**FORMULAS:**

 Median:     

z-score:  Pearson’s  

**STAT ESSENTIALS:** Be able to: 1) obtain the five-number-summary; 2) build a modified box plot – including identifying adjacent points (if any); 3) determine variability measures: skew (Pearson’s I), Coefficient of Variability, z-score. **[NOTE: See Orange sheets for additional problems and resources.]**

**PROBLEMS:**

**1)** Daily saturated fat intakes (in grams) of a sample of people are as shown below. Construct a box plot of these data. Determine the five-number-summary, IQR, Lower Limit, Upper Limit, and Adjacent Points (if any). Determine if these data are approximately normally distributed. Identify outlier values.

38, 32, 34, 39, 40, 54, 32, 17, 29, 33, 57, 40, 25, 36, 33, 24, 42, 16, 31, 33

[Larson 4th edition, page 53]



IQR: 10.251g

LL: 14.125g

UL: 55.125g

Adj. Pt: 54g

Outliers: 57g

Skew via Pearson’s I: .37 => approx.. normally dist.

(SPSS uses a different formula)

**2)** Pungencies (in 1000s of Scoville units) of 24 tobasco peppers are shown below. Construct a box plot of these data. Determine the five-number-summary, IQR, Lower Limit, Upper Limit, and Adjacent Points (if any)

35, 51, 44, 42, 37, 38, 36, 39, 44, 43, 40, 40, 32, 39, 41, 38, 42, 39, 40, 46, 37, 35, 41, 39

[Larson 4th edition, page 52]

All in Scoville Units

IQR: 4.75 LL: 30.125 UL: 49.125 Adj. Pt: 46 Outlier: 51

**3)**  Determine which of the previous two variables (saturated fat and pungency) demonstrates the greatest variability.

CVAR Fat = 29.56% CVAR Peppers = 10.02%

Sat Fat is more variable



**4)** Explain what this modified box plot is presenting. Why does it look like this?

**ANSWER:** Large number of data values located at the same point such that the minimum,Q1 and Q2 are at the same location.

**5)** Reaction times (in milliseconds) of a sample of 30 adult females to an auditory stimulus are shown below. Determine if these data are approximately normally distributed. How many standard deviations away from the mean is a reaction time of 359 milliseconds?

Data: 507, 389, 305, 291, 336, 310, 514, 442, 307, 337, 373, 428, 387, 454, 323, 441, 388, 426, 469, 351, 411, 382, 320, 450, 309, 416, 359, 388, 422, 413

[Larson 4th edition, page 52]

**Descriptive Statistics: 5) Reaction Times (milliseconds)**

Variable N Mean StDev Minimum Q1 Median Q3 Maximum

Reaction Time 30 388.3 61.2 291.0 332.8 388.0 431.3 514.0

SKEW: I = [3(388.3-388)]/61.2 = .0147

standard deviations below the mean

**6)** Trade winds are one of the beautiful features of island life in Hawaii. The following data represent total air movement in miles each day over a weather station in Hawaii as determined by a continuous anemometer recorder. The period of observation was January 1 to February 15, 1971. Create a box plot of these data and identify any outliers that exist in this data set.

****

In Miles/day

**IQR: 18.25**

**LL: -11.375**

**UL: 61.625**

**Adj Pt: 57**

**Outlier: 72**

**and**

**Extreme**

**Outliers: 100, 105, 113,,138**

**(more than Q3+3(IQR) out)**

**7)** Box plots divide a data set into four equal quarters. Review the trade winds box plot. The four quarters don’t appear “equal.” Why?

**ANSWER:** Each quarter contains 25 % of the data. However, in some sections (quarters) the data values are close together and in others they are farther apart. So, the range for each quarter varies.

**8)** The Jefferson Valley Bank wanted to determine if they could best serve their customers by having waiting queues at each teller's window or by having a single waiting queue from which patrons would go to the next available teller. Data were obtained for each queuing approach and are noted below. Review the two box plots and discuss which appears to be the “better” approach.

**Single Queue** Data (minutes): 6.5, 6.6, 6.7, 6.8, 7.1, 7.3, 7.4, 7.7, 7.7, 7.7

**Individual Queues** Data (minutes): 4.2, 5.4, 5.8, 6.2, 6.7, 7.7, 7.7, 8.5, 9.3, 10.0



**ANSWER:** Both have the same mean, median, and mode times. The single queue would get customers served within a small range of time, so there is a consistent flow, but the likelihood of quick service is not there. The individual lines could result in quick service, or a lengthy wait. The decision then moves to a corporate decision well above an adjunct’s lecturer’s pay grade.

**9)** For a recent year, the number of murders in 25 selected cities is shown. Construct a modified box plot for these data. Determine if these data are skewed and the number of standard deviations the minimum value is from the mean.

Data; 248, 348, 74, 514, 597, 270, 71, 226, 41, 39, 366, 73, 241, 46, 34, 149, 68, 73, 63, 65, 109, 598, 278, 69, 27

Variable N Mean StDev Minimum Q1 Median Q3 Maximum

Murder s 25 187.5 177.4 27.0 64.0 74.0 274.0 598.0

Source: *Pittsburgh Tribune Review*
[Bluman 6th edition, page 90]

Skew via Pearson’s I: 1.91 (SPSS uses a different formula)

Distance from mean in standard deviations: 

Murders in a given year

IQR: 210

LL: -251

UL: 589

Adj. Pt: 514

Outliers: 597, 598

