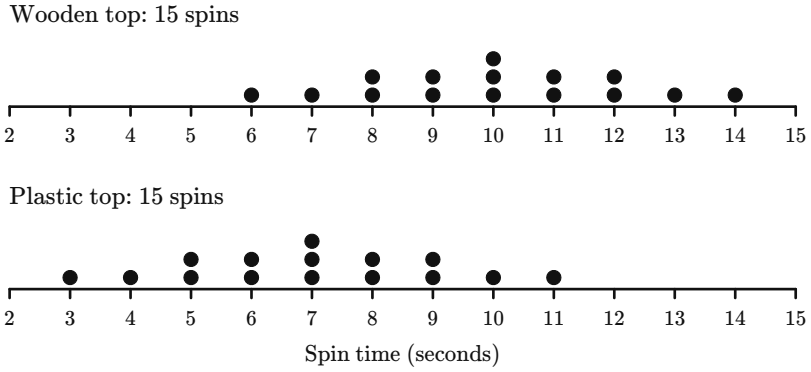
Statistics Unit 1 worksheet: Comparing two distributions (answer key)

ANSWERS AT A GLANCE

1. (a) Wooden top: median 10 seconds, mean 10 seconds, range 8 seconds. Plastic top: median 7 seconds, mean 7 seconds, range 8 seconds. (b) Both roughly symmetric with one peak, no gaps or outliers. (c) "In these spins, the mean wooden-top spin, 10 seconds, is 3 seconds longer than the mean plastic-top spin, 7 seconds." "Both tops have a range of 8 seconds, so their spin times are equally spread out." (d) 1 plastic-top spin and 1 wooden-top spin.
2. (a) P: roughly symmetric with one peak at 10 to 15 minutes. Q: skewed right, peak at 5 to 10 minutes and a long tail out to 35 minutes. (b) P: the 10 to 15 minute bin. Q: the 5 to 10 minute bin. (c) P: $\frac{6}{20} = 30\%$; Q: $\frac{8}{40} = 20\%$. Mia is wrong: that wait is more common at P. (d) Medians and interquartile ranges, for both restaurants, because Q's waits are skewed right.
3. (a) None. (b) The parent-and-child ages form two clusters with a gap from 8 to 31, and the median 19.5 lands in the gap, so it is no one's typical age and a comparison with it compares nothing real. (c) The parent-and-child class is 6 children aged 5 to 8 (median 6.5 years) and 6 adults aged 31 to 40 (median 35.5 years). Every child is younger than everyone in the evening class, and the adults are typically 7 years older than the evening class, whose median is 28.5 years.
4. (a) Means and ranges. Variety A: mean 6.5 peas, range 5 peas. Variety B: mean 7 peas, range 5 peas. (b) 8 pods of each variety. (c) No. The shift of 0.5 peas is a tenth of each range of 5 peas, and the two samples mostly sit on top of each other; another pair of samples could shrink or reverse the gap. (d) Both means are now 7 peas: three more pods each wiped out the gap of 0.5 peas, the kind of swing a gap that is small against the spread can show. It does not show the varieties are the same: more pods could open a gap again, and samples that show no difference do not show the populations are equal.
5. (a) Club X: mean 8 hours, median 4.5 hours. Club Y: mean 5 hours, median 5 hours. (b) The medians: Club X's quartiles are 3 and 6 hours, so 6 + 1.5 × 3 = 10.5, and the 42 hours is an outlier that pulls the mean up. (c) No. The medians differ by only 0.5 hours, and Club Y's is the higher one; that shift is small against the interquartile ranges of 3 and 2 hours. The 3-hour gap in means comes from one member.

WORKED SOLUTIONS

1. Ava has two spinning tops, one wooden and one plastic. She spun each top 15 times on the same table and timed each spin, to the nearest second. The dot plots show the times.



Wooden top: 6, 7, 8, 8, 9, 9, 10, 10, 10, 11, 11, 12, 12, 13, 14

Plastic top: 3, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9, 9, 10, 11

(a) Find the median, the mean and the range of each top's spin times.

(b) Describe the shape of each distribution.

(c) Write one comparison sentence about center and one about spread.

(d) How many plastic-top spins lasted longer than the median wooden-top spin? How many wooden-top spins were shorter than the median plastic-top spin?

ANSWER

Free sample from math.wiki for classroom use. Do not post. math.wiki

(a) Wooden top: median 10 seconds, mean 10 seconds, range 8 seconds. Plastic top: median 7 seconds, mean 7 seconds, range 8 seconds. (b) Both roughly symmetric with one peak, no gaps or outliers. (c) "In these spins, the mean wooden-top spin, 10 seconds, is 3 seconds longer than the mean plastic-top spin, 7 seconds." "Both tops have a range of 8 seconds, so their spin times are equally spread out." (d) 1 plastic-top spin and 1 wooden-top spin.

SOLUTION

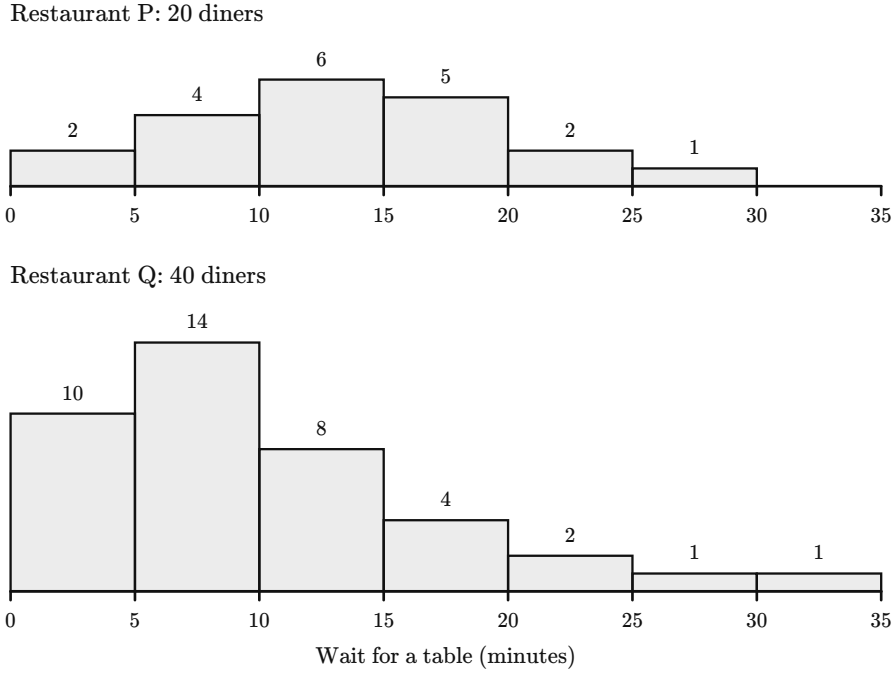
(a) Each list has 15 values, so the median is the 8th. Wooden top: the 8th value is 10, so the median is 10 seconds; the sum is 150, so the mean is $\frac{150}{15} = 10$ seconds; the range is 14 − 6 = 8 seconds. Plastic top: the 8th value is 7, so the median is 7 seconds; the sum is 105, so the mean is $\frac{105}{15} = 7$ seconds; the range is 11 − 3 = 8 seconds.

(b) Each plot has its tallest stack in the middle (at 10 seconds for the wooden top, at 7 seconds for the plastic top) and falls away evenly on both sides: roughly symmetric with one peak, with no gaps or outliers.

(c) "In these spins, the mean wooden-top spin, 10 seconds, is 3 seconds longer than the mean plastic-top spin, 7 seconds." "Both tops have a range of 8 seconds, so their spin times are equally spread out; both distributions are roughly symmetric with one peak." On one scale the plastic-top plot is the wooden-top plot slid 3 seconds to the left.

(d) Only the plastic top's 11 second spin lasted longer than 10 seconds, so 1 plastic-top spin. Only the wooden top's 6 second spin was shorter than 7 seconds, so 1 wooden-top spin. The bulk of the wooden-top spins lies to the right of the bulk of the plastic-top spins.

2. Mia asked 20 diners at Restaurant P and 40 diners at Restaurant Q how many minutes they waited for a table on a Saturday night. The histograms show the waits in 5-minute bins, a wait on a bin edge going into the bin to its right.



(a) Describe the shape of each distribution.

(b) Find the bin that holds each restaurant's median wait.

(c) Mia says waits of 10 to 15 minutes are more common at Restaurant Q, because its bar for that bin is taller, 8 against 6. Find the fraction of each restaurant's diners who waited 10 to 15 minutes, and say whether Mia is

Free sample from math.wiki for classroom use. Do not post. math.wiki

right.

(d) Which measures of center and spread should a comparison of the two restaurants use? Explain.

ANSWER

(a) P: roughly symmetric with one peak at 10 to 15 minutes. Q: skewed right, peak at 5 to 10 minutes and a long tail out to 35 minutes. (b) P: the 10 to 15 minute bin. Q: the 5 to 10 minute bin. (c) P: $\frac{6}{20} = 30\%$; Q: $\frac{8}{40} = 20\%$. Mia is wrong: that wait is more common at P. (d) Medians and interquartile ranges, for both restaurants, because Q's waits are skewed right.

SOLUTION

(a) Restaurant P's bars, 2, 4, 6, 5, 2, 1, rise to a peak from 10 to 15 minutes and fall again: roughly symmetric with one peak. Restaurant Q's bars, 10, 14, 8, 4, 2, 1, 1, peak from 5 to 10 minutes and then fall away in a long tail of short bars out to 35 minutes: skewed right.

(b) P has 20 waits, so the median is halfway between the 10th and 11th. 2 + 4 = 6 waits are shorter than 10 minutes and 6 + 6 = 12 are shorter than 15, so both the 10th and the 11th are in the 10 to 15 minute bin. Q has 40 waits, so the median is halfway between the 20th and 21st. 10 waits are shorter than 5 minutes and 10 + 14 = 24 are shorter than 10, so both are in the 5 to 10 minute bin.

(c) P: $\frac{6}{20} = 0.3$, or 30%. Q: $\frac{8}{40} = 0.2$, or 20%. Q's bar is taller only because Mia asked twice as many diners there. Out of each restaurant's own diners, a wait of 10 to 15 minutes is more common at P, so Mia is wrong. With different sample sizes, compare fractions, not bar heights.

(d) Q's waits are skewed right, so Q calls for the median, and one measure for both groups means the medians and the interquartile ranges for both restaurants. The mean of Q would be pulled up by its long tail of slow waits.

3. Omar recorded the ages, in years, of everyone in two made-up pottery classes at a community center.

Parent-and-child class: 5, 6, 6, 7, 7, 8, 31, 33, 35, 36, 38, 40

Evening class: 19, 22, 24, 25, 27, 28, 29, 31, 33, 34, 36, 40

Omar writes: "The median age in the parent-and-child class is 19.5 years, 9 years below the evening class's median of 28.5 years, so people in the parent-and-child class are typically 9 years younger."

(a) How many people in the parent-and-child class are within 5 years of 19.5?

(b) Explain what is wrong with Omar's comparison.

(c) Write a better description of the parent-and-child class, and use it to compare the two classes.

ANSWER

(a) None. (b) The parent-and-child ages form two clusters with a gap from 8 to 31, and the median 19.5 lands in the gap, so it is no one's typical age and a comparison with it compares nothing real. (c) The parent-and-child class is 6 children aged 5 to 8 (median 6.5 years) and 6 adults aged 31 to 40 (median 35.5 years). Every child is younger than everyone in the evening class, and the adults are typically 7 years older than the evening class, whose median is 28.5 years.

SOLUTION

(a) Within 5 years of 19.5 means from 14.5 to 24.5 years. The oldest child is 8 and the youngest adult is 31, so no one in the parent-and-child class is in that interval: none.

(b) The parent-and-child ages form two clusters, children from 5 to 8 and adults from 31 to 40, with no one in between. With 12 ages the median is halfway between the 6th and 7th, $\frac{8 + 31}{2} = 19.5$, which lands in the gap. One center needs one peak; this one describes nobody, so "typically 9 years younger" describes no one in the class either.

Free sample from math.wiki for classroom use. Do not post. math.wiki

(c) Describe each cluster. Children: 5, 6, 6, 7, 7, 8, median $\frac{6+7}{2} = 6.5$ years. Adults: 31, 33, 35, 36, 38, 40, median $\frac{35+36}{2} = 35.5$ years. The evening class has one cluster, from 19 to 40, median $\frac{28+29}{2} = 28.5$ years. Comparison: "Every child in the parent-and-child class is younger than the youngest evening student, 19. The adults in the parent-and-child class are typically older than the evening class: their median, 35.5 years, is 7 years above the evening median of 28.5 years."

4. A made-up farm grows two varieties of peas in two large fields. Jenny opened a random sample of 12 pods from each field and counted the peas in each pod.

Variety A: 4, 5, 5, 6, 6, 6, 7, 7, 7, 8, 8, 9

Variety B: 4, 5, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9

(a) Both distributions are roughly symmetric. Which measures of center and spread should a comparison use? Find them for each variety.

(b) How many pods of each variety hold from 6 to 8 peas?

(c) Jenny says, "Variety B has more peas per pod." Do the samples support that for all the pods in the two fields? Explain by comparing the shift with the spread.

(d) Jenny opens three more pods chosen at random from each field. Variety A's hold 8, 9 and 10 peas, and Variety B's hold 6, 7 and 8. Find each variety's mean for all 15 of its pods. What does the change show about the first gap between the means? Does it show that the two varieties have the same number of peas per pod?

ANSWER

(a) Means and ranges. Variety A: mean 6.5 peas, range 5 peas. Variety B: mean 7 peas, range 5 peas. (b) 8 pods of each variety. (c) No. The shift of 0.5 peas is a tenth of each range of 5 peas, and the two samples mostly sit on top of each other; another pair of samples could shrink or reverse the gap. (d) Both means are now 7 peas: three more pods each wiped out the gap of 0.5 peas, the kind of swing a gap that is small against the spread can show. It does not show the varieties are the same: more pods could open a gap again, and samples that show no difference do not show the populations are equal.

SOLUTION

(a) Both are roughly symmetric with no outliers, so compare means, with the range for spread.

$$\text{Variety A mean} = \frac{78}{12} = 6.5$$
$$\text{Variety B mean} = \frac{84}{12} = 7$$

The ranges are 9 − 4 = 5 peas for each.

(b) Variety A: three 6s, three 7s and two 8s, so 8 pods. Variety B: two 6s, three 7s and three 8s, so 8 pods.

(c) The means differ by 7 − 6.5 = 0.5 peas, while each variety's pods spread over a range of 5 peas: the shift is only a tenth of the spread. Two thirds of each sample sits in the same band from 6 to 8 peas. Two random samples from fields with the same typical pod could easily differ this much, so the samples do not support Jenny's claim for the whole fields. "In these samples, Variety B averaged 0.5 peas per pod more, but that is small against a range of 5 peas."

(d)

$$\text{Variety A mean} = \frac{78+8+9+10}{15} = \frac{105}{15} = 7$$
$$\text{Variety B mean} = \frac{84+6+7+8}{15} = \frac{105}{15} = 7$$

Free sample from math.wiki for classroom use. Do not post. math.wiki

Three more random pods from each field closed the gap of 0.5 peas completely. That is what (c) warned of: a shift that is small against the spread can shrink, vanish or reverse with the luck of the draw. It still does not show the varieties are the same. Another few pods could open a small gap again, and "the samples do not show a difference" is not "the populations are the same".

5. Kevin chose random samples of 10 members from each of two made-up community clubs and asked how many hours each member volunteered last month.

Club X: 2, 3, 3, 4, 4, 5, 5, 6, 6, 42

Club Y: 2, 3, 4, 4, 5, 5, 6, 6, 7, 8

Kevin says, "Club X members volunteered 3 hours more on average, so members of Club X tend to volunteer more."

(a) Find the mean and the median of each sample.

(b) Which pair, the means or the medians, should be compared? Explain, using the 1.5 IQR rule.

(c) Is Kevin's conclusion justified? Explain by comparing the shift with the spread.

ANSWER

(a) Club X: mean 8 hours, median 4.5 hours. Club Y: mean 5 hours, median 5 hours. (b) The medians: Club X's quartiles are 3 and 6 hours, so 6 + 1.5 × 3 = 10.5, and the 42 hours is an outlier that pulls the mean up. (c) No. The medians differ by only 0.5 hours, and Club Y's is the higher one; that shift is small against the interquartile ranges of 3 and 2 hours. The 3-hour gap in means comes from one member.

SOLUTION

(a) Club X: the sum is 80, so the mean is $\frac{80}{10} = 8$ hours; the 5th and 6th values are 4 and 5, so the median is 4.5 hours. Club Y: the sum is 50, so the mean is $\frac{50}{10} = 5$ hours; the 5th and 6th values are both 5, so the median is 5 hours.

(b) Club X's lower half is 2, 3, 3, 4, 4, so $Q_1 = 3$; its upper half is 5, 5, 6, 6, 42, so $Q_3 = 6$ and the IQR is 3 hours. Since 6 + 1.5 × 3 = 10.5 and 42 > 10.5, the 42 hours is an outlier. It pulls Club X's mean above 9 of its 10 values. Club X calls for the median, so compare medians and interquartile ranges for both clubs. Club Y's quartiles are 4 and 6, an IQR of 2 hours.

(c) The medians are 4.5 and 5 hours: a shift of 0.5 hours, in Club Y's favor, against interquartile ranges of 3 and 2 hours. The middle halves, 3 to 6 and 4 to 6 hours, almost coincide. The gap in means does not run through the samples: without the 42 hours, Club X's mean would be $\frac{38}{9} \approx 4.2$ hours, below Club Y's. So the samples give no reason to think either club's members tend to volunteer more.

Free sample from math.wiki for classroom use. Do not post. math.wiki