STAT 2112: Mock Final

1. In an American metropolitan area, the list of the known inhabitants indicates that out of 700,000 people, 46,000 of them are 17 to 26 years old, 400,000 of them are 27 to 65 years old, 54,000 of them are 66 years old and older, the rest are too young to attend college. The demographic unit of the local municipality wants to report to the common council the number of people who are attending school in a higher education institution like a technical college or a university. The three strata which were defined above are sufficient for the purposes of the common council to allocate money to the local technical college. The precision which the chairman of the subcommittee of the common council has requested is $CV = 3\%$ for stratum I, $CV = 10\%$ for stratum II, and $CV = 20\%$ for the stratum III.

   Based on the Census Bureau's American Community Survey for this metropolitan area, about 5,000 with a CV of 3% or a mean of $\frac{5000}{46000} = 10.86\%$ with a standard deviation $\sqrt{\frac{5000(.03)}{46000}} = .05710$ of those who are 17 to 26 years old are attending or have attended college; 4,000 with a CV of 6% or a mean of $\frac{4000}{400000} = 1\%$ with a standard deviation of $\sqrt{\frac{4000(.06)}{400000}} = .02449$ of those who are in the 27 to 65 years old stratum of the population and who have attended college since age 27 ; and of the retired group, 150 with a CV of 10% or a mean of $\frac{150}{54000} = .27\%$ with a standard deviation of $\sqrt{\frac{150(.10)}{54000}} = .0166$ are attending college or have taken a college level courses since age 66.

   From other surveys which this demographic unit has recently conducted, about 60% respond in the 17 to 26 years old group, about 40% respond in the 27 to 65 years old group, and about 80% respond in the 66 years old and older group.

   (a) Calculate the sampling sizes for each of the three strata.

   (b) In addition to discovering the proportion of people in each stratum who are attending or have attended college, the enumerator asked for the the respondent's monthly rent or mortgage payment or if no rent nor mortgage payment then the monthly real estate tax. The following descriptive statistics were obtained.

	Stratum I	Stratum II	Stratum III
n	500	1500	1110
Monthly Rent ($\bar{x}$)	750	1750	500
$\frac{s}{\sqrt{n}}$	30	105	50
1^{st} quartile	690	1405	400
2^{nd} quartile	700	1800	510
3^{rd} quartile	910	2100	650
maximum	1500	4200	5000
minimum	0	150	250

      i. Construct a box plot for each stratum.

      ii. Calculate the 95% confidence interval of the average monthly rent for each stratum.

iii. Test the hypothesis that

$$H_0 : \mu_I = \mu_{III}$$

iv. What assumptions need to be made about the survey?

2. Design of Experiment

   (a) Write a 2×2 factorial design of experiment of your conception.

      i. List the two factors and the two levels.
      ii. Explain your concept of the theory which underlies your proposed experiment and what you hope to find from the experiment.
      iii. What assumptions need to be made about the survey?
      iv. Be original and keep the design of the experiment simple.

3. A pediatrician claims that the weight of the second child is more than the weight of a woman's first child regardless of the sex of the child at birth. He asserts that there exists a linear relationship between the weights of the first and second child. The following table contains thirteen observations in which the weight of the first child is the explanatory variable and the weight of the second child is the response variable.

Child	1	2	3	4	5	6	7	8	9	10	11	12	13
First	5.13	5.25	6.71	4.71	3.77	5.81	8.29	6.36	6.08	4.91	3.19	5.64	6.37
Second	6.43	5.83	7.78	5.27	4.45	6.61	9.51	7.01	8.49	5.62	4.04	6.10	7.25

$\sum_{i=1}^{13} x_i = 72.22$ $\sum_{i=1}^{13} x_i^2 = 421.8674$ $\sum_{i=1}^{13} x_i y_i = 491.9085$ $\sum_{i=1}^{13} y_i = 84.39$. After the line was fitted to the data, the sum of squared errors was computed to be: SSE=2.839573.

   (a) Write the linear model:

   (b) What are the two principal reasons for using linear models:

   Suppose that the weight of your first child was 8 lbs 7 oz. What is the predicted weight of your second child? To that end, compute:

   (c) SS_{xx}=

   (d) SS_{xy}=

   (e) $\widehat{\beta}_0$=

(f) $\widehat{\beta}_1$=

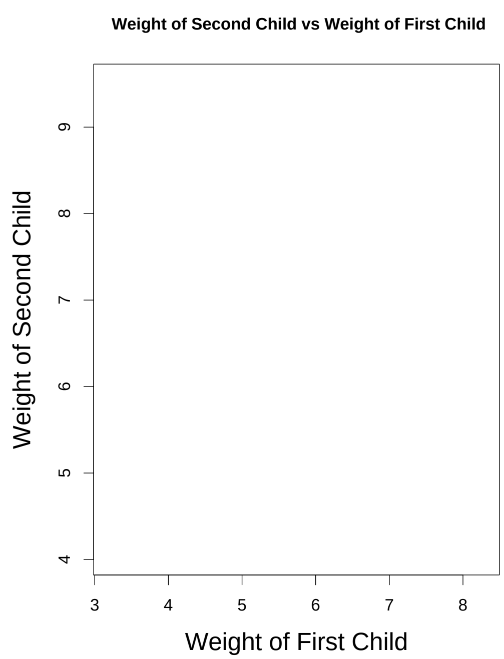
(g) When x_0=8.4375, find $\widehat{y}$=

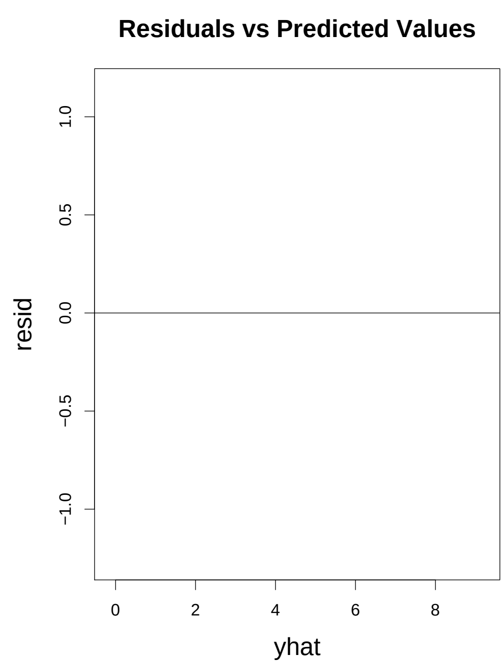
(h) The 95% for β_1 is: (0.87, 1.36). Is the model valid?

(i) What is the 95% confidence interval of $E[\widehat{y}]$, when x_0=8.4375.

Child	1	2	3	4	5	6	7	8	9	10	11	12	13
Predicted	6.01	6.15	7.78	5.54	4.49	6.77	9.54	7.39	7.07	5.77	3.84	6.58	7.40
Residual	0.410	-0.320	-0.002	-0.276	-0.045	-0.166	-0.038	-0.380	1.412	-0.150	0.192	-0.486	-0.152

(j) Plot the data and residuals versus predicted values on the following templates.





(k) Test the hypothesis $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ at $\alpha = .05$ by using the F test statistic.

(l) Interpret the QQ Plot: Test of Normality of the Residuals shown in Figure 1.

Table 1: Analysis of Variance Table

Source of Variation	df	Sum of Squares	Mean Sum of Squares	F statistic
Mean				
Regression				
Residual Error				
Total				

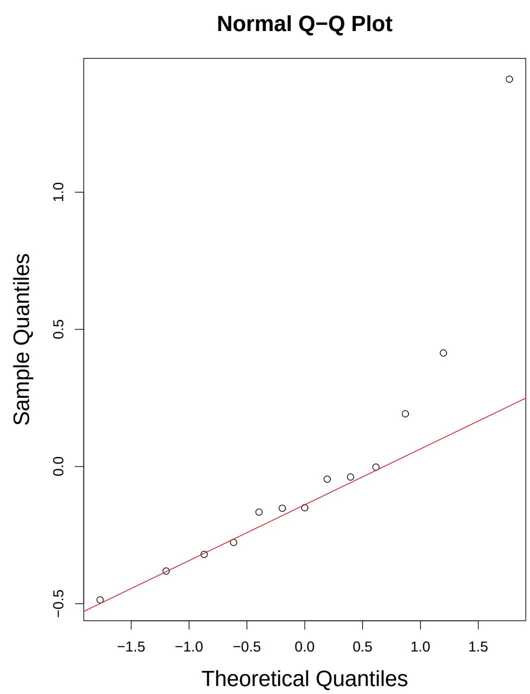


Figure 1: