Mathematics 132 – Calculus for Physical and Life Sciences
Discussion 4 – Distribution Functions
March 21, 2005

Background

We have now studied computing the mass and center of mass of physical objects with given mass density functions. Many of the same ideas are also used to study the *distribution of various quantities in populations* and this kind of approach gives the foundations for *mathematical 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 three 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 our class (using the data collected at the start of class last Friday). Here is the height data set:

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 our 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 WW 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:

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!

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.

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.

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.
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*.

If we had a large sample of people from a population (much larger than the 30-odd people in our 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 $a \leq h \leq b$ as an integral of $p(h)$? Why must it be true that $p(h) \geq 0$ for all h, and $p(h) = 0$ for all $h < 0$. What would be true about the value of

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

Assignment

Writeups due Tuesday, March 22.