MTH 324, Exam #2, At Home, Fall 2023 Name KEY

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted in the course for the at-home portion and you will answer questions about that work during the in-class portion of the exam; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use.

At home, prepare for questions in Part I using R. Open the data file entitled **324exam2data.xlsx** posted in Blackboard. (Note: this file has multiple sheets of data. You may want to separate the data into separate files to upload to R, or look up how to access multiple sheets in R from a single upload.) Complete the calculations noted below. You will be asked for additional analysis and interpretation of this data in the in-class portion of the test. Print out the results of your analysis and code, and bring the pages with you to the exam. You will submit all this work along with the in-class exam.

1. A sample of 81 students is selected and it is determined that their mean math ACT score is 24.2. If the true mean math score is actually 21.6 (with a standard deviation of 5.2), what is the power of the one-sample test to detect this 3-point difference?

From Sheet 1:

2. Customer purchases from a store are recorded and the type of card they use. Conduct an appropriate hypothesis test of the data to determine if the total sale is more for the store card than for other cards. Test your assumptions with normal probability plots.

From Sheet 2:

3. Conduct a two-way ANOVA test on whether promotion type or competitor affect sales (and any interactions). Apply Tukey's method to plot the differences of means for each set of effects. Which factors produce the highest sales? Be sure to check your data for normality. Create a comparative boxplot of both factors to confirm your analysis.

From Sheet 3:

4. Use the data to determine if salary type in the dataset had tried or has not tried the product (lasagna) the same rate. You'll need to count the number of salaried vs hourly in the data, and within each group, count the number of people who have tried or not tried the product in each group. Conduct a two-sample proportion test to determine if the difference is statistically significant. Check the assumptions of your test.

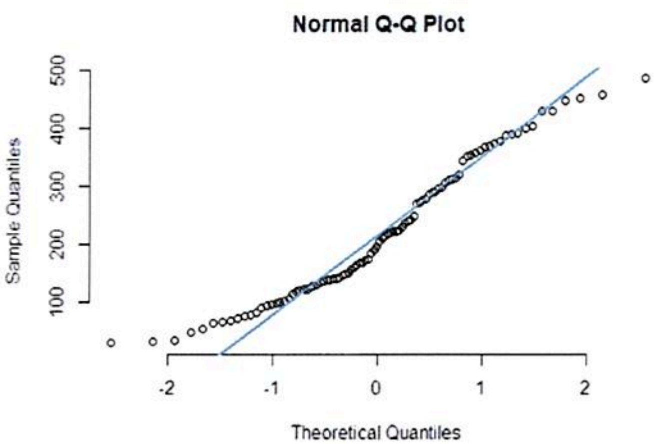
5. Using Neighborhood, Dwell Type and Live Alone, conduct a three-way ANOVA of ~~school~~ car value. Test main and interaction effects where possible. Test for the normality of car value. Apply Tukey's method.

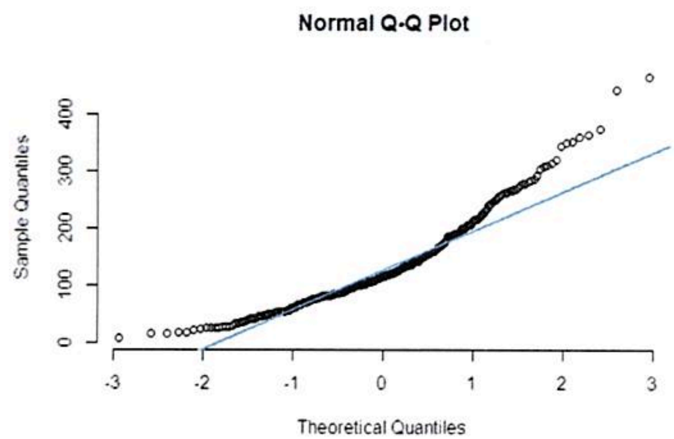
6. Conduct a one-sample hypothesis test of CC Debt (credit card debt) to see if there is significant reason to believe CC Debt is greater than $1,400 per person.

7. Build a sampling distribution of Income. Collect 1000 samples of 50 people each. Calculate the mean of the sample. Build a histogram of your sample mean data. Find the mean (of the means) and the standard deviation of your sample means (the standard error). Find the mean and standard deviation of the original data. Compare the results.

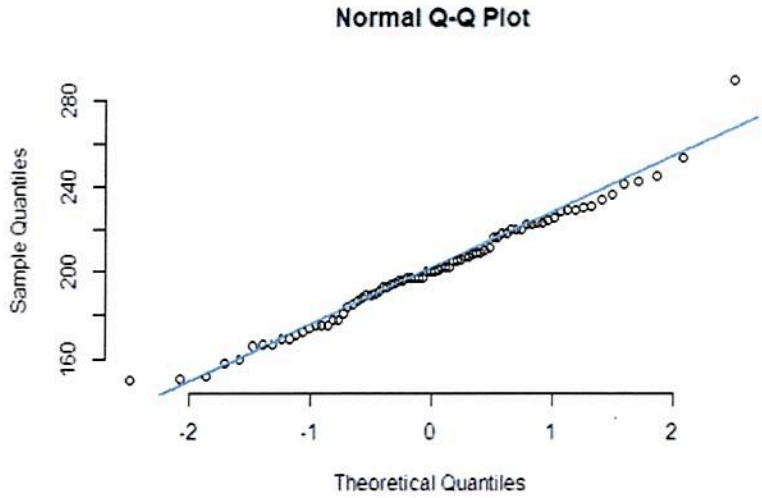
MTH 324, Fall 2023, Exam #2 At-home Analysis

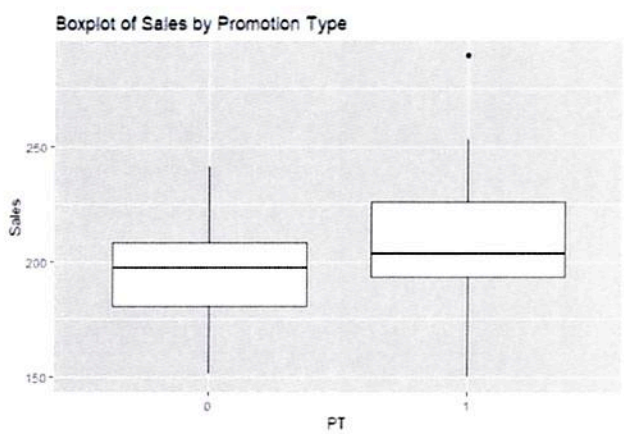
```
     One-sample t test power calculation

              n = 81
          delta = 0.5769231
             sd = 1
      sig.level = 0.05
          power = 0.9992355
    alternative = two.sided

        Welch Two Sample t-test

data:  x and y
t = 6.4611, df = 120.02, p-value = 2.325e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  58.64234 110.46197
sample estimates:
mean of x mean of y
 218.5004  133.9483
```

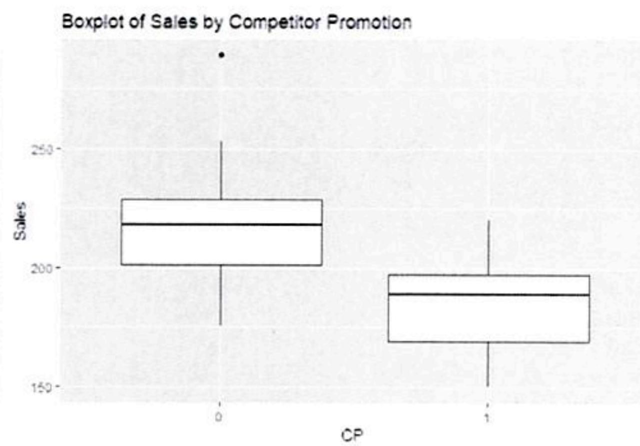


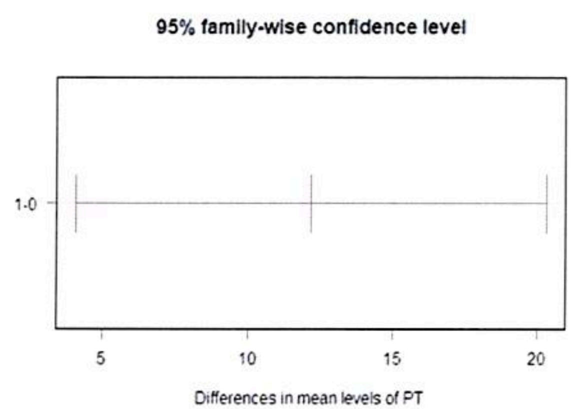


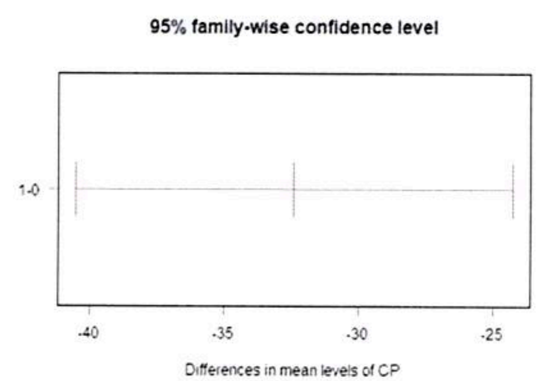

```
           Df Sum Sq Mean Sq F value    Pr(>F)    
PT          1   2965    2965   8.916   0.00380 ** 
CP          1  20903   20903  62.868  1.51e-11 ***
PT:CP       1   2774    2774   8.344   0.00504 ** 
Residuals  76  25269     332                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

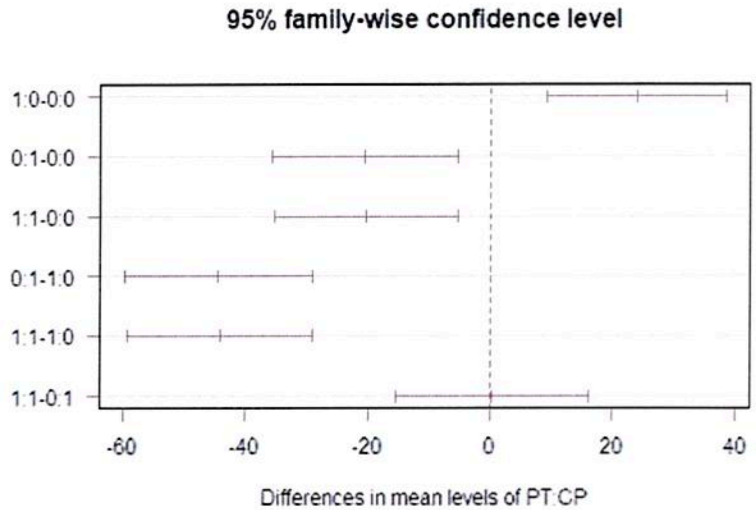

Normal Q-Q Plot
Sample Quantiles
160
200
240
280
-2
-1
0
1
2
Theoretical Quantiles


Boxplot of Sales by Promotion Type
Sales
PT


Boxplot of Sales by Competitor Promotion
Sales
CP


95% family-wise confidence level
1-0
5
10
15
20
Differences in mean levels of PT

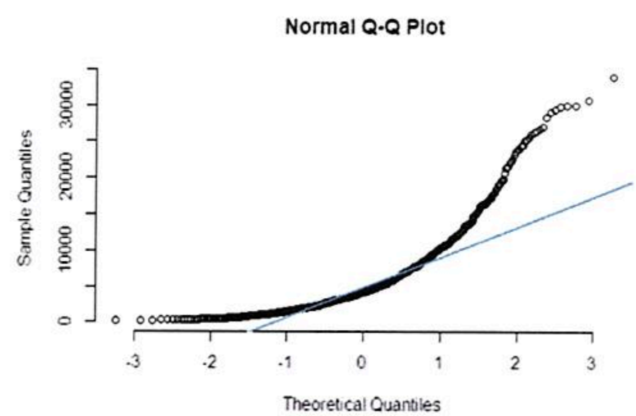

95% family-wise confidence level
1-0
-40
-35
-30
-25
Differences in mean levels of CP



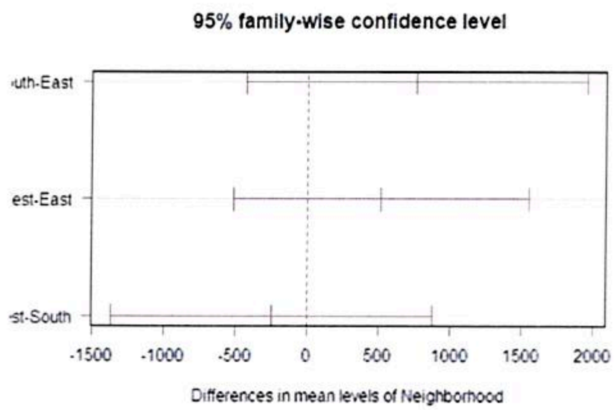
data3	856 obs. of 13 variables
data3_1	375 obs. of 13 variables
data3_1_1	153 obs. of 13 variables
data3_1_2	222 obs. of 13 variables
data3_2	481 obs. of 13 variables
data3_2_1	342 obs. of 13 variables
data3_2_2	139 obs. of 13 variables

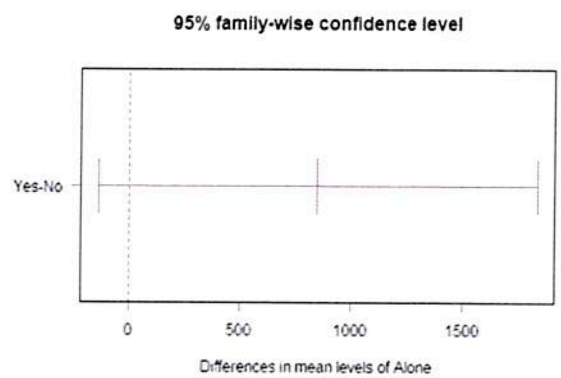
Order: Total, Pay Type=Hourly, then Have Tried =Yes, No, Pay Type=Salaried, Have Tried=Yes, No

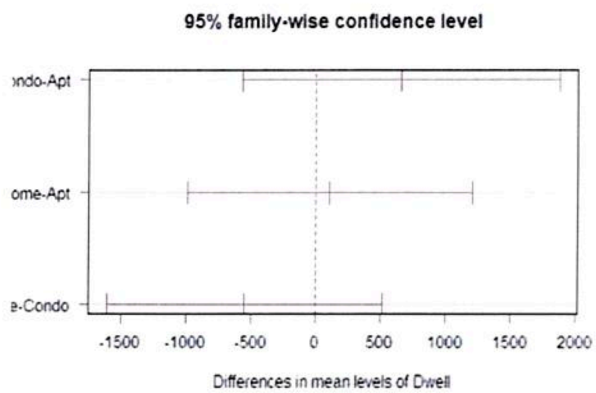
```
        2-sample test for equality of proportions with continuity correction

data:  c(153, 342) out of c(375, 481)
X-squared = 78.099, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.3695417 -0.2364957
sample estimates:
   prop 1    prop 2
0.4080000 0.7110187

Npq= 90.576
Npq= 98.8316
```

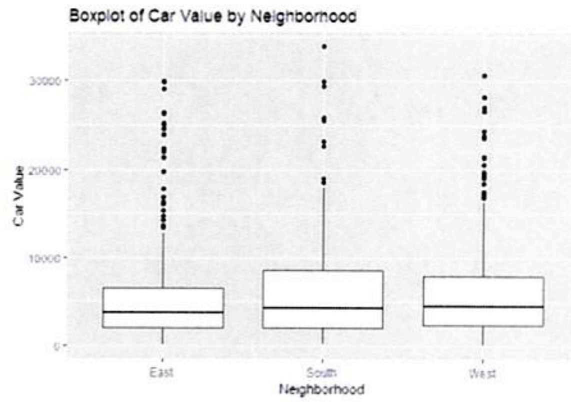


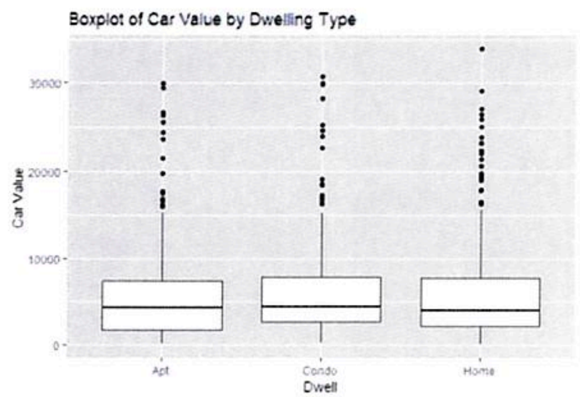
```
                                            Df    Sum Sq  Mean Sq F value Pr(>F)
Neighborhood                                 2 7.687e+07 38437312   1.256 0.2852
`Live Alone`                                 1 8.742e+07 87418534   2.857 0.0913 .
`Dwell Type`                                 2 6.122e+07 30607556   1.000 0.3682
Neighborhood:`Live Alone`                    2 2.841e+07 14205889   0.464 0.6287
Neighborhood:`Dwell Type`                    4 9.732e+07 24329205   0.795 0.5284
`Live Alone`:`Dwell Type`                    2 2.742e+07 13709948   0.448 0.6390
Neighborhood:`Live Alone`:`Dwell Type`       4 1.621e+08 40516355   1.324 0.2591
Residuals                                  838 2.564e+10 30595069
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

              Df    Sum Sq  Mean Sq F value Pr(>F)
Neighborhood   2 7.687e+07 38437312   1.259  0.285
`Live Alone`   1 8.742e+07 87418534   2.863  0.091 .
`Dwell Type`   2 6.122e+07 30607556   1.002  0.367
Residuals    850 2.595e+10 30533979
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = `Car Value` ~ Neighborhood + Alone + Dwell, data = data3)

$Neighborhood
               diff        lwr       upr     p adj
South-East  763.6047  -428.3350 1955.5444 0.2894723
West-East   519.1077  -510.4902 1548.7055 0.4632016
West-South -244.4970 -1369.4627  880.4687 0.8663651

$Alone
           diff       lwr      upr    p adj
Yes-No 848.2056 -140.0669 1836.478 0.092437

$Dwell
               diff        lwr       upr     p adj
Condo-Apt   654.7484  -567.7185 1877.2153 0.4197679
Home-Apt    105.2802  -993.2671 1203.8274 0.9724774
Home-Condo -549.4682 -1608.1672  509.2308 0.4425010
```

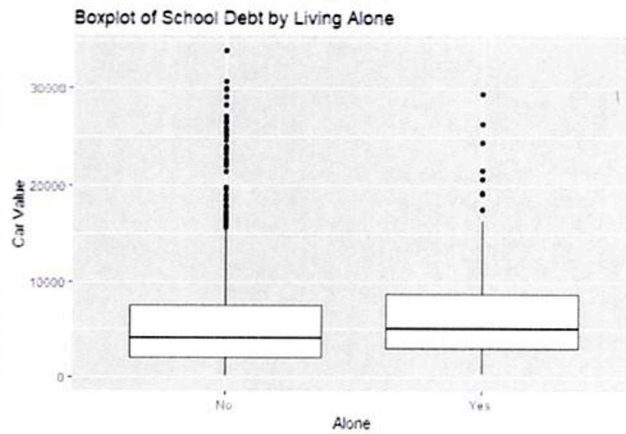






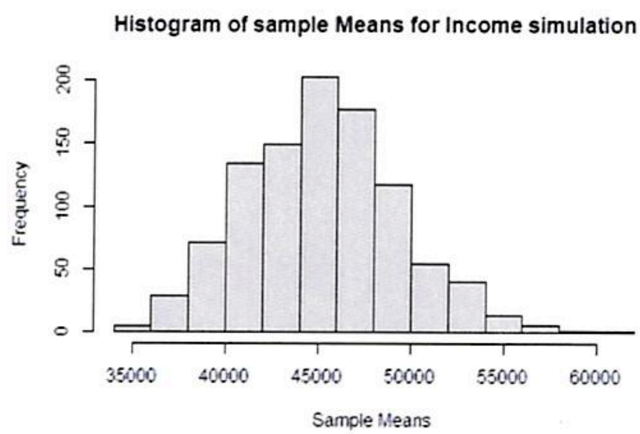







```
        One Sample t-test

data:  data3$`CC Debt`
t = 0.71432, df = 855, p-value = 0.2376
alternative hypothesis: true mean is greater than 1400
95 percent confidence interval:
 1359.274      Inf
sample estimates:
mean of x
 1431.203
```

```
mean(means)
[1] 45159.41
> mean(data3$Income)
[1] 45266.94
> sd(means)
[1] 4125.512
> sd(data3$Income)
[1] 28631.29
```



MTH 324, Exam #2, Fall 2023 Name KEY

Instructions: Answer each question thoroughly. For questions in Part 1, use the work you did at home to answer the questions. Be sure to answer each part of each question. In Part 2, report exact answers unless directed to round.

Part I:

1. A sample of 81 students is selected and it is determined that their mean math ACT score is 24.2. If the true mean math score is actually 21.6 (with a standard deviation of 5.2), what is the power of the one-sample test to detect this 10-point difference?

99.92% yes, it can be detected

Use the work you did at home to answer these questions about ~~tax paid~~ Sale and the ~~neighborhoods~~ Card type s in our dataset.

2. State your null and alternative hypotheses for the card type question.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

3. What kind of test did you conduct? What is the P-value for your tests?

2-sample t - independent   p-value: 2.325×10^{-7}

4. What do you conclude from your test? State your conclusion in plain language in context.

reject null

there is evidence the two groups have different sales values.

5. What is the null and alternative hypotheses for your two-way ANOVA?

A
$H_0: \mu_i = \mu_j \quad \forall\, i \neq j$
$H_a: \mu_i \neq \mu_j$ for some $i \neq j$

B
$H_0: \mu_i = \mu_j \quad \forall\, i \neq j$
$H_a: \mu_i \neq \mu_j$ for some $i \neq j$

AB
$H_0: \mu_i = \mu_j \quad \forall\, i \neq j$
$H_a: \mu_i = \mu_j$ for some $i \neq j$

6. What were the results of your test? Using Tukey's method and a box plot, which factors or combination of factors produced the most sales. Explain your reasoning.

All factors and interaction were significant

Only the 1:1-0:1 combination were not different

most sales when promotion type 1 and comp. promotion 0

7. For the salary type question, state the null and alternative hypotheses for your test of proportions. What did you conclude about the differences in likelihood to try the product given their salary type?

$H_0: p_1 = p_2$
$H_a: p_1 \neq p_2$

p-value 2.2×10^{-16}

reject null

Salary type does influence likelihood of trying product

8. What are the null and alternative hypotheses for your three-way ANOVA? List them all and label them clearly.

A $H_0: \mu_i = \mu_j \ \forall i \neq j$
$H_a: \mu_i \neq \mu_j$ for some $i \neq j$

ditto B, C, AB, AC, BC, ABC

9. Did any of the null hypotheses get rejected for your test? Which ones? State the form of your final model (linear model).

all of them

none of these variables or combination of variables have a statistically significant impact on car value

$\hat{y}$ = grand mean

10. Describe the normality of the data? Is it normal? Are there significant deviations from normal?

not very - yes, not similar to normal

11. Describe the Tukey intervals for the three main effects. (Explain what each one means.)

all overlap w/ 0

12. State the null and alternative hypotheses for your one-sample test of credit card debt levels. What is your P-value? What did you conclude?

Ho : $\mu_1 = 1400$

Ha : $\mu_1 > 1400$

p-value 0.2376

fail to reject null

There is not enough evidence to think CC debt is higher than ~~$1400~~ $1400

13. For your sampling distribution, describe the shape of the distribution.

approximately normal

very slight right skew

14. What is the mean of your means? What is the mean of the data? Are they similar?

45,159.41

45,266.94 yes

15. What is the standard deviation of your means? What is the standard deviation of the data? Given that you took samples of size 50, what is the predicted standard error from the central limit theorem? Is it similar to what you simulated?

4125.51

28,631.29

$\frac{28631.29}{\sqrt{50}} = 4049.076...$

yes

Part II:

16. Describe the main results of the Central Limit Theorem.

it describes the distribution of sample statistics

distributions tend toward normal as sample sizes increase

and variability is reduces

mean of sampling distribution centered around population value of same statistic

17. A sample of the weights of seven feral cats is collected and the data is found to be {7.8, 9.1, 6.4, 5.8, 7.3, 7.7, 8.2} pounds each. Assuming the data follows a normal distribution, use the method of moments to find estimates for the mean and variance.

$E(x) = \bar{x} = \frac{52.3}{7} = 7.47 = \hat{\mu}$

$E(x^2) = \frac{398.07}{7} = 56.867$

$V(x) = \hat{\sigma}^2 = 56.867 - 7.47^2 = 1.644897...$

$E(x^2) - [E(x)]^2 =$

18. At the beginning of the semester a representative sample of 342 students were surveyed and asked if they owned a dog. The sample proportion was 0.31. Use this information to construct a 95% confidence interval for the proportion of all college students who own a dog.

$SE = \sqrt{\frac{.31(.69)}{342}} = 0.025008...$

$ME = 1.96 \cdot SE = 0.0490...$

$(0.261, .359)$

19. If you want to determine the appropriate sample size needed to conduct a poll with just at 2.5% margin of error for a proportion, with a 95% level of confidence, use the formula $n = p(1-p)\left(\frac{z^*}{E}\right)^2$. Use this formula with $p = 0.5$ to estimate the sample size needed.

$$0.5(0.5)\left(\frac{1.96}{0.025}\right)^2 = 1536.64$$

$$\Rightarrow n = 1537$$

20. Describe what a Latin Square design is. Give an example of a Latin Square design for three levels of data, each with 4 levels each.

	Factor A			
Factor B	1	2	4	3
	2	1	3	4
	3	4	2	1
	4	3	1	2