STAT 515 fa 2023 Final Exam

Karl Gregory

- Do not open this exam until told to do so.
- You may have three handwritten sheets of notes out during the exam.
- You have 2.5 hours to work on this exam.
- You may NOT use any kind of calculator.
- If you are unsure what a question is asking for, please ask me for clarification.
- *Good luck, and may the odds be ever in your favor!*

Some survey data collected on the first day of class is appended to this exam.

1. Use the class survey data to investigate whether there is an association between being a CIS major and being a coffee drinker.

   (a) Use the class survey data appended to this exam to fill out the table of observed counts. Include the totals. One major is a missing value (NA); treat this as a non-CIS major.

	Coffee		
Major	Yes	No	Total
CIS			
not CIS			
Total			

   (b) Write down the null and alternate hypotheses which are of interest.

   (c) The R output for Pearson's Chi-squared test of association is the following:

```
        Pearson's Chi-squared test

data:  data
X-squared = 0.14514, df = 1, p-value = 0.7032
```

   From the R output, state your conclusion about the null hypothesis from part (b).

   (d) Suppose four of the CIS majors who drink coffee had answered that they do not drink coffee. What affect would this have on the p-value in the output of part (c)?

2. Students in this class reported on a survey the weight of their keys in grams and their student year —freshman, sophomore, junior, senior, or graduate student. Consider checking whether there is any difference in the mean weight of keys between these groups of students. The R code

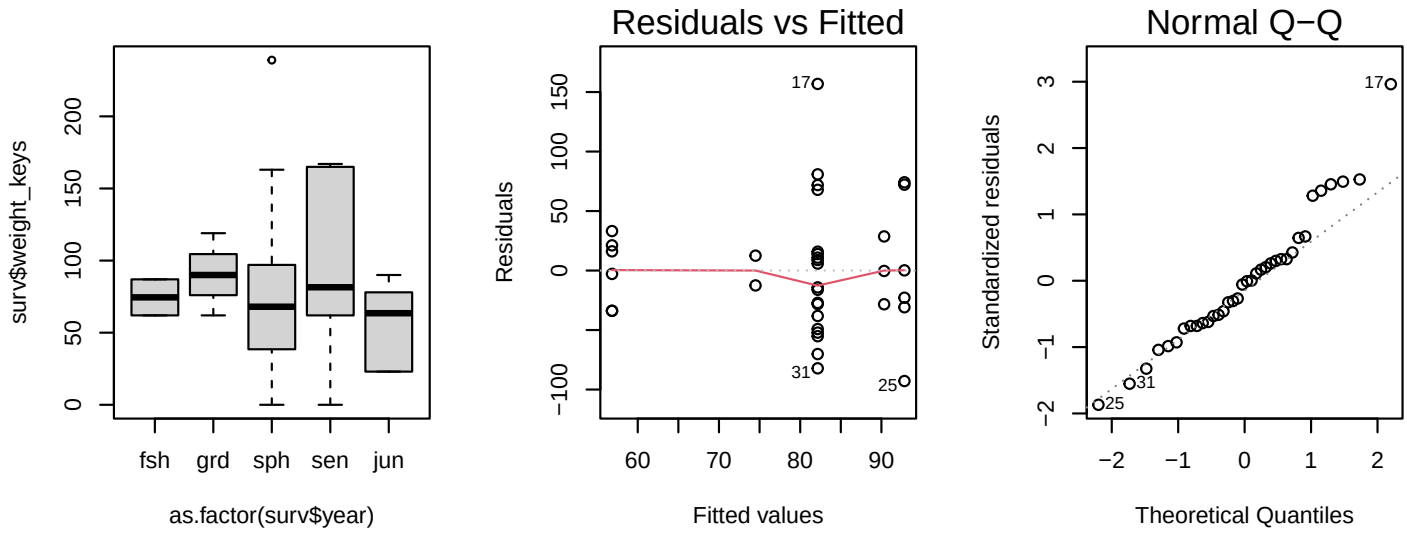
```
lm_out <- lm(surv$weight_keys ~ as.factor(surv$year))
anova(lm_out)
par(mfrow = c(1,3))
boxplot(surv$weight_keys ~ as.factor(surv$year),
        names = c("fsh","grd","sph","sen","jun"))
plot(lm_out,which=c(1,2))
```

prints to the console the output

```
        Analysis of Variance Table

Response: surv$weight_keys
                     Df Sum Sq Mean Sq F value Pr(>F)
as.factor(surv$year)  4   4678  1169.4  0.3953 0.8104
Residuals            31  91713  2958.5
```

and generates the plots



(a) Write down the hypotheses of interest in terms of the group means μ_{fsh}, μ_{grd}, μ_{sph}, μ_{sen}, and μ_{jun}.

(b) Do these data come from a comparative experiment or an observational study? Explain your answer.

(c) The word "Residuals" appears in the output. Explain what the residuals are.

(d) What assumption is one checking when one looks at the residuals versus fitted values plot?

(e) What assumption is one checking when one looks at the Normal QQ plot of the residuals?

(f) Give your own diagnosis about whether the assumptions of the analysis are satisfied.

(g) The F test statistic for testing whether there is a difference between the means of the groups is a ratio of quantities describing between-group variation and within-group variation. Give the numbers (i) and (ii) from the R output with which the F test statistic is computed as $F = (i)/(ii)$.

(h) Supposing (this may or may not be the case) that the assumptions of the analysis are satisfied, give your conclusion about your hypotheses in part (a).

3. Use the data in the appended class survey to answer this question. Consider the experiment of randomly sampling a single student from the class and letting X be the number of siblings the student has. The numbers of siblings reported by the students in the class are given in the `sibs` column of the survey data set.

   (a) Tabulate the probability distribution of X.

   (b) Compute the expected value of X. Express your answer as a fraction.

4. Suppose 2% of children in a population have a certain food allergy and suppose there is a test which, if a child does not have the allergy, will give a positive result with probability 0.20; on the other hand, if a child has the allergy, the test will give a positive result with probability 0.90.

   (a) If a child drawn from the population tests positive, give the probability that the child does not have the food allergy. You do not have to simplify your answer.

   (b) If a child drawn from the population tests negative, give the probability that the child does not have the food allergy. Again, you do not need to simplify your answer.

   (c) Are the events that a child tests negative and a that a child has the allergy mutually exclusive?

5. For a random sample $X_1, \ldots, X_n$, recall the formula for the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and for the sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

   (a) What does S_n^2 describe about a random sample?

   (b) Compute S_n^2 for a random sample with the three values: $3, 7, 11$.

   (c) Consider S_n^2 for a random sample with the values $6, 7, 8$. Will it be smaller or larger than the value of S_n^2 from part (b)?

6. Answer each question as TRUE or FALSE.

   (a) The larger the p-value the more implausible the null hypothesis is in light of the data.

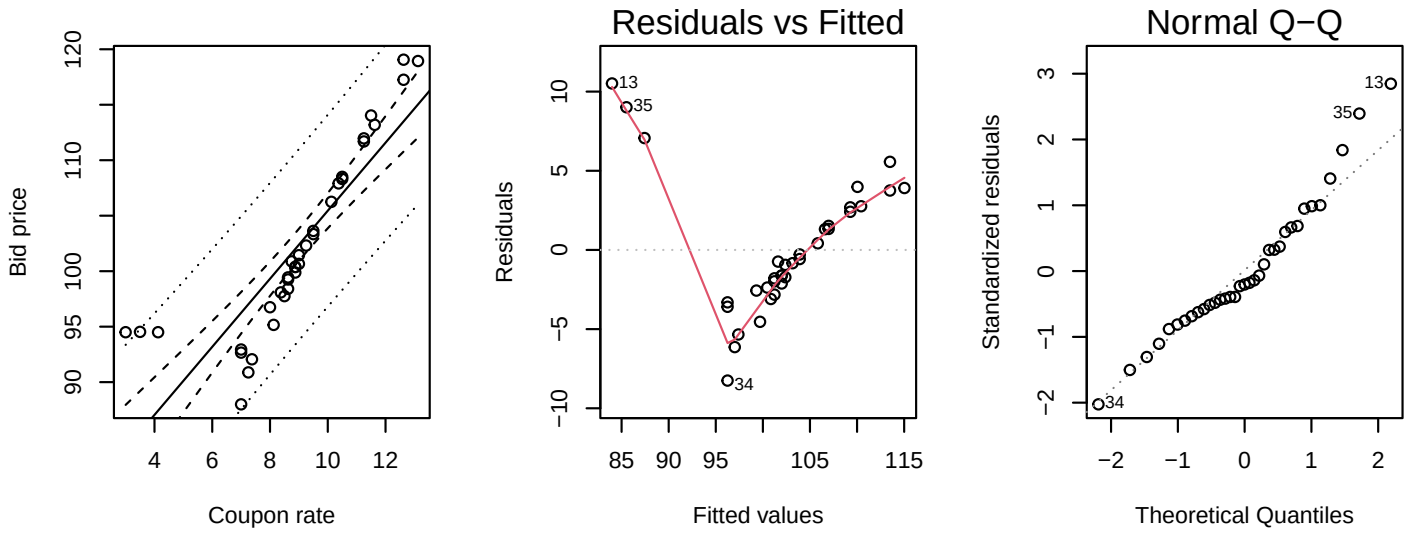
   (b) If a value lies within a 95% confidence interval then it will also lie within the 99% confidence interval computed on the same data.

   (c) We cannot make a Type II error if the null hypothesis is true.

   (d) If we increase the significance level α of a test of hypotheses, then we increase the probability of making a Type I error.

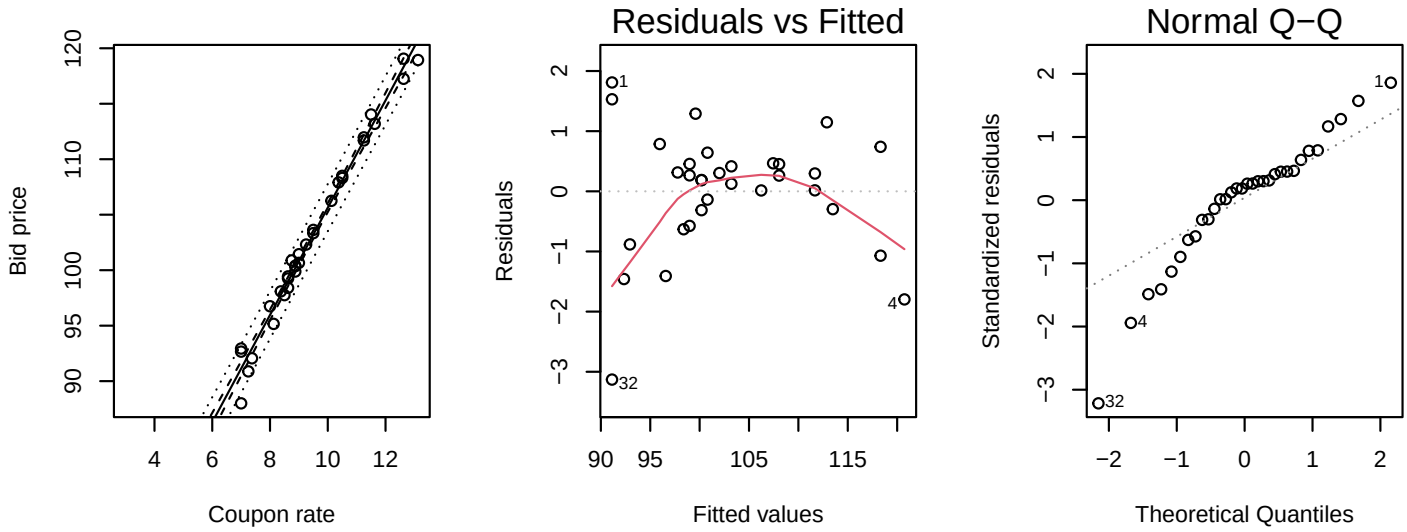
7. Consider two analyses of a data set of bid prices (selling price) versus the coupon rates (the payout to the holder every six months) on November 9, 1988, of US Treasury bonds maturing between 1994 and 1998. The leftmost panel in each analysis shows a scatterplot with the least-squares line overlaid, along with confidence and prediction intervals over the range of the observed data.

Analysis 1:




```
        Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  74.7866     2.8267  26.458  < 2e-16 ***
coupon_rate   3.0661     0.3068   9.994 1.64e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis 2:



```
        Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  57.2932     1.0358   55.31   <2e-16 ***
coupon_rate   4.8338     0.1082   44.67   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) What accounts for the difference between Analysis 1 and Analysis 2?

(b) For both Analysis 1 and Analysis 2, give the estimated increase in the expected bid price due to an increase by one unit in the coupon rate.

(c) Considering Analysis 2, give your conclusion about the null hypothesis H_0: $\beta_1 = 0$.

(d) In Analysis 1, what does the residuals versus fitted values plot reveal?

(e) In Analysis 1, what does the residuals versus fitted values plot reveal?

(f) Use Analysis 2 to predict the bid price of a US treasury bond which has a coupon rate of 10.

(g) Explain why the confidence and prediction intervals are much narrower in Analysis 2.

weight_keys	year	bm	ne	bmov	ie	major	app	coffee	car	plant	bike	sibs	cats	music	ppen	apple	grad	sumjob	school	camp
27	junior	m	n	m	i	CIS	messages	n	n	n	n	2	y	n	pen	y	y	y	private	y
62	senior	m	n	m	i	CIS	spotify	y	y	n	y	2	n	n	pencil	y	n	y	public	y
93	junior	b	n	m	i	CIS	NA	y	y	y	n	1	y	y	pen	y	y	y	public	y
88	junior	m	e	m	i	CIS	youtube	n	y	n	n	1	y	n	pen	y	n	y	public	n
73	sophomore	b	n	m	NA	Econ	X	y	y	n	y	1	n	n	pen	y	n	y	public	y
68	junior	m	e	b	e	Finance_Econ	CNN	y	y	n	n	1	y	n	pen	y	y	y	private	y
23	sophomore	b	e	b	i	Chemistry	instagram	n	n	y	y	2	y	n	pen	y	y	y	public	y
23	sophomore	b	n	m	e	CIS	NA	n	y	n	n	2	y	n	pencil	y	n	y	public	n
33	junior	b	n	m	e	Finance	snapchat	y	y	n	n	1	n	n	pencil	y	n	y	private	n
93	senior	m	n	m	i	CIS	spotify	n	y	n	n	1	y	y	pencil	y	n	y	public	n
87	freshman	m	n	m	e	Math	netflix	n	y	n	y	2	y	y	pencil	y	y	y	public	y
30	junior	b	n	m	e	CIS	snapchat	y	y	n	y	1	n	y	pen	y	n	y	public	n
91	junior	b	n	m	i	CIS	youtube	n	y	n	n	1	y	y	pen	y	n	y	public	y
96	junior	b	n	m	i	SEM	X	n	y	y	n	1	n	n	pen	y	n	y	public	n
NA	junior	b	n	m	e	Math	blackboard	n	y	n	n	0	n	n	pencil	y	n	y	public	y
12	junior	b	n	m	e	Marine_Science	snapchat	y	y	n	n	2	n	n	pencil	y	y	y	private	y
239	junior	b	e	b	e	Marine_Science	instagram	y	y	y	n	1	y	y	pencil	y	y	y	public	y
163	junior	b	n	b	i	CIS	tiktok	y	y	n	n	0	y	y	pencil	y	y	n	public	n
54	sophomore	m	e	m	e	CIS	safari	y	y	n	n	1	y	y	pencil	y	n	n	public	n
90	sophomore	m	n	m	e	Statistics	snapchat	n	y	n	y	3	y	n	pen	y	n	y	private	y
167	senior	b	n	m	i	Marine_Science	tiktok	n	n	y	y	4	y	n	pencil	y	y	n	public	y
66	junior	b	n	m	e	Marine_Biology	clash_of_clans	n	y	n	n	2	y	n	pencil	y	y	y	public	y
150	junior	m	n	m	i	CIS	youtube	y	y	n	n	1	y	n	pencil	y	y	y	public	y
98	junior	m	n	m	i	CIS	instagram	n	y	n	y	1	n	n	pencil	y	n	y	private	n
0	senior	b	n	m	i	Econ	X	y	y	y	n	1	y	n	pencil	y	n	y	public	n
54	junior	m	n	m	i	CIS	spotify	n	y	n	y	1	n	y	pencil	y	n	n	private	y
55	junior	m	n	m	e	CIS	instagram	y	y	y	y	1	y	n	pen	y	y	y	private	y
86	NA	m	e	m	NA	NA	google_sheets	n	y	y	y	2	n	n	pen	y	y	y	public	n
62	freshman	m	e	m	i	Math	delta_chat	n	n	n	n	2	y	n	pencil	n	n	n	public	y
119	grad	b	n	m	e	Econ	youtube	y	y	y	y	2	n	n	pen	y	y	y	public	y
0	junior	m	n	b	i	CIS	snapchat	y	n	n	n	2	n	n	pencil	n	n	y	public	y
62	grad	b	n	m	i	Biology	yahoo_email	n	n	n	n	1	n	n	pen	y	y	n	private	y
78	sophomore	b	n	b	i	Neuroscience	instagram	y	y	n	n	1	y	y	pen	y	y	y	public	n
154	junior	m	e	b	e	Environmental_Science	pinterest	n	y	y	n	2	n	n	NA	y	y	y	public	n
165	senior	b	n	m	e	Education	instagram	y	y	y	n	1	n	n	pencil	y	y	y	public	y
44	junior	b	n	m	i	Education	instagram	y	n	n	n	2	n	n	pencil	y	y	y	public	n
90	grad	b	n	m	e	SEM	instagram	y	n	n	n	1	n	n	pen	y	y	y	public	y
70	senior	b	n	m	i	CIS	discord	y	n	n	n	2	n	n	pen	y	y	y	public	y