1.25) - P(Z \leq -0.5) = 0.8944 - 0.3085$, which gives the desired answer.

Example 5

The height of hens are normally distributed with mean μ and standard deviation σ ($N(\mu, \sigma)$).

What proportion of hens will be within **one** standard deviation of the mean? Remember, **one** corresponds to the z-score.

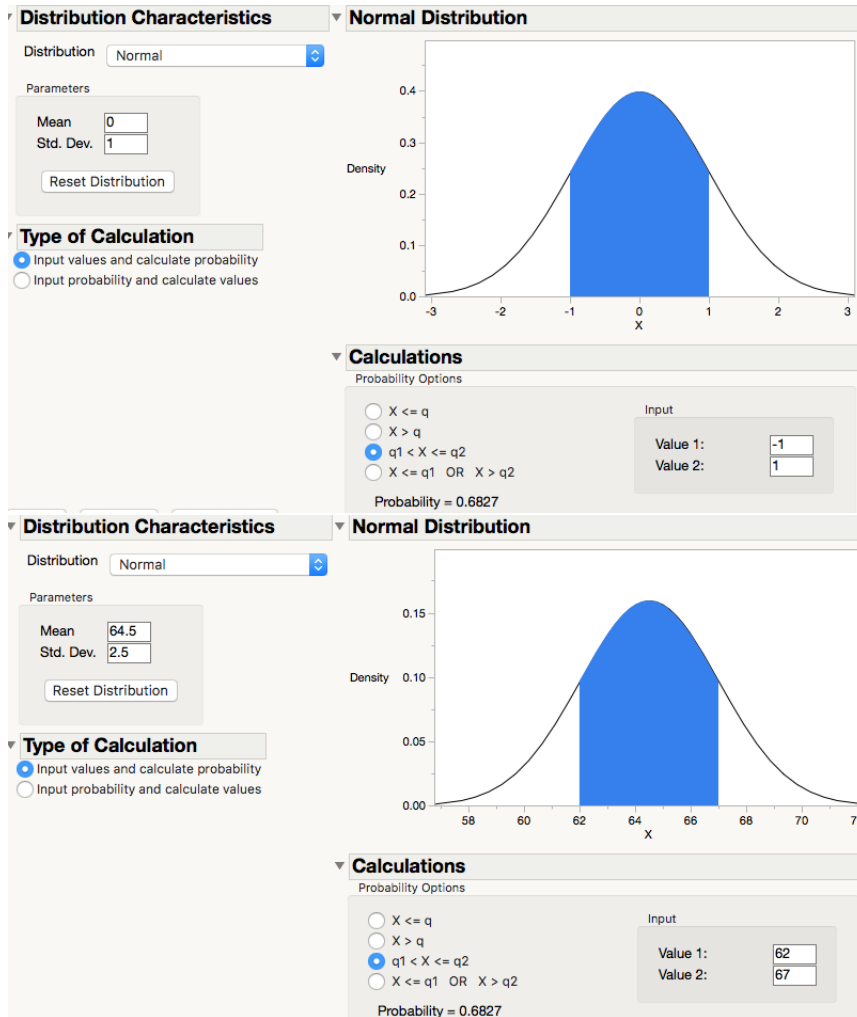
Solution

- The question asks what proportion of the heights lies within the interval $[\mu - \sigma, \mu + \sigma]$, which corresponds to the z-score $[-1, 1]$

In other words $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1)$ (since one standard deviation from the mean corresponds to the z-transforms -1 and 1).

Therefore, $P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z < -1) = 84.13 - 15.87 = \mathbf{68.26\%}$.

Lecture 9 (MWF) The normal distribution calculations



For a normal distribution within one standard deviation of the mean will always be the same proportion (the blue area). The same is true of within any z -standard deviations of the mean.

Reminder: Interpreting the z-transform

- If x is z -standard deviations to the right of the mean and the mean is μ and standard deviations σ (to be general), then

$$x = \mu + z \times \sigma.$$

- The z-score corresponding to x is

$$\frac{\mu + z \times \sigma - \mu}{\sigma} = z.$$

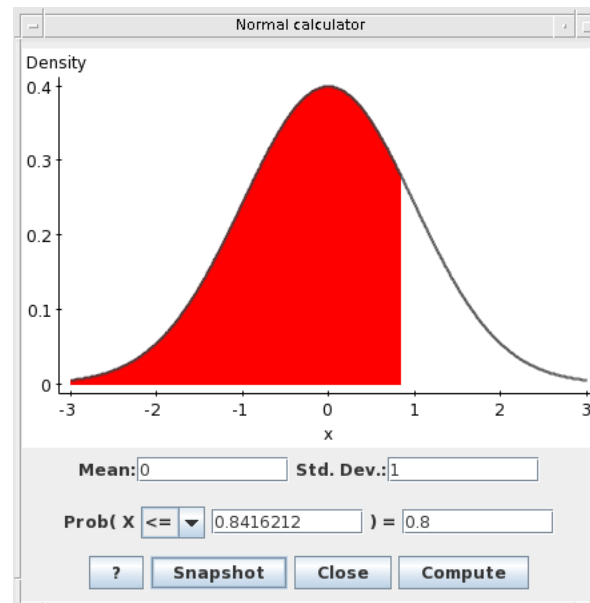
- Remember the z-score measures the number of standard deviations between the observation to the mean.

Example 6: Inverse transforms

- Let us return to the height example. Recall that we claim that female heights are normally distributed with mean 64.5 inches and standard deviation 2.5 inches.
 - (i) Jane is in the 80th percentile. How tall is Jane?
 - (ii) Judy is in the *top* 10th percentile. How tall is Judy?

Solution 6(i)

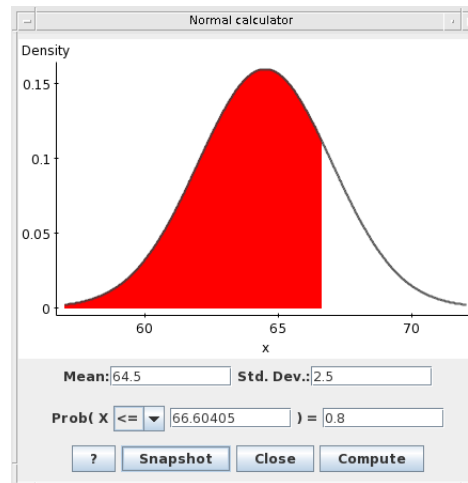
- Jane is in the 80th percentile. 80th percentile corresponds to the area below the graph. Looking from inside the tables out, this corresponds to a z-transform 0.84.



- We recall that the z-transform is a transformation of Jane's height

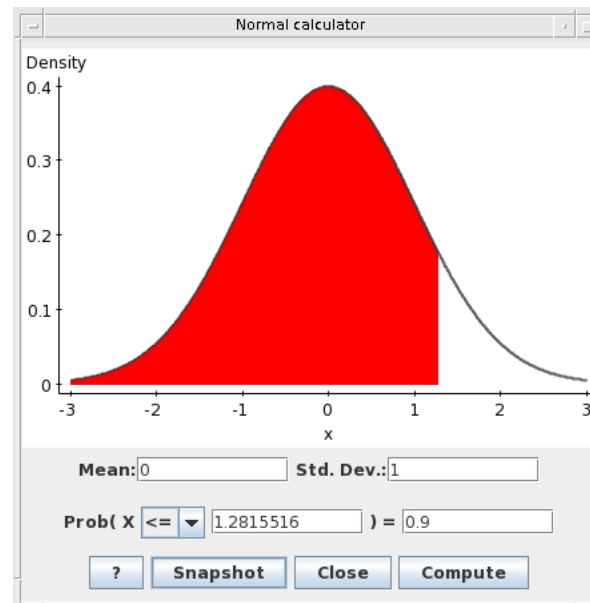
$$z = 0.84 = \frac{\text{Jane's height} - 64.5}{2.5}.$$

- Therefore, Jane's height is 0.84 standard deviations to the right of the mean. Solving for this gives Jane's height to be $64.5 + 0.84 \times 2.5 = 66.6$.



Solution 6(ii)

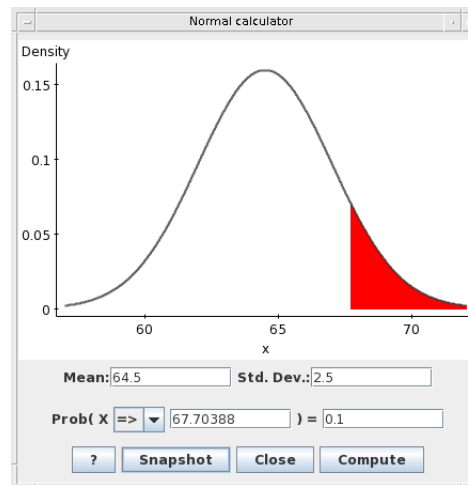
- Judy is in the top 10th percentile. This means her percentile is 90% (she is taller than 90% of females). 90th percentile corresponds to the area below the graph. Looking from inside the tables out, this corresponds to a z-transform 1.28.



- We recall that the z-transform is a transformation of Judy's height

$$z = 1.28 = \frac{\text{Judy's height} - 64.5}{2.5}.$$

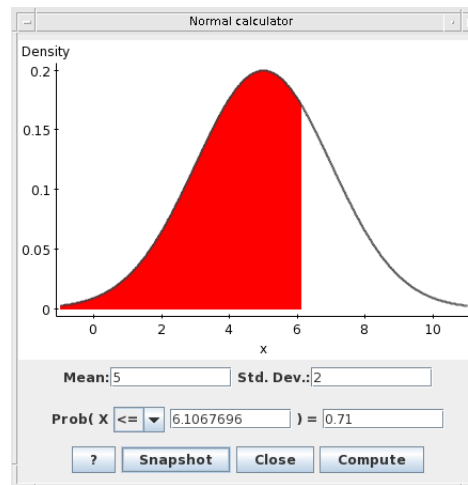
- Solving for this gives Judy's height to be $64.5 + 1.28 \times 2.5 = 67.7$



Question 7 (inverse transforms)

- Suppose $X \sim N(5, 2)$, find the value of x such that $P(X \leq x) = 0.71$.
- **Solution:** (can also be found at [xaxis_lecture9.pdf](#)) In the same way we try to evaluate probabilities, we transform into a standard normal:
- Recalling that $X \sim N(5, 2)$ (mean 5 and standard deviation 2) we have $0.71 = P(X \leq x) = P\left(\frac{X-5}{2} \leq \frac{x-5}{2}\right) = P\left(Z \leq \frac{x-5}{2}\right) = 0.71$, where $Z \sim N(0, 1)$.
- Now look up $P(Z \leq y) = 0.71$ in the tables. We find that $y = 0.55$.
- This means $P(X \leq x) = P\left(\frac{X-5}{2} \leq \frac{x-5}{2}\right) = P\left(Z \leq \frac{x-5}{2}\right) = P(Z \leq 0.55) = 0.71$.

- See $\frac{x-5}{2} = 0.55$. Solve this equation to give $x = 2 \times 0.55 + 5 = 6.1$.
- This means $P(X \leq 6.1) = 0.71$. You can check this!
- $P(X \leq 6.1) = P\left(\frac{X-5}{2} \leq \frac{6.1-5}{2}\right) = P(Z \leq 0.55)$, looking this up in the tables we get $P(Z \leq 0.55) = 0.71$. Hence it is correct. A pictorial depiction is given below:



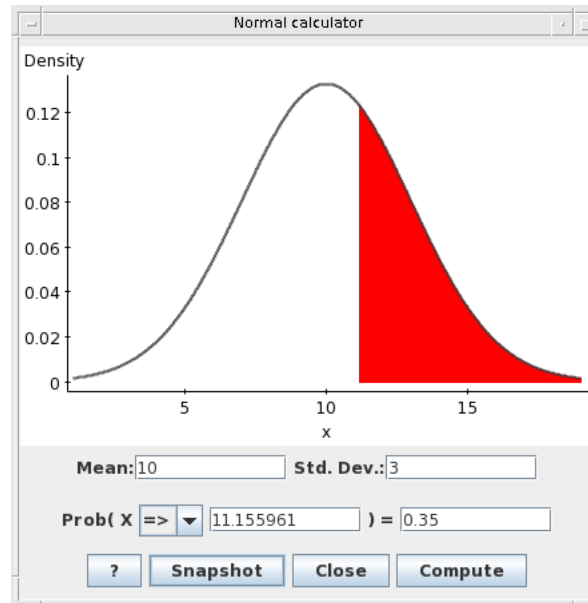
Question 8 (inverse transforms)

- Suppose $X \sim N(10, 3)$ (mean 10 and standard deviation 3).

Find the x such that $P(X \geq x) = 0.35$.

- **Solution:** We know from the density plot that $P(X < x) = 1 - P(X \geq x) = 1 - 0.35 = 0.65$. Hence it is easier, given the tables, to evaluate $P(X < x) = 0.65$, rather than $P(X \geq x) = 0.35$.
- $P(X < x) = P\left(\frac{X-10}{3} < \frac{x-10}{3}\right) = P\left(Z < \frac{x-10}{3}\right) = 0.65$.
- This gives, $P(Z < y) = 0.65$, looking up inside the tables we have $y \approx 0.39$.
- Hence $0.39 = y = \frac{x-10}{3}$. Solving this gives $x = 0.39 \times 3 + 10 = 11.17$.
 $P(X \geq 11.17) = 0.35$.

Lecture 9 (MWF) The normal distribution calculations

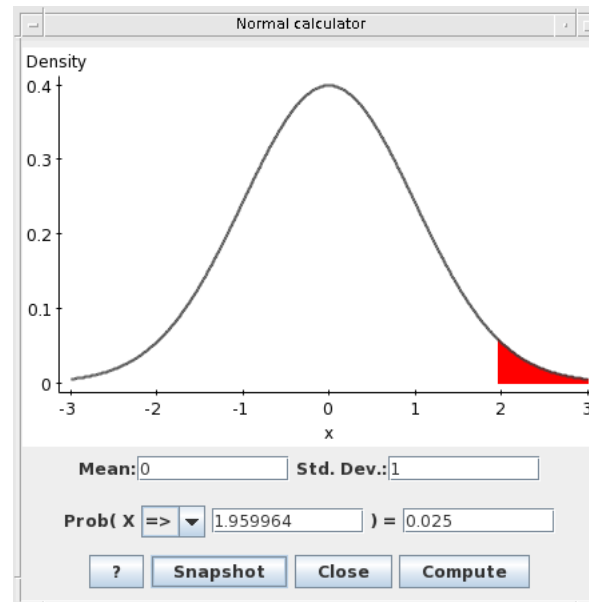
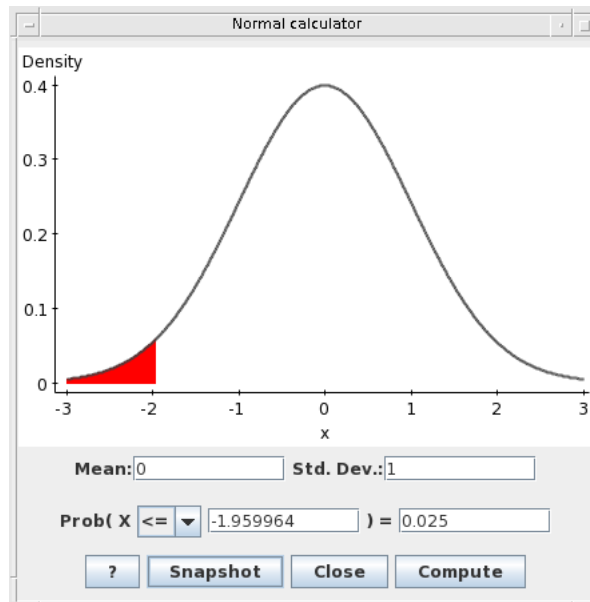


Above we 'see' what we mean by $P(X \geq 11.17) = 0.35$.

Question 9 (inverse transforms)

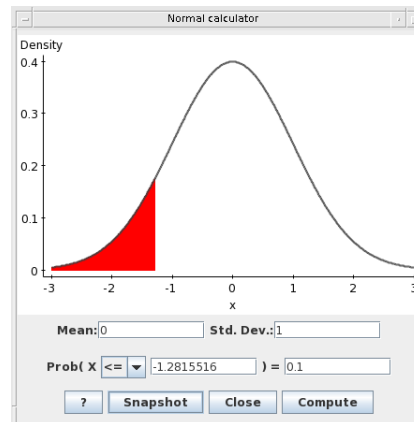
- Suppose $Z \sim N(0, 1)$. Find the x such that $P(-x \leq Z \leq x) = 0.95$.
- **Solution:** We know that $P(-x \leq Z \leq x) = P(Z \leq x) - P(Z \leq -x)$.
- Now we know that $P(Z \leq -x) = 1 - P(Z > -x)$.
- Now due to symmetry of the normal density we have $P(Z > -x) = P(Z \leq x)$.
- Altogether this gives $P(-x \leq Z \leq x) = P(Z \leq x) - (1 - P(Z \leq -x)) = 2P(Z \leq x) - 1 = 0.95$.
- This means $P(Z \leq x) = 1.95/2 = 0.975$. Looking up in the standard normal tables we have $P(Z \leq 1.96) = 0.975$.

- Looking up 0.975 in the table we have, $P(Z \leq 1.96) = 0.975$, hence $x = 1.96$.
- Altogether this means $P(-1.96 < Z < 1.96) = 0.975$.



Question 10 (inverse transforms)

- Suppose $X \sim N(5, 2)$ (mean 5 and standard deviation 2). Find an interval **centered** about the mean such that 80% of the population lie in that interval. Another, way of asking the same question, is find the x such that $P(5 - x \leq X \leq 5 + x) = 0.8$.
- Hint: You can use the picture:



Solution 10

- Since the interval is centered about the mean, the area (probability) to the left and right of $5-x$ and $5+x$ should be 0.1 (10%). Looking up 0.1 in the z-tables gives 1.28. This means that the boundary of the interval should be 1.28 standard deviations from the mean. Thus $x = 1.28 \times 2$ and in the interval is

$$[5 - 1.28 \times 2, 5 + 1.28 \times 2].$$

Question 11

- Suppose that it is known that women's heights roughly follow a normal distribution with $N(66, 3)$ (mean height 66 and standard deviation 3).
- Reality check We say roughly, because a normal distribution means the random variable (height in this example) can take any value from negative infinite to positive infinite, but I don't know that many women who have negative height. Thus the tails of the distribution of heights (this means the extremes - the probabilities of very small heights and very large heights are unlikely to be close to the probabilities of a normal, just something to keep in mind).
- Calculate an interval centered about the mean, where roughly 95% of women's heights should lie (in other words 95% of the population should lie in this interval).

Solution 11

- It is easiest to answer this question with a plot of the normal curve.
- We need to find an interval, centered about the mean $[66 - t, 66 + t]$, such that there is a 95% chance a women's height will lie there.
- In other words, how many standard deviations (how many z-transforms) from the mean do we have to be such that 95% of the heights are within that interval.
- To do this, we transform (via shifting and squidging) this interval to a standard normal, this gives $[-t/3, t/3]$.
- The interval $[-t/3, t/3]$ lies on the standard normal distribution and there is 95% chance a standard normal variable lies in this interval. By

using our solution at the end of Lecture 9, we know that by looking up tables $P(-1.96 \leq Z < 1.96) = 0.95$ (since $P(Z \leq -1.96) = 0.025$). Matching $[-t/3, t/3]$ with $[-1.96, 1.96]$, this means that $t/3 = 1.96$, thus $t = 1.96 \times 3$. Therefore 95% population of female heights lie in the interval $[66 - 1.96 \times 3, 66 + 1.96 \times 3]$.