The chirping crickets problem from Australia (by way of Popular Mechanics)

Phil Petrocelli, mymathteacheristerrible.com

November 19, 2020

Math problems periodically set the Internet alight, and here is a recent one. It's a classic final exam question where it tests multiple things, and asks us to jump between Math topics in order to correctly solve it. It appears that there has been one agreed-upon solution to the problem, but no definitive answer key? I got the agreed-upon answer, and here, I present how we can get that answer.

Question.

You can check out the problem here:

This High School Math Problem Has the Internet Dumbfounded. How Would You Solve It? [1]

Solution.

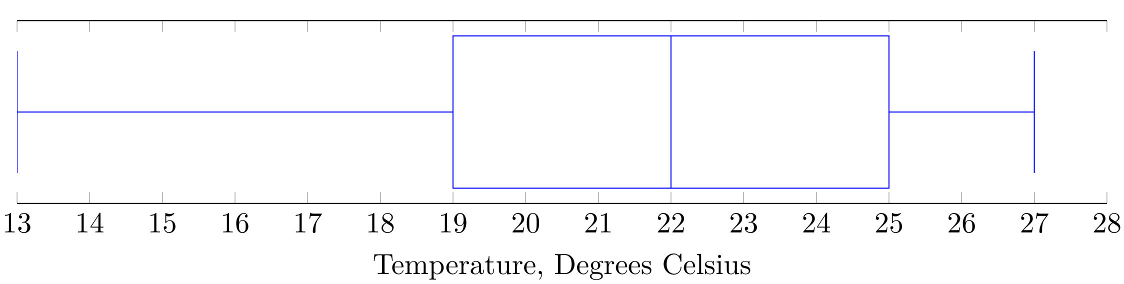
This was definitely a problem where I read through it and just wrote down all the information I thought I was explicitly being given, as well as information that was *implied* by the given information. Let's go through it, bit by bit.

Once a day, for 20 days, the scientist collects data. Based on the 20 data points, the scientist provides the information below.

The first sentence is a little ambiguous, but the second sentence clears it up and offers a key bit that we likely would be after for *any* statistics problem: the number of data points, N, also known as the *sample size*. So:

$$N = 20 \tag{1}$$

A box-plot of the temperature data is shown.



[1] https://www.popularmechanics.com/science/math/a34510782/viral-high-school-math-problem-australia-crickets/

This is telling us to read and interpret the graph they give us. In some books, this kind of graph is also referred-to as a *box and whisker plot.* The box tells us some key bits of data, as do the whiskers.

OK, so what does this graph tell us? There are *five* key bits of information:

$$T_{min} = 13^\circ \tag{2}$$
$$T_{max} = 27^\circ \tag{3}$$
$$Q_1 = 19^\circ \tag{4}$$
$$Q_3 = 25^\circ \tag{5}$$
$$T_{median} = 22^\circ \tag{6}$$

You may recall that a box and whisker plot gives us what's known as a *five number summary* for a set of data.

We can further calculate the inter-quartile range (IQR) from the given graph:

$$\begin{aligned} T_{IQR} &= Q_3 - Q_1 \\ &= 25^\circ - 19^\circ \\ T_{IQR} &= 6^\circ \end{aligned} \tag{7}$$

This is quite a bit of information! Let's continue.

> **The mean temperature in the dataset is** 0.525° **below the median temperature in the dataset.**

This is the first bit of information that actually draws on the previous information presented. The box and whisker plot does not show the mean, but it *does* show the median. Recall from (6) that $T_{median} = 22^\circ$. Finding the mean is, therefore, a simple subtraction:

$$\begin{aligned} T_{\overline{x}} &= T_{median} - 0.525^\circ \\ &= 22^\circ - 0.525^\circ \\ T_{\overline{x}} &= 21.475^\circ \end{aligned} \tag{8}$$

> **A total of 684 chirps was counted when collecting the 20 data points.**

There is a *subtle* shift here that's designed to throw us off. Note that the subject has switched from that of data about ***temperature*** to that of data about ***chirps***. We don't have to fall for the trick. Also note, that, from (1), above, we now have a bit of clarity about how we're supposed to use the $N = 20$ that we recorded a little earlier. Let's say:

$$\begin{aligned} N &= 20 \\ \Sigma_{chirps} &= 684 \end{aligned} \tag{9}$$

> **The scientist fits a least-squares regression line using the data** (x, y)**, where** x **is the temperature in degrees Celsius and** y **is the number of chirps heard in a 15-second time interval. The equation of this line is** $y = -10.6063 + bx$**, where** b **is the slope of the regression line.**

This is largely informational, but for the first time, we see what the relationship between temperature and chirps is supposed to be. That is, (x, y) is really (T, num_{chirps}). Then, we are given the equation of the least-squares regression line, also known as the *line of best fit*:

$$y = -10.6063 + bx \tag{10}$$

With this information, we see that, based on the data collected (because that's *how* the least-squares line is generated), we can calculate (i.e., the *dependent variable*) y, the number of chirps heard in a 15-second period, from (the *independent variable*) x, the temperature in degrees Celsius.

There is a *slight* confusion tactic here, but you're expected to be able to navigate it, based on having some statistics knowledge. For whatever reason(!), linear regression lines are stated in the form $y = a + bx$, whereas, in any algebra class, one of the forms in which a line is stated is $y = mx + b$. *Both* are lines, *both* are in *slope-intercept form*, but the slopes and y-intercepts can be confusing. It doesn't help that b is used in both and that b has different functions in each. Don't fall for the trick. Read on.

> **The least-squares regression line passes through the point $(\overline{x}, \overline{y})$, where $\overline{x}$ is the sample mean of the temperature data, and $\overline{y}$ is the sample mean of the chirp data.**

Now, of course, they switch up the notation on us, and we're expected to make some sense of it. Using $\overline{x}$ and $\overline{y}$ is ***sample*** *mean* notation. We use this when collecting data to approximate the underlying or ***population*** *mean*, which is unknown. Recall that for population means, we use μ. Subtle, but important.

So what of this $(\overline{x}, \overline{y})$? Well, it's simply saying that we know a point on the regression line; this is actually not statistics at all, it's strictly an application of algebra. We're supposed to remember that we can use this point to figure out the slope, b of the line we're given in (10). Recall that we did calculate $\overline{x}$, I called it $T_{\overline{x}}$ in (8), so we can say

$$\overline{x} = T_{\overline{x}} = 21.475^\circ$$

To calculate $\overline{y}$, *now* we use (1) and (9). We simply want to calculate the (sample) mean using:

$$\overline{y} = \frac{\Sigma_{chirps}}{N}$$

So:

$$\overline{y} = \frac{\Sigma_{chirps}}{N}$$

$$= \frac{684}{20}$$

$$\overline{y} = 34.2 \text{ chirps} \tag{11}$$

Making $(\overline{x}, \overline{y}) = (21.475, 34.2)$.

Using this information, we can solve for the slope, b of the least-squares regression line:

$$\begin{aligned}
y &= -10.6063 + bx \\
34.2 &= -10.6063 + b(21.475) && \text{(substitute)} \\
34.2 + 10.6063 &= 21.475b && \text{(subtract)} \\
44.8063 &= 21.475b && \text{(simplify)} \\
\frac{44.8063}{21.475} &= b && \text{(divide)} \\
b &\approx 2.0864 && \text{(simplify and approximate)} \qquad (12)
\end{aligned}$$

Making the equation of the least-squares regression line:

$$y = -10.6063 + 2.0864 \cdot x \tag{13}$$

Where the units of y is number of chirps and the units of x is degrees Celsius. Almost done!

Calculate the number of chirps expected in a 15-second interval when the temperature is 19° Celsius. Give your answer correct to the nearest whole number.

For this, simply recall that the purpose of generating a least-squares regression line for a set of sample data is to *predict* the value of the dependent variable, given a value for the independent variable. We are asked to calculate the number of chirps expected, y, for a given value of x, where x is 19:

$$
\begin{aligned}
y &= -10.6063 + 2.0864 \cdot x \\
&= -10.6063 + 2.0864(19) && \text{(subsitute)} \\
&= -10.6063 + 39.6416 && \text{(multiply)} \\
&= 29.0353 && \text{(simplify)} \\
y &\approx 29 && \text{(round)}
\end{aligned}
$$

Done.

Reporting errors and giving feedback

I am so pleased that you have downloaded this study guide and have considered the techniques herein. To that end, I am the only writer and the only editor of these things, so if you find an error in the text or calculations, please email me and tell me about it! I am committed to prompt changes when something is inaccurate. I also really appreciate it when someone takes a moment to tell me how I'm doing with these sorts of things, so please do so, if you feel inclined.

My email address is: phil.petrocelli@gmail.com.

Please visit `https://mymathteacheristerrible.com` for other study guides. Please tell others about it.

Please donate

I write these study guides with interest in good outcomes for math students and to be a part of the solution. If you would consider donating a few dollars to me so that these can remain free to everyone who wants them, please visit my PayPal and pay what you feel this is worth to you. Every little bit helps.

My PayPal URL is: `https://paypal.me/philpetrocelli`.

Thank you so much.