		Full Transcript
1	**Sonya:**	Another gutter ball. Hi there, it's Sonya. This afternoon, I'm at the bowling alley where my friend, Tim, just started a new job.
2	**Voice-Over Sonya:**	He does all sorts of things from polishing bowling balls to handing out shoes. He was impressed by how the manager can just look at someone and tell what size shoe they wear. I told Tim that it's no secret. His manager just understands the correlation between people's height and their shoe size.
3	**Sonya:**	Tim needs to review regression, and maybe you do too. Grab a ball, lace up your shoes, and let's get another *Problem Solved*.
4	**Sonya:**	Here at the bowling alley, they have bowling shoes of every size and color.
5	**Voice-Over Sonya:**	Tim's manager uses his experience to guess the bowler's shoe size based on the person's height.
6	**Sonya:**	He knows there is a positive correlation between the variables of height and shoe size. Typically, taller people wear larger shoes.
7	**Voice-Over Sonya:**	My friend Jamie needs shoes. We estimate her to be about 5 foot 7 or 67 inches tall. What size shoe is she likely to wear?
8	**Voice-Over Sonya:**	To answer this question, we can look at a scatter plot for the heights and shoe sizes for the last 15 women who rented shoes. For example, this person is 62 inches tall and wears a size 7; here is someone 64 inches tall who wears a size 8, and so on.
9	**Sonya:**	But how can we use this to predict the shoe size of Jamie who is 67 inches tall?
10	**Voice-Over Sonya:**	First, we need to look at the data to see if there is a trend.
11	**Sonya:**	The points don't quite form a line. Too bad, because if they did, we would be able use the equation of that line to predict the shoe size for any given height.
12	**Sonya:**	Hmmm, what if we find a line that is close to the data and use it to predict shoe size?

13	**Voice-Over Sonya:**	Let's begin by drawing a line that seems close to the data points. We can experiment by changing the slope of the line and by shifting it vertically. Notice as we move the line, the equation changes.
14	**Sonya:**	Here the line looks pretty good, but how close is it to the actual data? To find out, let's calculate the residuals. A residual is sort of like an error.
15	**Voice-Over Sonya:**	It's the difference between a shoe size predicted by the line and the actual shoe size. Notice for a woman 63 inches tall that our line predicts approximately a size 7 shoe. But hey, our data set included a woman of this height, and her shoe size was 8. So the residual at this point is 7 minus 8, or -1. We can calculate the residuals for each point in our data set. Now remember, the line we are looking at isn't necessarily the best fit. We haven't found that yet.
16	**Sonya:**	The most commonly used line of best fit is called the least squares regression line. So, why is it called the least squares regression line?
17	**Voice-Over Sonya:**	Because we're going to build squares whose sides are the lengths of the residuals. The least squares regression is the line when the sum of the areas of these squares is as small as possible. So, let's experiment. When we move the line, we can see the total area or the sum of the areas of the squares change as with the equation of the line. Here the total area is about seven. Here it's a little more than 10. When the total area is as small as possible, we have found the least squares regression line. Here it is. Our least squares regression line is $s = 0.4h - 18.4$.
18	**Sonya:**	What do the slope and y-intercept of the equation tell us? To figure it out, we need to keep in mind our situation and some common sense.
19	**Voice-Over Sonya:**	First, the slope is 0.4 or about ½ since we are working with shoe size. This means, if Jamie is 1 inch taller than her friend, she will wear about a half-size larger shoe. What does the y-intercept tell us? It is approximately the point (0, -18). This would mean that if a woman is 0 inches tall, she wears a size -18. Of course this doesn't make sense. The y-intercept for this regression line does not have a practical meaning. What about our prediction for Jamie who is 67 inches tall? We make our prediction by substituting 67 inches for the height in our equation.

		Calculating, we get about 8.8, or between a size 8½ and a 9.
20	**Sonya:**	The least squares regression line can be a useful tool to help you make predictions…
21	**Voice-Over Sonya:**	especially if the data is clustered about the line. The best predictions are those for the values that fall between the smallest and the largest values. If you try to make predictions beyond the data, or extrapolate, they may not be as reliable.
22	**Voice-Over Sonya:**	Well, were our predictions correct?
23	**Jamie:**	Size 9, you're right. Thanks!
24	**Sonya:**	*Problem Solved.*