

Lecture 10

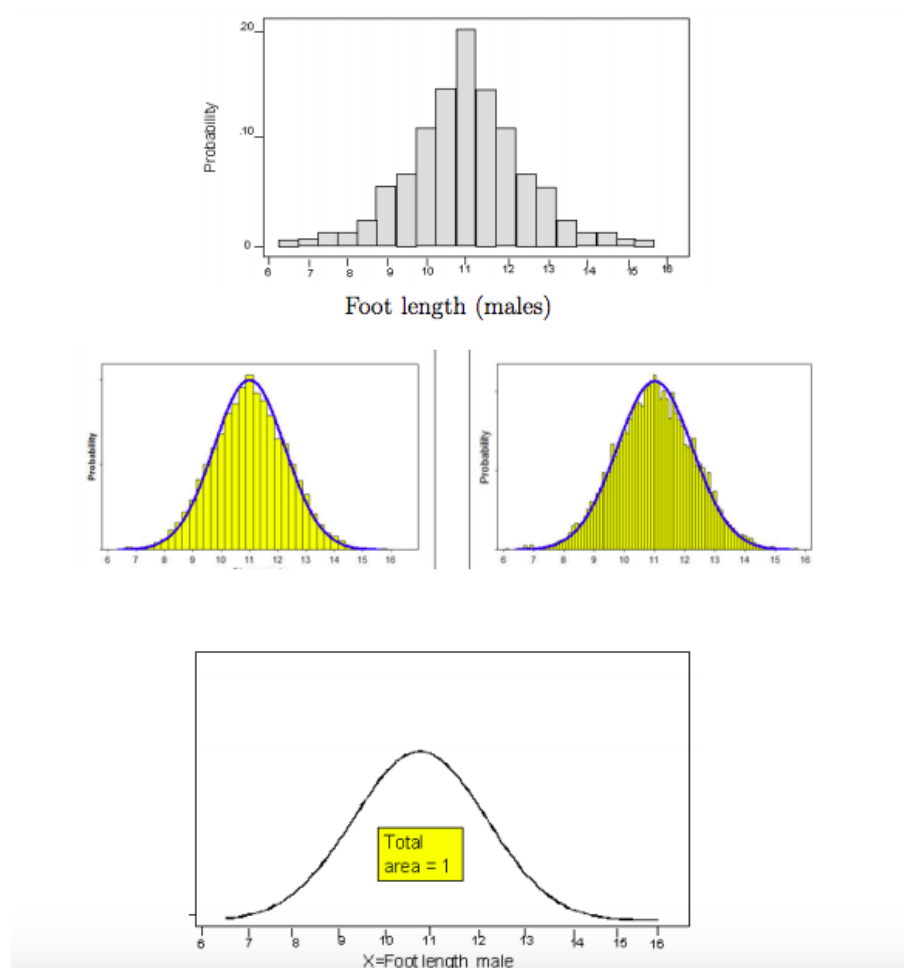
Continuous Random Variables

Continuous random variable →

When the random variable is continuous, the intervals of values for the bars of the probability histogram become very narrow.

The key here is that we want each bin to only have 1 possible value in a probability histogram.

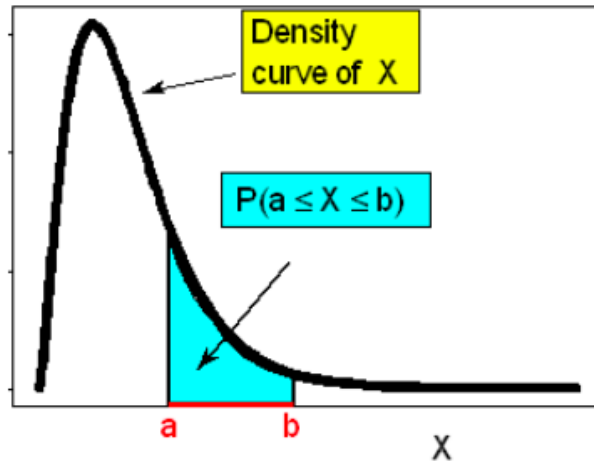
As the number of intervals increase with more possible values in the interval, the bin width narrows, and the shape of the histogram gradually approaches **a smooth curve**.



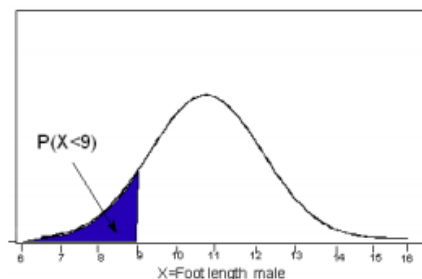
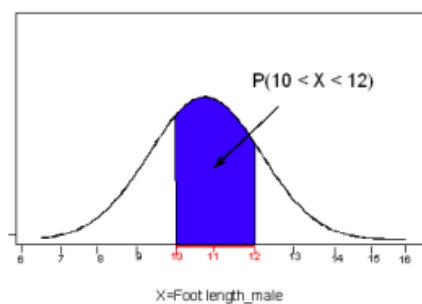
This smoothed curve is called a

Density curves portray the probability distributions of continuous variables.

So, how can we use the density curve to find probabilities? We take the area under the curve:



Based on the interval that we are interested in (from an inequality), we find the probability that X takes a value in that interval by finding the area from the density curve down to the x-axis.



Let's take a look now at 2 special cases of continuous random variables, the Uniform and Normal Distributions...

The Uniform Distribution

We first saw the term uniform when we discussed modality and peakedness. A dataset with 1 mode was termed unimodal, a dataset with 2 modes was termed bimodal, and a dataset with no mode is termed *uniform*.

Now, let's shift into the world of random variables. When a random variable X is said to have the uniform distribution, we use the notation:

Visually:

When a RV X is defined as $X \sim U(a, b)$, we can define the distribution formulaically:

Intuitively:

Example – Weight Gain

a) What is the probability that a person will gain between 10 and 20 pounds during the winter months?

b) What is the probability that a person will gain between 5 and 25 pounds during the winter months?

c) What is the probability that a person will gain between 0 and 30 pounds during the winter months?

Of course, as with every distribution, we can describe the center and spread.

For a RV X such that $X \sim U(a, b)$, we can formulaically define the *expectation* of X as

and the *standard deviation* of X as

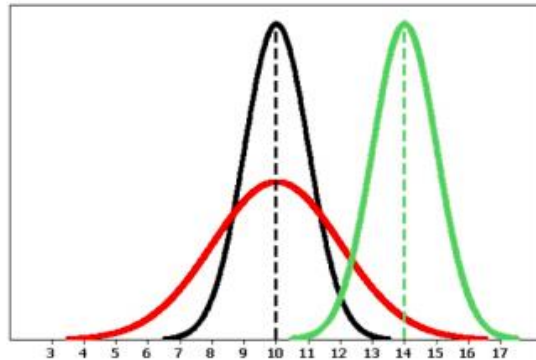
Example – Weight Gain (cont.)

Find the expected amount of pounds gained during the winter months along with the standard deviation.

The Normal Distribution

The Normal Distribution is Bell-shaped →

Consider the 3 following density curves for 3 different Normal Distributions:



What do we notice about the black and red distributions?

So, the _____ distribution has a larger standard deviation.

Why?

What about the black and green distributions?

So, the _____ distribution has a larger mean.

Why?

So, knowing only 1 does not fully specify a Normal distribution → we need both!

In practice, if we know the distribution of the random variable X is Normal, we use the notation

If we know that $X \sim N(\mu, \sigma^2)$, we can define the distribution formulaically as

So, how can we use this to find probabilities?

Well, in the beginning of the course, we discussed the Empirical Rule that stated:

Example – iPhone Area

Let the RV X represent the area of an iPhone. Suppose we know that the area of an iPhone follows a bell-shaped curve with a mean of 8 square inches and a standard deviation of 0.25 square inches.

- a) Draw the situation described above.

- What is the probability that a randomly chosen iPhone has an area between 7.75 and 8.25 square inches?
- What proportion of iPhones have an area above 8.25 square inches?
- What is the probability that a randomly chosen iPhone has an area above 7.75 square inches?
- 2.5% of iPhones have an area above what value?
- What proportion of iPhones have an area below 8.30 square inches?

The Normal Table

TABLE A Standard Normal Cumulative Probabilities (continued)										
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

What does this table find?

Example – Using the Normal Table

Consider the above table to answer the following questions:

- In any normal distribution, what is the probability of getting a value that is less than .5 standard deviations above the mean (i.e falls below $\mu + .5\sigma$)?

- b) In any normal distribution, the probability of getting a value that is less than 1.35 standard deviations above the mean (i.e falls below $\mu + 1.35\sigma$) is:

Example – iPhone Area (cont.)

Recall that the RV X is bell-shaped with a mean of 8 square inches and standard deviation of 0.25 square inches.

- a) What proportion of iPhones have an area below 8.30 square inches?

We need a way to determine the number of standard deviations each point is away from the mean!

Z-Score

Intuitively:

Formulaically:

This process of calculating a z-score for a given value of x is called **standardization**.

Visually, here is what is happening:

- c) What proportion of iPhones have an area between 7.70 square inches and 8.30 square inches?

- d) 20% of iPhones have an area below what value? (In other words, what is the **20th percentile** of the iPhone area distribution?)
- e) In order for Steve Jobs not to throw the iPhone into an Aquarium, the iPhone has to be in the highest 2% of the distribution. What is the actual area above which the iPhone is “safe”?
(**OR** 2% of iPhones have an area above what value?
OR What is the 98th percentile of the iPhone area distribution?)

Let's look at another application of z-scores...

Example – Comparing SAT and ACT Scores

When you applied to college, you scored 650 on the SAT exam, which had a mean of 500 and standard deviation of 100. Your friend took the comparable ACT, scoring 30. For that year, the ACT had a mean score of 21 with standard deviation of 4.7. How can we compare these scores to tell who did better?

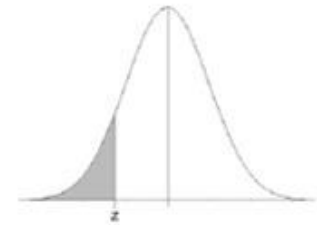
Visually:

Mathematically:

Note: If you get a z-score below -3.4, then the corresponding probability is assumed to be 0. Likewise, if you get a z-score above 3.4, then the corresponding probability is assumed to be 1.

PROBABILITY ENDS

Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-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
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

A normal distribution curve is shown with a vertical line at the center representing the mean. A vertical line is drawn to the right of the center, labeled z_1 on the horizontal axis. The area under the curve to the left of z_1 is shaded gray.

$$p(z \leq z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2} dz$$

[illegible]