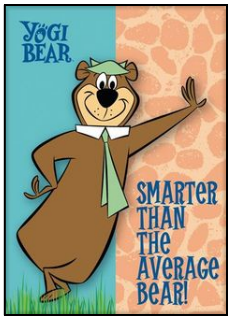
What's the Average?

by Stu Schwartz



Over the years, I taught both AP calculus and AP statistics. While many of the same students took both courses, rarely was there a problem with students confusing ideas from both courses. In fact, rarely was calculus mentioned in statistics class and vice versa. It was almost as if calculus skills and statistics skills came from a different part of the brain.

The one concept that is common to both courses is "average" or "mean." In calculus we have the average value of a function and the average rate of change of a function. In statistics, we have the arithmetic mean or average. Are there occasions when these concepts can be confused? Let's examine three problems.

Problem 1: Katie has been given a job of babysitting. She goes to the Hays home to babysit their son Micah. They will be gone from 6 PM to 10 PM and Mr. Hays agrees to pay Katie \$40. Following are the times that Katie got paid along with the amount. What was the average payment to Katie?

7 PM	8 PM	9 PM	10 PM
\$10	\$10	\$10	\$10

Problem 2: Jerry had an exam every week for 5 weeks as shown in the table below. What is his average over the 5-week period?

Week 1	Week 2	Week 3	Week 4	Week 5
55%	60%	65%	70%	95%

Problem 3: Alice's restaurant has the following number of tables in use over a 4-hour period. What is the average number of tables that were in use in the restaurant over the 4-hour period?

5 PM	6 PM	7 PM	8 PM	9 PM
50	40	30	40	50

The correct answers are: Problem 1: \$10   Problem 2: 69%   Problem 3: 40

My guess is that you had the first 2 problems correct and the 3rd problem incorrect. In this paper, we will see that the concept of average isn't quite as simple as you may think it is!

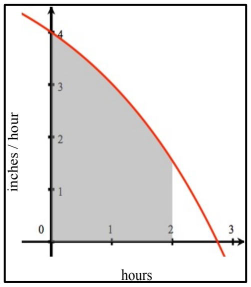
First, a review. In our statistics course, we defined the average or "arithmetic mean" as the sum of n distinct data points divided by n. Average = $\dfrac{(x_1 + x_2 + x_3 + ... + x_n)}{n}$.

In our calculus course, we dealt with average in two ways: the average rate of change and the average value of a function.

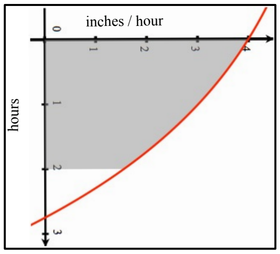
The average rate of change of a function f over time $[t_1, t_2]$ is given as $\dfrac{f(t_2) - f(t_1)}{t_2 - t_1}$.

The average value of a continuous function *f* over the interval [*a*, *b*] as $F_{avg} = \dfrac{\int_a^b f(x)\,dx}{b-a}$. Typically though, *f* is a function of time *t* and the average value of the function is taken between two times. So $F_{avg} = \dfrac{\int_{t_1}^{t_2} f(t)\,dt}{t_1 - t_2}$.

For instance, let *f* represent the rate of snowfall measured in inches per hour and *t* is measured in hours, as shown by the graph. Then $\int_{t_1}^{t_2} f(t)\,dt$ will be measured in $\dfrac{\text{inches}}{\text{hour}} \cdot \text{hours} = \text{inches}$ and F_{avg} will be measured in $\dfrac{\text{inches}}{\text{hour}}$. So if $\int_0^2 f(t)\,dt = 5.86$ inches, the average rate of snowfall was $\dfrac{5.86}{2} = 2.93$ inches per hour.

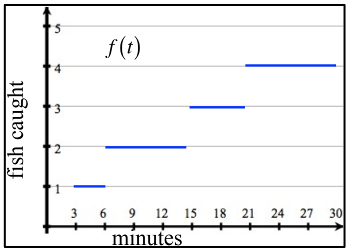


Suppose we switch the axes, with the horizontal axis measured in inches per hour and the vertical axis measured in hours. The function now will appear differently as it will measure time as a function of rate of snow instead of the usual rate of snow as a function of time. But the definite integral will still be the same 5.86 inches – the area under the curve. And the average rate of snow will remain the same too as 2.93 inches per hour. Keep this in mind later on.

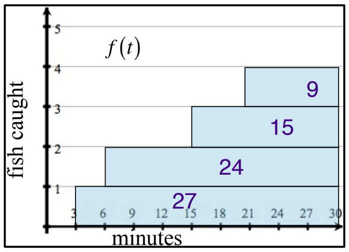


Suppose that that Matthew goes fishing for a period of 30 minutes. He puts his caught fish in a bucket. He catches a fish at the 3-minute mark, the 6-minute mark, the 15-minute mark, and the 21-minute mark.

In the graph, $f(t)$ represents the number of fish that are in the bucket at *t* minutes. So $\int_0^{30} f(t)\,dt$ is measured in fish-minutes. Fish-minutes? What in the world are fish-minutes? We will see that fish-minutes can be explained by the actual calculation of the definite integral. Rather than using traditional vertical rectangles, we will compute our area using horizontal rectangles.

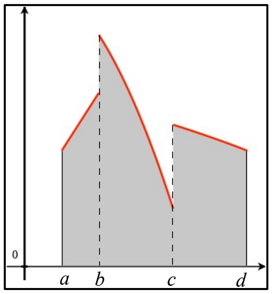


As shown by the graph, the first fish is caught at 3 minutes, so it was in the bucket for 27 minutes. The second fish is caught at the 6-minute mark and it was in the bucket for 24 minutes. The 3rd fish was caught at the 15-minute mark, so it was in the bucket for 15 minutes. And the 4th fish was caught at the 21-minute mark, so it was in the bucket for 9 minutes. So $\int_0^{30} f(t)\,dt = 75$. This represents the total number of minutes that fish were in the bucket. Hence: fish-minutes.



If we divide our 75 fish-minutes by 30 minutes, we end up with 2.5 fish. This is the average value of $f(t)$ over 30 minutes. It represents the average number of fish in the bucket over the 30-minute period. On average, the bucket contained 2.5 fish.

One difference between this problem and the snow problem is that the number of fish that are caught is not a continuous function. We jump from one fish caught to two fish caught at an instant in time which means that the function transition is not smooth. However, this is not an issue when calculating the definite integral of a non-continuous function. Even in the non-continuous function to the right, there are two jump discontinuities, but we can still calculate the area under the curve using $\int_a^d f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx + \int_c^d f(x)\,dx$.



Armed with that information, let's tackle problem 1. Katie has been given a job of babysitting. She goes to the Hays home to babysit their son Micah. They will be gone from 6 PM to 10 PM and Mr. Hays agrees to pay Katie $40. Suppose they don't quite trust each other. Mr. Hays doesn't want to give her the money upfront and Katie doesn't want to wait until they get home to receive her money. So Mr. Hays agrees to send money to her account at the end of each hour. Following are the times that Katie got paid along with the amount. What was the average payment to Katie?

7 PM	8 PM	9 PM	10 PM
$10	$10	$10	$10

Is this a calculus problem or a statistics problem? Yes, there is a time element in it, perhaps making you think that it is a calculus problem. But it isn't. These are 4 distinct payments and the time issue doesn't enter into it. The calculation is: $\text{Average} = \frac{10+10+10+10}{4} = \$10$. The 4 doesn't have anything to do with the 4-hour time gap between 6 PM and 10 PM; it's there because there are 4 data values.

Where it gets confusing is that we can also think of this as an average rate of change problem. At time $t = 0$, Katie has no money and at time $t = 4$, Katie has $40. 4 hours has elapsed. So her average rate of change is $\frac{40-0}{4-0} = \frac{\$10}{\text{hour}}$. Don't confuse that with the $10 we got above. The statistics gave us the <u>average payment</u> of $10. The calculus gave us the <u>average rate of change</u> of $10 per hour. They are not the same.

If the payments were in the table below, the average payment would be: $\text{Average} = \frac{8(5)}{8} = \$5$.

6:30 PM	7 PM	7:30 PM	8 PM	8:30 PM	9:00 PM	9:30 PM	10:00 PM
$5	$5	$5	$5	$5	$5	$5	$5

However, the average rate of change is the same. Katie still had no money at time $t = 0$, $40 at time $t = 4$ so the average rate of change of payment is still $10 an hour.

If the payments were in the table below, the average payment would be: $\text{Average payment} = \frac{40}{8} = \$5$. But the average rate of change of payment is still $10 an hour.

6:30 PM	7 PM	7:30 PM	8 PM	8:30 PM	9:00 PM	9:30 PM	10:00 PM
$15	$5	$0	$10	$0	$5	$5	$0

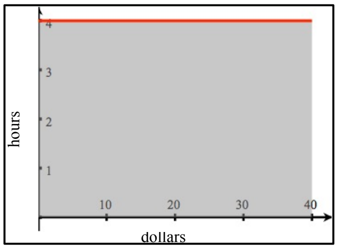
In general, since Katie always ends up with $40, we can say that her average payment is $\dfrac{40}{\text{number of payments}}$. So if Katie got paid an equal amount every 10 minutes her average payment would be $1.67 but her average rate of change of payment is $10 an hour. It doesn't matter how many payments Katie got paid, her average rate of change of payment is $10 an hour.

However, let's tweak the problem. Let us say that the definition of being paid is Katie actually having the money in her account. And let's also assume that when Katie starts work, she has no money in her account. We are interested in the average amount of money in her account over the 4 hours. Does this change the problem from a statistics problem to a calculus problem?

Since time is an issue now, the problem changes to a calculus problem and the average amount of money in her account will be $\dfrac{\int_0^4 f(t)\,dt}{4}$. Our job will be defining $f(t)$ and calculating $\int_0^4 f(t)\,dt$.

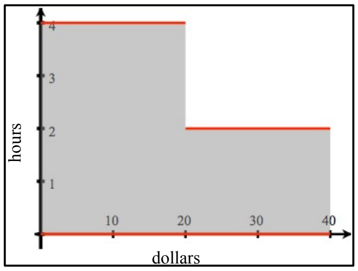
First, let's suppose that Mr. Hays agrees to pay Katie up-front, meaning that he pays the $40 at the start of the 4-hour period. Let's examine this from a graphical point of view. Since the $40 is in her account at all times for $t = 0$ to $t = 4$, the line $f(t) = 40$, measured in dollars represents the dollars in her account. So $\dfrac{\int_0^4 f(t)\,dt}{4}$ is measured in dollar-hours and this value is simply the area of the rectangle which is 160. If we divide this sum by 4 hours, we end up with the average value of *f* over the 4-hour period = 40. And this makes total sense. Since over the 4-hour period, there was always $40 in her account, it makes sense that she averaged $40 in her account over the 4-hours.

As before, let's switch the axes – making the *x* axis in dollars and the *y*-axis in hours. The reason for this will be clear later. The graph is to the right and the integral is still defined as the area under the curve which is 160 dollar-hours, and the average amount in her account still takes that sum and divides it by 4 hours to get $40. The calculation is the same if we switch the axes.

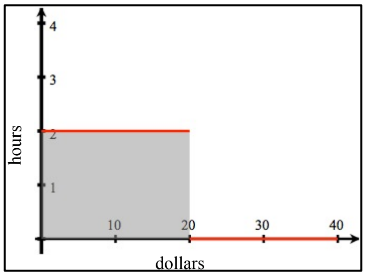


Now, let's suppose that Mr. Hays decides that he will pay Katie at the end of the 4-hour period. How does that change her average, as we have defined it – the total number of dollars in her account over the 4-hour period divided by 4? There is no need to graph it. Since her account has no money for the entire 4 hours, except for the magical instant in time that the $40 appears, the definite integral (area under the curve will be 0 dollar-hours and the average is 0 dollars. And this again makes total sense as for the entire 4 hours (except at $t = 4$), she was broke.

Next, let's suppose that Mr. Hays decided to pay Katie at the beginning of the first 2-hour period and then at the beginning of the second 2-hour period. So she receives \$20 at the 0-hour mark and \$20 at the 2-hour mark. How does that change her average, as we have defined it? The graph to the right shows the situation with the axes reversed. Since she got paid \$20 at the start, she had \$20 in her account for the full 4 hours. Since she got paid \$20 at the 2-hour mark, she had another \$20 in her account for 2 hours. So the amount in her account is \$20 (4 + 2) = \$120. So the average amount in her account was \$30 over the four-hour period.

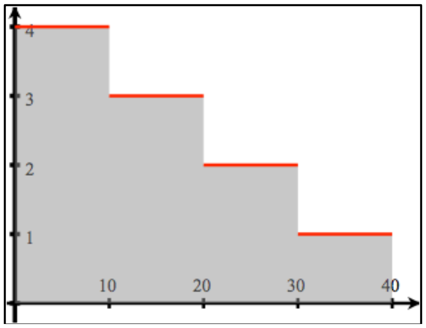


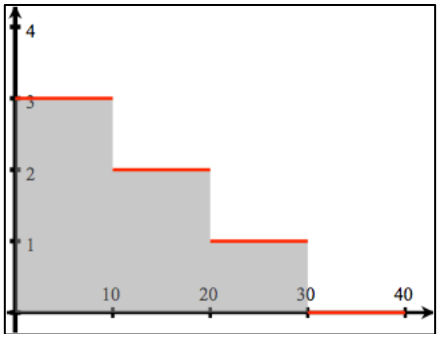
If Mr. Hays decided to pay Katie at the end of the first 2-hour period and then at the end of the second 2-hour period, the graph that describes this situation appears to the right, again with the axes reversed. Since she doesn't get paid her \$20 until the 2-hour mark, she only has that \$20 for 2-hours. And since she doesn't get her second \$20 until the 4-hour mark, she only has that \$20 for a split second in time. So the amount in her account is \$20 (2 + 0) = \$40. So the average amount in her account was \$10 over the four-hour period.



So it seems that the average amount in her account depends when she gets paid.

Let's look at what the graphs would look like if Mr. Hays decided to pay Katie every hour. On the left, she gets paid at the start of every hour and on the right, she gets paid at the end of every hour. Let's also concentrate on the definite integral, the area under the curve.

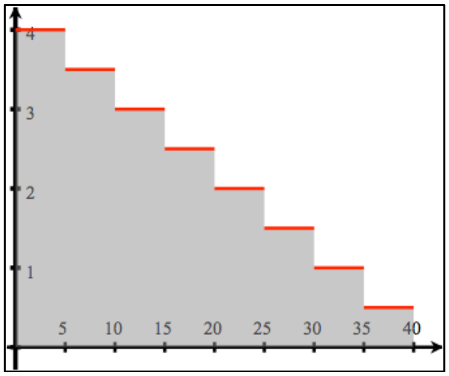


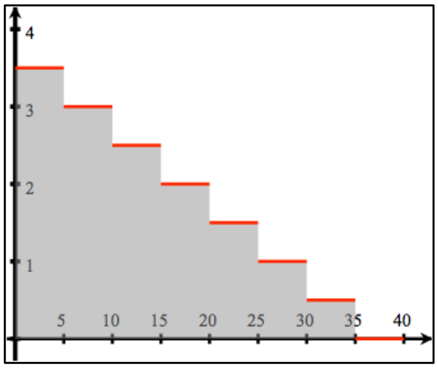


\$10 in the account for 4 hours
\$20 in the account for 3 hours
\$30 in the account for 2 hours
\$40 in the account for 1 hour
Area $= 10\left(4+3+2+1\right)=100$
Average in the account = \$25

\$10 in the account for 3 hours
\$20 in the account for 2 hours
\$30 in the account for 1 hours
\$40 in the account for 0 hour
Area $= 10\left(3+2+1+0\right)=60$
Average in the account = \$15

Let's look at what the graphs would look like if Mr. Hays decided to pay Katie every half hour. On the left, she gets paid at the start of every half hour period and on the right, she gets paid at the end of every half hour period. Let's also just concentrate on the definite integral, the area under the curve as we can easily compute the average amount of money in her account by dividing the integral by 4.





$5 in the account for 4 hours
$10 in the account for 3.5 hours
$15 in the account for 3 hours
$20 in the account for 2.5 hours
$25 in the account for 2 hours
$30 in the account for 1.5 hours
$35 in the account for 1 hours
$40 in the account for 0.5 hours
$\text{Area} = 5\left(4+3.5+3+2.5+2+1.5+1+0.5\right)=90$

$5 in the account for 3.5 hours
$10 in the account for 3 hours
$15 in the account for 2.5 hours
$20 in the account for 2 hours
$25 in the account for 1.5 hours
$30 in the account for 1 hours
$35 in the account for 0.5 hours
$40 in the account for 0.0 hours
$\text{Area} = 5\left(3.5+3+2.5+2+1.5+1+0.5+0\right)=70$

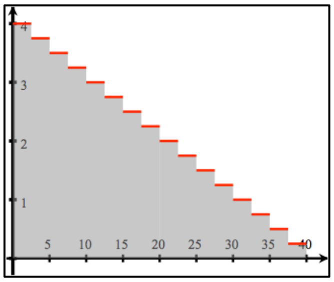
With the examples shown, it seems that paying Katie at the start of the time period is calculated using Left Riemann sums and that paying Katie at the end of the time period is calculated using Right Riemann sums. It also appears that the average of the left Riemann sums and right Riemann sums is 80. So for the remainder of this paper, we will use only left Riemann sums.

The question is: what is the limit of this definite integral as the number of times Katie gets paid approaches infinity? Put another way, we are interested in the limit of this definite integral if Katie gets paid continuously. So far, getting paid once gave her $160 dollar-hours. Getting paid twice gave her $120 dollar-hours. Getting paid 4 times gave her $100 dollar-hours and getting paid 8 times gave her $90 dollar-hours.

Look at the sequence, 160, 120, 100, 90. Do you see a pattern?

Notice that the difference between the first and 2nd number in the sequence is 40. Then 20. Then 10. It seems that each time, the difference is half the previous difference. So it would appear that the next number in the sequence should be 85.

So let’s check it out if Katie gets paid every quarter hour. Every quarter hour, she gets paid \$2.50.

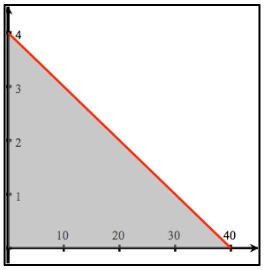


\$2.50 in the account for 4 hours
\$5 in the account for 3.75 hours
\$7.50 in the account for 3.5 hours
\$10 in the account for 3.25 hours
\$12.50 in the account for 3 hours
\$15 in the account for 2.75 hours
\$17.50 in the account for 2.5 hours
\$20 in the account for 2.25 hours
\$22.50 in the account for 2 hours
\$25 in the account for 1.75 hours
\$27.50 in the account for 1.5 hours
\$30 in the account for 1.25 hours
\$32.50 in the account for 1 hour
\$35 in the account for 0.75 hour
\$37.50 in the account for 0.5 hour
\$40 in the account for 0.25 hour

$$2.50(4+3.75+3.5+3+2.75+2.5+2.25+2+1.75+1.5+1.25+1+0.75+0.5+0.25)=85$$

The question is: As the number of payments that Mr. Hays makes approaches infinity, what value does our definite integral (and thus area under the curve) approach? That is, if Mr. Hays pays her continuously over the 4-hour period, what is the average amount of money that Katie will have in her account?

It might strike you that the more payments Mr. Hays makes to Katie, the more our piecewise function appears to be approaching a line. If so, our shape is approaching a right triangle with base 40 and height 4 as shown on the right. That would make the area 80. But how can we prove this?



We need to find an expression for the total number of dollar-hours that Katie is being paid in terms of the variable n. If we can do that, we can take $\lim\limits_{n\to\infty}\left[\sum\limits_{n=1}^{\infty}\left(\text{payment-dollars}_n\right)\right]$. We believe it approaches 80. Let’s see what we have so far:

n	Payment-dollars
1	$40(4+0)=160$
2	$20(4+2+0)=120$
3	$10(4+3+2+1+0)=100$
4	$5(4+3.5+3+2.5+2+1.5+1+0.5+0)=90$
5	$2.50(4+3.75+3.5+3+2.75+2.5+2.25+2+1.75+1.5+1.25+1+.75+.5+.25+0)=85$

If you look closely, you will begin to see a pattern here. The number in front represents the constant payment Mr. Hays makes to Katie at each payment time. The sum in the parentheses always starts at 4 and ends at 0. And as we go to the next row in the table, the increment between terms is halved. Let’s rewrite the table so that we can do some creative algebra to create our expression for any positive integer n.

n	Payment-dollars
1	$40(4)=160$
2	$20(4+2)=120$ $\frac{40}{2}\left[(2)(2+1)\right]=120$
3	$10(4+3+2+1)=100$ $\frac{40}{4}\left[(1)(4+3+2+1)\right]=100$
4	$5(4+3.5+3+2.5+2+1.5+1+0.5)=90$ $\frac{40}{8}\left[\left(\frac{1}{2}\right)(8+7+6+5+4+3+2+1)\right]=90$
5	$2.50(4+3.75+3.5+3+2.75+2.5+2.25+2+1.75+1.5+1.25+1+.75+.5+.25)=85$ $\frac{40}{16}\left[\frac{1}{4}(16+15+14+13+12+11+10+9+9+7+6+5+4+3+2+1)\right]=85$

If *n* were 6, according to our pattern, 160, 120, 100, 90, 85, …, we would expect the next number in the sequence to be 82.5. Let's see if we can generate it;

n	Payment-dollars
6	$1.25(4+3.875+3.75+3.25+3+...+0.75+0.5+0.25)=82.5$ $\frac{40}{32}\left[\frac{1}{8}(32+31+30+...+3+2+1)\right]=82.5$

If *n* were 7, according to our pattern, 160, 120, 100, 90, 85, 82.5, …, we would expect the next number in the sequence to be 81.25. Let's see if we can generate it;

n	Payment-dollars
7	$0.625(4+3.9375+3.875+3.8125+...+0.1875+0.125+0.0625)=81.25$ $\frac{40}{64}\left[\frac{1}{16}(64+63+62+...+3+2+1)\right]=81.25$

It should strike you that our expressions always contain the number 40 and also contains numbers which are powers of 2. Our expression will always be in the form of:

$$\boxed{\begin{gathered}\frac{40}{2^p}\left[\frac{1}{2^q}\left(2^p+\left[2^p-1\right]+\left[2^p-2\right]+...+3+2+1\right)\right]= \\ \frac{40}{2^{p+q}}\left[2^p+\left(2^p-1\right)+\left(2^p-2\right)+...+3+2+1\right]\end{gathered}}$$

Let us try to find p and q in terms of n.

n	p	q	$p+q$
1	0	-2	-2
2	1	-1	0
3	2	0	2
4	3	1	4
5	4	2	6
6	5	3	8
7	6	4	10

It appears that our p formula is $p=n-1$ and our q formula is $q=n-3$. So $p+q=2n-4$ which seems to be verified by the table.

So for any value n, our total number of payment-dollars is given by the formula:

$$\boxed{\frac{40}{2^{2n-4}}\left[2^p+\left(2^p-1\right)+\left(2^p-2\right)+...+3+2+1\right]}.$$

The expression in the parentheses is simply the sum of the integers from 1 to 2^p. We know that the sum of the integers from 1 to z is given by $\frac{z(z+1)}{2}$. So the sum of the integers from 1 to 2^p is given by $\frac{2^p\left(2^p+1\right)}{2}$. Using that $p = n - 1$, we find that for any value n, our total number of payment-dollars is given by:

$$\boxed{\begin{aligned}&\frac{40}{2^{p+q}}\left[2^p+\left(2^p-1\right)+\left(2^p-2\right)+...+3+2+1\right]=\\&\frac{40}{2^{2n-4}}\left[2^{n-1}+\left(2^{n-1}-1\right)+\left(2^{n-1}-2\right)+...+3+2+1\right]=\\&\frac{40}{2^{2n-4}}\left[\frac{\left(2^{n-1}\right)\left(2^{n-1}+1\right)}{2}\right]=\\&\frac{40\left(2^{2n-2}+2^{n-1}\right)}{2^{2n-3}}\end{aligned}}.$$

So we want $\lim\limits_{n\to\infty}\frac{40\left(2^{2n-2}+2^{n-1}\right)}{2^{2n-3}}$. We can factor the numerator into: $\lim\limits_{n\to\infty}\frac{40\left(2^{2n-2}\right)\left(1+2^{1-n}\right)}{2^{2n-3}}$. This becomes

$$\boxed{\begin{aligned}&\lim_{n\to\infty}\frac{40\left(2^{2n-2}\right)\left(1+2^{1-n}\right)}{2^{2n-3}}=\\&\lim_{n\to\infty}\left[40(2)\left(1+2^{1-n}\right)\right]=\\&\lim_{n\to\infty}\left[80\left(1+\frac{1}{2^{n-1}}\right)\right]=\\&80(1)=\\&80\end{aligned}}$$

So we have shown that for an infinite number of payments, the number of dollar-hours is 80, as surmised. I leave it to the reader to show that the right Riemann sum gives the same value. So if Katie gets paid continuously over the 4-hour period, the average amount of money that Katie has in her account is $20.

So we need to be careful how we define average. If we are just examining a bunch of data points, we simply average them in the statistics sense, even if they are camouflaged by given at specific times. But if we define earning as money in the account and pay attention to when she actually gets paid, the problem changes to a more complicated calculus one. We find that her average earnings can be anything from $0 to $40 and if she is paid continuously, the average amount of money in her account will be half of 40 or $20.

Going back to problem 2, Jerry had an exam every week for 5 weeks as shown in the table below. What is his average over the 5-week period?

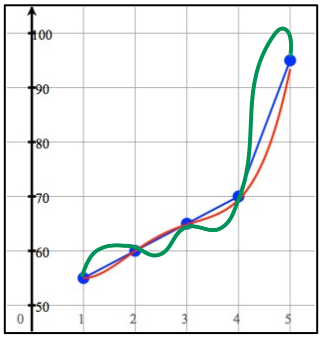
Week 1	Week 2	Week 3	Week 4	Week 5
55%	60%	65%	70%	95%

A 3rd grader can tell you that his average is 69% by adding the percentages and dividing by 5. Why don't we use calculus, using the Riemann sum method that we just illustrated?

The answer is that these numbers are discrete values – not part of a function. There is no time issue here, despite the data given to you at specific time intervals. When there is no time issue, you take the average the way you have always taken it, by adding the scores and dividing by the number of scores. This is a statistics problem, not a calculus problem.

However, let's tweak the problem slightly. Suppose Jerry is studying for the law exam and takes an exam every week that determines the percentage of knowledge he has on the subject. Does that change the problem?

On the right we see the 5 points. We have no idea what the values are between these points, but we do know that there is a continuous function that represents Jerry's knowledge of the material. The function could increase linearly from week to week as shown by the blue curve. Or it could change according to the red curve or more dramatically as shown by the green curve. All we know though that Jerry's knowledge of the material is a function of time.



For instance between week 4 and week 5, his knowledge changes from 70% to 95%. We are not sure if his knowledge always increases during the time interval, but we do know that because of the Intermediate Value Theorem, at some time, between week 4 and week 5, Jerry must have 80% knowledge.

So his average knowledge has to be defined as $\dfrac{\int_1^5 f(t)\,dt}{5-1}$. With no knowledge of what is occurring between the weekly exams, we use a trapezoidal rule to find $\int_1^5 f(t)\,dt \approx \frac{1}{2}\left[55+2(60)+2(65)+2(70)+95\right]=270$.

So on average, Jerry has 270 / 4 = 67.5% knowledge of the material between week 1 and week 5. This is not the same as the statistical answer (69%) when each exam was not a function of time. Just the manner in which the problem was defined changed it from a statistical problem to a calculus one.

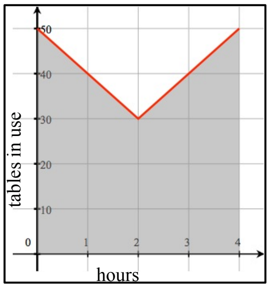
Note that the average rate of change of Jerry's knowledge is given by $\frac{f(5)-f(1)}{5-1}=\frac{0.95-0.55}{4}=\frac{0.4}{4}=\frac{10\%}{\text{week}}$.
This formula holds no matter whether we interpret average as a calculus or statistic problem.

In problem 3, shown are the number of tables that are in use one night at Alice's restaurant for the following times. We want to know the average number of tables that are in use from 5 PM to 9 PM.

5 PM	6 PM	7 PM	8 PM	9 PM
50	40	30	40	50

One school of thought treats the data as a statistics problem: $\boxed{\text{Average: } \frac{50+40+30+40+50}{5}=42}$

The second school of thought uses our calculus definition of average value of the function: $f(t)=\frac{\int_0^4 f(t)\,dt}{4}$. The function f represents the number of tables in use at time t. It is important to understand that we can think of f as a continuous function. We compute the average number of people for the restaurants by first approximating $\int_0^4 f(t)\,dt$ using a trapezoidal rule and is measured in table - hours.



$$\boxed{\text{Average: } \frac{\frac{1}{2}(1)\left[50+2(40)+2(30)+2(40)+50\right]}{4}=40}$$

So does Alice's restaurant have an average number of 42 tables in use or 40 tables in use? Is it a statistics problem or a calculus problem?

Calculating the average as a statistics problem by adding the data up and dividing by 5 neglects the issue of time. But there is a function of time describing the number of tables in use $f(t)$, which is in effect from $t=0$ to $t=4$. For instance at 5:00 PM there are 50 tables in use and at 6 PM, there are 40 tables in use. Assuming this function is continuous, there must be some specific time between 5:00 PM and 6:00 PM when there are 45 tables in use.

So we need to use the $\frac{\int_{t_1}^{t_2} f(t)\,dt}{t_2-t_1}$ to compute the average number of tables in use where $\int_{t_1}^{t_2} f(t)\,dt$ will be measured in table-hours and t_2-t_1 is measured in hours. When time is an issue, the definition of the average value function $\frac{\int_{t_1}^{t_2} f(t)\,dt}{t_2-t_1}$ must be used.

You might argue that the function that represents the number of tables being used at the restaurant is not continuous. And you would be correct. It is possible that the people at two tables leave simultaneously leaving the number of people in the restaurant jump from 48 to 46. Also, unlike Jerry, whose knowledge of his law material could be 79.5%, it doesn't make sense that there would be 47.5 tables in use.

Suppose we went into more depth and found that these were the number of tables being used at Alice's between 5 PM and 6 PM.

5 PM	5:15 PM	5:30 PM	5:45 PM	6 PM
50	49	43	31	40

We see between 5 and 5:15 PM, a table was vacated. We have no idea when this occurred. But no matter when it occurred, the function would have a jump discontinuity – jumping from 50 to 49. However, we have found that in calculating a definite integral, jump discontinuities cause no problem. So we conclude that Alice's restaurant had an average number of 40 tables in use based on the original given information.

Finally, let's change the problem slightly. Suppose the table below now represents the number of *reservations made* for the following times at Alice's. What is the average number of reservations at Alice's restaurant that evening?

5 PM	6 PM	7 PM	8 PM	9 PM
50	40	30	40	50

It turns out that changing the word from tables to reservations makes a world of difference. If Alice had 50 reservations for 5 PM and 40 reservations for 6 PM, would there be some number like 45 reservations for 5:30 PM? Or 49 reservations for 5:01 PM? Surely not, unless this restaurant had a huge number of seats.

As far as we know, there were only reservations taken for times on the hour, not at 5:30 or 6:45. With the tables, we know that if there were 50 in use one at 5 PM and 40 in use at 6 PM, the number of tables in use had to fluctuate and change in the times between 5 and 6 PM. That is not true with the reservations. Tables in use are a function of time elapsed. That is not true with the reservations.

So as far as we know, these are discrete numbers and there is no function of time that describes the reservations. So like Jerry's grades, as originally presented, our average reservations value is found by adding them up and dividing by 5. Alice had 42 reservations for each reservation time. This is a statistics problem, not a calculus problem.

The moral of the story is to pay careful attention when you are asked to find an average. If you are given a series of data points that are not functions of time and want their average, you use the statistics definition of average by adding them up and dividing by the number of data. If you are asked to find an average of a function over time, you will use the formula $\dfrac{\int_{t_1}^{t_2} f(t)\,dt}{t_2 - t_1}$. And you also have to pay careful to the words. A statistics problem can change into a calculus problem and vice versa at the change of a word.