

Unit 1 – Review of BIOSTATS 540  
Practice Problems

Due Thursday September 12, 2024

Last Date for Late Submission with –20 points: Thursday September 12, 2024

**Reminder – The homeworks are a “participation” component of your grade, not an assessment. You will get a score of 100 for attempting each question, regardless of whether or not your answer is correct.**

Dear BIOSTATS 640 Fall 2024,

I hope you will really work with these practice questions as a way to learn the material. Give each question a try! Then, if you're stuck, by all means, look at the solutions! In fact, look at the solution whenever you get stuck! Please do not merely copy my solution and turn these in.

**Tip.** As you go along, please feel welcome to write your notes and questions in the margins. We will be on the look-out for these and will do our best to clear up any confusion.

**Have you read the syllabus?**

*If you're like me, often, you do not read the manuals that come with various purchases (e.g., microwave, lawn mower, pretty new blender, etc.). So, possibly, you don't read course syllabi either. But please. Read the syllabus! It is actually worth doing. So, to ensure that you do just that, this homework includes some questions about the syllabus. The answers are all in the syllabus.*

- \_\_\_ a) How are homeworks graded? \_\_\_\_\_
- \_\_\_ b) What is the policy on late homework submissions? \_\_\_\_\_
- \_\_\_ c) What are the posting dates of the exams? \_\_\_\_\_
- \_\_\_ d) What are the due dates of the exams? \_\_\_\_\_
- \_\_\_ e) Are the exams timed? \_\_\_\_\_
- \_\_\_ f) Will I need to use on R on any exam? \_\_\_\_\_
- \_\_\_ g) What is the policy on late exam submissions? \_\_\_\_\_
- \_\_\_ h) Is it possible for exams to be posted ahead of schedule? \_\_\_\_\_
- \_\_\_ i) Can I get postings early? \_\_\_\_\_
- \_\_\_ j) What is the last date of class? \_\_\_\_\_
- \_\_\_ k) When is the final due? \_\_\_\_\_

**For your reference**

**BIOSTATS 640, Intermediate Biostatistics**

<https://people.umass.edu/~biep640w/>

**BIOSTATS 540, Introductory Biostatistics**

<https://people.umass.edu/~biep540w/>

Dear BIOSTATS 640 Fall 2024,

This assignment has a lot of questions. Don't panic! Future homeworks will not have so many questions, I promise. The reason there are so many questions in this assignment is because the goal is to refresh your memory of an entire course (possibly BIOSTATS 540 or whatever your introductory course might have been).

**Tip – I strongly encourage you to make use of the website for BIOSTATS 540.**

Each question includes a reference to the BIOSTATS 540 unit, here: <https://people.umass.edu/~biep540w/>

**#1. (Reviews BIOSTATS 540 Unit: 1 - Numerical Summarization).**

The following table lists length of stay in hospital (days) for a sample of 25 patients.

5	10	6	11	5	14	30	11	17	3
9	3	8	8	5	5	7	4	3	7
9	11	11	9	4					

By any means you like: a) create a data set; and b) construct a frequency/relative frequency table for these data using 5-day class intervals. Include columns for the frequency counts and relative frequencies.

**#2. (Reviews BIOSTATS 540 Unit: 2 - Data Visualization).**

Depending, download from the course website:

R Users: [BIOSTATS640\\_hw01\\_q2.Rdata](#)

If you are using artofstat.com or anything else: [BIOSTATS640\\_hw01\\_q2.xlsx](#)

These data are fasting cholesterol levels (mg/dl) for two groups of men.

<u><b>Group 1:</b></u>									
233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325
<u><b>Group 2:</b></u>									
344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

By any means you like, construct the following graphical comparisons of the two groups:

2a. Side-by-side box plot

2b. Side-by-side histograms with same definitions (starting value, ending value, tick marks, etc) of the horizontal and vertical axes.

In 1-2 sentences, compare the two distributions. What conclusions do you draw?

**#3. (Reviews BIOSTATS 540 Unit 6 – Bernoulli and Binomial Distributions).**

Consider the following setting.

Seventy-nine firefighters were exposed to burning polyvinyl chloride (PVC) in a warehouse fire in Plainfield, New Jersey on March 20, 1985. A study was conducted in an attempt to determine whether or not there were short- and long-term respiratory effects of the PVC. At the long-term follow-up visit at 22 months after the exposure, 64 firefighters who had been exposed during the fire and 22 firefighters who were not exposed reported on the presence of various respiratory conditions. Eleven of the PVC exposed firefighters had moderate to severe shortness of breath compared to only 1 of the non-exposed firefighters.

Calculate the probability of finding 11 or more of the 64 exposed firefighters reporting moderate to severe shortness of breath if the rate of moderate to severe shortness of breath is 1 case per 22 persons. **Show your work.**

**#4. (Reviews BIOSTATS 540 Unit 7 – Normal Distribution).**

The Air Force uses ACES-II ejection seats that are designed for men who weigh between 140 lb and 211 lb. Suppose it is known that women's weights are distributed Normal with mean 143 lb and standard deviation 29 lb.

- 4a. What proportion of women have weights that are *outside* the ACES-II ejection seat acceptable range?
- 4b. In a sample of 1000 women, how many are expected to have weights below the 140 lb threshold?

**#5. (Reviews BIOSTATS 540 Unit 8 - Statistical Literacy and Unit 9 - One Sample Inference).**

Consider the setting of a single sample of  $n=16$  data values that are a random sample from a normal distribution. Suppose it is of interest to perform a type I error  $\alpha = 0.01$  statistical hypothesis test of  $H_0: \mu \geq 100$  versus  $H_A: \mu < 100$ ,  $\alpha = 0.01$ . Suppose further that  $\sigma$  is unknown.

- 5a. State the appropriate test statistic
- 5b. Determine the critical region for values of the sample mean  $\bar{X}$ .

#6. (Reviews BIOSTATS 540 Unit 9 - One Sample Inference).

An investigator is interested in the mean cholesterol level  $\mu$  of patients with myocardial infarction. S/he drew a simple random sample of  $n=50$  patients and from these data constructed a 95% confidence interval for the mean  $\mu$ . In these calculations, it was assumed that the data are a simple random sample from a normal distribution with known variance. The resulting width of the confidence interval was 10 mg/dl.

How large a sample size would have been required if the investigator wished to obtain a confidence interval width equal to 5 mg/dl?

#7. (Reviews BIOSTATS 540 Unit 9 - One Sample Inference and Unit 10 - Two Sample Inference).

In (a) – (e) below, you may assume that the data are a simple random sample (or samples) from a normal distribution (or distributions). Each setting is a different setting of confidence interval estimation. In each, state the values of the confidence coefficients (*recall – these will be the values of specific percentiles from the appropriate probability distribution*).

7a.

For a single sample size of  $n=15$  and the estimation of the population mean  $\mu$  when the variance is unknown using a 90% confidence interval, what are the values of the confidence coefficients?

7b.

For a single sample size  $n=35$  and the estimation of a variance parameter  $\sigma^2$  using a 95% confidence interval, what are the values of the confidence coefficients?

7c.

For a single sample size of  $n=25$  and the estimation of the population mean  $\mu$  when the variance is known using a 80% confidence interval, what are the values of the confidence coefficients?

7d.

For the setting of two independent samples, one with sample size  $n_1 = 13$  and the other with sample size  $n_2 = 22$ , it is of interest to construct a 90% confidence interval estimate of the ratio of the two population variances,

$[\sigma_1^2/\sigma_2^2]$ . What are the values of the confidence coefficients?

### #8. (Reviews BIOSTATS 540 Unit 10 - Two Sample Inference).

A study was investigated of length of hospital stay associated with seat belt use among children hospitalized following motor vehicle crashes. The following are the observed sample mean and sample standard deviations for two groups of children: 290 children who were *not* wearing a seat belt at the time of the accident plus 123 children who *were* wearing a seat belt at the time of the accident.

Group	Sample size, n	Sample mean	Sample standard deviation
Seat belt = no	$n_{\text{NO}} = 290$	$\bar{X}_{\text{NO}} = 1.39 \text{ days}$	$S_{\text{NO}} = 3.06 \text{ days}$
Seat belt = yes	$n_{\text{YES}} = 123$	$\bar{X}_{\text{YES}} = 0.83 \text{ days}$	$S_{\text{YES}} = 2.77 \text{ days}$

You may assume normality. You may also assume that the unknown variances are equal. Construct a 95% confidence interval estimate of the difference between the two population means. In developing your answer, you may assume that the population variances are unknown but EQUAL.

8a.

What is the value of the point estimate?

8b.

What is the value of the estimated standard error of the point estimate?

8c.

What is the value of the confidence coefficient?

8d.

What are values of the lower and upper limits of the confidence interval?

8e.

Write a clear interpretation of the confidence interval.

### #9. (Reviews BIOSTATS 540 Unit 6 - Binomial Distribution).

A test consists of multiple-choice questions, each having four possible answers, one of which is correct. What is the probability of getting exactly four correct answers when six guesses are made?

### #10. (Reviews BIOSTATS 540 Unit 6 - Binomial Distribution).

After being rejected for employment, woman “A” learns that company “X” has hired only 2 women among the last 20 new employees. She also learns that the pool of applicants is very large, with an approximately equal number of qualified men and women. Help her address the charge of gender discrimination by finding the probability of getting 2 or fewer women when 20 people hired under the assumption that there is no discrimination based on gender. Does the resulting probability really support such a charge?

**#11. (Reviews BIOSTATS 540 Unit 7 - Normal Distribution).**

Suppose the length of newborn infants is distributed normal with mean 52.5 cm and standard deviation 4.5 cm. What is the probability that the mean of a sample of size 15 is greater than 56 cm?

**#12. (Reviews BIOSTATS 540 Unit 7 - Normal Distribution).**

Suppose that 25-year old males have a remaining life expectancy of an additional 55 years with a standard deviation of 6 years. Suppose further that this distribution of additional years life is normal. What proportion of 25 year-old males will live past 65 years of age?