QUARTER 2 PROJECT

[redacted]

Math Question – Different Z-Score Problems Using NormalCDF, InvNorm, and Percentiles

At Largo Central Park, the distance walked by visitors during their stay is approximately normally distributed, with a mean of 500 meters and a standard deviation of 100 meters.

Z - Score Formula

$$\text{Z - Score} = \frac{\text{X - Mean}}{\text{Standard Deviation}}$$

or

$$\text{Z - Score} = \frac{X - \bar{X}}{SD}$$

NormalCDF 5 Steps

1. List N(Mean, Standard Deviation)
   N = Normal Distribution
2. Create and Shade Normal Curve
3. Find Z - Score (Show calculations)
4. Use NormalCDF (2nd > Vars > NormalCDF)
   Calculator Steps
5. Answer With Context

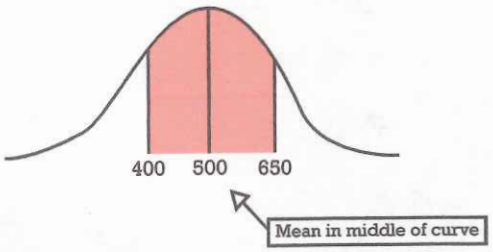
Questions

A. What percentage of people at Largo Central Park walk between 400 and 650 meters during their stay? (NormCDF 5 steps)

1. List N(Mean, Standard Deviation) = N(500,100)

2. Create and Shade Normal Curve

Shaded part is what we are trying to find



3. Find Z - Score (Show calculations)

Place the given numbers into the Z - Score formula. Since we have 2 numbers, we must find Z - Score twice.

$$Z = \frac{400 - 500}{100} = -1 \qquad Z = \frac{650 - 500}{100} = 1.5$$

4. Use NormalCDF (2nd > Vars > NormalCDF)
   Calculator Steps

NormalCDF (Lower : -1 , Upper : 1.5) = .7745375117

5. Answer With Context

.7745375117 of people at Largo Central Park are predicted to walk between 400 and 650 meters during their stays.

B. A person at the 90th percentile walked how far? (InvNorm)

When the problem asks you to find what is at a percentile, you use InverseNorm.
We lack information to solve for Z (We are missing X). InverseNorm finds Z for us.

90th Percentile = .90 Place this into InvNorm

InvNorm(Area: .90) = 1.281551567

Now you have enough information to use the formula →

Answer With Context

Calculator Steps
(2nd > Vars > InvNorm)

Simple Algebra

$$1.281551567 = \frac{X - 500}{100}$$

$$128.1551567 = X - 500$$

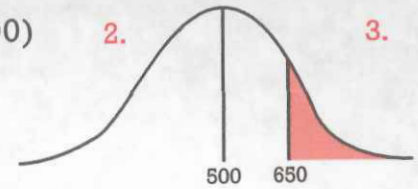
$$X = 628.1551567$$

A person at the 90th percentile is predicted to walk 628.1551567 meters during their stay at Largo Central park

C. What percentage of people walked over 650 meters during their stay at Largo Central Park? (NormCDF 5 steps) -

1. N(500,100)

2.



3.

$$Z = \frac{650 - 500}{100} = 1.5$$

4. NormalCDF (Lower : 1.5 , Upper : 10)
   = 0.0668072287

5. 0.0668072287 of people at Largo Central park are predicted to walk over 650 meters during their stays.

SEMESTER 2 GROUP PROJECT 125/100

[redacted]

AP STATS PROJECT





Who did what?

[redacted]

- Clearwater and St. Pete Pier Data Collection
- Conditions Page
- Chapter 8
- Video Editing

[redacted]

- Clearwater Pier Data Collection
- Chapter 1
- Chapter 2
- Chapter 4
- Graphics

[redacted]

- Sprouts and Publix Data Collection

[redacted]

- Graphics

Conditions Page:

Conditions for Approximately Normal (1 Proportion z Test/Interval):

I. Random selection stated
II. $50 < 10\%$ of people at the pier
III. $50(13/50) \geq 10$
$50(37/50) \geq 10$

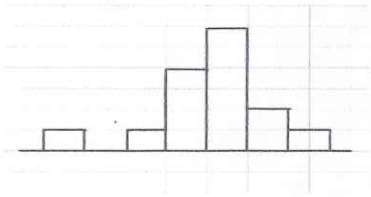
Conditions for Approximately Normal (2 Proportion z Test/Interval):

I. Random selection stated for Clearwater Pier
II. $50 < 10\%$ of people at Clearwater Pier
III. $50(7/50) \geq 10 \rightarrow$ **Proceed with caution**
$50(43/50) \geq 10$

I. Random selection stated for St. Pete Pier
II. $50 < 10\%$ of people at St. Pete Pier
III. $50(13/50) \geq 10$
$50(37/50) \geq 10$

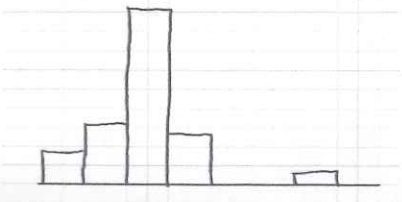
Conditions for Approximately Normal (Matched Pairs Mean t Test/Interval):

I. Random selection stated
II. $25 < 10\%$ of organic products
III. May be outliers, proceed with caution



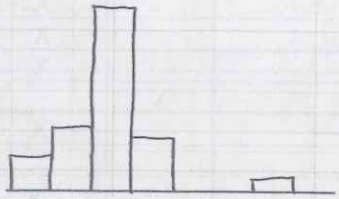
Conditions for Approximately Normal (1 Means t Test/Interval)

I. Random Selection Stated
II. $25 < 10\%$ of organic foods
III. May be outliers, proceed with caution

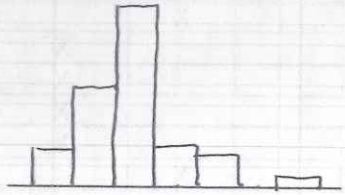


Conditions for Approximately Normal (2 Means t Test/Interval)

I. Random Selection Stated for Sprouts
II. 25 < 10% of organic foods
III. May be outliers, proceed with caution



I. Random Selection Stated for Publix
II. 25 < 10% of organic foods
III. May be outliers, proceed with caution



Matched Pairs Data:

Item	Sprouts Price	Publix Price
Banana	$0.36	$0.39
Honey Crisp Apple	$1.68	$1.99
Navel Orange	$1.05	$1.37
Annie's Organic Cheddar Squares	$4.99	$5.99
Nature's Path Wild berry Acai Toaster Pastier	$5.49	$5.49
Annie's Organic Chocolate Chip Chewy Granola Bars	$5.49	$4.49
Lesser Evil Organic Himalayan Pink Salt	$3.49	$3.99
Bear Snacks Strawberry Fruit Rolls	$3.99	$3.89
Chomps Mini Original Beef Sticks	$7.64	$8.99
Simple Mills Chocolate Brownie Seed and Nut Sweet Thins	$5.49	$5.89
Good Culture Organic Whole Milk Cottage Cheese	$6.29	$6.99
Almond Milk	$7.49	$6.69
Siete Chocolate Chip Cookies	$4.99	$3.89
Crafter's Premium Spread Strawberry	$6.49	$7.49
Miyako's Organic European Style Vegan Butter	$4.79	$5.49
Santa Cruz Organic Creamy Dark Roasted, No Stir Peanut Butter	$8.49	$8.99
Organic Valley Shredded Mozzarella	$6.99	$5.99
Tillamook Medium Cheddar Snack Portion	$5.99	$5.59
Red's Burrito, Grilled Chicken & Cheese	$2.79	$2.99
Dave's Killer Bread Organic 21 Whole Grain Thin Sliced	$5.99	$6.69
Applegate Naturals Sausage Patties Chicken and Maple	$13.99	$13.99
Theo Chocolate Sea Salt 70% dark	$4.29	$4.89
Unreal Dark Chocolate Peanut Butter Cups	$6.49	$6.69
Once Upon a Farm, Dairy Free	$8.99	$10.99
Oat Haus Organic Original Granola Butter	$6.99	$10.99

Clearwater Pier Data:

#	Description	Dog	No Dog
1	Female with tie dye shirt, blonde hair, and beige hat	X	
2	Male with neon green shirt with a matching green hat and sunglasses		X
3	Female with pink shirt and a pink backpack		X
4	Female with brown hair, yellow shirt, denim shorts		X
5	Male with red shirt, black shorts, black hat black sunglasses		X
6	Male with white shirt black shorts brown slides		X
7	Female with a baby, jeans, black sandals, black tank top		X
8	Female with purple tank top, blue-green shorts, white shoes	X	
9	Female black tank top black shorts white/grey shoes		X
10	Female green tank top, black leggings, black/white shoes		X
11	Male, beige shirt, plaid shirt, brown sandals		X
12	Male with red shirt, guitar, beige shorts, only one sock on		X
13	Male, older, white sunhat, dark green shirt, grey shoes		X
14	Female with a "Where's Waldo" shirt, black leggings		X
15	Female with blue sundress, white hat, black sunglasses		X
16	Male with red tank top, pink pineapple shorts		X
17	Male with blue hat, yellow shirt, grey sweatpants		X
18	Female with coral shirt, blue floral shorts, black and beige sunhat	X	
19	Male. Bald with white shirt, beige cargo pants		X
20	Female with a black and white Poka dot dress, brown hair		X
21	Female with blue tank top, black shorts, white shoes		X
22	Male with white shirt, dark blue shorts, flip flops		X
23	Female on a bike with dog on one of those baby holder harnesses, black shirt	X	
24	Blonde female in a green sundress and grey shoes		X
25	Female with a green and yellow pineapple(?) dress		X
26	Male with a "dad tax" shirt and hat and glasses		X
27	Man with pink shirt and beige shorts brown		X
28	Lady with pink shirt, denim shorts, long hair		X
29	Lady with green cardigan, black tank top and shorts		X
30	Lady with green eagles shirt and sunglasses and brown hair		X
31	Lady with orange shirt, beige shorts, orange hat		X
32	Lady with floral side bag, purple flowers, white shirt		X
33	Lady running with white tank top, pink shorts, bright blue shoes		X
34	Man. White shirt, black shorts, sunglasses		X
35	Man. Green hat, red shoes, black shirt and shorts		X
36	Lady. Red tube top, black pants, no shoes		X
37	Lady. Blue sweatshirt, blue jeans, light pink shoes		X
38	Man with collar shirt, two sweaters(??), sunglasses, hat		X
39	Shirtless guy, with USA swim shorts		X
40	Older lady, blue jeans, grey shirt		X
41	Female with pink sundress, black shoes		X
42	Guy with a white shirt, sunset swim shorts		X
43	Dude with white, blue shirt, grey shorts, white dog on a pink harness	X	
44	Small white dog with a lady in a blue tank to black sweatshirt around waist	X	
45	Girl scout with cookies and scrunchie in her hair		X
46	Big dog, guy with black shirt, buzzcut	X	
47	Guy with white shirt, beard, and colorful galaxy color shorts		X
48	Lady with pink and black bikini		X
49	Man pushing baby stroller, dark grey shirt blue shorts		X
50	Lady black tank top and blue floral skirt		X

St. Pete Pier Data:

#	Description	Dog	No Dog
1	Man in green shirt gray cargo pants, balding at the top		X
2	Lady with magenta shorts, black tank top, black sunglasses, white shoes		X
3	Lady Blue shirt, black shorts, grey shoes		X
4	Black poodle, older man w grey sweater, US army hat	X	
5	Male with blue Rays shirt, blue denim shorts		X
6	Woman with floral dress and red hair		X
7	Lady with magenta tank top, long hair in a ponytail, black shorts, white-ish gold dog	X	
8	Lady with a white shirt, white hat, black pants, golden dog	X	
9	Man with long grey hair, black shirt, blue jeans		X
10	Man with red shirt, dark blue hat. I don't know what dog it is, but it was cute and jumping around	X	
11	Man with tan shirt, brown pants, black hat		X
12	Man with white shirt, grey hat and shorts		X
13	Man with fishing rod, blue shirt, lighter blue shorts		X
14	Man with black/green tank top, blue shorts and earbuds in		X
15	man with little boy on his shoulders, blue jeans		X
16	Man with grey shirt with a flower on it, black shoes		X
17	Older man with grey hair, red shirt, shorts		X
18	Man with grey beard, Rays shirt (Bradley 45) and grey shorts		X
19	Man with backwards hat, grey shirt and jeans		X
20	Man with brown sunhat, collar shirt, tan shorts	X	
21	Man with longer hair, green shirt, black shorts		X
22	Man with tan shirt, black shorts, sunglasses		X
23	Woman with Rays t-shirt, blue jeans. Black dog with a Rays handkerchief around their neck	X	
24	Man pulling(?) a stroller (I have no clue why) with grey shirt and jeans		X
25	Guy in grey sweatpants and a red sweater		X
26	Man with one of those iconic fisherman hats. Like sunhats but for men. Blue shirt.		X
27	Lady with purplish shirt, sunglasses and a hat		X
28	Two BIIIG dog. Man with green hat, blue shirt, darker blue shorts	X	
29	Man with black hair, those glasses that have colorful lenses.		X
30	Girl with braids, off white cardigan, ripped blue jeans		X
31	Female with pale purple shirt, blue jeans, white shoes	X	
32	Lady with coral-like orange shirt, black pants		X
33	Lady brownish hair, Orange long sleeve, dark blue jeans		X
34	Woman in black hoodie, black shorts, black hat, black shoes		X
35	Man with red/white sweater, sandals with socks, headphones		X
36	Guy with grey hair, white shirt, beige shorts, black shoes		X
37	Man with red shirt, black shorts, black hat, sunglasses	X	
38	Lady with red hair, a black tank top and shorts		X
39	Lady with a floral white shirt, dark muted green pants		X
40	Man with dark grey Bucs shirt and lighter grey shorts and white hat with a black stripe		X
41	White fluffy dog with brown/black spots, Lady with ponytail, sunglasses, black purse, coral shirt	X	
42	Guys with blue Tampa Bay rays shirt and white hat, black shorts		X
43	Bald guy with beard. Bright red shirt, blue denim shorts		X
44	Guy. Headphones pink shirt, black shorts, ankle socks	X	
45	Guy on electric scooter, helmet, black shirt, and shorts		X
46	White hat, black sunglasses man, white fluffy dog, big dog	X	
47	Man in white polo, long pink pants, white shoes		X
48	Lady with green tank top, black purse, black shorts, black hair		X
49	Lady with white button up, black skirt, black combat boots. Handheld pizza hut fan		X
50	Lady with pink/white tank top, dark red shoes, small fluffy dog	X	

Chapter 1

A. Describe Each Distribution (CUSS)

Sprouts Prices (L1)

Center: The median price of Sprouts items is **$5.49**.

Unusual: (Outlier Test)

Q1: **$3.99** Q3: **$7.49** IQR Calculation: IQR = Q3 − Q1= 7.49 − 3.99 = 3.50

- Lower Bound Calculation: Q1 − (1.5 × IQR) = 3.99 − (1.5 × 3.50) = −1.26
- Upper Bound Calculation: Q3 + (1.5 × IQR) = 7.49 + (1.5 × 3.50) = 12.74
- Outliers: **$13.99** (above the upper bound)

Shape: The distribution is approximately skewed right, with a concentration of lower-priced items but a few higher values pulling the mean upward.

Spread

- Range Calculation: Range = Max – Min =13.99 − 0.36 = 13.63
- IQR: **$3.50**

Publix Prices (L2)

Center: The median price of Publix items is **$5.89**.

Unusual: (Outlier Test)

Q1: **$3.99** Q3: **$7.49** IQR Calculation: IQR = Q3 − Q1= 7.49 − 3.99 = 3.50

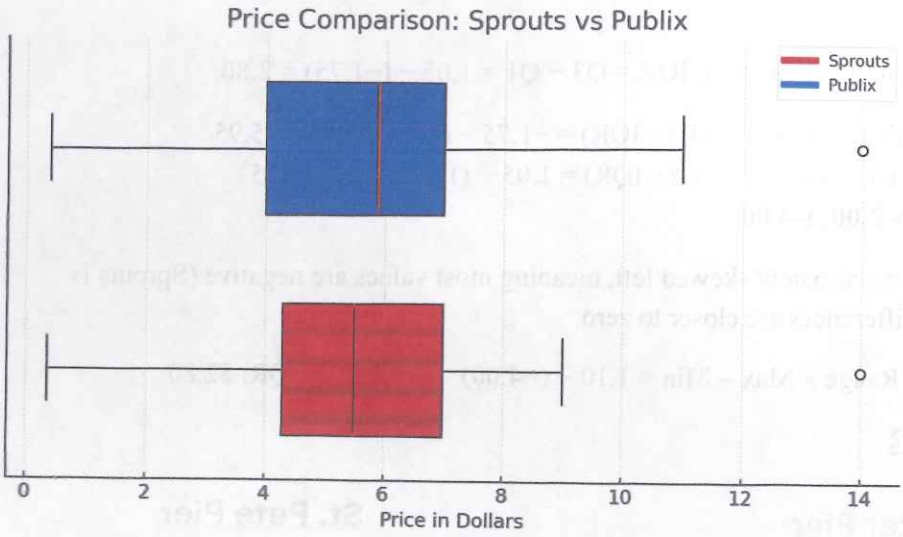
- Lower Bound Calculation: Q1− (1.5 × IQR) = 3.99 − (1.5 × 3.50) = −1.26
- Upper Bound Calculation: Q3 + (1.5 × IQR) = 7.49 + (1.5 × 3.50) = 12.74
- Outliers: **$13.99** (above the upper bound)

Shape: The distribution is approximately skewed right, similar to Sprouts, with most prices on the lower end but a few higher-priced items pulling the mean upward.

Spread:

- Range Calculation: Range = Max – Min = 13.99 − 0.39 = 13.60
- IQR: **$3.50**

B. Compare Graphs (Side by side Boxplots)



Sprouts (L1) – Five Number Summary:

Min: $0.36, Q1: $3.99, Median: $5.49, Q3: $7.49, Max: $13.99

Publix (L2) – Five Number Summary:

Min: $0.39, Q1: $3.99, Median: $5.89, Q3: $7.49, Max: $13.99

C. Back-To-Back Stem & Leaf Plot Rounded to the Nearest Dollar

Key: 6 | 0 ← = $6 Key: 0 | 5 → = $5

Sprouts		Publix
1,0	0	0,1
3,3,2	0	2,3
5,5,5,5,5,5,4,4	0	4,4,5,5,5,5
7,7,7,7,6,6,6,6	0	6,6,6,6,7,7,7,7,7
9,8,8	0	8,9,9
	1	1,1
	1	
4	1	4
	1	
	1	

D & F. CUSS/BS for L3 (Sprouts – Publix)

Center: Median: -$0.32

Unusual (Outlier Test):

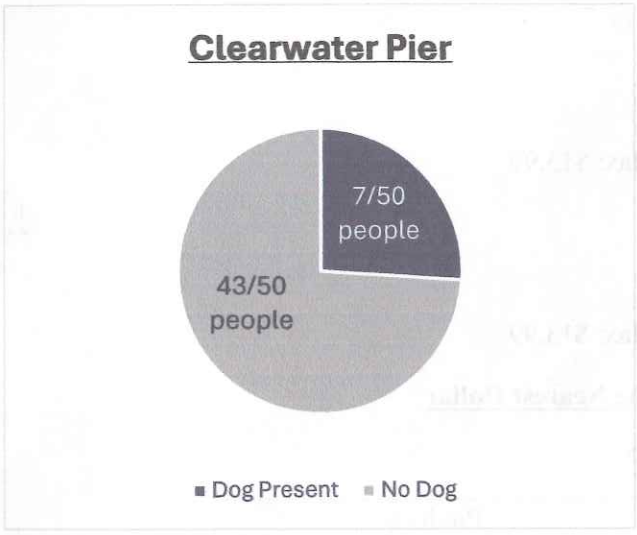
Q1: **$-1.75** Q3: **$1.05** IQR Calculation: IQR = Q3 − Q1 = 1.05 − (−1.75) = 2.80

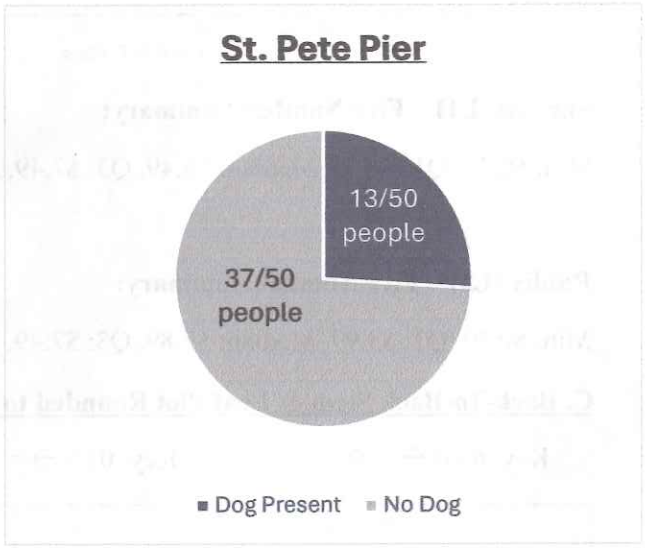
- Lower Bound Calculation: Q1 − (1.5× IQR) = −1.75 − (1.5 × 2.80) = −5.95
- Upper Bound Calculation: Q3 + (1.5× IQR) = 1.05 + (1.5 × 2.80) = 4.35
- Outliers: **$1.10, $-2.00, $-4.00**

Shape: The distribution is approximately skewed left, meaning most values are negative (Sprouts is usually cheaper), but some differences are closer to zero.

Spread: Range Calculation: Range = Max – Min = 1.10 − (−4.00) = 5.10 IQR: $2.80

E. Pie Graphs for $\hat{p}1$ and $\hat{p}2$



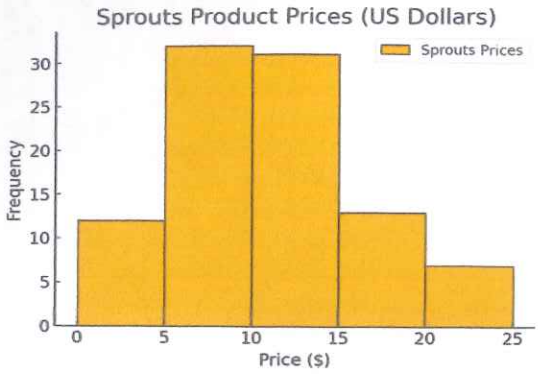


Chapter 2: Percentile Calculations (All units are in US dollars: $)

Sprouts (L1)

Sample Statistics

- **Sample Mean ($\bar{X}$):** 10.9122
- **Standard Deviation (Sx):** 6.2545



The distribution is approximately normal but approximately right skewed.

35th Percentile

InverseNorm(0.35,0,1) = −0.3853

X = 10.9122+ (−0.3853 × 6.2545) = 8.50

88th Percentile

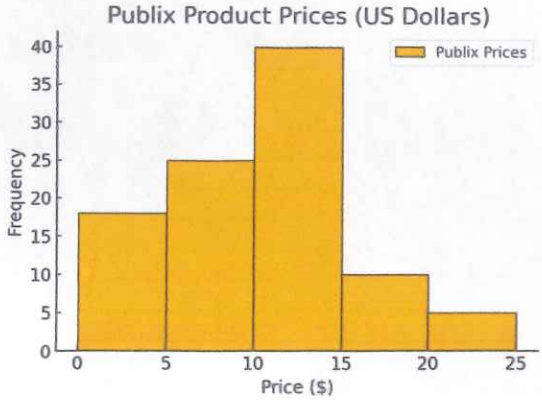
InverseNorm(0.88,0,1) = 1.1750

X =10.9122 + (1.1750 × 6.2545) = 18.26

Publix (L2)

Sample Statistics

- **Sample Mean ($\bar{X}$):** 10.2177
- **Standard Deviation (Sx):** 5.8237



The distribution is approximately normal but approximately right-skewed.

35th Percentile

InverseNorm(0.35,0,1) = −0.3853

X = 10.2177 + (-0.3853 x 5.8237) = 7.97

88th Percentile

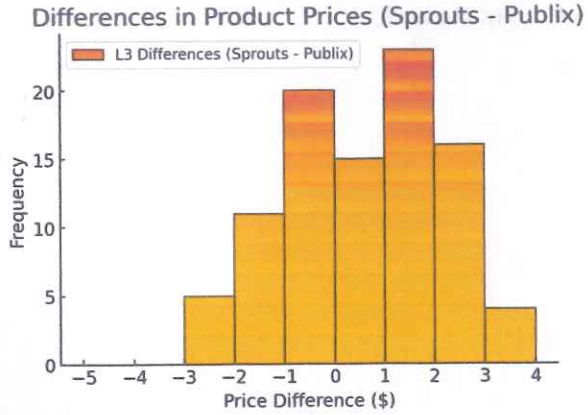
InverseNorm(0.88,0,1) = 1.1750

X= 10.2177 + (1.1750 × 5.8237) = 2.82

Difference (L3 = L1 - L2)

Sample Statistics

- **Sample Mean ($\bar{X}$):** 0.6944
- **Standard Deviation (Sx):** 1.8128



The distribution is approximately symmetric but approximately right-skewed.

35th Percentile

InverseNorm(0.35,0,1) = −0.3853

X = 0.6944 + (−0.3853 × 1.8128) = −0.0041

88th Percentile

InverseNorm(0.88,0,1) = 1.1750

X = 0.6944 + (1.1750 × 1.8128) = 2.82

Chapter 4: Write-Up/Math Analysis

Understanding pricing differences between grocery stores and studying public behavior at local piers are crucial topics in statistics. This project analyzes organic food prices at Sprouts and Publix and examines observational data collected at Clearwater Pier and St. Pete Pier. By employing **sampling methods, confidence intervals, and hypothesis tests**, we ensure that our conclusions are statistically valid. Furthermore, we explore **biases** in data collection, the reliability of conclusions, and how statistical significance affects real-world decision-making.

Data Collection Methods and Their Strengths and Weaknesses

The data for this project was collected using two distinct **sampling methods**, one for grocery store prices and another for pier observations. For the grocery store analysis, we selected random items from Sprouts and Publix, aiming for a broad representation of organic food products. This method allows for a general comparison but lacks true **random sampling** since the selection process depended on availability and researcher choice. One **pro** of this approach is its feasibility—data could be collected quickly and efficiently. However, a major con is the potential for **selection bias**, as certain products may be more or less expensive depending on store inventory on a given day.

For the pier data, an **observational study** was conducted, where one researcher stood in a fixed location and counted every other person who passed by at both Clearwater and St. Pete Piers. This method ensures **random selection**, as the sample is not pre-determined and represents a natural flow of foot traffic. A key advantage of this method is that it eliminates **response bias** since individuals were unaware they were being recorded. However, a **limitation** is the presence of **time-based variability**, meaning that traffic patterns could change depending on the time of day or the day of the week.

Statistical Analysis and Interpretation

To analyze the collected data, we examined **center, spread, and shape** using descriptive statistics and graphical representations. The **mean** price at Sprouts was found to be $10.91 with a **standard deviation** of 6.25, while Publix had a **mean** of $10.22 and a **standard deviation** of 5.82. Visual inspections using **histograms and boxplots** showed that both price distributions were **approximately normal but approximately right-skewed**, indicating a tendency toward higher-priced organic items.

For the pier data, we calculated the **proportion** of people bringing dogs to each location. At St. Pete Pier, 26% of people had a dog, while at Clearwater Pier, only 14% had a dog. These figures were used to compare proportions between the two locations using **hypothesis testing**.

Confidence Intervals and Hypothesis Testing

A **matched pairs t-interval** was used to estimate the true **mean difference** in organic prices at a **99% confidence level**. The resulting confidence interval (-2.642, 1.833) suggests that the true **mean difference** in prices could be either positive or negative, meaning we lack strong evidence that one store is consistently cheaper.

A **matched pairs t-test** was then conducted to determine whether Sprouts' organic prices were significantly higher than Publix's. The **null hypothesis (H_0)** assumed that there was no difference in mean prices, while the **alternative hypothesis (H_a)** suggested that Sprouts had higher prices. The calculated **p-value** of 0.3149 was greater than our **significance level ($\alpha = 0.05$)**, leading us to **fail to reject H_0**. This means we do not have enough statistical evidence to conclude that Sprouts charges significantly higher prices for organic food than Publix.

For the pier data, a **two-proportion z-test** was conducted to determine whether more people bring their dogs to St. Pete Pier than to Clearwater Pier. The **null hypothesis (H_0)** stated that there was no difference in the proportion of dog owners at each pier, while the **alternative hypothesis (H_a)** suggested that more people brought their dogs to St. Pete Pier. The **p-value** was calculated to be 0.06681, which is greater than 0.05, meaning we again **fail to reject H_0**. This result indicates that we lack strong statistical evidence that more people bring their dogs to St. Pete Pier than to Clearwater Pier.

Understanding Type I and Type II Errors

A **Type I error** occurs when we incorrectly reject a true **null hypothesis**, while a **Type II error** occurs when we fail to reject a **false null hypothesis**. For the grocery price analysis, a **Type I error** would mean concluding that Sprouts is more expensive when there is actually no difference. A **Type II error** would mean failing to detect a true difference in prices, which could lead consumers to make suboptimal shopping decisions. Similarly, for the pier data, a **Type I error** would mean incorrectly concluding that more people bring their dogs to St. Pete Pier, while a **Type II error** would mean failing to recognize a real difference in dog ownership rates at the two locations. The **consequences** of these errors can impact consumer decisions and public space policies, making it essential to minimize the risk of misinterpretation through careful study design and robust **statistical testing**.

Biases and Areas for Improvement

Several **biases** may have influenced our findings. **Observational bias** in the pier data collection may have affected the results if the sample was not truly representative of all visitors. Additionally, **undercoverage bias** is a concern in the grocery pricing study, as only two stores were included, potentially overlooking price variations at other retailers such as Whole Foods or Walmart. **Time-based**

variability is another potential issue, as grocery prices fluctuate, and pier traffic may be different at various times of the day or week.

To improve future studies, we recommend conducting **stratified random sampling** for organic product prices, where products are grouped by category before being randomly selected. Expanding the pier study to multiple days and time periods would help reduce **systematic bias**, leading to more reliable results.

Relevance and Broader Implications

The findings of this study are relevant to both consumers and policymakers. Organic food prices continue to be a key concern for shoppers, with research showing that organic items can cost 7% to 82% more than conventional alternatives (USDA, 2022). Our study provides valuable insight into the pricing differences between retailers, allowing consumers to make more informed decisions. Additionally, understanding dog ownership trends at public spaces like piers can help local governments develop policies for pet-friendly areas and infrastructure improvements.

Future Research and the Impact of a $1 Million Grant

If we were awarded a $1 million research grant, we would expand the study significantly. The pier observations would be conducted at multiple locations and at various times of the day to capture a more complete picture of visitor demographics. Grocery price comparisons would be extended to include major retailers like Whole Foods, Walmart, and Trader Joe's. Additionally, we would explore the impact of organic food pricing, investigating how affordability affects purchasing behavior among different income groups. With better data, we could provide stronger insights into consumer trends and policy recommendations, ultimately contributing to improved accessibility and affordability of organic food.

Through **statistical analysis**, we found no significant evidence that Sprouts consistently offers lower organic food prices than Publix. While our **confidence intervals** suggest possible differences, our **hypothesis tests** indicate that these differences are not statistically significant. Similarly, we found **no strong evidence** that more people bring their dogs to St. Pete Pier than Clearwater Pier. Future research should focus on **reducing bias, expanding the dataset, and improving data collection methods** to strengthen these findings. The results of this study provide valuable insights into consumer behavior and public space usage, emphasizing the importance of data-driven decision-making in everyday life.

Citations:

Organic Trade Association. (2023). U.S. Organic Industry Survey. Retrieved from www.ota.com

United States Department of Agriculture (USDA). (2022). Organic Food Pricing and Market Trends. Retrieved from www.usda.gov

Chapter 8-10:

1 Proportion z Interval

Find the 95% confidence interval for the percent of people who bring their dog to St. Pete Pier with them.

Parameter: p = population percent of people at the pier with their dog

Name: 1 proportion z confidence interval

See conditions page (1 Proportion z Test/Interval)

(0.1384, 0.3816)

$\hat{p} = .26$ $n = 50$

We are 95% confident our interval from 0.1384 to 0.3816 captures the actual population percent of people at the pier with their dog.

1 Proportion z Test

Is there convincing evidence that less than 45% of people bring their dog to St. Pete Pier?

Parameter: p = population percent of people at the pier with their dog

Name: 1 proportion z test

Hypothesis:

$H_O: p = 0.45$

$H_A: p < 0.45$

See conditions page (1 Proportion z Test/Interval)

$z = -2.7005$ $p = 0.0035$

$\hat{p} = .26$ $n = 50$

Is $p < \alpha$?

Is $0.0035 < 0.05$?

Yes

Our data is statistically significant. We reject the null. There is convincing evidence that the population percentage of people at piers that bring their dog is less than 0.45.

Matched Pairs Means t Confidence Interval

Find the confidence interval at the 95% of the average prices of organic food at Sprouts to Publix.

Parameter: $\mu_{Sprouts} - \mu_{Publix}$ = Population mean prices of organic foods at Sprouts to Publix

Name: Matched Pairs t Confidence Interval for means

See conditions page (Matched Pairs Mean t Test/Interval)

(-0.231, -0.0315)

$\bar{x} = -0.13125$

$Sx = 0.11934$

$n = 8$

We are 95% confident our interval from -0.231 to -0.0315 captures the true population mean difference of organic food prices at Sprouts to Publix.

Matched Pairs Means t Test

Is there convincing evidence that Sprouts costs more than Publix?

Parameter: $\mu_{Sprouts} - \mu_{Publix}$ = Population mean prices of organic foods at Sprouts to Publix

Name: Matched Pairs t test for means

Hypothesis:

$H_O: \mu_{Sprouts} - \mu_{Publix} = 0$

$H_A: \mu_{Sprouts} - \mu_{Publix} > 0$

See conditions page (Matched Pairs Mean t Test/Interval)

$t = -1.9234$

$p = 0.96682$

$\bar{x} = -0.4044$

$Sx = 1.0513$

$n = 25$

2 Proportion z Interval

Find the 95% confidence interval for the proportion of people who bring their dog with them to St. Pete Pier versus Clearwater Pier.

Parameter:

$P_{St.\ Pete}$ = Population percent of people at St. Pete pier with dogs

$P_{Clearwater}$ = Population percent of people at Clearwater pier with dogs

Name: 2 proportion z test

See conditions page (2 Proportion z Test/Interval)

(-0.0350, 0.2750)

$\hat{p}_{St.\ Pete} = 0.26$ $n_{St.\ Pete} = 50$

$\hat{p}_{Clearwater} = 0.14$ $n_{Clearwater} = 50$

We are 95% confident that our interval captures the true population percent difference of people who bring their dog with them to St. Pete Pier versus Clearwater Pier.

2 Proportion z Test

Is there convincing evidence that the population percent of people who bring their dog with them is greater at St. Pete Pier over Clearwater Pier?

Parameter:

$P_{St.\ Pete}$ = Population percent of people at St. Pete pier with dogs

$P_{Clearwater}$ = Population percent of people at Clearwater pier with dogs

Name: 2 proportion z test

Hypothesis:

H_O: $P_{St.\ Pete} = P_{Clearwater}$

H_A: $P_{St.\ Pete} > P_{Clearwater}$

See conditions page (2 Proportion z Test/Interval)

$z = 1.5$ $p = 0.06681$

$\hat{p}_{St.\ Pete} = 0.26$ $n_{St.\ Pete} = 50$

$\hat{p}_{Clearwater} = 0.14$ $n_{Clearwater} = 50$

$\hat{p} = 0.2$

Is $p < \alpha$?

Is $0.06681 < 0.05$?

No

We fail to reject the null. There is not convincing evidence that the population percent of people who bring their dog to the pier is greater at St. Pete Pier over Clearwater Pier.

Type II Error

If we made this error, then it means we failed to reject the null even though the alternate is correct. In specific terms, it means we said that the proportion of people who bring their dog to St. Pete Pier is equal to those who bring their dog to Clearwater Pier, even though the proportion of people who bring their dog to St. Pete Pier is greater than the proportion of people who bring their dog to Clearwater Pier.

Calculate N for Proportion

Find the minimum number of people needed to be surveyed with a $\pm 2\%$ margin of error at the 99% confidence level.

$N \geq (z^*/error)^2 pq$

$N \geq (2.576/0.02)^2\ (0.5)(0.5)$

$N \geq 4147.36$

We must round up, so the minimum number of people we need to survey is 4148.

1 Means T Interval

Find the 90% confidence interval of mean organic food prices at Sprouts.

Parameter: μ = Population mean prices of organic foods at Sprouts

Name: 1 Means t Interval

See conditions page (1 Means t Test/Interval)

(4.6767, 6.5809)

$\bar{x} = 5.6288$

$Sx = 2.7824$

$n = 25$

We are 90% confident that our interval captures the true population mean prices of organic foods at Sprouts.

1 Means T Test

Is there convincing evidence that the mean price of organic foods is greater than $7?

Parameter: μ = Population mean prices of organic foods at Sprouts

Name: 1 Means t Interval

Hypothesis:

$H_O: \mu = \$7$

$H_A: \mu < \$7$

See conditions page (1 Means t Test/Interval)

$t = -2.4641$

$p = 0.0106$

$\bar{x} = 5.6288$

$Sx = 2.7824$

$n = 25$

Is $p < \alpha$?

Is $0.0106 < 0.05$?

Yes

We reject the null. There is convincing evidence that the population mean prices of organic foods at Sprouts are less than $7.

2 Means T Interval

Find the 99% confidence interval of mean organic food prices at Sprouts vs Publix.

Parameter: $\mu_{Sprouts}$ = Population mean prices of organic foods at Sprouts

μ_{Publix} = Population mean prices of organic foods at Publix

Name: 2 Means t Interval

See conditions page (2 Means t Test/Interval)

(-2.6420, 1.8331)

df = 47.4348

$\bar{x}_{Sprouts} = 5.6288$ $\bar{x}_{Publix} = 6.0332$

$Sx_{Sprouts} = 2.7824$ $Sx_{Publix} = 3.1046$

$n_{Sprouts} = 25$ $n_{Publix} = 25$

We are 99% confident that our interval captures the true population mean difference of organic food prices at Sprouts versus Publix.

2 Sample T Test

Is there convincing evidence that the mean prices of organic foods are less at Sprouts versus Publix?

Parameter: $\mu_{Sprouts}$ = Population mean prices of organic foods at Sprouts

μ_{Publix} = Population mean prices of organic foods at Publix

Name: 2 Means t Test

Hypothesis:

Ho: $\mu_{Sprouts} = \mu_{Publix}$

$H_A: \mu_{Sprouts} < \mu_{Publix}$

See conditions page (2 Means t Test/Interval)

$t = -0.4850$

$p = 0.3149$

$\bar{x}_{Sprouts} = 5.6288$ $\bar{x}_{Publix} = 6.0332$

$Sx_{Sprouts} = 2.7824$ $Sx_{Publix} = 3.1046$

$n_{Sprouts} = 25$ $n_{Publix} = 25$

Is $p < \alpha$?

Is $0.3149 < 0.05$?

No

We fail to reject the null. There is not convincing evidence that the population mean difference of organic food prices are less at Sprouts versus Publix.

Calculate N for Means

Find the minimum number of organic foods needed to be surveyed to find the true mean prices of organic foods at Publix with a ±5% margin of error at the 90% confidence level.

$N \geq (z^*\sigma/error)^2$

$N \geq (1.645(3.1046)/0.05)^2$

$N \geq 10404$

The minimum number of organic foods we'll need in our samples is 10404.

A+

[redacted]

Cumulative AP test 2

☆ #1

$IQR = 40 - 18 = 22$ $Q_1 - (1.5 \times IQR)$

$1.5 \times 22 = \boxed{33}$

answer: a

notes: 5 number summary = [min Q_1 med Q_3 max], outlier test = $Q_1 - (1.5 \times IQR)$, upper = $Q_3 + (1.5 \times IQR)$, $IQR = Q_3 - Q_1$

☆ #2

answer: d

notes: $.2500 + .3570 + .2500 + .0625 = \boxed{.9375}$

probability is any outcome of a random process is a number between 0 and 1. The compliment of an outcome is the opposite, meaning 1 - outcome = compliment.

☆ #3

answer: E

: parameter: μ = population mean of IQ scores

1) random selection stated
2) $200(10) < 10\%$ of pop. of adults
3) $200 > 30$ CLT applies

$\boxed{N(122, \frac{5}{\sqrt{200}})}$ 35316

notes: when determining if a distribution is approx. normal you have to follow a flow chart of parameter:, 3 conditions, N( ), z-score.

✓ #4

answer: b

notes: a simulation imitates a random process in such a way that simulated outcomes are consistent with real-world outcomes

✓ #5

answer: C

notes: when calculating consecutive probability you multiply the probability of event 1 by the probability of event 2 by the # of ways possible.

X #6

answer: ~~c~~ → d

notes. when generalizing a population, you can only generalize to those who took the poll, but you generalize to the largest population possible

✓ #7

answer: C

invnorm(.05) = -1.6449

$-1.6449 = \frac{x-450}{20} \rightarrow -32.898 = x - 450 \rightarrow x = 417.102$

↳ 418

notes: invnorm is used to work backwards from a z-score problem, you input the percentile which is the percent equal to or worse than your score.

★ #8

answer: a

notes: a block design is creating groups based on a similarity that are known prior to the study that is expected to impact the response to the treatment.

X #9

answer: ~~b~~ → d

notes: conditional probability is the probability one event happens given that another event that is known to have happened

equation = $\frac{P(\text{both events occur})}{P(\text{given event occurs})}$

☆ #10
answer: C
notes: The least square regression line (LSRL) is the line that makes the sum of the squared residuals as small as possibe
Equation = $\hat{y} = a + bx$

☆ #11
answer: b
notes: a distribution is skewed if one side of the graph is much longer (has a tail) compared to the other side. You can also look at the mean compared to the median.

☆ #12
answer: C
notes: if the mean = median the data is symetric, mean > median the data is skewed right, mean < median data is skewed left

✓ #13
answer: a $11/177 = .0621$
notes: percentile is the percent of people whose score is equal to or worse than yours. This can be calculated on a dot plot by doing dot # / total # of dots.

X #14
answer: b → C
notes: correlation (r) is the linear association between 2 quantitative variables which measures the strength and direction of the association.

✓ #15
answer: d
notes: A treatment is what is being applied to an individual, experimental unit is what object the treatment is applied to, a response variable measures the outcome of the study.

☆ #16
answer: C
notes: Z-score can be calculated using the equation z = value - mean / standard deviation. Sd can be calculated using $\sqrt{\text{proportion(compliment / sample size}}$

☆ #17
answer: E
notes: The z-score for an individual in a distribution tells us how many standard deviations from the mean the value falls and which direction it is.

☆ #18
answer: a $233 + 252 - 116 = 369$
notes: The general addition rules states that if it is an "or" problem we can use the equation: P(A or B) = P(A) + P(B) - P(A and B), these values can be found on the two-way table.

✓ #19
answer: c $.3^2 + .1^2 + 3^2 = 9.34 \rightarrow \sqrt{} \rightarrow 3.02$
notes: There is no equation to add the sd's, but we can use the variance equation of $\sigma^2 = \sigma_x^2 + \sigma_y^2$, and then square root that answer.

X #20

answer: ~~D~~ → E

notes: The expected value = $x(p) + y(p) + z(p) \ldots$, while the sd $= \sqrt{\theta^2}$, the variance = $x^2(p) + y^2(p) + z^2(p) \ldots - \mu^2$, the sd does not have a specific formula so you have to get variance first.

☆ #21

answer: @

notes: A cumulative frequency graph plots a point corresponding to the percentile of a given value in a distribution of quantitative data.

mcq: 17/21

☆: 11

✓: 6

X: 4

frq

#22

a) This is an observational study because there was no treatment added, there was only an examination of the health records

b) Confounding can happen due to another variable causing the outcome, meaning theres no way to determine if the treatment or cause is actually working. In this case a confounding variable could be adult diet or location of residence.

c) No, nothing can be determined as this was observational so no treatment was added and there is no control to base this statement off of.

#23

a) $\hat{y} = a + bx$

Estimated # of wins = 25.166 − .003131(yards allowed)

b) Estimated # of wins = 25.166 − .003131(4666)

= 11.0445

residual = actual − predicted

10 − 11.0445

= −1.0445, overpredicted

c) This point is an outlier meaning if we were to remove this point it would increase the correlation. This would cause the slope to increase and the y-int. to increase in terms of the LSRL equation.

#24

a) Median - The median for Cicada plants and control plants are exactly equal at .25 milligram seed mass

Unusual - Cicada: IQR = .28 − .22 = .06

.17 > .22 − (1.5 · .06) = .13

.35 < .28 + (1.5 · .06) = .37

no outliers

Control: IQR = .26 − .19 = .07

.14 > .19 − (1.5 · .07) = .085

.29 < .26 + (1.5 · .07) = .365

no outliers

Shape - The shape of Cicada plants is approx. symmetric, while the shape of the control is approx skewed left.

Spread - Cicada: IQR = .28 − .22 = .06 < Control: IQR = .26 − .19 = .07

range = .35 − .17 = .18 > range = .29 − .14 = .15

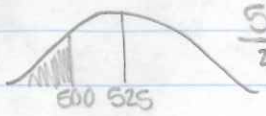
b) The Cicada plants distribution likely has a larger mean since the control group is approx skewed left, the tail would pull the mean down.

c) random assignment in this study is used to decrease the bias as much as possible to create a completely randomized study.

d) One benefit is that since only one plant was used we can for sure make a conclusion about cicadas impact as we have a control. But we cant determine the impact of cicadas on other plants because other plants may react differently.

#25

a) $N(525, 200/\sqrt{50})$



$\frac{500-525}{200/\sqrt{50}} = -.8839 \longrightarrow .1884$, probability the average # of pages is less the 500

b) Binomial $\leq$ (more / fewer)

Independant

$N = 50$

$P = .30$

$X = 20$

1-binomcdf(50, .3, 19)

$= .0848$, probability atleast 20 novels have fewer than 400 pages.

CHRIS STAPLETON



		seconds
1	Tennessee Whiskey	271.2
2	You should probably leave	199.8
3	millionare	197.4
4	Starting over	215.4
5	White horse	256.2
6	Cold	305.4
7	Broken Halos	180
8	Joy of my life	260.4
9	Fire away	242.4
10	What are you listening to?	243
11	What am I gonna do	180.6
12	Loving ~~you~~ on my mind	197.4

TRIBAL SEEDS

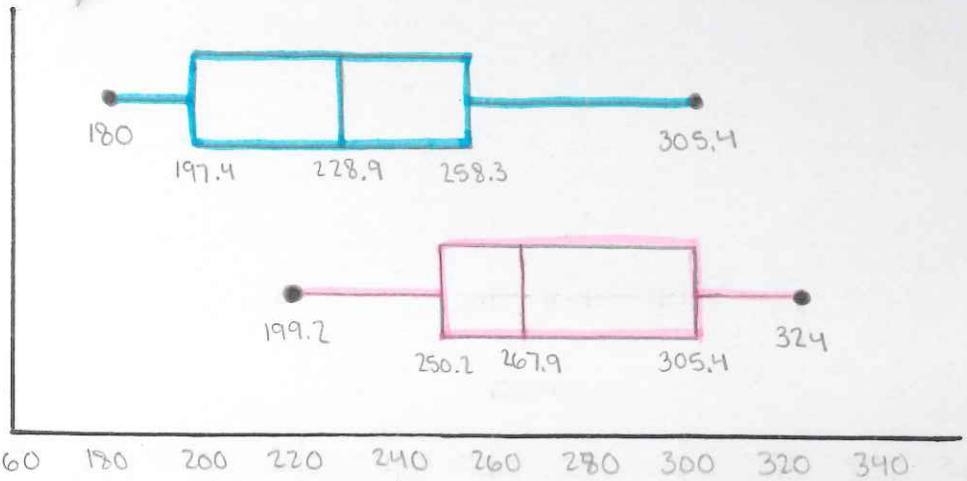


		seconds
1	In your eyes	304.2
2	Dawn of time	249.6
3	The garden	274.2
4	Ruined	199.2
5	moonlight	324
6	Island Girl	250.8
7	Rock the Night	274.2
8	Rude girl	258
9	Deep blue sea	207
10	Tempest	309.6
11	Aroma	261.6
12	The Harvest	306.6

BOX PLOTS

Chris Stapleton

Tribal Seeds



MUSS

Tribal seeds median for length of songs in seconds is 267.9, greater than Chris S., which had a median of 228.9.

Chris S.
$180 < 197.4 - (1.5 \times 60.9)$
$180 < 106.05$ NOT
$305 > 258.3 + (1.5 \times 60.9)$
$305 > 349.65$ NOT

tribal seeds
$199.2 < 250.2 - (1.5 \times 55.2)$
$199.2 < 167.4$ NOT
$324 > 305.4 + (1.5 \times 55.2)$
$324 > 388.2$ NOT

Neither distributions have any outliers.

Chris S.
Sum of all # - 3219
$3219 / 12 = 268.25$
$268.25 > 228.9$
mean median

tribal seeds
Sum of all # - 2749
$2749 / 12 = 229.08$
$229.08 < 267.9$
mean median

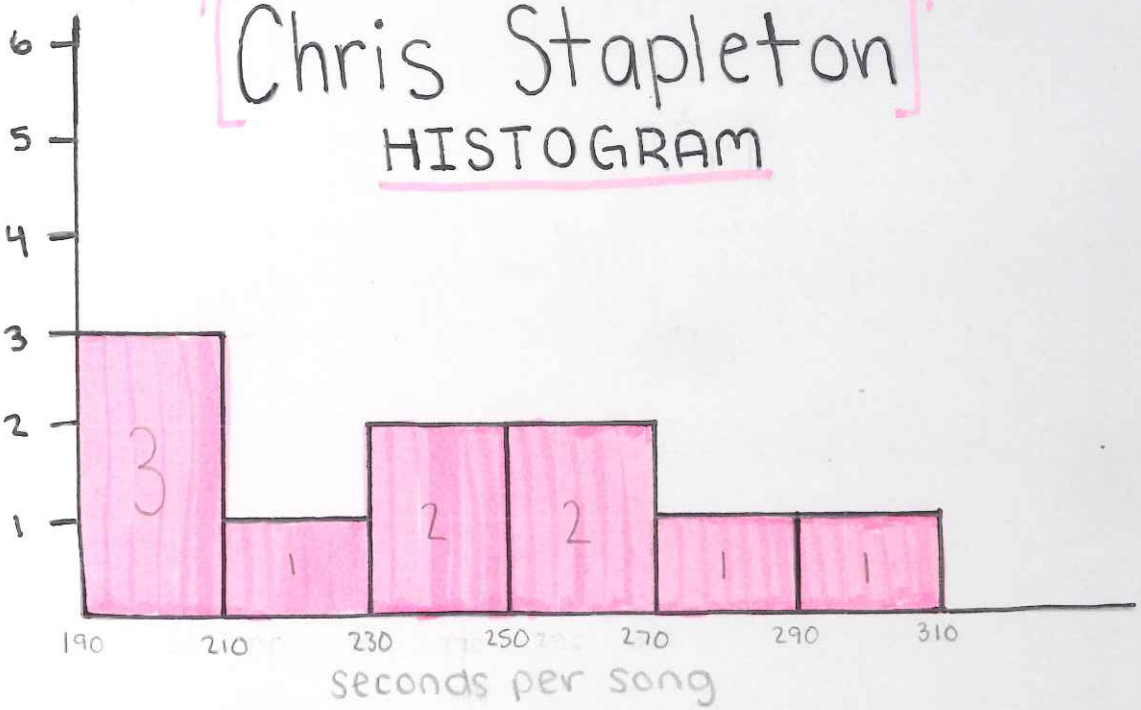
Tribal S. data is approximately skewed left, while Chris S. distribution is approximately skewed right.

Tribal seeds IQR = 55.2
(305.4 - 250.2)

Chris S. IQR = 60.9
(258.3 - 197.4)

Chris Stapletons range of song length is greater than tribal Seed's range. $60.9 > 55.2$ seconds

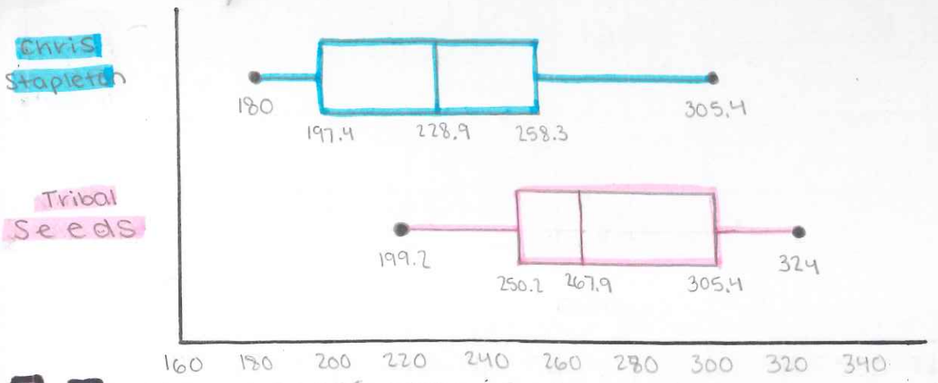
Chris Stapleton
HISTOGRAM



Tribal Seeds fun fact-
I saw them in concert over summer. 10/10 show.

Chris Stapleton fun fact- 8 total Grammy awards.

BOX PLOTS



	Chris Stapleton		Tribal Seeds		
mean	268.25	$\frac{3219}{12}$	229.08	$\frac{2749}{12}$	add all #, ÷ 12
median	228.9		267.9		middle #
Range	125.4	305.4 − 180	124.8	324 − 199.2	(highest − lowest)
IQR	60.9	258.3 − 197.4	55.2	305.4 − 250.2	($Q_3 - Q_1$)
SD	39.883		39.280		
variance	1590.65		1542.92		(s^2)

MUSIC

Tribal seeds median for length of songs in seconds is 267.9, greater than Chris S., which had a median of 228.9.

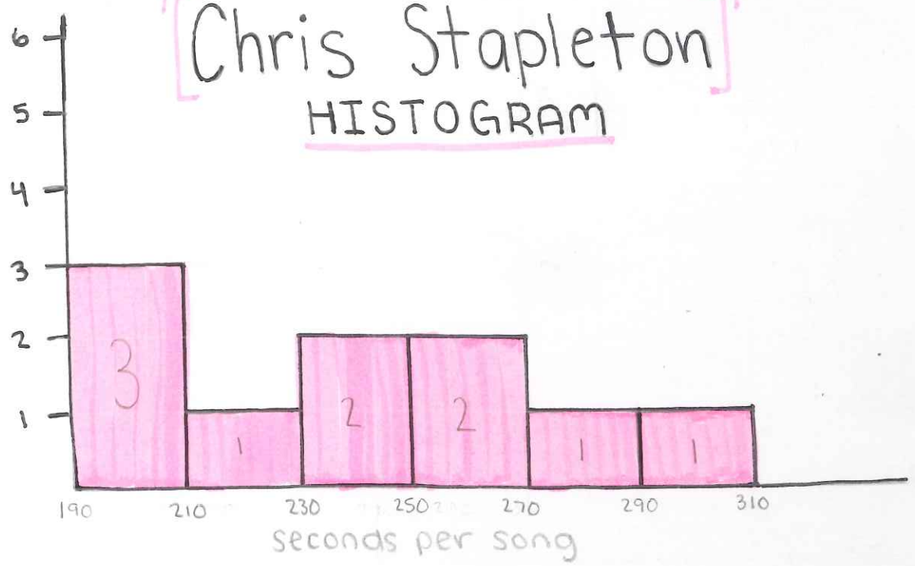
chris S.	tribal seeds	
$180 < 197.4 - (1.5 \times 60.9)$	$199.2 < 250.2 - (1.5 \times 55.2)$	Neither distributions have any outliers.
$180 < 106.05$ NOT	$199.2 < 167.4$ NOT	
$305 > 258.3 + (1.5 \times 60.9)$	$324 > 305.4 + (1.5 \times 55.2)$	
$305 > 349.65$ NOT	$324 > 388.2$ NOT	

Chris S.	tribal seeds	
Sum of all # − 3219	Sum of all # − 2749	Tribal S. data is approximately skewed left, while Chris S. distribution is approximately skewed right.
3219 / 12 = 268.25	2749 / 12 = 229.08	
268.25 > 228.9	229.08 < 267.9	
mean median	mean median	

Tribal seeds IQR = 55.2 (305.4 − 250.2)

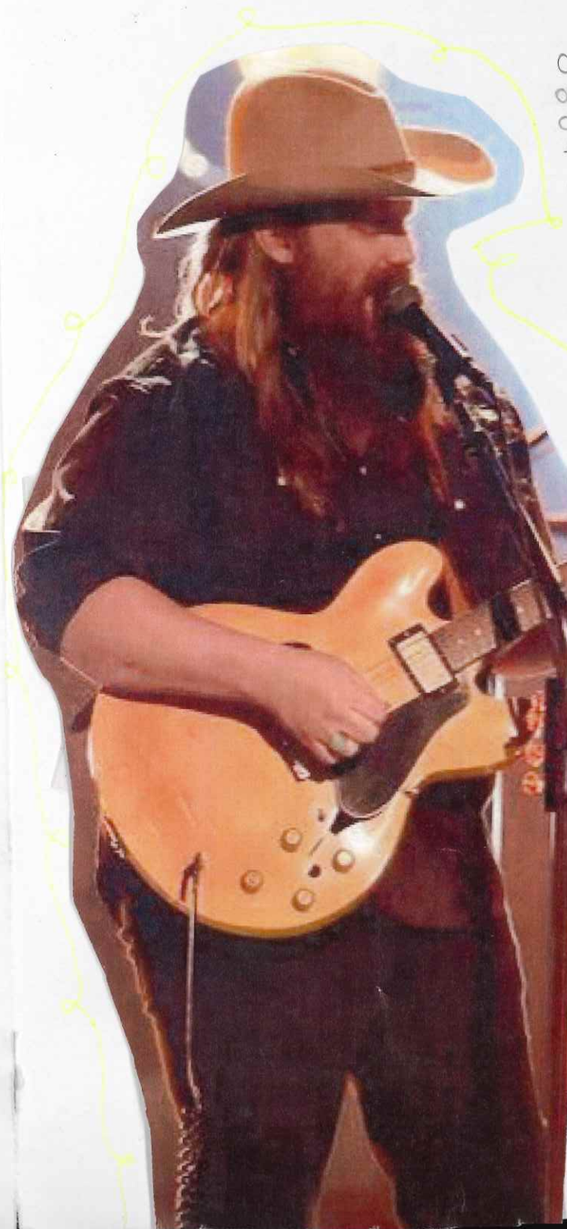
Chris S. IQR = 60.9 (258.3 − 197.4)

Chris Stapletons range of song length is greater than tribal Seed's range. 60.9 > 55.2 seconds



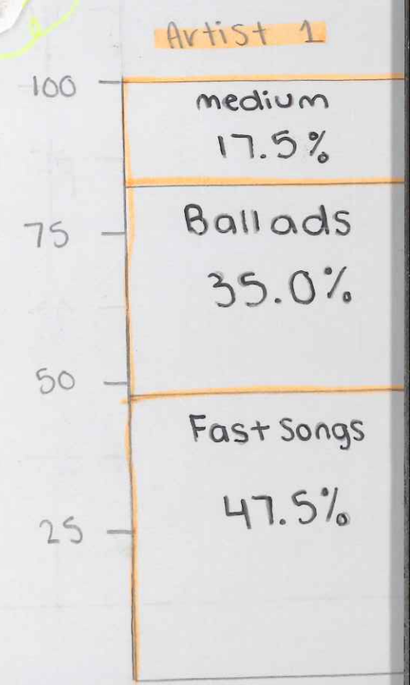
Tribal Seeds fun fact - I saw them in concert over summer. 10/10 show.

Chris Stapleton fun fact - 8 total Grammy awards.



INTERPRET

Chris Stapleton - The range of the middle 50% of values is 60.9, in song length (seconds). Standard deviation - 39.88 is the average distance from the mean of 268.25 seconds in terms of song length.



Tribal Seeds		
229.08	$\frac{2749}{12}$	add all #, ÷ 12
267.9		middle #
124.8	324 − 199.2	(highest − lowest)
55.2	305.4 − 250.2	($Q_3 - Q_1$)
39.280		
1542.92		(S^2)

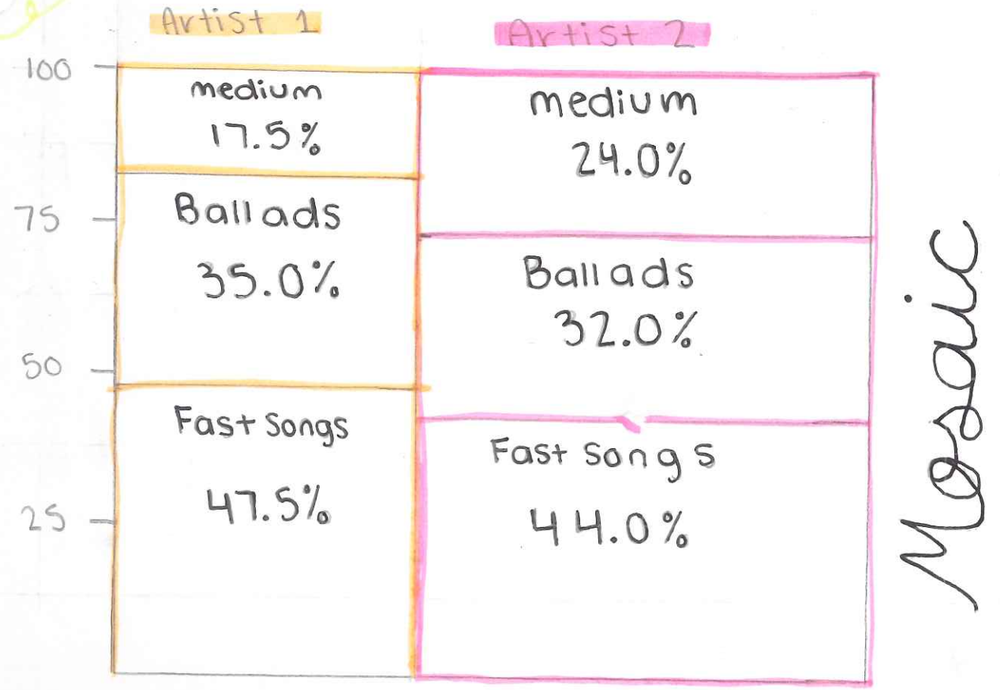
INTERPRET

Chris Stapleton - The range of the middle 50% of values is 60.9, in song length (seconds). Standard deviation - 39.88 is the average distance from the mean of 268.25 seconds in terms of song length.



Tribal Seeds - The range of the middle 50% of values is 55.2, in terms of song length in seconds (IQR). Standard deviation - 39.28 is the average distance from the mean of 229.08 in terms of song length in seconds.





Mosaic

STEM PLOT

Key:
9/1 = 190 secs.
* All in hundreds

88	1	9
999	1	
441	2	4550
765	2	677
0	3	00
	3	02

TRIBAL SEEDS

Key:
2/6 = 260 sec

CHRIS STAPLETON

OUTLIER TEST

Chris Stapleton
lower: $180 < 197.4 - (1.5 \times 60.9)$
$180 < 106.05$ NOT TRUE
higher: $305 > 258.3 + (1.5 \times 60.9)$
$305 > 349.65$ NOT TRUE

Tribal Seeds
lower: $199.2 < 250.2 - (1.5 \times 55.2)$
$199.2 < 167.4$ NOT TRUE
higher: $324 > 305.4 + (1.5 \times 55.2)$
$324 > 388.2$ NOT TRUE

NO OUTLIERS

