

Stat 1040

Review 1

1. What is a randomized, controlled, double-blind experiment?

See Chapter 1.

2. The 2008 Women's Health Study followed a group of women who took either aspirin or a placebo (without knowing which) every other day for 10 years. The women who took aspirin had no difference in heart attack occurrence but they were 17 percent less likely to have a stroke than the women who took the placebo. Give brief answers to the following questions about this study:

a) Was the study controlled? How do you know?

Controlled, the treatment & control groups were determined by the investigators.

b) Was the study blind? How do you know?

Yes, a placebo was administered.

c) Was this study longitudinal or cross-sectional? How do you know?

Longitudinal, 10 years.

3. A few years ago, a researcher looked at a large, representative group of men. He found that, on average, the older these men were, the less meat they consumed. True or false and explain: "The data show that on average, men eat less meat as they get older."

False, healthier men generally eat less meat throughout their adult lives. Healthier men live longer.

4. A class of 400 students is divided into two sections of 200 each. Both sections are given a common final exam. The following numerical summaries are observed.

Section 1: $AV = 86$ $SD = 10$

Section 2: $AV = 56$ $SD = 10$

Suppose that all 400 scores are combined as one list. What is the average of the combined scores? Will the SD of the combined scores be equal to 10, less than 10, or greater than 10? Explain.

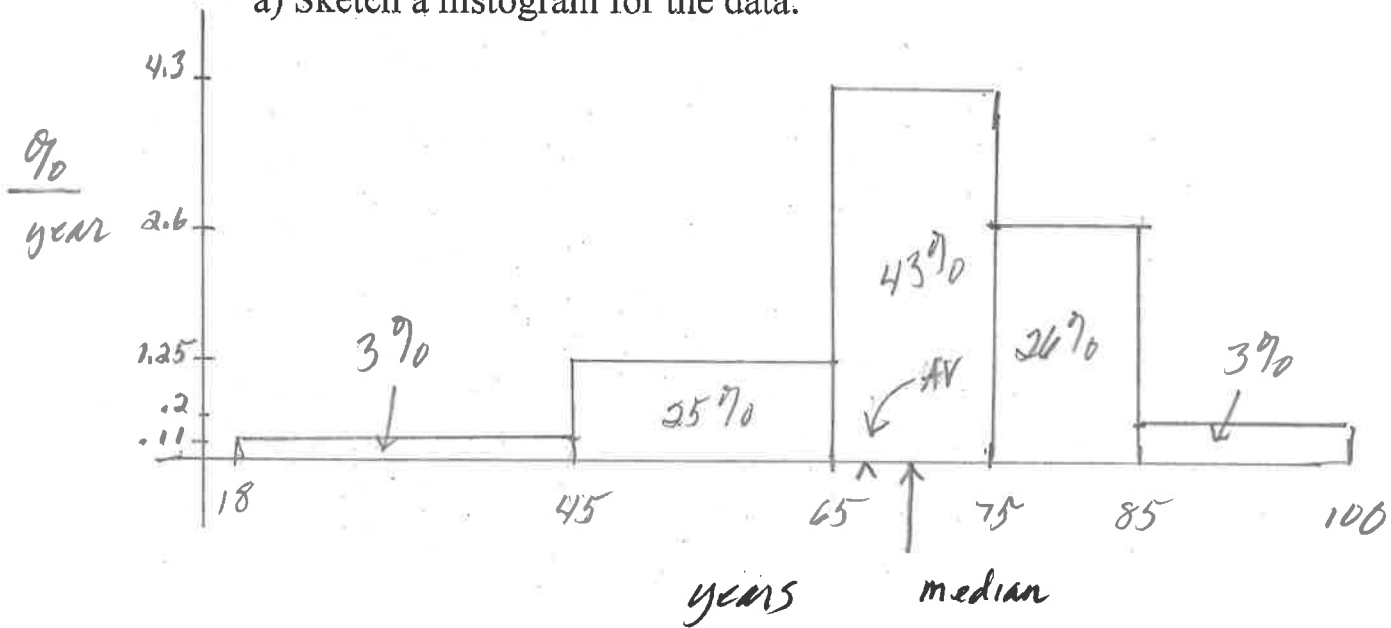
$$\text{Combined average} = \frac{86 + 56}{2} = 71.$$

The SD is greater than 10 because the data is more spread out.

5. According to USA Today, the age of getting knee replacements is distributed according to the following table.

Age (in years)	Percentage	Height = % / width
18-45	3	$3/27 = .11$
45-65	25	$25/20 = 1.25$
65-75	43	$43/10 = 4.3$
75-85	26	$26/10 = 2.6$
85-100	3	$3/15 = .2$

a) Sketch a histogram for the data.



- b) In which interval is the median age for these individuals? $[65, 75]$
- c) About what percentage of individuals were 70 or older when they had the surgery? $21.5 + 26 + 3 = 50.5\%$
- d) Indicate, approximately the average and median on the histogram.

6. Adult-American women's heights have a distribution that is approximately normal, with an average of 63.5 inches and an SD of 2.5 inches.

a) What is the percentage of women between 61 and 68.5 inches tall?

$$\frac{61 - 63.5}{2.5} = -1, \quad \frac{68.5 - 63.5}{2.5} = 2$$



$$\frac{A(1)}{2} + \frac{A(2)}{2}$$

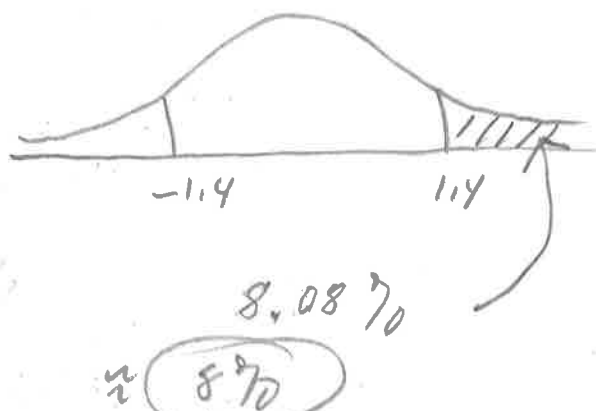
$$= 34 + 47.5 = 81.5\%$$

b) What is the percentage of women taller than 67 inches?

$$\frac{67 - 63.5}{2.5} = 1.4$$

$$A(1.4) = 83.85$$

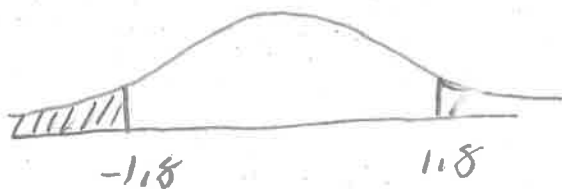
$$100 - 83.85 = 16.15\%$$



$$\approx 8\%$$

c) If a woman is 59 inches tall, in what percentile is she?

$$\frac{59 - 63.5}{2.5} = -1.8$$



$$A(1.8) = 92.81$$

$$100 - 92.81 = 7.19$$

$$\frac{7.19}{2} = 3.6\%$$

d) What height is the 90th percentile?

$$z = 1.3$$

$$\frac{\text{height} - 63.5}{2.5} = 1.3$$

$$66\frac{3}{4} \text{ inches}$$

7. For a Stat 1040 class with 380 students, the first midterm average was 78 and the SD was 10. The second midterm averaged 70 and the SD was also equal to 10. The correlation between the two midterms was $r = 0.8$. True or false and explain.

- a) If 10 points are added to the 1st midterm score of each student, the SD for the 1st midterm will increase slightly.

False, it is still = 10

- b) If 10 points are added to the 2nd midterm score of each student, the correlation between the two midterms will increase slightly.

False, r is still = .8

- c) Those students who scored one SD above average on the first midterm averaged less than one SD above average on the second midterm. This data supports the professor's claim that students tend to lose interest during the latter part of the semester.

False, regression effect.

8. The first year professional baseball allowed free agency (allowing players the right to negotiate contracts with any team), a few of the wealthiest clubs paid enormous sums to obtain a few players who had very high batting averages the previous year. During the course of the year, sports writers had lots of fun pointing out what bad judgment the owners had shown, for almost none of the high-priced players did as well that year as they had the year before. Is there a statistical explanation for this result, which does not assume the (now rich) players had grown fat and lazy?

Regression effect

9. A large group of students takes two successive IQ tests. The correlation between the two IQ tests is 0.8. Both tests have an average of 100 and an SD of 15. If someone scores 75 on the first test, which is the most likely:

- i) The person will score 75 on the second test.
- ii) The person will score somewhat less than 75 on the second test.
- iii) The person will score somewhat more than 75 on the second test.

regression effect

10. Reading comprehension is tested for all of the third-grade students in a large school district, once at the beginning of the school year and once at the end of the school year. The following results are obtained:

x - Beginning of year: average score = 75 SD = 15
 y - End of year: average score = 80 SD = 17 $r = 0.6$

a) What does r (the correlation) measure?

The strength of the linear relationship, or the degree of clustering about the SD line.

b) Describe how the correlation is computed.

AV of the products $\left(\frac{x - AV_x}{SD_x} \right) \left(\frac{y - AV_y}{SD_y} \right)$

c) Find the slope of the SD line.

$$+ \frac{SD_y}{SD_x} = \frac{17}{15}$$

d) Find the slope of the regression line.

$$r \cdot \frac{SD_y}{SD_x} = (.6) \frac{17}{15} = .68$$

e) Associate with each increase of 1 SD in x there is an increase of only .68 SDs in y .

↑ .68

f) Find the equation of the regression line for predicting end-of-year score from beginning-of-year score.

$$y - y_1 = m(x - x_1) \quad y - 80 = .68(x - 75)$$

$$y = .68x + 29$$

g) Use the regression method to estimate end-of-year score for a student who scored 65 on the initial test.

$$\text{When } x = 65, \quad y = (.68)(65) + 29 = 73.2$$

h) Find the RMS error. What does the RMS error measure?

$$\sqrt{1 - r^2} \cdot SD_y = 13.6$$

It measures the overall size of the error in regression.

i) For those students who scored 65 on the beginning-of-year score, what percentage scored 75 or higher on the end-of-year score?

$$\text{New } \mu = 73.2$$

$$\text{New } SD = 13.6$$

$$\frac{75 - 73.2}{13.6} = .13$$



11. Revisit the following review exercises:

Chapter	Exercise
2	12
3	4
4	6
5	4
8	1
9	8, 10
10	4, 6
11	9

STATISTICS 1040
Test 1, Spring 2012

Name _____
Recitation Instructor _____ Time _____

1. A recent issue of the New York Times included the following article:

Two widely used nutritional supplements for arthritis pain do not effectively relieve patients' aching arthritic knees, a large federal study has found. The study of two supplements, glucosamine and chondroitin sulfate, was published in The New England Journal of Medicine. The study was meant to provide a definitive answer as to whether these supplement work. In the 24-week study, 1583 patients with osteoarthritis of the knee were randomly assigned to one of five groups. Some patients took chondroitin and some took both. Others, serving as comparison subjects, took a placebo.

a) Is this an observational study? Briefly explain. (3 points)

No, the investigators decided who would be in the treatment group (randomly).

b) Why do they randomly assign the patients to the groups? (4 points)

Randomization makes the groups as much alike as possible and reduces bias.

c) What is a placebo and why is it used? (4 points)

A placebo is neutral but resembles the treatment. The response should be to the treatment itself and not the idea of treatment.

2. In 2006, a researcher claimed that elderly people who regularly indulged in "recreational gambling" were healthier, on average, than people who did not gamble at all. Provide a confounding factor that sheds doubt on the claim. (10 points)

Wealth could be a confounding factor. Wealthy people tend to be healthier and can also afford to gamble, to take trips to Vegas and Atlantic City.

3. Jeff took measurements of his diastolic blood pressure each day for a week. The values obtained were: 102, 95, 108, 103, 92, 97, and 117.

- a) Compute the average and median of these values. (3 points)

$$AV = 102, \text{ Median} = 102$$

Compute the SD of the 7 measurements. (3 points)

$$SD = 7.89$$

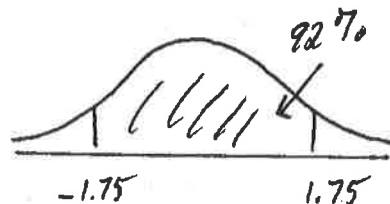
- b) Kristin also measured her diastolic blood pressure once a day for a week. Her measurements averaged 64 with an Sd of 8.2. If we combined our two sets of measurements, what would the SD of the combined group be equal to? (3 points)

- i) about 8.2 ii) less than 8.2 **iii) more than 8.2**

4. Many universities (including USU) require students to take the *Graduate Records Examination (GRE)* before being admitted to graduate school. The *Verbal GRE* is measured on a scale that runs from 200 to 800 points. Over the last 4 years the average Verbal GRE score was 460, the standard deviation was 120, and the histogram of Verbal GRE scores looked like the normal curve.

- a. Approximately what percentage of students got 670 or higher on the Verbal GRE? (12 points)

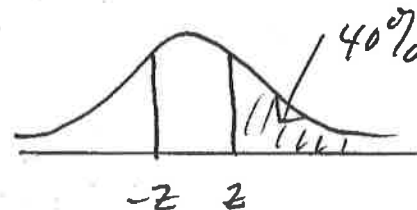
$$\frac{670 - 460}{120} = 1.75 \quad A(1.75) = 92\%$$



$$\frac{100 - 92}{2} = 4 \quad \boxed{4\%}$$

- b. The Graduate School at a large state university requires students to score above the 60th percentile on the Verbal GRE. What is the score that represents the 60th percentile? (10 points)

$$A(z) = 20\% \quad \text{so} \quad z = .25 \quad \frac{\text{score} - 460}{120} = .25$$

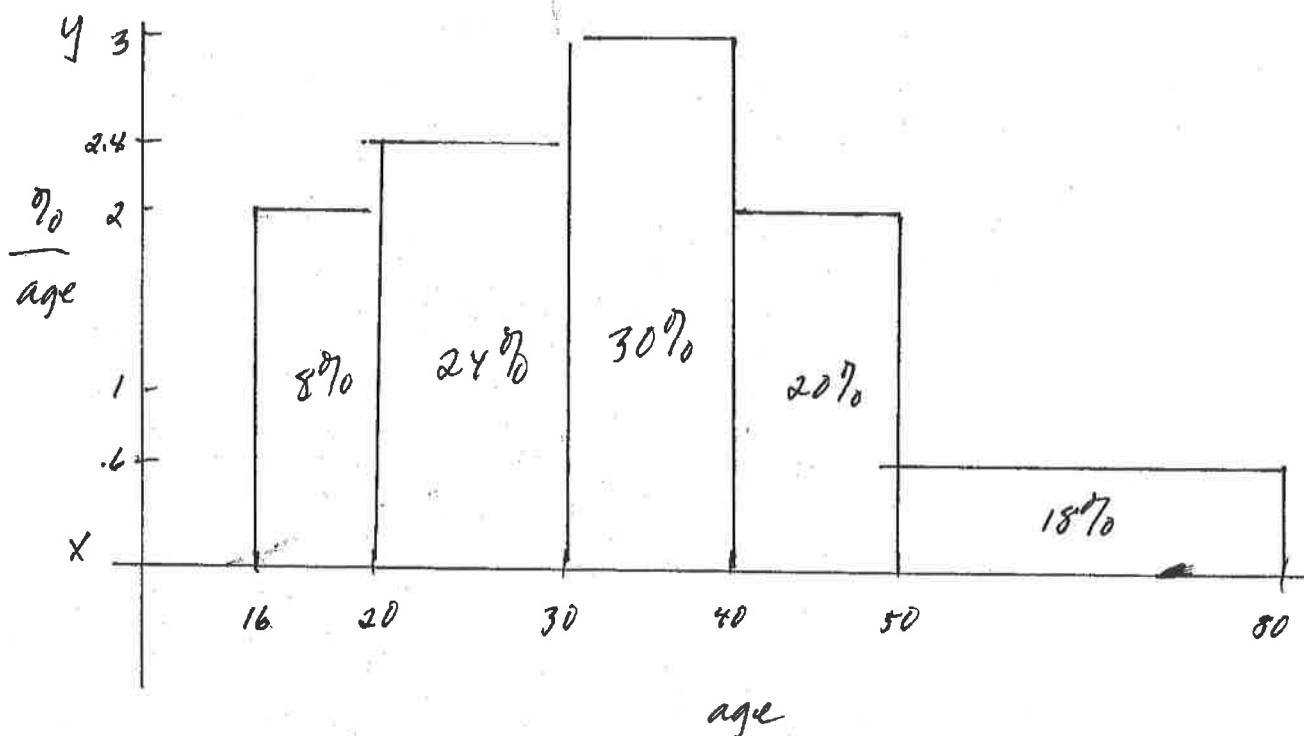


$$\text{score} = (120)(.25) + 460 = \boxed{490}$$

5. A study in a neighboring state showed that 4,750 drivers were stopped by police for a variety of reasons. The following table gives the ages of these drivers. Age intervals include the left endpoint but not the right.

Age	Number	%	Width	Height
16 - 20	380	8 %	4	2
20 - 30	1140	24%	10	2.4
30 - 40	1425	30%	10	3
40 - 50	950	20%	10	2
50 - 80	855	18%	30	.6

a) Draw the histogram. (15 points)



b) Estimate the percentage of drivers who were 35 years or older and stopped by the police. (3 points)

$$15\% + 20\% + 18\% = \boxed{53\%}$$

6. More than 5000 high-school students in California recently completed the SAT exam. The verbal SAT scores follow the normal curve with an average of 600 and a standard deviation of 100. The Math SAT scores for these students also had an average of 600 and an SD of 100. The correlation coefficient was .7 and the scatter diagram was football-shaped.

- a) Find the equation of the regression line for predicting the math SAT from the verbal SAT. (10 points)

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = (.7) \frac{100}{100} .7, \text{ point} = (600, 600)$$

$$y - 600 = .7(x - 600) = .7x - 420, \quad y = .7x + 180$$

- b) Use the regression method to predict the math SAT score for a student who scored 750 on the verbal SAT. (10 points)

Use the equation above or the 6-step method.

$$1. \text{ Verbal score} \quad 2. 750 \quad 3. \frac{750 - 600}{100} = 1.5$$

$$4. 9.7(1.5) \quad 5. (100)(.7)(1.5) = 105 \quad 6. 600 + 105 = \boxed{705}$$

- c) The r.m.s. error for predicting the math SAT from the verbal SAT is equal to $\sqrt{1-r^2} \times SD_y = \sqrt{1-(.7)^2} \times 100 \approx 71$. What does the r.m.s. error measure? (5 points)

It measures the overall size of the error in regression.

- d) If a student scored at the 80th percentile on the verbal SAT, do you expect them to score at about the 80th percentile on the math SAT? Explain. (5 points)

No, you would expect the student to score less than the 80th percentile because of the regression effect.

3. Jeff took measurements of his diastolic blood pressure each day for a week. The values obtained were: 102, 95, 108, 103, 92, 97, and 117.

a) Compute the average and median of these values. (3 points)

b) Compute the SD of the 7 measurements. (3 points)

c) Kristin also measured her diastolic blood pressure once a day for a week. Her measurements averaged 64 with an Sd of 8.2. If we combined our two sets of measurements, what would the SD of the combined group be equal to? (3 points)

i) about 8.2

ii) less than 8.2

iii) more than 8.2

4. Many universities (including USU) require students to take the *Graduate Records Examination (GRE)* before being admitted to graduate school. The *Verbal GRE* is measured on a scale that runs from 200 to 800 points. Over the last 4 years the average Verbal GRE score was 460, the standard deviation was 120, and the histogram of Verbal GRE scores looked like the normal curve.

a. Approximately what percentage of students got 670 or higher on the Verbal GRE? (12 points)

b. The Graduate School at a large state university requires students to score above the 60th percentile on the Verbal GRE. What is the score that represents the 60th percentile? (10 points)

5. A study in a neighboring state showed that 4,750 drivers were stopped by police for a variety of reasons. The following table gives the ages of these drivers. Age intervals include the left endpoint but not the right.

Age	Number	%		
16 - 20	380	8 %		
20 - 30	1140	24%		
30 - 40	1425	30%		
40 - 50	950	20%		
50 - 80	855	18%		

a) Draw the histogram. (15 points)

b) Estimate the percentage of drivers who were 35 years or older and stopped by the police. (3 points)

6. More than 5000 high-school students in California recently completed the SAT exam. The verbal SAT scores follow the normal curve with an average of 600 and a standard deviation of 100. The Math SAT scores for these students also had an average of 600 and an SD of 100. The correlation coefficient was .7 and the scatter diagram was football-shaped.

- a) Find the equation of the regression line for predicting the math SAT from the verbal SAT. (10 points)
- b) Use the regression method to predict the math SAT score for a student who scored 750 on the verbal SAT. (10 points)
- c) The r.m.s. error for predicting the math SAT from the verbal SAT is equal to $\sqrt{1-r^2} \times SD_y = \sqrt{1-(.7)^2} \times 100 \approx 71$. What does the r.m.s. error measure? (5 points)
- d) If a student scored at the 80th percentile on the verbal SAT, do you expect them to score at about the 80th percentile on the math SAT? Explain. (5 points)