

# Exam I

## STAT 305, Section D FALL 2019

### Instructions

- The exam is scheduled for 80 minutes, from 3:40 to 5:00 PM. At 5:00 PM the exam will end.
- A formula sheet is attached to the end of the exam. Feel free to tear it off.
- You are allowed to use a self-produced one-page (front and back) formula sheet during this exam.
- 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: \_\_\_\_\_

Student ID: \_\_\_\_\_

1. (2 points) Circle the **bold face** term that makes the following statement true:

A measurement device that reports the measurements which are close to each other when repeatedly measuring the same thing is (**precise** or **accurate**).

2. (2 points) When dealing with qualitative data, the best way to describe the common value is using the (**mode** or **mean** or **median**).

3. A sample of size 5 was drawn from a population and the resulting observations are reported below.

110, 100, 105, 103, 105, 115

Using these observed values, report the following:

(a) (2 points) the mean

(b) (2 points) the mode

(c) (2 points) the median

(d) (2 points) the variance

(e) (2 points) the standard deviation

(f) (2 points) the value of  $Q(.75)$

(g) (2 points) the interquartile range

4. In the field of psychology, the Dunning–Kruger effect is a cognitive bias in which people of low ability overestimate their ability and students with high ability underestimate their ability. In order to study this phenomenon, a team of psychologists at Iowa State asked students from the College of Liberal Arts and Sciences to estimate the score they would receive on a standardized grammar test (as a percentage from 0 to 100). They recorded the student's sex, age, and major. The students were then given the standardized grammar test with 50 questions and the number of correct answers was recorded.

(a) (2 points) Is this an experiment or an observational study?

(b) (2 points) What is the population under study?

(c) For each of the following variables,

- Identify whether it is qualitative or quantitative variable, and
- If it is qualitative, provide the values it can take (if there are more than three, list at least three).
- If it is quantitative, is it continuous or discrete?

i. the number of questions the student answered correctly on the standardized test.

ii. The score the student predicted they would receive on the standardized test.

iii. The student's sex.

iv. The student's age.

v. The student's major.

**5. Twin Cinema**

A professor of theoretical sociology is convinced that horror movies contribute to anxiety. In order to test his hypothesis, the professor arranged to have to students in his class go to a movie showing on campus for extra credit. However, he randomly told half the students to see a movie in one location (location A) showing a scary movie! The other half of his students, he sent to a second location (location B) showing a romantic comedy. In order to fund this experiment through the National Institute of Health, the professor added a nutrition component: in order to examine the effect of diet on anxiety, he gave the students a voucher which could be redeemed for either a high caffeine soda, a tub of popcorn, or a candy snack. At the end of the movies, he recorded which movie the student had seen, which snack they had chosen, and then gave the student a questionnaire that scored the student as having low, medium, or high anxiety.

(a) (2 points) Is this an experiment or an observational study? Explain.

(b) Identify the following (if there was not one, simply put "not used").

i. (2 points) Response variable(s):

ii. (2 points) Experimental variable(s):

iii. (2 points) Blocking variable(s):

(c) (2 points) Was replication used in this experiment? If so, where was it applied? If not, how could we have applied it?

## 6. From Blown Speakers

In car audio speakers, large changes in the intensity of sound produced by the speaker may cause a speaker component known as a cone to vibrate non-uniformly. The non-uniform vibrations of the cone cause the cone to add stress to the materials surrounding the cone. The damage to the surrounding materials results in the sound produced by the speakers becoming crackly - a state in which the speaker is "blown out".

Audio engineers working on ways to stabilize the cone during large changes in intensity are reviewing data that a new "smart" audio system collects and reports information when a blown speaker is detected. The following is a list of the **change in intensity** (as measured in decibels) from the first 35 audio systems that sent a blown speaker report.

```
[1] 25.4 22.2 13.7 18.6 15.5 38.4 23.3 14.9 18.7 25.5 22.6 25.8 37.5 22.0
[15] 23.2 20.4 22.6 16.8 17.1 23.5 21.0 23.7 24.8 14.3 20.3 26.7 20.8 38.2
[29] 15.3 27.9 3.8 2.7 16.5 23.5 15.4
```

In this case, the range is 35.7, Q1 is 16.575, Q2 is 22, and the IQR is 2.525.

(a) (5 points) Fill in the values for the 5-number summary table:

Minimum	First Quartile	Median	Third Quartile	Maximum

(b) (10 points) Complete the following frequency table using 8 equal length intervals for the value range (it should include all the data):

Value Range	Frequency	Relative Frequency	Cumulative Relative Frequency

(c) (5 points) Create a **frequency histogram** to summarize the data. Carefully label the axes.

(d) (10 points) Create a box plot to summarize the data. Carefully label the axes.

(e) (4 points) Are there any unusually observations? If so what were the speeds for those observations?

(f) (10 points) The engineers also get information about how far the base of the cone was displaced during the blowout. Here are some of the displacements:

0, 0.3, 0.4, 0.6, 0.7, 0.7, 0.8, 1, 1.3, 3.5

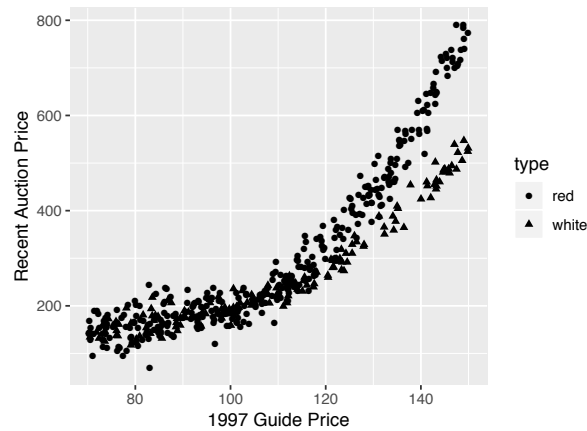
Create a theoretical Q-Q plot using the following quantiles from the normal distribution as the theoretical quantiles. Carefully label your axes. What does this graph tell us about the upload speeds?

	1	2	3	4	5	6	7	8	9	10
$p$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$Q(p)$	-1.64	-1.04	-0.67	-0.39	-0.13	0.13	0.39	0.67	1.04	1.64

## 7. 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.
  - ii. (5 points) Find and interpret the value of  $R^2$  for the fitted linear relationship.
  - iii. (5 points) Using the fitted line, provide a predicted Recent Auction Price when the 1997 Guide Price was \$120.
  - iv. (5 points) Sketch what you believe the plot of residuals vs 1997 Guide Price would look like. Why would this be a problem?
  - v. (3 points) The points in the plot have different shapes based on whether the wine is "Red" or "White." Is there a reason to believe that white wines and red wines should not be modeled together? Explain why.



- (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).

Response auction type=red						Response auction type=white					
Effect Summary						Effect Summary					
Lack Of Fit						Lack Of Fit					
Summary of Fit						Summary of Fit					
RSquare		0.966845				RSquare		0.972108			
RSquare Adj		0.96665				RSquare Adj		0.971748			
Root Mean Square Error		32.17645				Root Mean Square Error		19.18487			
Mean of Response		307.4038				Mean of Response		263.6046			
Observations (or Sum Wgts)		342				Observations (or Sum Wgts)		158			
Analysis of Variance						Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio		Source	DF	Sum of Squares	Mean Square	F Ratio	
Model	2	10234985	5117493	4942.891		Model	2	1988322.8	994161	2701.091	
Error	339	350975	1035	Prob > F		Error	155	57049.2	368	Prob > F	
C. Total	341	10585960		<.0001*		C. Total	157	2045372.0		<.0001*	
Parameter Estimates						Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob> t		Term	Estimate	Std Error	t Ratio	Prob> t	
Intercept	1165.806	43.53581	26.78	<.0001*		Intercept	591.88351	36.65328	16.15	<.0001*	
guide_sq	0.1421051	0.003729	38.11	<.0001*		guide_sq	0.0713926	0.003075	23.22	<.0001*	
guide	-24.00553	0.81914	-29.31	<.0001*		guide	-11.13206	0.682587	-16.31	<.0001*	

- (5 points) Write the equation of the fitted quadratic relationship for *Red Wines*.
- (5 points) Write the equation of the fitted quadratic relationship for *White Wines*.
- (5 points) Interpret the value of  $R^2$  for the two fitted quadratic relationships.
- (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.



# STAT 305 Exam I

## Reference Sheet

### Numeric Summaries

mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
population variance	$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
population standard deviation	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
sample variance	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
sample standard deviation	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

### Linear Relationships

Form	$y \approx \beta_0 + \beta_1 x$
Fitted linear relationship	$\hat{y} = b_0 + b_1 x$
Least squares estimates	$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$
	$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$
	$b_0 = \bar{y} - b_1 \bar{x}$
Residuals	$e_i = y_i - \hat{y}_i$
sample correlation coefficient	$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$
	$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}}$
coefficient of determination	$R^2 = (r)^2$

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

### Multivariate Relationships

Form	$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$
Fitted relationship	$\hat{y} \approx b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$
Residuals	$e_i = y_i - \hat{y}_i$
Sums of Squares	$SSTO = \sum_{i=1}^n (y_i - \bar{y})^2$
	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$
	$SSR = SSTO - SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
coefficient of determination	$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
	$R^2 = \frac{SSTO - SSE}{SSTO}$
	$R^2 = \frac{SSR}{SSTO}$
	$\frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

### Functions

**Quantile Function**  $Q(p)$  For a univariate sample consisting of  $n$  values that are ordered so that  $x_1 \leq x_2 \leq \dots \leq x_n$  and value  $p$  where  $0 \leq p \leq 1$ , let  $i = \lfloor n \cdot p + 0.5 \rfloor$ . Then the quantile function at  $p$  is:

$$Q(p) = x_i + (n \cdot p + 0.5 - i)(x_{i+1} - x_i)$$