BIOSTATS 640 Intermediate Biostatistics
Fall 2024
Examination I

Unit 2 – Discrete Distributions
Unit 3 - Introduction to Nonparametrics

Due: Monday October 7, 2024 (midnight)

48-hour late submission for credit with (-10 points) due: Wednesday October 9, 2024 (midnight)
last date for late submission for credit (-20 points): Monday October 14, 2024 (midnight)

Before you begin:

This is a "take-home" exam. You are welcome to: 1) use any reference materials you wish; 2) use the computer as you wish (including online apps); and 3) contact me with questions.

However, you MUST work this exam **by yourself.** You may **not** use artificial intelligence (AI)/Chat GPT, and you may **not** consult with anyone except me (Carol Bigelow).

Please:

__**1. Name your file as instructed below; and**
__**2. Complete the signature page**

How to name your exam submission

__a) Please be sure your name is somewhere on your submission.
__b) Next, save it as a SINGLE FILE pdf (please do not submit a word file)
__c) Suggestion: Use the following naming convention: lastname_exam1.pdf.

How to submit your exam in Canvas.

At left, click **ASSIGNMENTS** > Upcoming Assignments > *scroll down* > UPLOAD > Drag folder here or choose a file to upload > Submit Assignment

Questions?

Again, questions are always welcome; email me at **cbigelow@schoolph.umass.edu**

BIOSTATS 640 Intermediate Biostatistics
Fall 2024
Examination I

Units 1, 2 & 3 – Review of Introductory Biostatistics, Discrete Distributions & Introduction to Nonparametrics

Due: Monday October 7, 2024 (midnight)
48-hour late submission for credit with (-10 points) due: Wednesday October 9, 2024 (midnight)
last date for late submission for credit (-20 points): Monday October 14, 2024 (midnight)

Kindly, please provide 2 signatures.

Signature 1.

This is to confirm that **I did not utilize artificial intelligence (AI, Chat GPT, or any other artificial intelligence utilities)** in completing this exam.

Name: _____

Date: _____

Signature 2.

This is to confirm that in completing this exam, **I worked independently and did not consult with anyone (except Carol Bigelow)**.

Name: _____

Date: _____

Thank you!

Before You Begin

This test has 10 questions, with points that total = 115. This gives you "make up credit" of 15 points as you work these questions. I will grade your test up to a maximum earned score of 100 points.

Thus, the maximum earned score possible on this test is 100.

1. (10 points total)

This question is about numeric literacy and the observation that, sometimes, were are poor at judging probabilities!

> *"A certain town is served by two hospitals. In the larger hospital, about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are male sex at birth. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of one year, each hospital recorded the days on which more than 60% of the babies born were male sex at birth."*

Which hospital do you think recorded more such days? The larger hospital, the smaller hospital, or were they about the same (that is within 5% of each other)?

In developing your answer to this question, you are asked to work with two binomial probability distributions, once for the larger hospital and once for the smaller hospital. ***Hint: What is the expected value of each of the two binomial distributions and how do they compare?***

2. (10 points total)

Sodium polystyrene (brand name Kayexelate) in sorbitol is a drug used to reduce serum potassium levels in hyperkalemic patients. Hyperkalemia is a condition in which there is too much potassium in the blood. Unfortunately, it is suspected of causing an adverse reaction leading to colonic necrosis.

Suppose a study compared the incidence of colonic necrosis in 117 Kayexelate-exposed and 862 non-exposed post-surgical patients. Suppose further that two cases of colonic necrosis occurred in the Kayexelate-exposed group and none occurred in the non-exposed group.

Carry out the appropriate statistical hypothesis test to determine if there is statistically significant evidence of increased risk of colonic necrosis associated with Kayexelate exposure. In your answer, please answer the following.

2a. (2 points)

Using the information provided, what are the following 2x2 table of counts? Please complete the table below, including row and column totals.

	Necrosis	Not	
Kayexelate exposed	??	??	??
Not-exposed	??	??	??
	??	??	??

2b. (2 points)

State the null and alternative hypotheses, taking care to define all terms. If the alternative hypothesis is one sided, state the direction of the alternative hypothesis.

2c. (2 points)

What is the ***name*** of the appropriate statistical test? Note – I am ***not*** asking for its value here.

2d. (2 points)

What is the value of the significance level, the ***p-value***?

2e. (2 points)

In ***1-2 sentences***, what is your opinion regarding the risk of colonic necrosis in this population?

3. (10 points total)

Background for questions #3a and #3b.
Suppose it is known that 25% of 11 year-old children have no decayed, missing, or filled (DMF) teeth.

3a. (2 points)
What is the probability that in a random sample of 20 (independent) 11-year old children, there will be exactly 3 with no DMF teeth?

3b. (2 points)
(same set up as in question #3a)
What is the probability that in a random sample of 20 (independent) 11-year old children, there will be fewer than 3 with no DMF teeth?

Background for questions #3c and #3d.
Next, suppose that among all persons 17 years of age and over, half the males and one-third of the females are current smokers.

3c. (3 points)
What is the probability that a random sample of 10 males and 15 females includes exactly 4 male and 6 female current smokers? ***Hint: In developing your answer you will want to make use of the fact that if 2 outcomes are independent, then the probability that they both occur is the product of the separate probabilities (e.g., the probability that 2 independent tosses of a fair coin yields 2 heads is (.50)x(.50) = .25***

3d. (3 points)
(same set up as in question #3c)
In a random sample of 10 males and 15 females, what is the probability that there are no current smokers?

4. (10 points total)

Suppose it is known that a certain genetic mutation occurs in an insect population, on average, in 20 out of 10,000 insects.

Next, suppose that 8,000 insects are sampled and a count is obtained of the number of insects that have the genetic mutation. Let **X** be the appropriately defined Binomial distribution for this setting and let **Y** be the appropriately defined Poisson distribution for this setting. Thus, the count of insects with the genetic mutation is either ***X distributed Binomial*** or it is ***Y distributed Poisson.***

4a. (2 points)

What are the values of the population distribution mean and standard deviation of the random variable X?

4b. (2 points)

What are the values of the population distribution mean and standard deviation of the random variable Y?

4c. (6 points)

(same set up as in questions #4a and 4b)
Using the appropriately defined Binomial and Poisson distributions, complete the following table of probabilities:

X distributed Binomial	**Y distributed Poisson**
Pr [X = 0] = ??	Pr [Y = 0] = ??
Pr [X $\geq$ 1] = ??	Pr [Y $\geq$ 1] = ??
Pr [X < 5] = ??	Pr [Y < 5] = ??

5. (10 points total)

An insurance company checks police records on 582 accidents selected at random and notes that teenagers were at the wheel in 91 of them.

Give a 95% confidence interval estimate of the percentage of all auto accidents that involve teenage drivers at the wheel.

6. (25 points total)

In 1889, a study was conducted by Geissler. Investigated were 53,680 families, all with 8 children. Of interest was an investigation of the probability of male sex at birth. Is it .50? Does the probability of male sex at birth change with birth order?

Geissler's observations are presented in the table below. As an example of its interpretation, the first row of this table is telling you that there were 215 families of 8 children in which the number of male sex at birth children was zero.

# of male sex at births among 8 children	# Families	% of Families
0	215	0.4
1	1485	2.77
2	5331	9.93
3	10649	19.84
4	14959	27.87
5	11929	22.22
6	6678	12.44
7	2092	3.90
8	342	0.64
Total	**53680**	**100%**

6a. (5 points)

Estimate π, the probability that a randomly selected birth is male sex at birth. Assume that all births have the same probability of being male sex at birth.

6b. (5 points)

Using your answer for π in problem #6a, compute the binomial probabilities for number of male sex at birth children in a family of eight equal to X = 0, 1, 2, 3, 4, 5, 6, 7, and 8. In reporting your answers, complete the following table.

# Male sex at birth Outcomes X = x	Prob [X = x]
0	??
1	??
2	??
3	??
4	??
5	??
6	??
7	??
8	??
Total	??

6c.(5 points)

Using your answers to problem #6b, obtain the expected **frequencies** of 0, 1, 2, 3, 4, 5, 6, 7, and 8 male sex at births. In reporting your answer, complete the table below.

# of male sex at births among 8 children	Observed # Families	Expected # of Families
0	215	??
1	1485	??
2	5331	??
3	10649	??
4	14959	??
5	11929	??
6	6678	??
7	2092	??
8	342	??
Total	**53680**	??

6d. (5 points)

Next, using your answers to problem #6c, obtain the expected **relative frequencies** of 0, 1, 2, 3, 4, 5, 6, 7, and 8 male sex at births. In reporting your answer, complete the following table.

# of male sex at births among 8 children	Observed proportion % Families	Expected proportion % of Families
0	0.4	??
1	2.77	??
2	9.93	??
3	19.84	??
4	27.87	??
5	22.22	??
6	12.44	??
7	3.90	??
8	0.64	??
Total	**100%**	??

6e. (5 points)

In 1-3 sentences, what is your interpretation of Geissler's data with respect to the question of whether or not the probability of male sex at birth is 0.50 and whether or not it is the same, regardless of birth order?

In reporting your answer, consider the a reader who is intelligent but who is not familiar with statistical jargon.

7. (10 points total)

Twelve ants and eighteen flies were placed in a container with insecticide and observed. After sixteen insects had died, there were nine ants alive and five flies alive.

Peform a Fisher's exact test to test the null hypothesis that ants and flies are equally susceptible to the insecticide. In developing your answer, please answer the following questions.

7a. (2 points)

Using the information provided, complete the corrsponding 2x2 table of counts, including row and column totals.

	Died	Alive	
Ants	??	??	??
Flies	??	??	??
	??	??	??

7b. (2 points)

State the null and alternative hypotheses, taking care to define all terms. If the alternative hypothesis is one sided, state the direction of the alternative hypothesis.

7c. (2 points)

What is the ***name*** of the appropriate statistical test? Note – I am ***not*** asking for its value here.

7d. (2 points)

What is the value of the achieved level of significance (p-value)?

7e. (2 points)

In 1-2 sentences, provide an interpretation of your findings in terms that a layperson can understand.

8. (10 points total)

An advertising company is willing to renew its advertising contract with a local radio station only if the station can (convincingly) show that ***more than 20%*** of the residents of the city have heard the ad and recognize the company's product.

The radio station conducts a random phone survey of 600 people. Of these, only 133 remember the ad and recognize the company's product.

Do these data provide statistically significant evidence to warrant the advertising comany to renew its contract with the local radio station? Perform the appropriate statistical hypothesis test by answering the following.

8a. (2 points)
State the null and alternative hypotheses, taking care to define all terms. If the alternative hypothesis is one sided, state the direction of the alternative hypothesis.

8b. (2 points)
What is the ***name*** of the appropriate statistical test? Note – I am ***not*** asking for its value here.

8c. (2 points)
What is ***value*** of your test statistic for these data?

8d. (2 points)
What is the value of the achieved level of significance (p-value)?

8e. (2 points)
In 1-2 sentences, provide an n interpretation of your findings in terms that a layperson can understand.

9. (10 points)

You have been hired as the quality control office of a pharmaceutical company that manufactures aspirin tablets. According to the quality assurance protocol, your job is to do the following. For each shipment of possibly thousands and thousands of aspirin tablets, randomly select and test 24 tablets, then accept the batch if there is only one or no tablet that doesn't meet the test specification.

If a particular shipment of aspirin tablets actually has a 4% rate of defects, what is the probability that this shipment will be accepted?

10. (10 points total)

Consider a randomized trial of treatment "Active" versus treatment "Standard" for a given disease. At one clinic, eight out of nine patients were randomized to receive treatment "Active".

Based on these assignments, a complaint is made that randomization must not be working, that π = Pr[treatment assignment is "ACTIVE"] cannot be equal to 0.50.

10a. (3 points)

Under the assumption that randomization *is working*, calculate the probability of obtaining 8 or more assignments to treatment "Active" out of 9 assignments ***at one clinic***.

10b. (5 points)

Next, consider that this is a multi-site randomized trial with **15 sites, each with 9 assignments**.

Under the assumption that randomization is working at all 15 sites, what is the probability that there will be more than one site with 8 or more assignments to "Active"?