Midterm 1 - STAT 301
Fall 2017

Name:

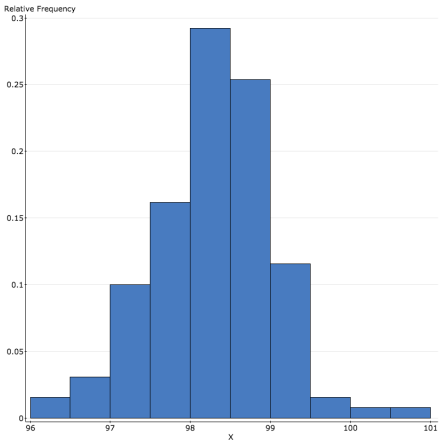
UIN:

Signature:

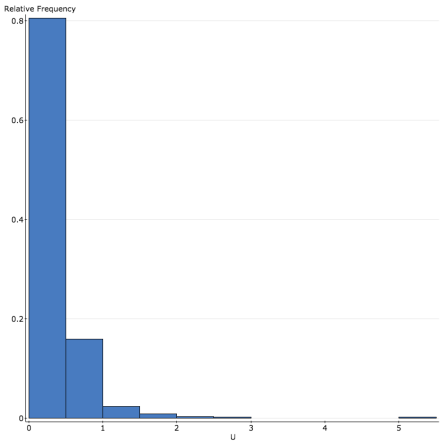
1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You can only bring the normal z-tables to this exam. A cheat sheet will be provided.
4. You have 60 minutes to work on this exam. There are 15 multiple choice questions. If you cannot do one question move on to the next.
5. You may use a calculator in the exam (but not a phone).
6. Partial credit is only given to answers which are correct (not partially correct).
7. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
8. I use sd to denote standard deviation.
9. Good Luck!!!

(1-2) The relative frequency histograms of three data sets are given in the plots below (since it is hard to see the range of the x-axis, it is stated below each plot).

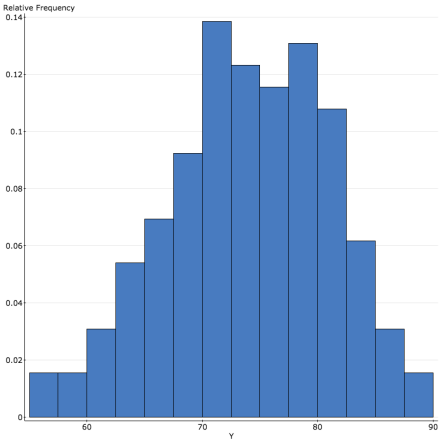
(1) Match the histogram to the data.



Plot (1). x-axis = 96 to 101.



Plot (2) x-axis = 0 to 5.



Plot (3) x-axis = 55 to 90

HOUSE House prices in California (in millions).

TEMP The temperature of a person (in Fahrenheit).

HEART Resting heart rate of a person (in beats per minute).

	(1)	(2)	(3)
A	HOUSE	TEMP	HEART
B	HOUSE	HEART	TEMP
C	HEART	TEMP	HEART
D	TEMP	HEART	HOUSE
E	TEMP	HOUSE	HEART

(2) Match the plot to the mean and standard deviation:

	(1)	(2)	(3)
A	mean = 98.2, sd = 0.73	mean = 0.38, sd = 0.12	mean = 73, sd = 7
B	mean = 98.2, sd = 5	mean = 0.38, sd = 5	mean = 73, sd = 40
C	mean = 5, sd = 98.2	mean = 5, sd = 0.38	mean = 40, sd = 73
D	mean = 98.2, sd = 5	mean = 2.5, sd = 5	mean = 73, sd = 40
E	mean = 98.2, sd = 0.1	mean = 0.38, sd = 0.1	mean = 73, sd = 1

(3) During the 1960s gastric freezing was used to reduce ulcers in people.

The procedure was introduced after the results of clinical trial involving 100 patients with ulcers was published. **All** the patients in the trial were given the gastric treatment. After treatment, 60% of the treatment these patients observed a decrease in pain.

Based on your understanding of how data should be collected. Which of the comment(s) about the trial are the most reasonable?

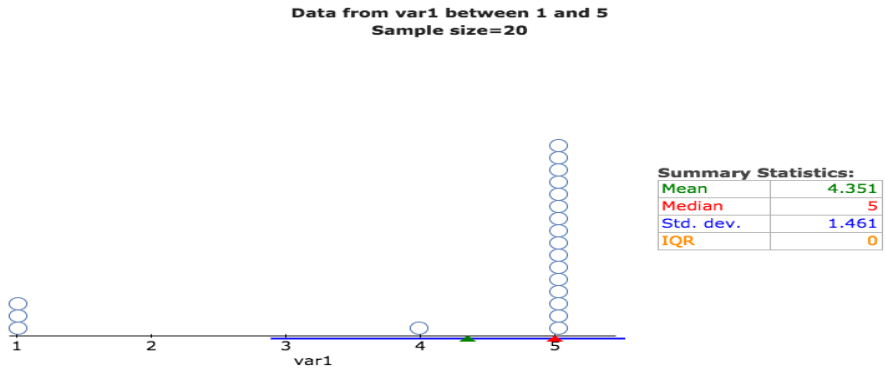
(A) All the patients were told they were given the treatment, which may have biased their response.

(B) The effect of the gastric freezing cannot be truely judged without comparison to a placebo treatment, which should have been included in the trial.

(C) The trial seemed reasonable and justifies the subsequent treatment.

(D) (A) and (B) (E) None of the above.

(4) Below, 20 observations are plotted on a time line.



What happens if **one** ball on number **5** is moved to number **1**.

(A) The *mean* and *standard deviation* stay the **same**, but the *median* and *IQR* **change**.

(B) The *mean* and *median* stay the **same**, but the *standard deviation* and *IQR* **change**.

(C) The *standard deviation* and *IQR* stay the **same**, but the *mean* and *median* **change**.

(D) The *median* and *IQR* stay the **same**, but the *mean* and *standard deviation* **change**.

(E) None of the above.

(5) The median of a data set is **6**, the first quartile is **4** and the third quartile is **6**. What can we say about this data set (making a boxplot can help)?

(A) 25% of the data set is the same. (B) 50% of the data set is the same.

(C) All of the data set is the same. (D) The standard deviation is zero.

(E) C and D.

(6) Suppose the distribution of a variable is normally distributed. What proportion of the data will be within 0.75 standard deviations of the mean?

(A) 51% (B) 54.7% (C) 22.7% (D) 77.3% (E) 75%

(7) In an airport in the U.S the following variables are measured.

1	2	3	4
Airline	The type of shops	The delay time of a flight	The number of passangers on a flight

Answer:

	Airline	Shop	Delay Time	Passanger Numbers
A	Numerical continuous	Categorical	Numerical discrete	Numerical continuous
B	Numerical continuous	Categorical	Numerical discrete	Numerical discrete
C	Numerical discrete	Categorical	Categorical	Numerical continuous
D	Categorical	Numerical continuous	Numerical continuous	Numerical Discrete
E	Categorical	Categorical	Numerical continuous	Numerical discrete

(8) Suppose male heights are normally distributed with mean **68** inches and standard deviation σ =**3** inches. Suppose female heights are normally distributed with mean **63** inches and standard deviation σ =**2** inches.

Jane is 59 inches tall. By matching the percentile, how tall would Jane be if she were male?

(A) 64 inches (B) 62 inches (C) 74 inches (D) 72 inches 71 inches.

(9) Suppose female heights are normally distributed with mean 65 inches and standard deviation 2.5 inches. Construct an interval **centered** about the mean where 80% of the female heights lie.

(A) 67.125 (B) [62.9,67.1] (C) [61.8, 68.2] (D) [64.2,65.8] (E) [64.47,65.53]

(10) The boxplot of baseball salaries are plotted over a 30 year period in Figure 1

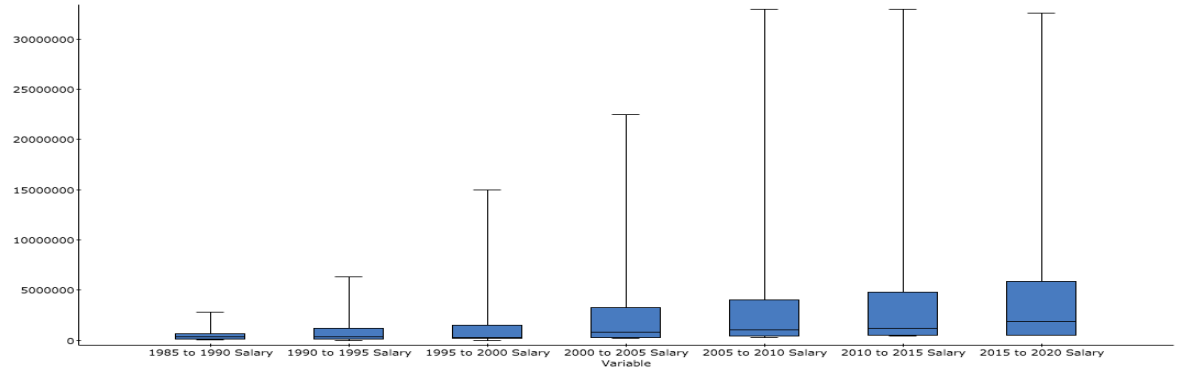


Figure 1: The boxplot of baseball salaries. In 5 year blocks from 1985 until now.

(A) The standard deviation of salaries tend to increase over time.

(B) The salaries are **right skewed**.

(C) The salaries are **symmetric** (D) [A] and [B] (E) [A] and [C].

(11) Below is the relative frequency histogram of old Faithful Geyser eruption times. The mean is 3.5 minutes and the standard deviation 1 minute.

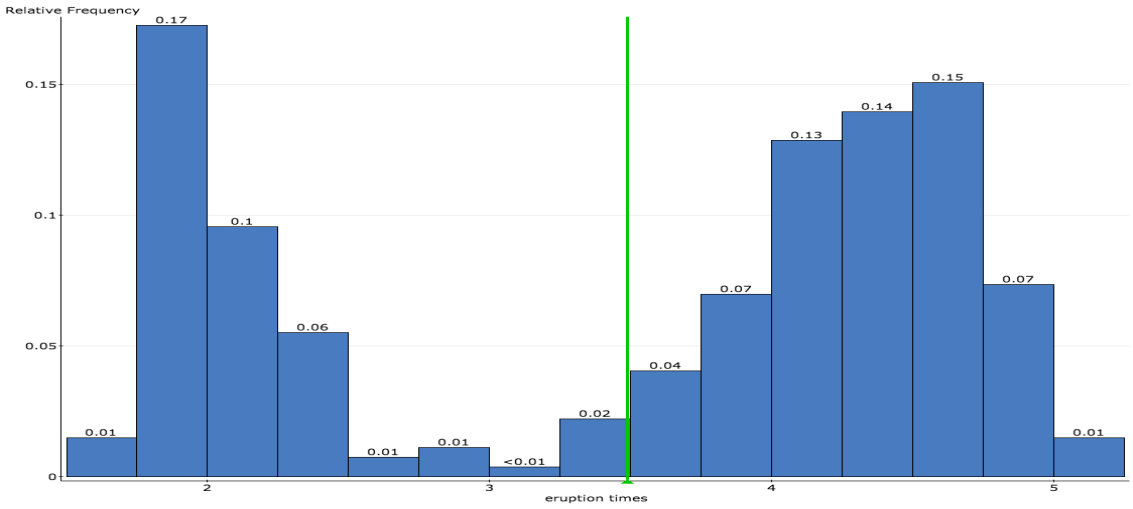


Figure 2: Old Faithful Eruption Times

Calculate the proportion of times which are **within** one standard deviation of the mean.

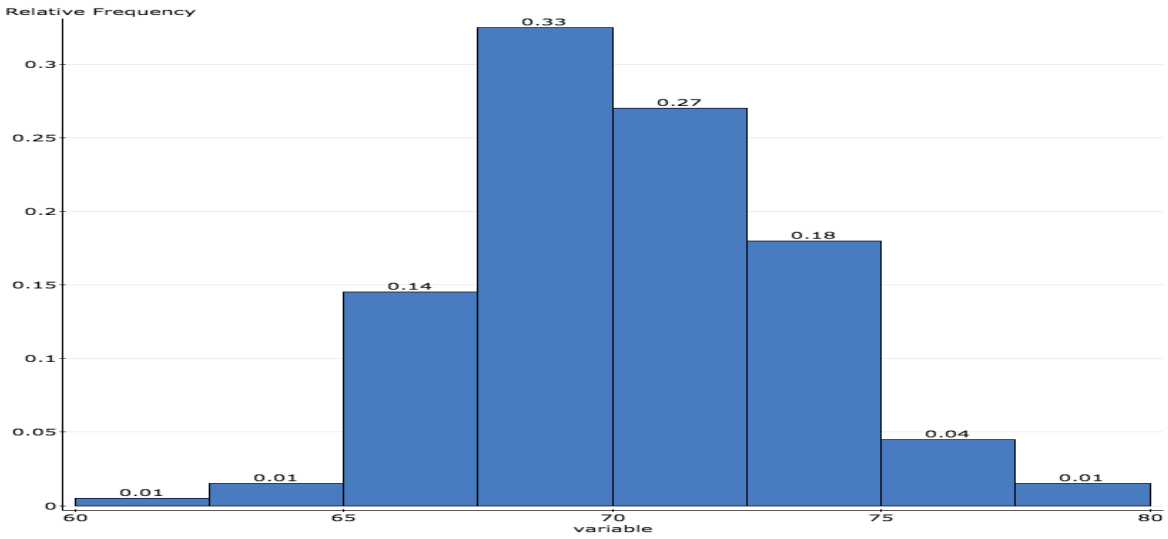
(A) 68% (B) 43% (C) 100% (D) 50% (E) 95%.

(12) It is believed that the height of cows are normally distributed with mean 70 inches and standard deviation 3 inches.

What proportion of cows have a height between 65 to 72.5 inches?

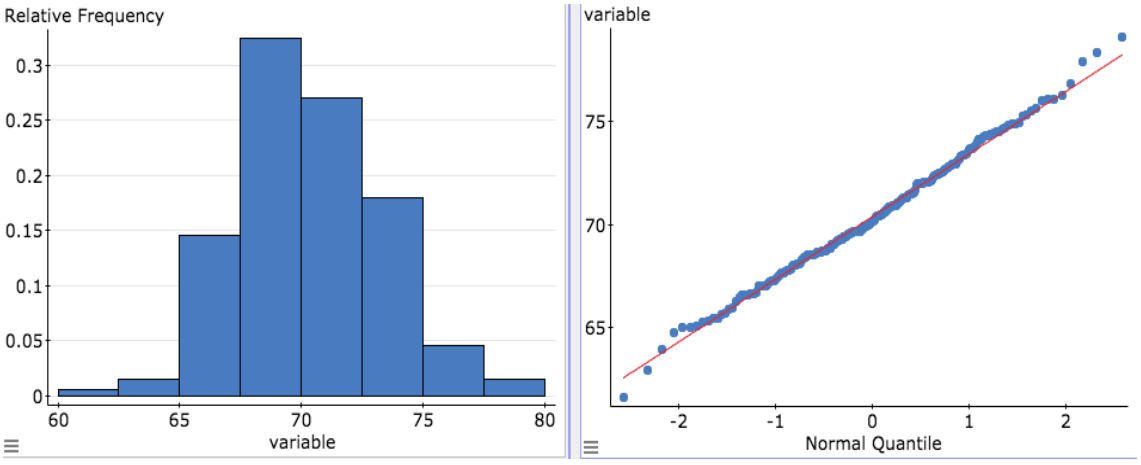
(A) 79.7% (B) 20.3% (C) 74.9% (D) 83.3% (E) -83.3%

(13) A sample of one thousand cows was collected and their height measured. The data is summarized in the relative frequency histogram below. Using the relative frequency histogram what proportion of the heights are between 65 to 72.5 inches.



(A) 100% (B) 68% (C) 95% (D) 78% (E) 74%

(14) The proportions calculated in question (12) and (13) were calculated in different ways. The histogram of the data and QQplot are given below.



Question overleaf.

Based on these plots, what statement below best describes what you observe.

(A) The distribution of the heights of cows is heavily right skewed.

(B) The distribution of the heights of cows is heavily left skewed.

(C) The proportions in (Q12) and (Q13) are close because the data is heavily right skewed.

(D) The proportions in (Q12) and (Q13) are close because the data is close to normal.

(E) (A) and (C).

(15)

1. A study of 2050 children in first and second grade childnre in 30 schools recordered their behaviours at lunch. One of the conclusions of the study was that first grader tend to throw away more food than second graders.
2. A class of students were randomly assigned to two groups. Group A was asked to do power weights everyday whereas Group B was asked to run one mile everyday. After one month it was reported that the bone strength in Group A was 30% more than the bone strength in Group B.
3. A magazine article tells the story of an elderly couple (male and female) who took Pilates classes. After taking the classes, the chronic back pain the man was suffering from had gone whereas the chronic back pain the female was suffering from remained. The article concludes than Pilates benefits males more than females.

What type of studies are the above?

	1	2	3
(A)	Biased	Anecdotal	Observational
(B)	Biased	Observational	Confounding
(C)	Observational	Anecdotal	Experimental
(D)	Observational	Experimental	Anecdotal
(E)	Observational	Experimental	Biased