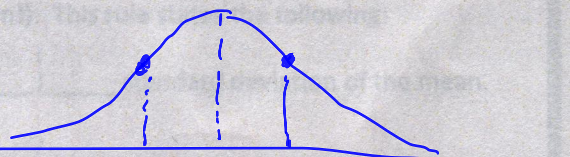
Packet 1: Normal Distributions

After completing this material, you should be able to:

- describe what a normal distribution is.
- use the Empirical rule to describe normal distributions.
- calculate a z-score and interpret its meaning.
- find probabilities (or percentages) of normal curves above a given observation, below a given observation, and between two observations.
- determine the value of a normal distribution corresponding to any given percentage (or probability).
- determine when observations are unusual based upon probabilities and explain the reasoning.



What exactly is a **normal distribution**? In the space below, briefly define a normal distribution and draw several normal curves.

A normal distribution is a symmetric bell-shaped distribution of values. A distribution describes how values vary over all possible values

A normal distribution is controlled by two features:

1. mean - center
2. standard deviation - width

The mean of the distribution controls the center of the distribution (see Figure 1 below).

The standard deviation of the distribution controls how spread out the curve is, which in turn controls its scale (see the Figure 2 below).

Notation Alert

(You *must* remember this notation!!)

mean = μ (mu)

std dev = σ (sigma)

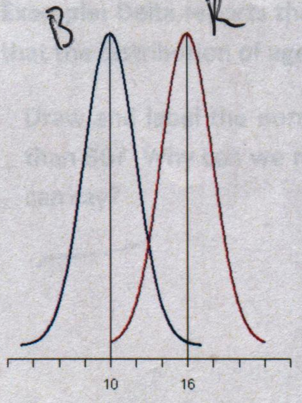


Figure 1.

Blue Curve: $\mu = 10$

Red Curve: $\mu = 16$

$\sigma_R = \sigma_B$

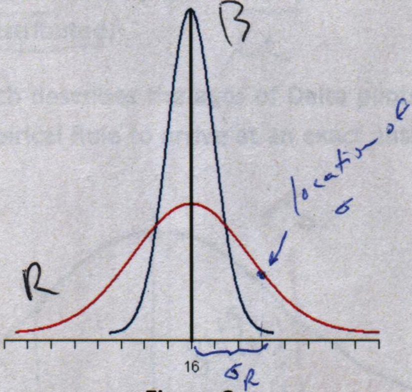


Figure 2.

Blue Curve: $\mu_B = 16$

Red Curve: $\mu_R = 16$

$\sigma_R > \sigma_B$

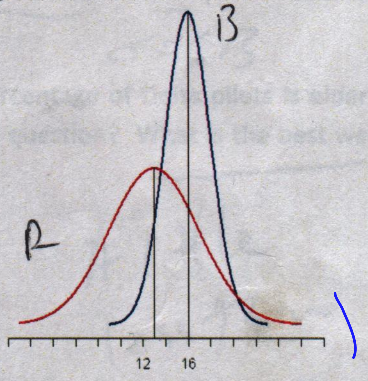


Figure 3.

Blue Curve: $\mu_B = 16$

Red Curve: $\mu_R = 13$

$\sigma_R > \sigma_B$

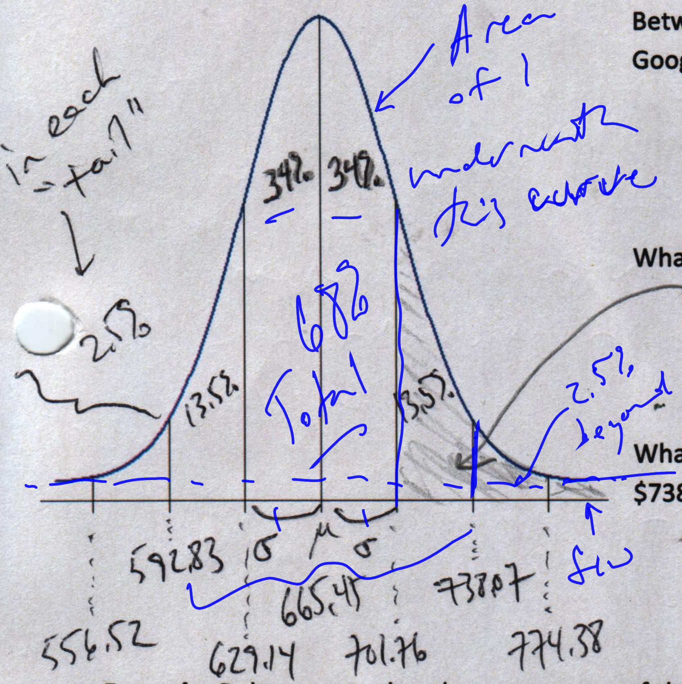
To help understand how normal distributions work, we can use the **Empirical Rule** (which is also referred to as the 68-95-99.7 Rule – I'll let you figure out where this name comes from!). This rule states the following:

- 68 % of the values in a normal distribution fall within 1 standard deviation of the mean.
- 95 % of the values in a normal distribution fall within 2 standard deviations of the mean.
- 99.7 % of the values in a normal distribution fall within 3 standard deviations of the mean.

Example: Last year, the opening price for one share of Google stock has been approximately normally distributed with a mean price of \$665.45 and a standard deviation of \$36.31.

μ σ

Draw and label the normal distribution corresponding to opening prices. Use the distribution to answer the following:



Between what two values do we expect the opening price of one share of Google stock to fall **95%** of the time?

Between $\mu+2\sigma$ and $\mu-2\sigma$ (592.83 to 738.07)

What percentage of the time was the opening price more than \$701.76?

About 16% (= 50% − 34% or = 13.5% + 2.5%)

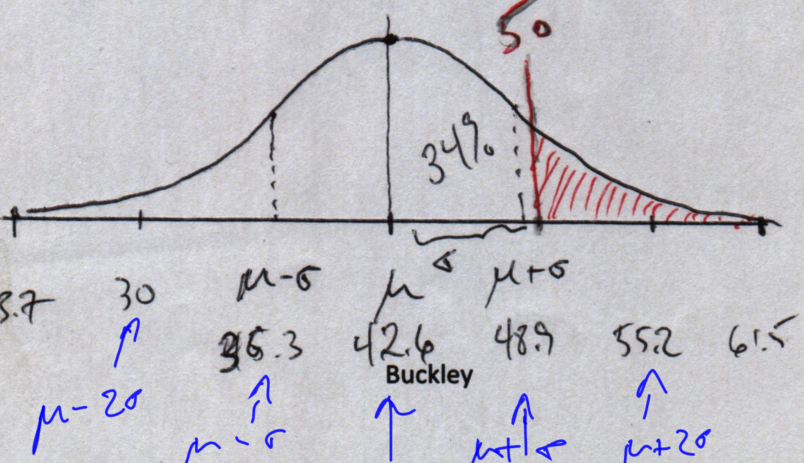
What percentage of the time is an opening price between \$629.14 and \$738.07?

81.5% = 34% + 34% + 13.5%

Example: Delta reports that the average age of their pilots is 42.6 years with a standard deviation of 6.3 years. Assume that the distribution of ages is normally distributed.

μ $\sigma = 6.3$

Draw and label the normal model which describes the ages of Delta pilots. What percentage of Delta pilots is older than 50? Why can we not use the Empirical Rule to arrive at an exact answer to this question? What is the best we can say?



A little less than 16%

Finding Normal Probabilities (also known as "working forwards")

In the Delta pilot example, we ran into a problem – the age we were interested in wasn't *exactly* 1, 2, or 3 standard deviations from the mean. This meant the Empirical Rule could no longer be used to find the percentage of the curve below the observation. We need a way to find **probabilities (or percentages)** associated with any observation.

In order to find probabilities, we first need a way to determine the exact number of standard deviations above or below the mean an observation falls. This process is known as **standardizing** the observation.

Formula Alert!!
This formula will be given on the formula sheet.

Let y represent an observation. y = 50

$$\boxed{z = \frac{y-\mu}{\sigma}}$$

standard normal: center $\mu = 0$, spread $\sigma = 1$

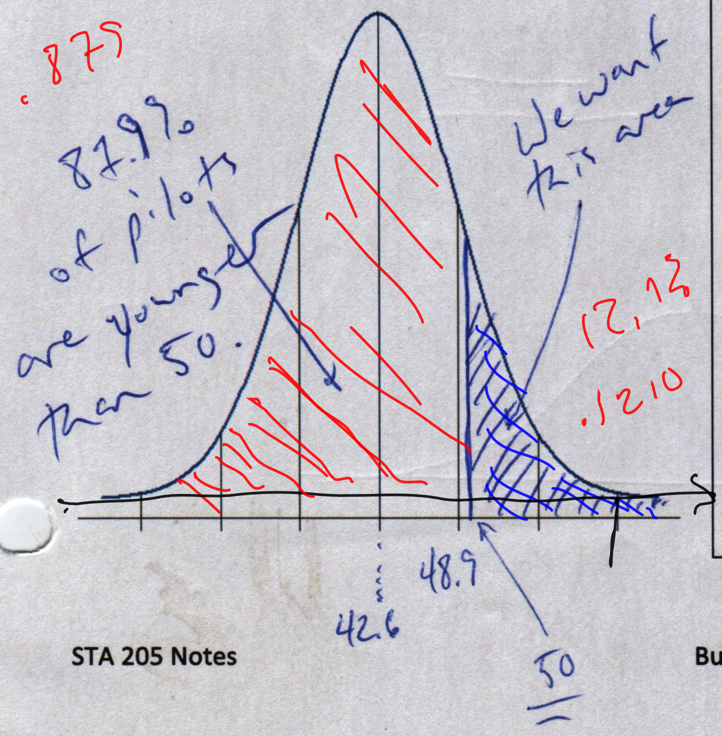
A z-score represents the # of standard deviations above or below the mean.
above $(z>0)$ below $(z<0)$

Let's return to the pilot example we looked at last class meeting. The age of Delta pilots is normally distributed with a mean of 42.6 years and a standard deviation of 6.3 years. We were interested in an age of 50 – calculate **and interpret** the z-score associated with this quantity.

$$z = \frac{y-\mu}{\sigma} = \frac{50-42.6}{6.3} = 1.17$$

So 50 is 1.17 std dev. above the mean 42.6. $(z>0)$

How will standardizing the observation by calculating its z-score help us find the percentage of pilots older than 50? To answer that question, let's fill in the normal model below:



What information would we like to know?

The area under the curve to the right of 50.

Using the normal table, what percentage of the curve is below an age of 50?

$z = 1.17 \rightarrow$ area of 0.8790 to the left (below)

$\boxed{87.9\%}$

Area to the right: $1 - 0.8790 = 0.1210$

$\boxed{12.1\%}$ (above)

In order to find shaded areas of the normal curve, we will need to use a **Normal probability table**. How does the table work?

1. Turn your observation into a z-score.
2. z-scores are given to the tenths place along the edge
3. columns represent the hundredths place.
4. At the intersection of the row + column are found the areas to the left of the observation.

Example: According to a January 2010 survey by Smith Travel Research, the daily rate for a luxury hotel in the United States was normally distributed with a mean of \$237.22 and a standard deviation of \$21.45.

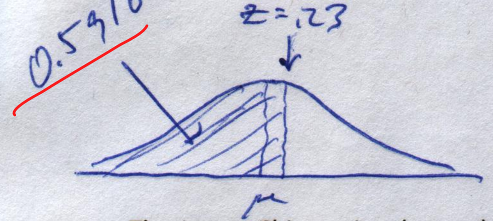
Hotel Giraffe is located on Park Avenue in New York City. Rooms in this hotel have an average rate of \$377.56. Standardize this observation **and interpret** the value.

$\mu = 237.22$, $\sigma = 21.45$, $y = 377.56$

$z = \frac{(377.56 - 237.22)}{21.45} = 6.54$; This price is 6.54 std dev above the mean! Extremely expensive!

The Cincinnatian is a 4-star hotel located in downtown Cincinnati. A standard hotel room in this hotel has a rate of \$242.10. What percentage of hotels has a rate less than this hotel?

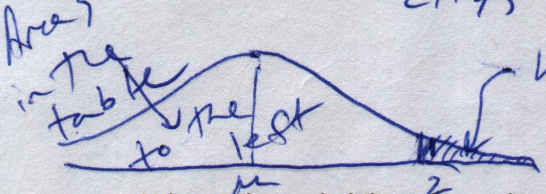
$z = \frac{y - \mu}{\sigma} = \frac{(242.10 - 237.22)}{21.45} = 0.23$ → area to the left of 0.5910



Convert to %: 59.10% of hotels have a rate less than this hotel.

The James Chicago is a luxury hotel located on the Magnificent Mile. A hotel room here can be reserved for \$279 per night. What proportion of hotels has a daily rate greater than this?

$z = \frac{(279 - 237.22)}{21.45} = 1.95$ → area to the left is 0.9744

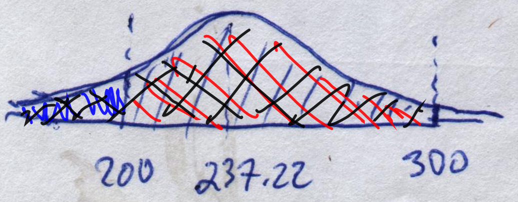


$= 1 - 0.9744 = \boxed{0.0256}$

What is the probability a luxury hotel's daily rate is between \$200 and \$300?

Two z-scores! $z_1 = \frac{200 - 237.22}{21.45} = -1.74$

$z_2 = \frac{300 - 237.22}{21.45} = 2.93$



Red & black − blue.

$z_1 \to 0.0409$

$z_2 \to 0.9983$

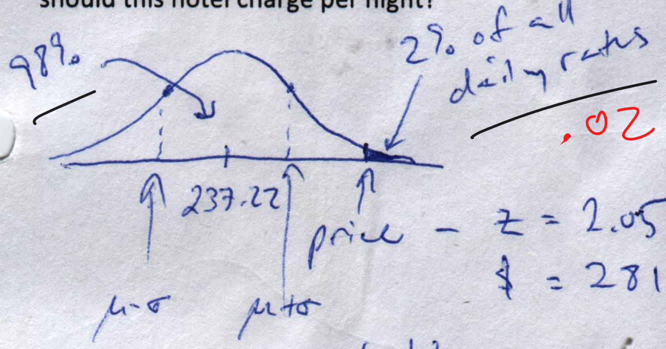
So we subtract to get the area between: $0.9983 - 0.0409 = 0.9574$

When working forwards to find probabilities, there are three types of probabilities which can be found:

Less than probabilities:	**Greater than probabilities:**	**Between probabilities:**
1. Calculate z $z = \frac{y - \mu}{\sigma}$ 2. Look up that z in the normal table! A	1. Calculate z 2. Look up the left probability in the table & subtract it from 1: 1-A	1. Calculate two z values, z_1 & z_2 2. Look up the areas in the table (A_1 & A_2) 3. Subtract the smaller from the larger $\max(A_1, A_2) - \min(A_1, A_2)$

Finding Observations (also known as "working backwards")

Example: According to a January 2010 survey by Smith Travel Research, the daily rate for a luxury hotel in the United States was normally distributed with a mean of $237.22 and a standard deviation of $21.45. Suppose a new luxury hotel is opening in Atlanta and the owners would like the daily rate to correspond to the *top 2%* of all daily rates. How much should this hotel charge per night?



.02

price - z = 2.05
$ = 281.19

How does this question differ from those we have considered previously?

Before:
obs. → z → area
Now
area → z → obs.

Find z in the table corresponding to .98 - closest is z = 2.05.

$z = \frac{y - \mu}{\sigma} \rightarrow \boxed{y = z\sigma + \mu}$

$y = 2.05 * 21.45 + 237.22$
$= \boxed{\$281.19}$

When working backwards to find the value of an observation from a normal distribution, use the following steps:

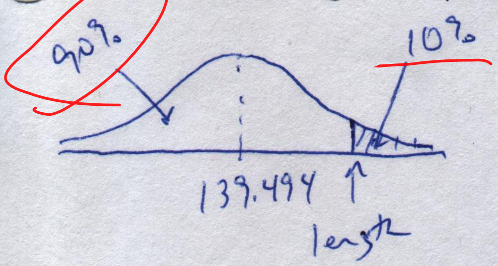
1. Draw the area in the normal distribution
2. Phrase it in terms of "area to the left"
3. Search for that area in the normal table, & take the closest corresponding z-score.
4. Solve for y: $\boxed{y = z\sigma + \mu}$

Formula Alert!! This formula will be given on the formula sheet.

Example: Assume that the lengths of the Best Picture Oscar winners are normally distributed with a mean of 139.494 minutes and a standard deviation of 31.656 minutes.

$\sigma = 31.656$ $\mu = 139.494$

What movie length corresponds to the *longest* 10% of all Best Picture Oscar winners?



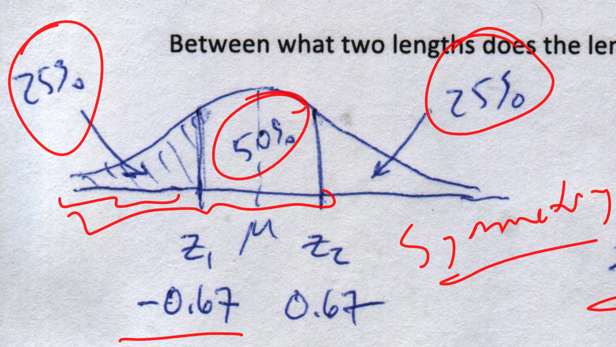
Find z corresponding to 90% (.9000)

$\boxed{z = 1.28}$

$y = z\sigma + \mu = 1.28 \cdot 31.656 + 139.494$

$= 180.014$ minutes About 3 hrs.

Between what two lengths does the length of the middle 50% of all Best Picture Oscar winners fall?



z_1 corresponds to an area of .2500 from the table, $z_1 = -0.67$

z_2 is 0.67 (by symmetry!)

Between

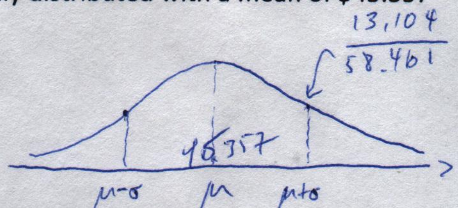
$y_1 = -0.67 * 31.656 + 139.494 = 118.284$ minutes

$y_2 = 0.67 * 31.656 + 139.494 = 160.704$ minutes

Let's tie everything we've learned about normal distributions together with the following **video example**:

Suppose that the cost for a one-day ticket to a theme park is approximately normally distributed with a mean of \$45.357 and a standard deviation of \$13.104.

Draw and label the normal model which describes theme park admissions.



If a newspaper wants to report an interval where 95% of all admission costs will fall, what interval should be reported?

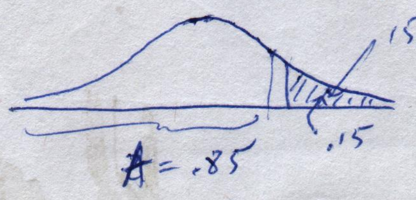
Probably centered – but there are certainly two others of interest 95% > low, + 95% < high. $\mu \pm 2\sigma$

Suppose that you are creating a budget for tickets to a theme park this year, and you decide that the most you will spend on a ticket is \$60. What is the probability that the park you choose to attend has an admission cost greater than this amount?

A little less than .16 (16%).

$z = 1.12 \rightarrow .8686$ $\therefore$ $P(\text{cost} > \$60) = 1 - .8686 = .1314$

Back to the budget ... if the amount you set aside has to cover the cost to all but the top 15% of theme park admissions, how much money should you set aside? How much money should be budgeted for four tickets?



$z\sigma + \mu = y$

$1.035 * 13.104 + \cancel{58.461}\ 45.357 = 58.92$

$z = 1.035$

$58.92 \rightarrow 4 * 58.92 = \$235.68$