**FORMULAS: See prior handouts**

**STAT ESSENTIALS: See prior handouts. [NOTE: see Orange sheets for additional problems and resources.]**

**­­­**

**PROBLEMS:**

**1)** The following table contains statistics about 200 eruptions of the Old Faithful geyser. Assume these data to be approximately normally distributed.



1A) The time between the most recent two eruptions was 93 minutes. Determine the proportion of times between eruptions in the sample that occurred in less than 93 minutes.

1B) What time separates the quickest 10% of times between eruptions from the others?

1C) What time separates the slowest 1% of times from the rest?

1D) What proportion of the recorded times fall within +/- 2 minutes of the mean?

**1A)**  = 1.25; area to left = **.8944 DRAW A PICTURE!**

**1B)**  = 74.27 + (-1.28)(15.03) = **55.03 minutes**

WHY? THE fastest times between eruptions are the shortest times - those at the left end of the distribution. So how did we get this when we do not know x or z? You are given an area from which you can obtain z. Then solve for x.

**1C)**  = 74.27 + (2.33)(15.03) = **109.28 minutes (**Same process as 1B))

**1D)** Need to find the area between two values.

 = = .13 => area = .5517

 = = -.13 => area = .4483 => difference = **.1034**

**2)** Old Faithful revisited:

2A) Given the information provided for question #1, what is the probability that a randomly selected eruption time would exceed 100 minutes?

2B) Given the information provided for question #1, what is the probability that a randomly selected sample of 10 eruption times would exceed 100 minutes?

2C) Calculate the mean and standard deviation for all samples of size 200.

**2A)** Draw a picture.  =  = 1.72 => area = .9573 to the left 1 - .9573 = **.0427,** the slowest times

**2B)**  =  = 5.41 => area to **RIGHT** is essentially = **0**

**2C)** mean = population mean = 74.27 min.; standard deviation = 1.06 min.

**3)** Good grief, Old Faithful revisited, yet again: Referring to the data table presented for question #1, determine a 95% confidence interval for these data.

**3)** Confidence Interval: **GIVEN**: n = 200; mean = 74.27 min.; s.d. = 15.03 min.,  .05; 95% Conf. Level.

 From the given we find: /2 = .025; z = 1.96; large sample with sample s.d.

  =  = 74.27  2.08 minutes

 **95% C.I. = (72.19, 76.35) minutes**

**4)** Tennis Ball Manufacturing: A company manufactures tennis balls. When its tennis balls are dropped onto a concrete surface from a height of 100 inches, the company wants the mean height of a ball’s bounce upward to be 55.5 inches. This average is maintained by periodically testing random samples of 25 tennis balls. If a 99% confidence interval contains the desired bounce height (55.5 inches), the company will be satisfied that it is manufacturing acceptable tennis balls. A sample of 25 tennis balls is randomly selected and tested. The mean bounce height of the sample is 56 inches and the standard deviation is .25 inches. Is the company making acceptable tennis balls? Explain your reasoning

**4) GIVEN:** n = 25; mean = 56.0 in.; s.d. = .25 in, 99% C.L. As this is a small sample, use the t formula.

 From the given t = 2.797

  =  =  => **99% C.I. = (55.86, 56.14)**

So is the production meeting the desired bounce height of 55.5 inches? NO. Why not?

**5)** The information provided in the following tables and charts pertain to the age of cash registers and their cost of repair. Determine the level at which the correlation is significant. Identify the equation of the regression line. Assume that a cash register is 4.5 years old. Estimate the cost of repair for this cash register. Assume that the cash register is 12 years old. Estimate the cost of repair for this cash register. (“It cannot be calculated” is an incorrect answer because there is a value that can be used as the estimate.) [NOTE: In the Coefficients table under the “B,” the Constant is the y-intercept and the Age value is the slope.]







**5) Cash Registers:** Correlation is significant at the = .01 level.

Regression:  =>  =>**Cost (est.) = 58.489 + 19.915(Age)**

**Cost (est.) = 58.489 + 19.915(4.5)** = **$148.11**

A 12 year old cash register falls outside the range of available data. Therefore one cannot use the regression equation. The next best estimate is the mean for the cost of repair = **$150.10**

**![MCAN04184_0000[1]]()6)** Tuna: A buyer for the *Let’s Eat a Fish* seafood company is interested in purchasing a catch of 1,030 albacore tuna. Small fish are more expensive to process, yield less edible tuna, and thereby reduce profits. To avoid purchasing a catch of small tuna the buyer decides that he will buy the catch only if no more than 5% would weigh less than 18 pounds. The buyer knows that the weight of mature albacore tuna is approximately normally distributed with an average weight of 20 pounds and a standard deviation of 1.25 pounds. Based upon the information provided to the buyer, what should be his purchasing decision? Show how you came to your recommendation.

Proportion of tuna weighing less than 18 pounds: \_\_\_\_\_\_\_\_\_\_\_

Estimated number of tuna weighing less than 18 pounds: \_\_\_\_\_\_\_\_\_\_\_\_\_

DECISION: BUY \_\_\_\_\_\_ DO NOT BUY \_\_\_\_\_\_\_

**6) Tuna:** z-score = -1.6 with associated area of .0548; # Tuna = 1030\*.0548 = 56.44; as .0548 represents more than 5% (.05) do not buy.

**7)** The following table lists the average number of hours per year that a driver is delayed by road congestion in selected cities. Create a modified box plot of these data.

CITY HOURS CITY HOURS CITY HOURS

Los Angeles 56 Dallas 46 St. Louis 44

Atlanta 53 Washington 46 Orlando 42

Seattle 53 Austin 45 Oneonta 1

Houston 50 Denver 45

**7) Traffic Delays (hours):** Min: 1 Q1: 44 Q2: 46 Q3: 53 Max: 56

IQR: 9 L. Limit: 30.5 U. Limit: 66.5 Adj. Pt. 42

 BOX PLOT TITLE: Road congestion delays in selected cities

**8)** Hypothesis test for Lodging Costs: A travel association says the daily lodging costs for a family in the United States is $152. You work for a tourist publication and want to test this claim. Ten U.S. families are randomly selected and it is found that for an overnight trip they had a mean expenditure of $142.80 with a standard deviation of $37.52. At = .02, can you reject the travel association’s claim?



The mean lodging costs/day = $152 CLAIM

Reject H0

Reject H0

 ≠ 152 The mean lodging costs/day not equal $152

Given: = $142.80 = $152 s = $37.52 n = 10  = .02

Retain H0

2-tailed test

 = -0.775 (Test Stat.) Critical Value = -+/- 2.821

T.S.-0.775

C.V.+/-2.821

Conclusion: Retain null. At the = .02 sig. level the data do not provide sufficient evidence to conclude that the lodging costs/day differ from $152.

**9) Dive Depth:** An oceanographer claims that the mean dive depth of a North Atlantic right whale is 115 meters. A random sample of 34 dive depths has a mean of 121.2 meters and a standard deviation of 24.2 meters. Is there enough evidence to reject the claim at  = 0.10? [Larson 6th ed., p. 385, #25]



The mean dive depth = 115 meters. CLAIM

Reject H0

Reject H0

 ≠ 115 The mean dive depth ≠ 115 meters.

Given: = 121.2 m = 115 m s = 24.2 m n = 34  = 0.10

Retain H0

2-tailed test

 = 1.49 (Test Stat.) Critical Valu+e = -+/- 1.645

T.S.= -1.49

C.V.+/-1.645

Conclusion: Retain null. At the = .10 sig. level the data do not provide sufficient evidence to reject the claim that the mean dive depth of a North Atlantic right whale is 115 meters.