**QMB-6305**

**Take-Home Mid-Term Exam**

**Fall, 2018**

***Dr. Ronald K. Satterfield***

1. You have the option of working in a small group on this exam. However, groups may be no larger than three.

2. The deliverable for this assignment is a **single** MS Office-compatible file, spreadsheet or document. (No pdf or zip files, please.)

3. Each group can turn in one submission. Make certain the names of all group members deserving credit appears on the opening page of your submission.

4. Your submission will be made via Canvas. Please do not email exams directly to your professor.

5. This exam is due via electronic submission by **September 23, 2018 at 11:59 PM.**

1. Parking in New Orleans can be challenging, but no time of year is more challenging for drivers in the city than during Mardi Gras. Mardi Gras in New Orleans is not a single day but rather a season running for roughly a month prior to “Fat Tuesday”, the day before Ash Wednesday and the start of the Lent. Mardi Gras parades occur on designated routes through the city with the most well-attended along routes through the famous French Quarter and the Garden District. While French Quarter parades can be raucous and attract a large number of tourists, Garden District crowds are much more local, sedate, family-oriented, and well-behaved.

Let’s say the New Orleans Police Department gathered data for Mardi Gras 2018 on the time required to tow away illegally parked vehicles discovered in the French Quarter and the Garden District. (In 2018 Mardi Gras day fell on February 13.) This data is given in the embedded spreadsheet below and shows the intervals of time in minutes between the issuance of a citation by a NOPD officer and the time the vehicle is picked up by a tow truck and removed to impound. The data was gathered for the three-day period covering the day before Mardi Gras, Mardi Gras day itself, and the day after Mardi Gras. The data was gathered from records kept by the NOPD and derived from the time the original citation was written and the time a tow truck seized the offending vehicle.



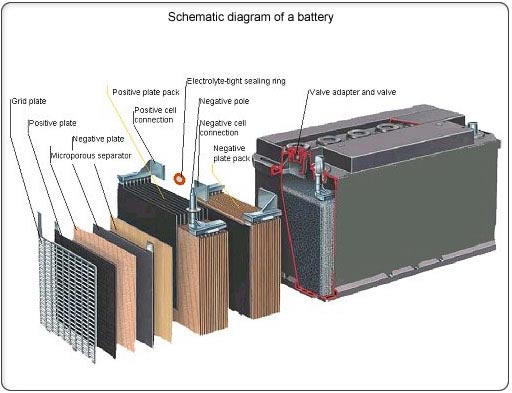
a. There are six different sets of data given, accounting for the three days and two neighborhoods in question. Using methods discussed in our course, do you believe any of these six data sets shows skewed data? Are there other anomalies in the data you think are important? **(5 pts.)**

b. Identify any points you believe may be outliers among the six data sets. **(5 pts.)**

c. Combine the data by neighborhood so you have only two variables, one for the French Quarter and one for the Garden District. Can you say with statistical certainty the population mean time intervals are different between the two neighborhoods? **(5 pts.)**



d. Recombine the data such that you have three variables, one for each day under study. Using the tools of statistical analysis can you say the data among the three different samples follows different statistical distributions? **(5 pts.)**

1.  A manufacturer of automobile batteries believes the average length of life for its entire population of Grade A battery is 77 months. However, the full-replacement guarantee on this brand is for just 36 months. Suppose the standard deviation of the life length of the population is known to be 21 months and the population distribution for life length is normal. The sales forecast is for 7,500,000 batteries next year.

a. Approximately how many of the 7,500,000 batteries would we expect to fail between their fifth and eighth year of life? (This would be between month 60.000 and month 96.000.) **(8 pts.)**

b. Of the 7,500,000 forecasted to be sold next year (2019) how many batteries would we expect to be returned for a full replacement during the 36-month guarantee period? **(7 pts.)**

c. Competition is forcing this manufacturer to increase its guarantee from 3 years to 4 years. If the direct cost to the company is $85 to replace a defective battery in the field, how much would we expect warranty costs to increase for a group of 7,500,000 batteries sold with the new warranty? **(7 pts.)**

d. The engineering department of the battery company proposes a redesign of this model of battery to decrease the standard deviation of the battery lifespan distribution from 21 months to 16 months. For technical reasons the average lifespan, however, would decrease to 75 months. If this were realized by how much would the warranty cost on a group of 7,500,000 batteries change? (Note: Assume a 48-month warranty.) **(7 pts.)**

e. The CEO of this battery manufacturer enjoys classic cars and lovingly maintains a fully-restored 1958 Packard Hawk, a rare vehicle instantly recognizable to collectors. When the car isn’t under cover in its climate controlled garage it is usually being trailered to shows, parades, and other special events. The Packard is driven under its own power no more than a few hundred miles a year and the car is started, driven a short distance, and allowed to idle at least 1 hour every 2 weeks. A single battery was taken directly from the production line and placed in the Packard on July 27, 2009. The battery is still performing well to this day. Statistically, on what date will/was this battery’s lifespan be 3 standard deviations from the mean lifespan and thus be considered an outlier? **(7 pts.)**

f. The battery company is working on an experimental model which would offer a lower mean lifespan but an uncertain but tighter lifespan distribution. The mean of lifespan of this battery would be only 54 months with a standard deviation of just 14 months. The CEO gave an experimental unit to a fellow antique car enthusiast to install in his 1957 Ford Skyline Retractable Hardtop automobile. These cars pushed the limits of auto engineering in the late 50s and today are extraordinarily complex to maintain. The experimental battery was installed in the Ford on August 12, 2018. The CEO has a friendly bet with his friend that the battery will still be working well in the Ford on January 1, 2024. Based on this information what do you estimate is the likelihood the CEO will win his bet? **(6 pts.)**

3. The automobile battery manufacturer in the previous problem has another issue it would like to evaluate through elementary statistics. They’re concerned with whether the lifespan of their batteries is impacted by extremes in climate. For that battery model the population life expectancy is 77 months under normal use. The firm has taken a set of identical batteries and done testing to simulate use of the batteries in an environment with a constant ambient temperature of -20o F (informally labeled “Minnesota Tundra”) and simulated testing of another identical set batteries in an environment with a constant ambient temperature of +130o F (labeled “[Sonoran Desert](http://en.wikipedia.org/wiki/Sonoran_Desert)”). Simulated life spans (in months) for each battery in the tests is given in the embedded Excel spreadsheet shown here.

a. Using the data provided, would you be willing to say the Minnesota Tundra test or the Sonoran Desert test shows evidence the lifespan of these batteries would be different in these extreme environments from the “normal” environment? **(8 pts.)**

b. Can you reasonably say batteries operating under “Minnesota Tundra” conditions have a different expected lifespan than those operating under “Sonoran Desert” conditions? **(8 pts.)**

Note: For this question your deliverables for parts a and b will be a block of R script which you have written to conduct the desired analyses, the output from your script file, and interpretations of your results.





****4. Nestled on the eastern shore of Tampa Bay since 1958, the waterfront community of Apollo Beach offers its residents a unique lifestyle. Apollo Beach has miles of man-made canals which open onto Tampa Bay, making it highly desirable to anyone interested in boating and fishing.

|  |
| --- |
| **Apollo Beach residents enjoying a pleasant evening on the docks at Circles Restaurant.** |

As in many communities in Florida and the US, real estate values in Apollo Beach have rebounded from bad days of 8-12 years ago. Properties today are moving at higher prices and spending less time on the market.

Let’s say a local realtor has pulled data concerning recent sales of properties in Apollo Beach. The data, given here in the embedded Excel spreadsheet, includes the initial asking price, the final selling price, the date of initial listing for sale, and the date of closing. This represents a sample of Apollo Beach homes sold in 2017. The realtor has forwarded the data to you and is asking you to answer a few questions.

|  |
| --- |
| Image result for glengarry glen ross |
| “Have I got your attention now? Good. 'Cause we're adding a little something to this month's sales contest. As you all know, first prize is a Cadillac Eldorado. Anyone wanna see second prize? Second prize is a set of steak knives. Third prize is you're fired. Get the picture?”  -- Alec Baldwin  Glengarry Glen Ross |

a. Of the four types of data discussed in class, which type is the Asking Price variable? **(5 pts.)**

b. Use R/R Studio to Conduct a hypothesis test (alpha = .05) to determine if the mean selling price is greater than $400,000. Clearly state your null and alternate hypotheses. Provide the text of the R script file you write and the R output in your deliverable. Explain the results of the test in a 1-3 sentence statement easily understandable by your client, the supervising agent shown at right. **(8 pts.)**

c. Use the information provided to analyze the number of days on the market for each home in the sample. Calculate the 98% confidence interval on this variable. Also explain this result in language easily understandable by your client.Do any homes appear to be outliers when reviewing days on market? **(8 pts.)**

d. For each home sale the difference between the Asking and Selling price is the ultimate concession made by the seller to move the property. Using what you know about statistics and R/R Studio can you say the average concession is greater than $120,000? (For this question your deliverable should include the R script you write to conduct the analysis, the R output, and an interpretation of your analysis results.) **(8 pts.)**

****

5. Ladner Photography specializes in low-cost photographic portraits. Ladner sends traveling photographers from its base in Edwardsville, Illinois throughout the Midwest and Mid-South taking portraits at churches, schools, and civic organizations located mostly in small towns, as well as special events such as weddings and holiday gatherings. Photographers travel for 14 straight days, then return to work at the Ladner home office for 7 days working reduced hours. A typical session involves a photographer coming to a church, school, or other public building in a small or mid-sized rural location (by prior arrangement), setting up lights, backgrounds, and props and taking portraits of an organization’s members or the general public after they pay a sitting fee. Final poses and portrait packages are selected by the subjects later online and the portraits are mailed once final payment is made. The organization sponsoring the session does advance promotion and receives part of the revenue.



A key issue for Ladner is controlling for the time a photographic sitting takes. A sitting is the actual time needed to pose the subjects and any props, set lighting, and take several different shots. Sittings for individuals or couples can be short but unique poses, props, children, animals, or special lighting can take much longer. A shorter sitting means more sittings during a session and greater revenue for Ladner as well as the sponsoring organization. To better understand sitting times Ladner randomly sampled 845 sitting times for three of its traveling photographers. The times for those sittings (in minutes) are given in the embedded spreadsheet below.



a. Considering all four photographers together, what if anything is noteworthy about the data? **(6 pts.)**

b. Considering only photographer Rick Blaine, what are the mean and standard deviation of the sampling distribution for this sample set of data? **(6 pts.)**

c. Again considering only photographer Rick Blaine, if other samples of sittings with the same sample size were taken, what proportion of those samples would have means greater than 3 minutes? **(7 pts.)**

d. Considering only photographer Harry “Steve” Morgan, conduct a hypothesis test to determine if the true population mean sitting time is greater than 3 minutes and 45 seconds. Clearly state the null and alternate hypotheses as well as whether you are rejecting or failing to reject your null hypothesis as well as the p value for the test. **(7 pts.)**



e. For photographer Fred C. Dobbs, what is the 85% confidence interval for his population mean sitting time? **(8 pts.)**

f. Considering again only photographer Charlie Allnut, a confidence interval on the population mean has been calculated (using an appropriate Z value) as (6.593648673, 6.472044758). What level of confidence was used to calculate this interval? **(8 pts.)**

g. Determine if there is a statistically significant difference between the mean sitting times of each of the four photographers. If there is a statistically significant difference among the four, state where the difference lies. **(7 pts.)**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Offender Picture | Description: Offender Picture | Description: Offender Picture | Description: Offender Picture | Description: Offender Picture | Description: Offender Picture |
| 497 months | 360 months | 300 months | 115 months | 180 months | 180 months |

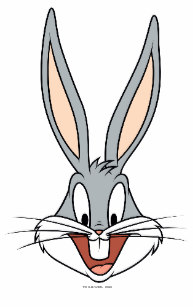
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Description: Offender Picture | Offender Picture | Description: Offender Picture | Description: Offender Picture | Description: Offender Picture | Description: Offender Picture |
| 18 months | 189 months | 60 months | 24 months | 96 months | 120 months |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Offender Picture | Offender Picture | Offender Picture | Offender Picture | Offender Picture | Offender Picture |
| 212 months | 8 months | 341 months | 11 months | 346 months | 241 months |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Offender Picture | Offender Picture | Offender Picture | Offender Picture | Offender Picture | Offender Picture |
| 27 months | 172 months | 333 months | 9 months | 25 months | 168 months |

6. Above are pictures of 24 randomly selected inmates taken from the website of the [Florida Department of Corrections](http://www.dc.state.fl.us/index.html). The first two rows shown are inmates serving a sentence for one or more sex offenses against a minor, and the second two rows are inmates serving a sentence for one or more drug offenses. The length of sentence (in months) is given for each.

1. Determine and write a clear interpretation of the 95% confidence interval on the population mean sentence length of all Florida inmates serving a sentence for a drug offense. **(4 pts.)**
2. Using this data can you say with statistical certainty that the population mean sentence length for drug offenses is different from the population mean sentence length for sex offenses against minors? **(5 pts.)**
3. If you combined this sample data on inmates sentenced for both sex and drug offenses and conducted a hypothesis test on whether the population mean is different from 15 years (180 months) you’d fail to reject the null hypothesis. What would be the maximum average sentence length you could test at which you’d reject the null hypothesis with p = alpha = .05? **(6 pts.)**

7. It has been reported that entertainment habits in America are changing and people today are watching less television than in the past (traditionally defined as programming delivered via broadcast, cable including fiber optic, and satellite). In its place many people are surfing the Internet for diversion as well as playing video games and watching on-demand services such as Amazon, Hulu and Netflix.

Let’s say in 2008 at least one television was turned on in the average American home 9 hours and 45 minutes per day. Given here in the embedded Excel spreadsheet is data representing a random sample of households. The data shows the number of minutes in a typical day at least one TV is turned on in each sampled household. This sample was collected recently. Given this data use an appropriate statistical test to show whether the population mean TV “on” time has declined since 2008. **(10 pts.)**



 8. Following the previous question, an independent researcher has decided to study variations in TV viewing habits among four American cities. She has managed to randomly sample households in Detroit, Los Angeles, Miami, and Pittsburgh and gather data on the daily “on” times for households on a typical day. The data is given in the embedded spreadsheet given here. Can we say with statistical certainty there are differences in TV “on” time between these cities? If so, which cities are statistically different?

**(12 pts.)**

 9. During January 28-30, 2018 The Economist commissioned a poll of US adults to learn their opinions on a variety of issues. For several questions polling results were reported for the sample at large and also split into subgroups. One question asked related to respondents’ beliefs as to whether President Trump considered Russia to be a friend or enemy. Responses given to this question fell into five categories: Ally, Friendly, Unfriendly, Enemy and Not Sure.

1. Responses of 1,481 individuals were aggregated and 40% stated the opinion that the President considered Russia to be either “Ally” or “Friendly”. What would be the 95% Confidence Interval on this analysis result? **(3 pts.)**
2. These responses were further subdivided by age. Of those in the 18-29 age category (n = 276) 16% expressed the opinion that the President considered Russia to be an “Ally”. Of those in the 45-64 age category (n = 548) 10% believed the President considered Russia to be an “Ally”. Can we say there is a statistically significant difference in the proportions of these two age categories? **(3 pts.)**