**Math 222: Exam 2**
**Fall – 2019**
**11/07/2019**
**80 Minutes**

Name: _____

| Question | Points | Score |
|----------|--------|-------|
| 1        | 16     |       |
| 2        | 16     |       |
| 3        | 27     |       |
| 4        | 31     |       |
| 5        | 10     |       |
| Total:   | 100    |       |

| Race | White | Hispanic/ Latino | African American | American Indian/ Alaska Native | Asian | Hawaiian/ Pacific Islander | Other | Two or More Races |
|---|---|---|---|---|---|---|---|---|
| Percentage of US Population | 56.1 | 16.3 | 12.6 | 0.9 | 4.8 | 0.2 | 6.2 | 2.9 |

1. (16 points) Is the Syracuse University a microcosm of the United States population, or is population at SU different, i.e. more/less diverse? According to the 2010 US Census, the population of the United States, broken down by race, is given in the table at the top of the page.[1] On the other hand, according to the 2018 Syracuse University Fall Census,[2] the breakdown of the 13,175 domestic undergraduate students by race is as in the table below. To determine if the distribution of races at SU is representative of the US race distribution, perform a Chi-Square Goodness of Fit Test. Be sure to state your null and alternative hypotheses, the test statistic, $p$-value, and conclusion at the 1% significance level.

| Race | White | Hispanic/ Latino | African American | American Indian/ Alaska Native | Asian | Hawaiian/ Pacific Islander | Other | Two or More Races |
|---|---|---|---|---|---|---|---|---|
| Number of SU Students | 8662 | 1393 | 987 | 80 | 1043 | 9 | 503 | 498 |

---

2. The Kaiser Family Foundation regularly polls Americans to track opinions on the Affordable Care Act (ACA). They ask the following "As you may know, a health reform bill was signed into law in 2010. Given what you know about the health reform law, do you have a generally favorable or generally unfavorable opinion of it?" According to their October 2019 poll,[3] if one took a survey of 100 Americans, you would obtain data as in Table 1 on the next page.

   (a) (2 points) Write the null and alternative hypotheses for a Chi-Square Test for Association in the context of the problem.

$$\begin{cases} H_0 : \\ \\ \\ H_a : \end{cases}$$

   (b) (4 points) Fill in the missing values on the tables on the next page.

   (c) (3 points) What are the assumptions for a Chi-Square Test for Association? Does this test meet these requirements?

   (d) (5 points) Write the statistics for the Chi-Square Test at the 5% significance level below:

   Degrees of Freedom: _____

   Critical Value: _____

   Test Statistic: _____

   $p$-value range: _____

   (e) (2 points) Write your conclusion for the Chi-Square Test in the context of the problem, using $\alpha = 0.05$.

Table 1: Counts of responses to the survey, broken down by age and opinion

| Age/Opinion | Favorable | Unfavorable | Don't Know | Total |
|---|---|---|---|---|
| 18–29 | 51 | 37 | 12 | 100 |
| 30–49 | 54 | 38 | 7 | 99 |
| 50–64 | 53 | 41 | 7 | 101 |
| 65+ | 43 | 44 | 11 | 98 |
| Total | 201 | 160 | 37 | 398 |

Table 2: Expected counts for the survey, assuming no association.

| Age/Opinion | Favorable | Unfavorable | Don't Know |
|---|---|---|---|
| 18–29 | 50.50 | 40.20 | 9.30 |
| 30–49 | 50.00 | | 9.20 |
| 50–64 | 51.01 | 40.60 | 9.39 |
| 65+ | | 39.40 | 9.11 |

Table 3: Contribution to $\chi^2$

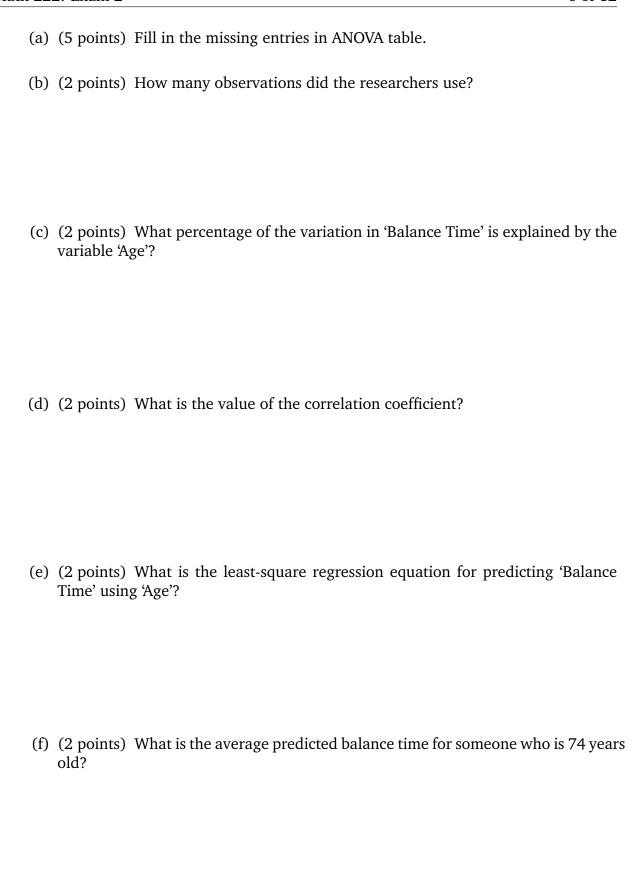| Age/Opinion | Favorable | Unfavorable | Don't Know |
|---|---|---|---|
| 18–29 | 0.00 | 0.25 | 0.79 |
| 30–49 | | 0.08 | 0.53 |
| 50–64 | 0.08 | 0.00 | 0.61 |
| 65+ | 0.85 | | 0.39 |

3. A Geriatric Rehabilitation Facility works with elderly patients that have suffered in-
   juries from falls. As part of the rehabilitation, the patients perform balancing exer-
   cises. To understand how long the 'healthy' elderly patient should be able to perform
   the exercise, researchers at the facility administer the exercise to a number of people
   from a variety of ages and try to determine if one can predict how long one should
   be able to hold the balance pose using the patient's age. The results from their simple
   linear regression are found below.

```
Analysis of Variance

        Source      DF   Adj SS   Adj MS  F-Value  P-Value
        Regression  __   10514.8  10514.8  380.96   0.000
          Age       __   _____  _____  _____   0.000
        Error       __   _____   27.6
        Total       26   _____


Model Summary

              S    R-sq  R-sq (adj)  R-sq (pred)
        5.25361  93.84%     93.60%       92.73%


Coefficients

        Term       Coef  SE Coef  T-Value  P-Value   VIF
        Constant  _____    2.81    30.43    0.000
        Age      -1.0135  0.0519   _____    0.000  1.00
```

(a) (5 points) Fill in the missing entries in ANOVA table.

(b) (2 points) How many observations did the researchers use?

(c) (2 points) What percentage of the variation in 'Balance Time' is explained by the variable 'Age'?

(d) (2 points) What is the value of the correlation coefficient?

(e) (2 points) What is the least-square regression equation for predicting 'Balance Time' using 'Age'?

(f) (2 points) What is the average predicted balance time for someone who is 74 years old?

(g) (3 points) Use $SE_{b_1}$ to show $\sum (x_i - \bar{x})^2 = 10246.6$.

(h) (4 points) Construct a 95% confidence interval for the average predicted balance time for someone who is 74 years old. [Note: the average age of the participants was 50.5.]

(i) (5 points) At the 10% significance level, test the hypothesis $H_0 : \beta_1 = 0$ against $H_a : \beta_1 < 0$. For this test, state your critical value, test statistic, $p$-value, and conclusion. From this test, are 'Balance Time' and 'Age' positively or negatively correlated, or neither?

4. A university is reviewing the types of students that it admits to try to accept the best possible students. They examine whether a student's HS average, SAT Reading/Writing score, SAT Math score, and SAT Essay scores can be used to predict a student's success, measured by their final college GPA. They examine 22 students averages and create a multilinear regression, the model summary of which is found below.

```
Analysis of Variance

        Source              DF    Adj SS     Adj MS  F-Value  P-Value
        Regression          __    2.47957   _____   _____    0.000
          HS Average        __    _____  0.022345    0.91    0.354
          Reading/Writing   __    0.00426   _____     0.17    0.683
          Math              __    0.01442   0.014421    _____    0.455
          Essay             __    0.00296   0.002964     0.12    0.733
        Error               __    _____    0.024636
        Total               __    2.89838
```

```
Model Summary

                  S       R-sq  R-sq (adj)  R-sq (pred)
          _____   _____%     82.15%       75.45%
```

```
Coefficients

        Term                Coef   SE Coef  T-Value  P-Value     VIF
        Constant                      1.86    -0.47    0.642
        HS Average        0.0356    _____    0.95    0.354   34.71
        Reading/Writing   0.00076   0.00182    _____    0.683   22.33
        Math              0.000715  0.000934    0.77    0.455    5.07
        Essay             0.042     0.122       0.35    0.733    8.47
```
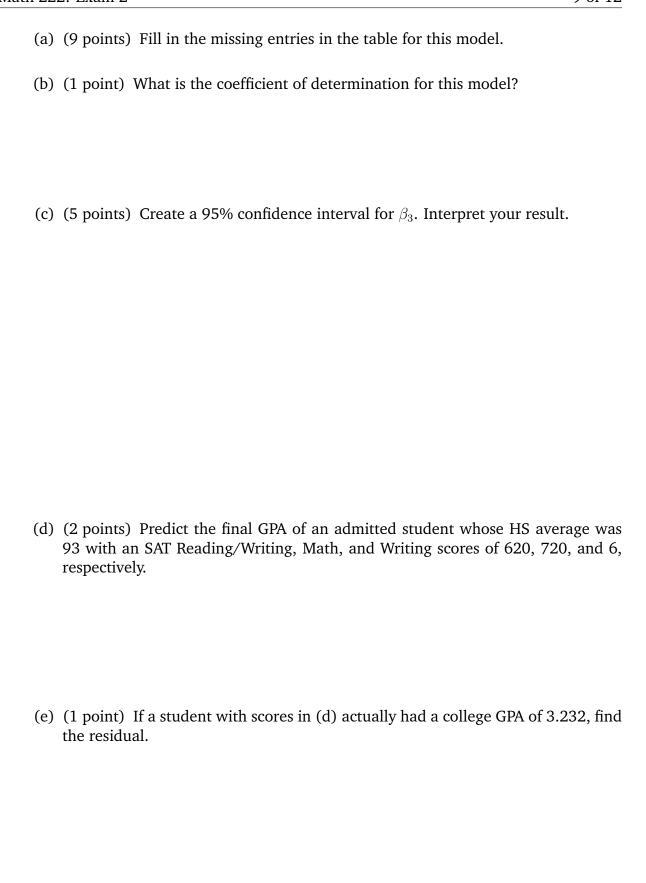
The regression equation is

GPA $= -0.88 + 0.0356$ HS Average $+ 0.00076$ Reading/Writing $+ 0.000715$ Math $+ 0.042$ Essay

(a) (9 points) Fill in the missing entries in the table for this model.

(b) (1 point) What is the coefficient of determination for this model?

(c) (5 points) Create a 95% confidence interval for $\beta_3$. Interpret your result.

(d) (2 points) Predict the final GPA of an admitted student whose HS average was 93 with an SAT Reading/Writing, Math, and Writing scores of 620, 720, and 6, respectively.

(e) (1 point) If a student with scores in (d) actually had a college GPA of 3.232, find the residual.

(f) (7 points) Perform the ANOVA $F$-test for the regression. Be sure to state the null and alternative hypotheses, test statistic, $p$-value, and conclusion at the 5% significance level.

(g) (3 points) What variables in the model are significant in this model? Which variables are insignificant in this model?

(h) (3 points) Does the answer to (g) conflict with the answer to (f)? Explain why or why not.

5. (10 points)  Mark the following statements T (True) or F (False):

   (a) _____   In a simple linear regression, a 95% confidence interval for the mean response at $x^*$ has the largest width when $x^* = \bar{x}$.

   (b) _____   In a multilinear regression, it is impossible to have the $p$-value for the $F$-statistic be less than 0.05 but have the $p$-values for every $t$-statistic for the coefficients be larger than 0.05.

   (c) _____   If one rejects the null hypothesis in an ANOVA $F$-test for a multilinear regression, then all the model coefficient parameters $\beta_i$ are nonzero.

   (d) _____   A residual plot helps assess the fit of a regression line.

   (e) _____   In a simple linear regression, the ANOVA $F$-statistic is always equal to the square of the $t$-statistic for $b_1$.

   (f) _____   If the residual for one of the data points in a simple linear regression is negative, then the point lies below the regression line.

   (g) _____   If one performs a linear regression and the model is not significant, that means there is no relationship between the response and explanatory variables.

   (h) _____   The value of $s$ is the estimate of the standard deviation about the regression line.

   (i) _____   Adding more variables to a linear regression does not necessarily improve the model.

   (j) _____   If one of the coefficients in a multivariable linear regression is insignificant, then removing it from the regression will improve the model.

**BONUS.** (10 points) Below is a partial ANOVA table for a linear regression model.

| Source | DF | Adj SS | Adj MS | F-Value |
|--------|-----|---------|---------|---------|
| Regression | ___ | 11583.6 | _____ | 3.884 |
| Error | ___ | 6700.18 | _____ | |
| Total | 224 | | | |

Complete the table above. For credit, you must show all your computations in the space below.