

Quiz II

STAT 305, Section 3 FALL 2019

Instructions

- The quiz is scheduled for 80 minutes, from 09:30 to 10:50 AM. At 10:50 AM the exam will end.
- Total points for the exam is 60. Points for individual questions are given at the beginning of each problem. Show all your calculations clearly to get full credit. Put final answers in the box at the right (except for the diagrams!).
- A formula sheet is attached to the end of the exam. Feel free to tear it off.
- Normal quantile table is attached to the end of the exam. Feel free to tear it off.
- You may use a calculator during this exam.
- Answer the questions in the space provided. If you run out of room, continue on the back of the page.
- If you have any questions about, or need clarification on the meaning of an item on this exam, please ask your instructor. No other form of external help is permitted attempting to receive help or provide help to others will be considered cheating.
- **Do not cheat on this exam.** Academic integrity demands an honest and fair testing environment. Cheating will not be tolerated and will result in an immediate score of 0 on the exam and an incident report will be submitted to the dean's office.

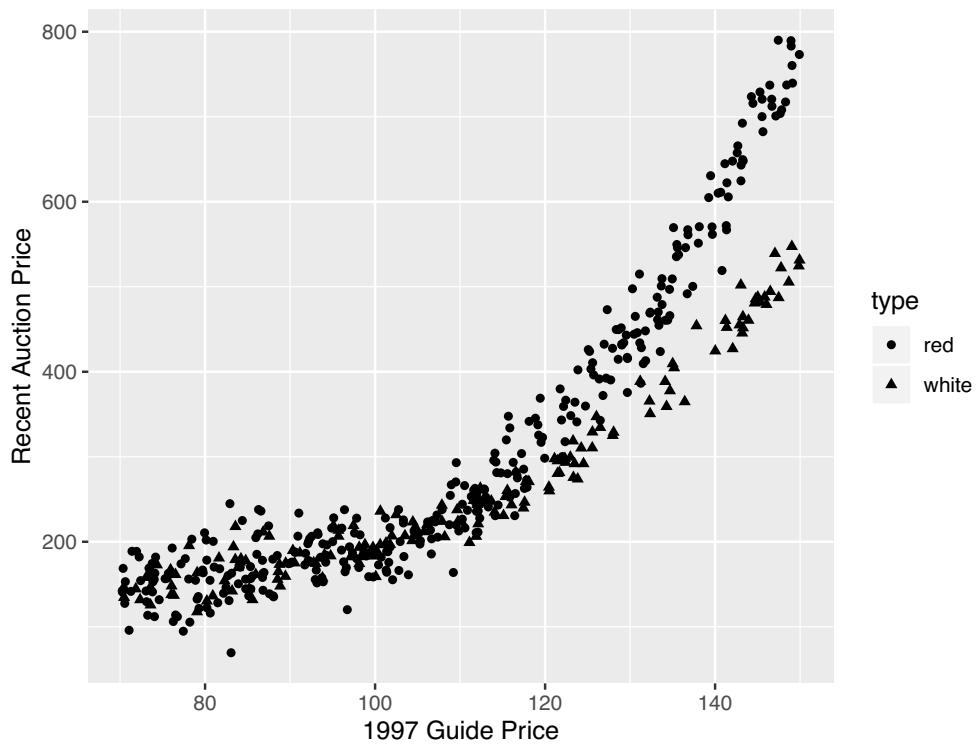
Name: _____ *Ke J*

Student ID: _____

1. Champions Red Wine

Oenophiles, or connoisseurs of fine wines, have been benefitted by sort of open information sharing we have in the days of online auctions. By monitoring the prices for which sought after bottles actually sell for, wine enthusiasts are able to both determine whether or not they are getting a fair price and whether or not they might make a nice profit by selling off a few of the bottles they have in storage. Before this era of easy and free data though, there was one source of information that stood above all the others: *Le Champions de vin rouge*, a guide to the world of fine wines published in 1997. While the information contained in it is largely still relevant in general, the price index has become somewhat outdated.

Or has it? Is it possible that the prices from the guide in 1997 could help us understand the market today? A certain statistician decided to do a deep analysis of the relationship between the cost of wine bottles in the 1997 book and the actual prices the same bottles fetched on open auction sites recently. The dataset he created consists of the last 500 bottles with publically available auction prices that were also listed in *Le Champions de vin rouge* price index.



Here are some summaries of the dataset with the 1997 Guide Price as x and the Recent Auction Price as y .

$$\sum_{i=1}^{500} x_i = 54348$$

$$\sum_{i=1}^{500} x_i^2 = 6164828$$

$$\sum_{i=1}^{500} y_i = 146782$$

$$\sum_{i=1}^{500} y_i^2 = 55928348$$

$$\sum_{i=1}^{500} x_i y_i = 17573001$$

- (a) Using the summaries, fit a linear relationship between **1997 Guide Price (x)** and **Recent Auction Price (y)**.

i. (5 points) Write the equation of the fitted linear relationship.

To write down the fitted relationship, we need to find b_0, b_1 . Then $\hat{y} = b_0 + b_1 x$.

$$b_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{17573001 - 500 \left(\frac{54348}{500}\right) \left(\frac{146782}{500}\right)}{(6164828 - 500 \left(\frac{54348}{500}\right)^2)} = 6.28$$

$$b_0 = \bar{y} - b_1 \bar{x} = \frac{146782}{500} - 6.28 \left(\frac{54348}{500}\right) = -389.0469$$

$$\text{Fitted relationship: } \hat{y} = -389.0469 + 6.28x$$

- ii. (5 points) Find and interpret the value of R^2 for the fitted linear relationship.

First find sample correlation (r) & then $R^2 = (r)^2$.

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}} = \frac{17573001 - 500 \left(\frac{54348}{500}\right) \left(\frac{146782}{500}\right)}{\sqrt{6164828 - 500 \left(\frac{54348}{500}\right)^2} \sqrt{55928348 - 500 \left(\frac{146782}{500}\right)^2}} = 0.89 \Rightarrow R^2 = (0.89)^2 = 0.79$$

79% of the variation in the response (auction price) can be explained by the linear relationship between the auction price and the 1997 Guide price

- iii. (5 points) Using the fitted line, provide a predicted Recent Auction Price when the 1997 Guide Price was \$120.

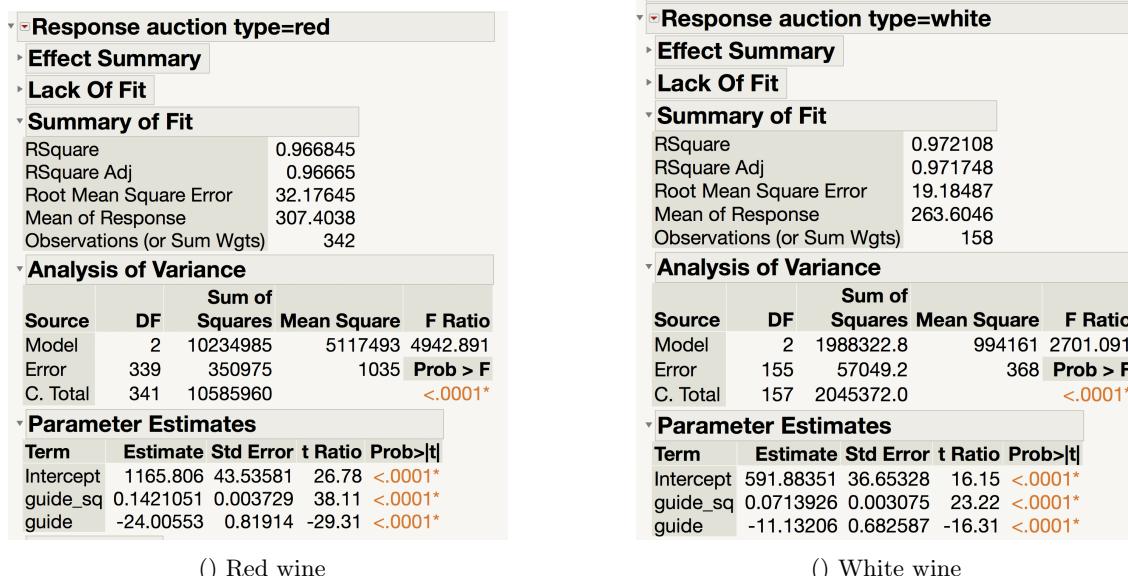
$$\hat{y} = -389.0469 + 6.28(120)$$

$$= \$364.6$$

- (b) The JMP output below comes from fitting a quadratic model using x and x^2 for *Red Wine Only* (left) and *White Wine Only* (right).

- i. (5 points) Write the equation of the fitted quadratic relationship for Red Wines i.e the polynomial to degree two.

$$\hat{y} = 1165.806 + 0.142(\text{Red wine}) - 24.00553(\text{Red wine})^2$$



() Red wine

() White wine

Figure 1: JMP output of fitting a quadratic model using x and x^2 for *Red Wine Only* (left) and *White Wine Only* (right).

- ii. (5 points) Write the equation of the fitted quadratic relationship for *White Wines* i.e the polynomial to degree two. .

$$\hat{y} = 591.88351 + 0.7139(\text{white wine}) - 11.13206(\text{white wine})^2$$

- iii. (5 points) Interpret the value of R^2 for the two fitted quadratic relationships.

- 96% of the variation in the auction price (response) can be explained by the quadratic relationship of the response and red wine price.
- 97% of the variation in the auction price (response) can be explained by the "quadratic" relationship between the auction price & white wine price.

- iv. (5 points) Using the fitted quadratic relationship, provide a predicted value of Recent Auction Price for a red wine with 1997 Guide Price of \$120. 4cm]

$$\begin{aligned}\hat{y} &= 1165.806 + 0.142(120) - 24(120)^2 \\ &= -345600\end{aligned}$$

Note that the result is a generated data set, and achieving a negative value for the recent auction price is acceptable in this case.

2. Suppose X is a discrete random variable with following probability function:

$$f(x) = \begin{cases} 0.1 & x = -2, 0, 2 \\ 0.35 & x = -1, 1 \\ 0 & o.w. \end{cases}$$

- i. (2 points) Find $P(X = 0)$

$$P(X=0) = 0.1$$

- ii. (2 points) Find $P(X \leq 0)$

$$\begin{aligned} P(X \leq 0) &= P(X = 0 \text{ or } X = -1 \text{ or } X = -2) = P(X = 0) + P(X = -1) + P(X = -2) \\ &= 0.1 + 0.35 + 0.1 = 0.55 \end{aligned}$$

- iii. (2 points) Find $P(|X| > 1)$

$$\begin{aligned} P(|X| > 1) &= P(X = -2 \text{ or } X = 2) = P(X = -2) + P(X = 2) \\ &= 0.1 + 0.1 = 0.2 \end{aligned}$$

- iv. (3 points) Find the CDF of X .

X	-2	-1	0	1	2
$P(X)$	0.1	0.35	0.1	0.35	0.1
$F(x)$	0.1	0.45	0.55	0.9	1

- v. (3 points) Find the expected value of X .

$$\begin{aligned} E[X] &= \sum_{x \in \{-2, -1, 0, 1, 2\}} x P(x) = -2P(-2) + (-1)P(-1) + (0)P(0) + (1)P(1) + (2)P(2) \\ &= -2(0.1) - 1(0.35) + 0 + 0.35 + 2(0.1) \\ &= 0 \end{aligned}$$

- vi. (3 points) Find the variance of X .

You can use any formulas.

$$\begin{aligned} \text{Var}(X) &= \sum (x_i - E[X])^2 P(x) = (-2 - 0)^2 P(-2) + (-1 - 0)^2 P(-1) + (0 - 0)^2 P(0) + \\ &\quad (1 - 0)^2 P(1) + (2 - 0)^2 P(2) \\ &= 4(0.1) + 1(0.35) + 0 + 1(0.35) + 4(0.1) \\ &= 1.5 \end{aligned}$$

3. Suppose a standup comedian plans to give a total of $n = 5$ jokes in an entire 2-hour performance. Call a joke a success if at least one audience member laughs. If no audience member laughs, the joke is a failure. Assume that all the jokes are equally funny, with $p = p(\text{success}) = 0.2$. Let X be the random variable that denotes the number of jokes out of the total 5 were successes.

i. (3 points) Precisely state the distribution of X , giving the values of any parameters necessary.

$$X \sim \text{binomial}(n=5, p=0.2)$$

ii. (3 points) Calculate the probability that the whole night is a failure: i.e., $P(\text{no laughs})$

$$\begin{aligned} P(X=0) &= \binom{5}{0} (0.2)^0 (1-0.2)^{5-0} \\ &= (0.8)^5 \end{aligned}$$

iii. (3 points) Calculate the probability that the comedian tells at least 4 successful jokes.

$$\begin{aligned} P(X \geq 4) &= P(X=4) + P(X=5) \\ &= \binom{5}{4} (0.2)^4 (0.8)^1 + \binom{5}{5} (0.2)^5 (0.8)^0 \\ &= 0.0064 + 0.00032 = 0.00672 \end{aligned}$$

iv. (3 points) Calculate the expected number of successful jokes.

In Binomial distribution

This means it is expected that

$$E(X) = n \cdot p = 5(0.2) = 1 \quad \text{one of his jokes is successful}$$

v. (3 points) Calculate the standard deviation of X

$$\text{in Binomial, } \text{var}(X) = np(1-p) = 5(0.2)(1-0.2) = 0.8$$

$$\text{SD}(X) = \sqrt{\text{var}(X)} = \sqrt{0.8} = 0.89$$