1.4.3 Quiz: Probability Density Functions to Model Data

Bookmark this page

Question 1

1/1 point (graded)

Suppose that $f(x)$ is the probability density function for the random variable $X =$ household income in the US.

Which integral measures $P(14500 < X < 15500)$, the probability of a US household having income between $14,500/year and $15,500/year?

$$\int_{14{,}500}^{1000} f(x)\,dx$$

$$\int_{14{,}500}^{15{,}500} f(x)\,dx$$

$$\frac{1}{1000}\int_{14{,}500}^{15{,}500} f(x)\,dx$$

$$\frac{1}{22{,}378}\int_{14{,}500}^{15{,}500} f(x)\,dx$$

None of these.

Answer: We know that $P(a \le X \le b) = \int_a^b f(x)\,dx$. *In this case, we're finding:*

$$P(14{,}500 \le X \le 15{,}500) = \int_{14{,}500}^{15{,}500} f(x)\,dx.$$

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 2

1/1 point (graded)

Which integral measures the probability of a US household having income above $15,000/year? This is denoted as $P(X > 15,000)$.

- $$\int_{15{,}000} f(x)\,dx$$
- $$1 - \int_{15{,}000} f(x)\,dx$$
- $$\int_{15{,}000}^{450{,}000} f(x)\,dx$$
- $$\int_{15{,}000}^{\infty} f(x)\,dx$$ ✔
- None of these.

Answer: We have a formula for the probability of a US household income lying between a lower and upper bound. In this problem we're given only a lower bound and no upper bound. This means the upper bound is infinity.

The probability of a household having income greater than $15,000/year is

$$P(15,000 < X < \infty) = \int_{15{,}000}^{\infty} f(x)\,dx.$$

Note that computing this integral is an application of improper integrals to deal with the infinite upper bound. The first two answers are not valid integrals. Integrals must have two bounds, though they can be infinite.

The third answer is incorrect because it only includes households making up to $450,000/year. This is most of the households but there still is a 'long tail' of households making more than this amount and thus more that $15,000/year.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Question 3

1/1 point (graded)

Nina said that $P(X = 15,000)$ equals 0 because the area under curve on the interval from 15000 to 15000 is 0. Another way to say this is that the chance that a household makes exactly $15,000/year (not rounded) is zero.

Using the fact that $P(X = 15,000)$ equals 0, give an explanation of why $P(X > 15,000) = P(X \geq 15,000)$.

```
since P(X>=15000) = P(X>15000) + P(X=15000)
```

Thank you for your response.

Explanation

Answer: The value $P(X > 15,000)$ is the probability of a household in our sample having an income greater than $15,000/year, while $P(X \geq 15,000)$ is the probability of a household having an income greater than *or equal to* $15,000/year.

The difference between the two values is the probability of a household having an income exactly equal to $15,000/year. We have said that this probability is 0. If the difference between the two quantities $P(X > 15,000)$ and $P(X \geq 15,000)$ is 0, they must be equal.

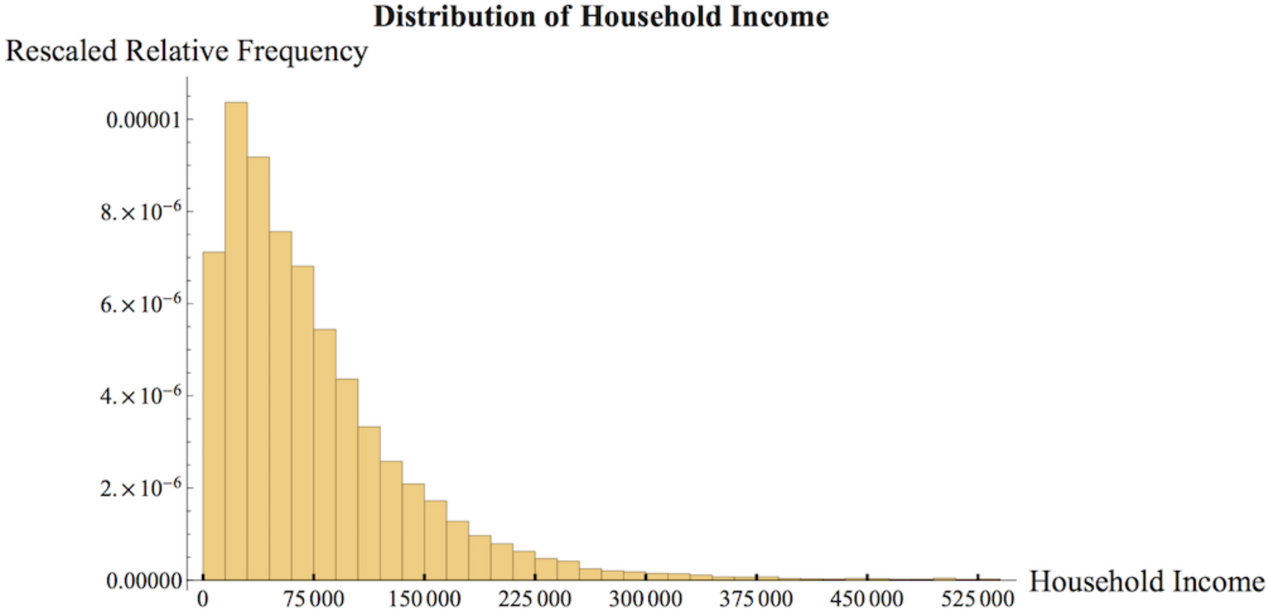
Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 4

1/1 point (graded)

Using the probability density histogram (the rescaled relative frequency histogram), which of the following computations will estimate the percentage of US households earning income whose income between $30,000/year and $60,000/year?



View Larger Image
Image Description

- [x] Add the areas of the bars above the range 30,000 - 45,000 and 45,000 - 60,000.
- [ ] Add the heights of the bars above the range 30,000 - 45,000 and 45,000 - 60,000.
- [ ] Add the heights of the bars above the range 30,000 - 45,000 and 45,000 - 60,000 and multiply by 30,000.
- [ ] Add the heights of the bars above the range 30,000 - 45,000 and 45,000 - 60,000 and divide by 15,000.
- [x] Add the heights of the bars above the range 30,000 - 45,000 and 45,000 - 60,000 and multiply by 15,000.
- [ ] None of the above.

Explanation

Remember that in the rescaled relative frequency histogram the area of each bar equals the percentage of households from our sample belonging to that bin. Therefore, we can find the percentage of US households earning income whose income is between $30,000/year and $60,000/year by adding the areas of the bars above the two ranges.
Because the area of each bar is the product of its height and width, this is equivalent to adding the heights of the bars and multiplying by 15,000. Here “$30-45$ bar” refers to the bar for the bin $[30000, 45000)$, and so on.

$$\begin{aligned}\text{Area}_{\text{30-45 bar}} + \text{Area}_{\text{45-60 bar}} &= 15,000 \cdot \text{height}_{\text{30-45 bar}} + 15,000 \cdot \text{height}_{\text{45-60 bar}} \\ &= 15,000(\text{height}_{\text{30-45 bar}} + \text{height}_{\text{45-60 bar}})\end{aligned}$$

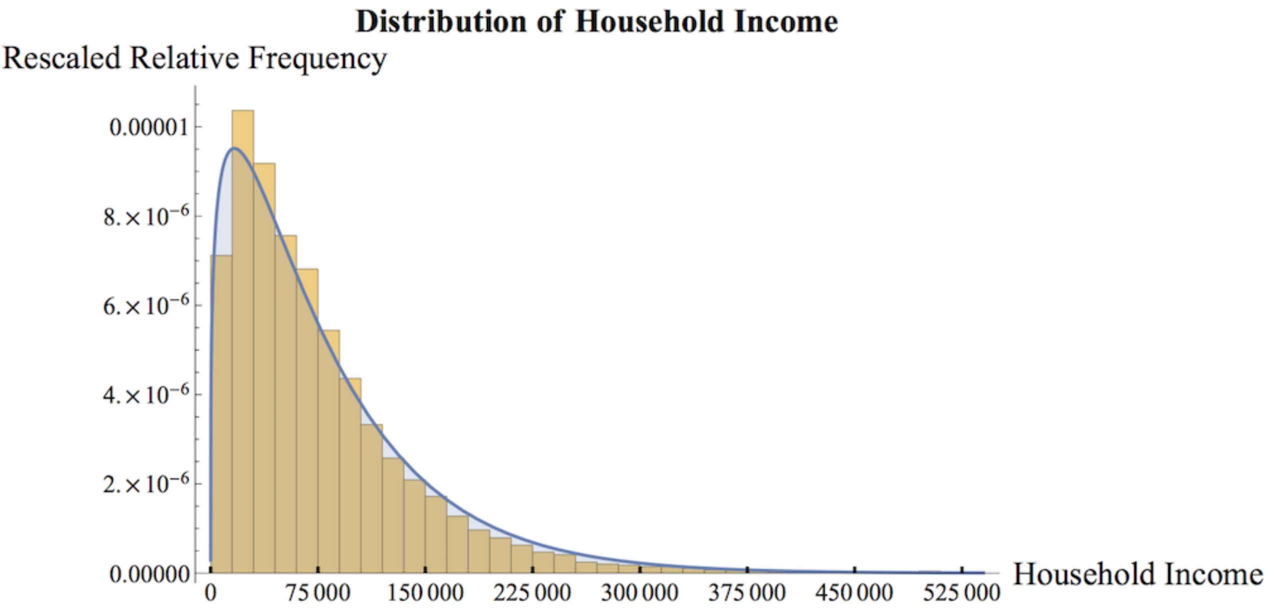
Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Question 5

1/1 point (graded)

Here is the probability density function that Nina showed along with the rescaled relative frequency histogram.



View Larger Image
Image Description

As you can see, the graph of the function $f(x)$ looks almost like the histogram but does not exactly match the data in our sample. Because we can apply more powerful mathematics to a continuous function than to a table of data, we would like to use this function $f(x)$ to model our data.

In this question we explore one difference between the model and the actual data.

On which intervals will the probability $P(a < X < b) = \int_a^b f(x)\,dx$ be greater than the actual fraction of households in the sample with income in that range?

- [x] $[0, 15000]$ ✔
- [ ] $[15000, 30000]$
- [ ] $[30000, 45000]$
- [x] $[150000, 165000]$ ✔
- [ ] None of these.

✔

Explanation

The probability estimated by integrating $f(x)$ will be greater than the corresponding fraction of households in the sample when the area under the curve on that interval is greater than the area of the bars of the histogram on that interval. In other words, this happens when the blue area (PDF) is greater than the yellow area(Histogram) over an interval. We can see from the graph that this occurs for $0 \leq X \leq 15,000$ and for any interval on which $120,000 \leq X \leq 300,000$.

It is not necessary that the graph of the function must always lie above the bars of the histogram, just that the total area under the curve must be greater than the total area of the bars of the histogram.”

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

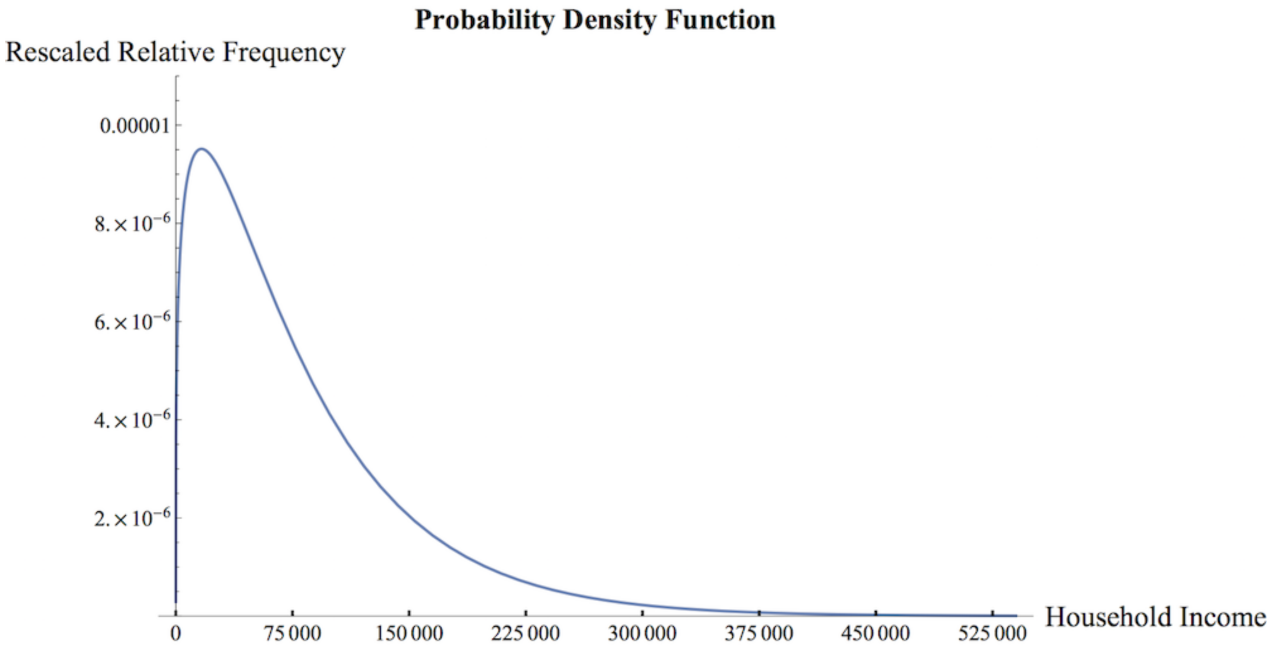
Question 6

1/1 point (graded)

According to the probability density function we are using to model the data, if we choose a random US household earning income, it is more likely that the household makes less than $75,000/year than that it makes between $75,000/year and $150,000/year.

In other words, $P(0 < X < 75,000)$ is greater than $P(75,000 < X < 150000)$.

Which of the following is an explanation of why this is true?



View Larger Image
Image Description

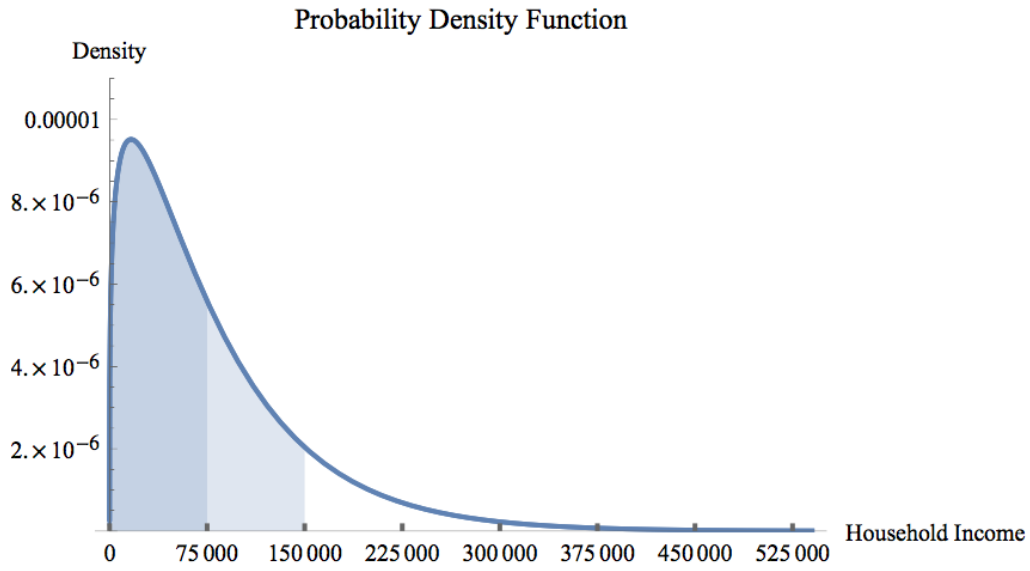
- The values of the function $f(x)$ on $(0, 75000)$ are always bigger than the values of $f(x)$ on $(75000, 150000)$.
- $f(75000) > f(150000)$

- (selected) The area under the probability density function $f(x)$ above the interval $[0, 75000]$ is greater than the area under the function on $(75000, 150000)$. ✔

- The function $f(x)$ has a higher maximum value on $(0, 75000)$ than on $(75000, 150000)$.

Explanation

Answer: By our definition of probability density function, the probability is the area under the curve. Thus $P(0 < X < 75,000)$ *is equal to the area under the curve on the interval* $(0, 75000)$*, and this is greater than the area under the curve on the interval* $(75000, 150000)$ *which is equal to* $P(75,000 < X < 150,000)$.



The second and fourth answers are true statements but do not explain why $P(0 < X < 75,000)$ is greater than $P(75,000 < X < 150000)$.

Submit   You have used 1 of 2 attempts

Answers are displayed within the problem

Question 7

1/1 point (graded)

Nina mentioned that she used a gamma function to model the data.

In the following Desmos dynamic graph, you'll see the graph of a function with parameters a and b. (This function is roughly of the Gamma function form.) What happens when you increase a or increase b? How does the graph change? Estimate values of a and b that make the graph a good model for the histogram data from our small sample of 100 households. (We are not expecting you to look at the function form or find the 'best fit', just make a guess.)

Desmos Graph: Scaled Relative Frequency Histogram - Small Sample (opens in a new window/tab) a is scale (changes the height of the peak) and b is shape parameter (changes the shape of the curve)

Thank you for your response.

Explanation

One good fit possibility is $a = 1.3$ *and* $b = 54000$. *We notice that values of* a *are relatively small while a good value for* b *is quite large.*

If $a = 1$ we notice the graph is an exponential graph.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Learn About Verified Certificates