Mathematics 136 – Calculus 2
Discussion Day – Distribution Functions
March 16, 2020

Background

We now look at applications of integration describing the *distribution of various quantities in populations.* This kind of approach gives the foundations for *mathematical probability and statistics.* Statistical reasoning is used to analyze data from experiments in almost all of the physical and social sciences, and is also used extensively in business and political decision-making. Over the next several days we will look at a few of the basic ideas here and see how calculus enters into these questions.

Our starting point will be today's discussion, where we will look at the distribution of heights in a hypothetical college class. Here is the heights data set (all heights in inches):

Women ($N = 16$): 71, 69, 68, 67, 67, 66, 66, 66, 66, 64, 63, 63, 63, 63, 63, 62

Men ($N = 10$): 72, 72, 72, 71, 70, 70, 69, 68, 68, 67

One way to represent the distribution of values in numerical data like this is a graph called a *relative frequency histogram.* To make a relative frequency histogram, we:

- First decide on a finite subdivision of the range of possible values of the quantity we are measuring. For instance for the *women's heights* in the class, we can use intervals of length $\Delta h = 1$ inch on the range $61.5 \le h \le 71.5$, *centered at the integer values.*

- On each interval, we plot a rectangle whose height is *the fraction of the total population* whose measurement (here the height) falls into that range. For instance, 5 of the women have height 63 inches, so the rectangle in the relative frequency histogram for the women's heights centered at $h = 63$ has length $5/16 = .3125$.

- We do this for all the intervals and assemble the rectangles (see plot on next page).

- Note that this makes the sum of the areas of all of the rectangles *equal to one.*

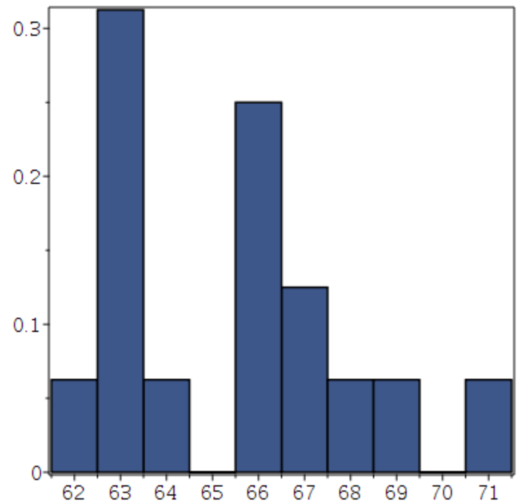

Figure 1: Relative frequency histogram for heights of the women.

Discussion Questions

A) Using the data given above, construct relative frequency histograms for the men's heights and then for the whole class (women and men). First decide on appropriate ranges of h-values for each. Your histograms should include everyone in the appropriate categories, of course!

Solution: See the graphs on the next page.

B) Using these histograms, determine the fraction of the whole class that had heights between 64 and 69 inches. Determine the fraction of the women that had heights 63 inches or less. Determine the fraction of the men that had heights 70 inches or more.

Solution: $13/26 = .5$ of the class had heights between 64 and 69 inches (inclusive). This is the sum of the heights of the bars for heights $64, 65, 66, 67, 68, 69$. $6/16 = .375$ of the women had heights 63 inches or less. $6/10 = .6$ of the men had heights 70 inches or more.

C) Thinking about what you did in question B, describe in general how the fraction of a population with heights h in a range $a \leq h \leq b$, or $h \leq b$, or $h \geq a$ relates to *areas* of rectangles in that histogram.

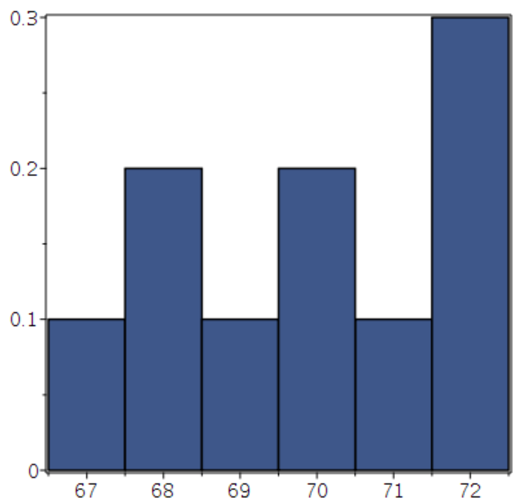

Figure 2: Relative frequency histogram for heights of the men.

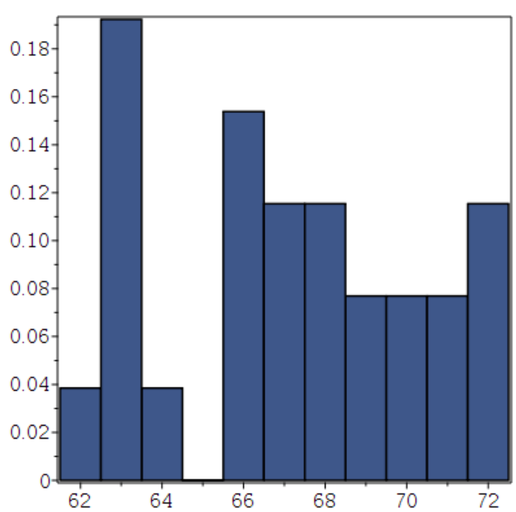

Figure 3: Relative frequency histogram for heights of the whole class.

Solution: You would add the heights of the rectangles for h in any one of these ranges to get the fraction of the population.

D) It might not be the first way you would describe what is happening with the histogram, but we can think of the histogram as a *graph* of a *"height density function"* $p(h)$. (Think of the "tops" of the rectangles – horizontal line segments – as portions of the graph.)

1) Give "split-domain" formula(s) for the height density function for the men in the class.

   Solution: The height density function for the men is

$$p(h) = \begin{cases} .1 & 66.5 < h < 67.5 \\ .2 & 67.5 < h < 68.5 \\ .1 & 68.5 < h < 69.5 \\ .2 & 69.5 < h < 70.5 \\ .1 & 70.5 < h < 71.5 \\ .3 & 71.5 < h < 72.5 \end{cases}$$

   and 0 outside the interval $[66.5, 72.5]$. The endpoints of the intervals could also be included as the upper limits of these bins.

2) Taking your answer from question C one step farther, describe how you could compute the fraction of a population with heights h in a range $a \leq h \leq b$, or $h \leq b$, or $h \geq a$ using *integrals*.

   Solution: These would be computed by

$$\int_a^b p(h)\ dh, \quad \int_{-\infty}^b p(h)\ dh, \quad \int_a^\infty p(h)\ dh$$

   Note that for the second integral, the lower limit could also be given as 66.5, since $p(h) = 0$ for all $h < 66.5$. Similarly, the upper limit in the third integral could be written as 72.5.

If we had a large sample of people from a population (much larger than the 26 people in this class), and we measured the heights more precisely (say to the 1/10) of an inch, then the height relatively frequency histogram would look "smoother", with lots more thinner rectangles.

Mathematically speaking, we can even imagine taking a limit with rectangles with $\Delta h \to 0$, and when we did this, we would obtain the graph of what is called the population *height density function* $p(h)$.

E) How would you express the fraction of the population with heights $a \le h \le b$ as an integral of $p(h)$? Why must it be true that $p(h) \ge 0$ for all h, and $p(h) = 0$ for all $h < 0$. What would be true about the value of

$$\int_{-\infty}^{\infty} p(h)\ dh = \int_{0}^{\infty} p(h)\ dh?$$

Solution: The fraction of the population with $a \le h \le b$ would be computed by

$$\int_{a}^{b} p(h)\ dh.$$

This last integral should equal 1, since the whole population would have heights in the range $[0, \infty)$.

Assignment

Writeups due Wednesday, March 18.