MATH 136 - Statistics **Spring 2020**

Exam 2

Ch 4, 5, 6 **Dr. Jorge Basilio**

April_20 **jbasilio@glendale.edu**

Solutions

Directions

1. **At the top of your first page:** Please hand-write the statement provided below in quotes; print your name; put the date; and sign your name below it that acknowledges the **honor code**:

   "On my honor, by printing and signing my name, I vow to neither receive nor give any unauthorized assistance on this examination. I understand what my professor has deemed appropriate and inappropriate for this test and vow to follow these rules."

2. The exam is written to last 65 minutes from 10:45 am to 11:50 am, however, you have until 12:10 pm to submit this exam without penalty.

3. How to submit: upload a **single PDF file** of your solutions to Canvas no later than 12:10 pm to avoid penalties.

4. You will need to be logged in to class via Zoom. You will also need to have your camera and microphone both functioning and turned ON for the duration of the test. Ask family members, roommates, to not disrupt you during the test. Please disable any virtual backgrounds.

5. Write your solutions to the exam on one side of the page (front side only, do NOT write double-sided). You do NOT need to copy the questions on your piece of paper. However, you must submit the test problems in the order given and you must clearly label each problem and part. If I cannot identify which problem you are working on, no points will be given.

6. Make sure to ALWAYS SHOW YOUR WORK; you will not receive any partial credits unless work is clearly shown unless told otherwise. *If in doubt, ask for clarification.* Correct answers with little to no work will receive no points. Students might be randomly selected to have a 1-1 conference where you are asked to defend your work and explain to me all your steps on certain questions or problems that are similar to a test question.

7. **Penalties for late submissions:** Exams received between 12:11 and 12:20 pm will have 10 points deducted from their score. Exams received between 12:21 and 12:35 pm, I will deduct an additional point from your score for each additional minute the exam is late. Exams received after 12:35 pm will not be graded and be given a score of 0.

8. **Allowed Materials**
   - You may use your calculator during the test (TI-83, 84, 84+, or 84+CET)
   - Blank pieces of paper to write your solutions. Writing utensils, erasers, etc

9. **Materials NOT Allowed**
   - Do not use your cell phone (for any reason: do not send or receive texts or calls or use the internet, etc)
   - Do not use your textbook (either digital or physical)
   - Do not use digital or printed out notes: the slides, the study guides, etc
   - Do not consult your HW
   - Do not give or receive any outside help (no getting help from a family member, friend, or any person either in person, via chat, message board, text message or any form of communication–again– you will be on camera the entire time so I will be looking for suspicious behavior)
   - Do not use your computer to look up anything using the internet (don't google; don't consult homework help websites, etc)

Continued on the next page...

Directions Continued:

10. The exam totals **75 points**.
11. There are 7 problems; many of them with multiple parts.
12. Handwriting should be neat and legible. If I cannot read your writing, zero points will be given.
13. Some questions contain multiple-parts which you must do individually and the parts are denoted by (a), (b), (c), etc. Some questions are multiple-choice and the choices are denoted with (A), (B), (C), (D), and (E). For True/False questions, you must spell out the entire word "true" or "false" in your answer.
14. Leave answers in exact form (as simplified as possible), unless told otherwise.
15. Put a $\boxed{\text{box around your final answer}}$ where applicable.
16. **PLEASE INCLUDE UNITS** where applicable
17. **PLEASE CHECK YOUR WORK!!!**
18. GOOD LUCK!!!!

Score	Grade

Problem 1: 8 pts (2 pts each)

The American black bear (Ursus americanus) is one of eight bear species in the world. It is the smallest North American bear and the most common bear species on the planet. In 1969, Dr. Michael R. Pelton of the University of Tennessee initiated a long-term study of the population in the Great Smoky Mountains National Park. One aspect of the study was to develop a model that could be used to predict a bear's weight (since it is not practical to weigh bears in the field). One variable thought to be related to weight is the length of the bear. The following data represent the lengths and weights of 12 American black bears.

Length (cm)	139.0	138.0	139.0	120.5	149.0	141.0	141.0	150.0	166.0	151.5	129.5	150.0
Weight (kg)	110	100	90	60	113	100	95	125	155	140	105	110

(a) Which variable is the explanatory variable based on the goals of the research?

(b) Determine the regression line (line of best fit) show the calculator function used.

(c) Does a linear relation exist between the weight of the bear and its length? *Use the correlation coefficient r and the table in the formula sheet to justify your answer.*

(d) Predict the weight of a black bear with a length of 156 cm.

See Next Page

Problem 2: 6 pts (1 pts each)

TRUE or **FALSE** (please spell out/write the entire word for credit).
(Hint: you might find it helpful to use Venn Diagrams to help you arrive at an answer)

(a) FALSE The equation of a linear regression equation always gives the best predicted value of the response variable (y) for a given explanatory variable (x).

(b) TRUE If A and B are mutually exclusive then $P(A \text{ and } B) = 0$.

→ no corr (chaos!)

(c) FALSE If the correlation coefficient is $r = 0.0001$ then there's a strong positive linear correlation.

(d) FALSE If $P(A) > P(B)$ then event B is less ~~more~~ likely than event A.

(e) TRUE If our population is 40,000 students, then a sample of 35 can be treated as being independent even if we sample without replacement. $n = 35$ $N = 40{,}000$ is $n \le 0.05N$? "5% Rule" ✓

(f) TRUE If $P(A) = 0.00081$ then A is an unusual event. is less than 0.05

+6

Problem 3: 16 pts

Answer each question for the following table. For probability, state your answer as **un-simplified** fractions AND as decimals (following our rounding rule).

	Low Obesity	Average Obesity	High Obesity	Total
Hypertension	24	33	46	103
No Hypertension	109	101	87	297
Total	133	134	133	400

Let $\mathbf{T}$ = Hypertension, $\mathbf{N}$ = No Hypertension, $\mathbf{L}$ = Low Obesity, $\mathbf{A}$ = Average Obesity, $\mathbf{H}$ = High Obesity

(2 pt) (a) Find the probability that a single randomly selected person has hypertension, $P(T)$:

(2 pt) (b) Find the probability that a single randomly selected person has average obesity and no hypertension, $P(A \text{ and } N)$:

(2 pt) (c) Find the probability that a single randomly selected person has high obesity or no hypertension, $P(H \text{ or } N)$:

Problem 1: 8 pts (2 pts each)

The American black bear (Ursus americanus) is one of eight bear species in the world. It is the smallest North American bear and the most common bear species on the planet. In 1969, Dr. Michael R. Pelton of the University of Tennessee initiated a long-term study of the population in the Great Smoky Mountains National Park. One aspect of the study was to develop a model that could be used to predict a bear's weight (since it is not practical to weigh bears in the field). One variable thought to be related to weight is the length of the bear. The following data represent the lengths and weights of 12 American black bears.

L1	Length (cm)	139.0	138.0	139.0	120.5	149.0	141.0	141.0	150.0	166.0	151.5	129.5	150.0	x
L2	Weight (kg)	110	100	90	60	113	100	95	125	155	140	105	110	y

(a) Which variable is the explanatory variable based on the goals of the research?

(b) Determine the regression line (line of best fit) show the calculator function used.

(c) Does a linear relation exist between the weight of the bear and its length? *Use the correlation coefficient r and the table in the formula sheet to justify your answer.*

(d) Predict the weight of a black bear with a length of 156 cm.

Solutions +2

(a) Length is explanatory variable (& Weight is response variable)

rounding rule: 4 places

(b) Enter Length into L1 & Weight into L2

STAT → CALC → 8. LINReg(a+bx): XList: L1, YList: L2 ⇒ $y = a + bx$

$a = -158.5702$
$b = 1.8698$

Regression Line is +2 $\boxed{\hat{y} = -158.5702 + 1.8698x}$

- x = length
- y = weight

(c) Correlation coefficient: $r = 0.896$

STAT → CALC → 8. LINReg

Remember must have Diagnostics Turned ON

- critical value is $CV = 0.576$

Since the correlation coefficient r is greater than the critical value we conclude that there is a positive linear relationship between the weight of the bear and its height. +2

(d) Predict weight of bear with length $x = 156$ cm.

$y = -158.5702 + 1.8698(156) = 133.1186$ kg → 133.1 kg

use stats law rounding

We predict that a bear of length 156 cm weighs 133.1 kg +2

Problem 3: 16 pts

11/16

Answer each question for the following table. For probability, state your answer as **un-simplified** fractions AND as decimals (following our rounding rule).

	Low Obesity	Average Obesity	High Obesity	Total
Hypertension	24	33	46	103
No Hypertension	109	101	87	297
Total	133	134	133	400

Let **T** = Hypertension, **N** = No Hypertension, **L** = Low Obesity, **A** = Average Obesity, **H** = High Obesity

(2 pt) (a) Find the probability that a single randomly selected person has hypertension, $P(T)$:

(2 pt) (b) Find the probability that a single randomly selected person has average obesity and no hypertension, $P(A \text{ and } N)$:

(2 pt) (c) Find the probability that a single randomly selected person has high obesity or no hypertension, $P(H \text{ or } N)$:

(2 pt) (d) Find the probability that a person has hypertension given that the person is of Average Obesity.

(2 pt) (e) Are the events "Hypertension" and "Low Obesity" independent? Explain why or why not.

(2 pt) (f) Are the events "No Hypertension" and "Average Obesity" mutually exclusive? Explain why or why not.

(2 pt) (g) Two adults are randomly selected without replacement, what is the probability both are of Average Obesity?

(2 pt) (h) Three adults are randomly selected with replacement, what is the probability at least one has Hypertension?

(a) $P(T) = \boxed{\frac{103}{400} = 0.258}$ +2

(b) $P(A \text{ and } N) = \frac{\# A \text{ and } N}{\# S} = \boxed{\frac{101}{400} = 0.253}$ +2

(c) $P(H \text{ or } N) = P(H) + P(N) - P(H \text{ and } N) = \frac{133}{400} + \frac{297}{400} - \frac{87}{400} = \boxed{\frac{343}{400}} = 0.858$ +2

(d) $P(T|A) = \frac{P(T \text{ and } A)}{P(A)} = \frac{\# T \text{ and } A}{\# A} = \boxed{\frac{33}{134} = 0.246}$ +2

(e) No, they are **NOT independent** because they overlap. There are 24 people who are both T and L which can influence probabilities. +2

(f) No, they are **NOT mutually exclusive** because they overlap. There are 101 people who are both N and A. +2

(g) $P(\overset{1st}{A} \text{ and } \overset{2nd}{A}) = P(A) * P(A) = \boxed{\frac{134}{400} * \frac{133}{399} = 0.112}$ +2

(h) $P(\text{at least one } T) = 1 - P(\text{none } T) = 1 - P(N) = 1 - \left(\frac{297}{400}\right)^3 = \boxed{0.591}$ +2

3 selections with replacement

+16

(2 pt) (d) Find the probability that a person has hypertension given that the person is of Average Obesity.

(2 pt) (e) Are the events "Hypertension" and "Low Obesity" independent? Explain why or why not.

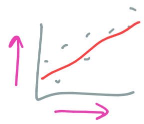
(2 pt) (f) Are the events "No Hypertension" and "Average Obesity" mutually exclusive? Explain why or why not.

(2 pt) (g) Two adults are randomly selected without replacement, what is the probability both are of Average Obesity?

(2 pt) (h) Three adults are randomly selected with replacement, what is the probability at least one has Hypertension?

Problem 4: 6 pts (1 pts each)

Fill in the blanks:



(a) Two variables are positively correlated means that an increase in the explanatory variable produces an increase +1 in the response variable.

(b) ______ $\le P(A) \le$ 1 +1

(c) $P(A) = 0$ means event A is impossible +1

(d) $P(A) = 1$ means event A is certain +1

Multiple-choice. Select the correct answer:

(d) The shaded area in the Venn Diagram represents

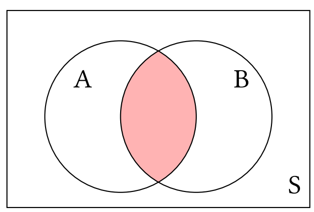
(A) $A|B$

B $\boxed{\text{(B) } A \text{ and } B}$ +1

(C) A or B

(D) $B|A$

(E) the entire sample space S



(e) In the scatter plot, a possible correlation coefficient is

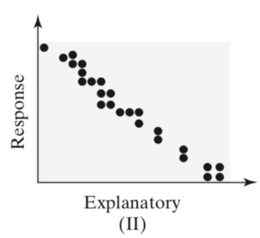
~~(A) $r = 0.01$~~

~~(B) $r = -0.01$~~

~~(C) $r = 0.95$~~

D $\boxed{\text{(D) } r = -0.95}$ +1

~~(E) $r = 0.50$~~



+6

Problem 5: 8 pts

Let X be a random variable for a game determined by chance and x be the amount won or lost.

(2 pt) (a) The table below gives the probability distribution for X. What does p have to be to guarantee X is a **probability distribution**? Show your mathematical reasoning.

x	$P(x)$
-30	0.65
10	0.20
25	p
50	0.05

(4 pt) (b) Find the **mean** and the **standard deviation** of the probability distribution. Show your mathematical reasoning.

(2 pt) (c) Would you play this game? *Justify your answer with statistical reasoning*

Problem 6: 12 pts

A fair, **12-sided die** is rolled. Let $\mathbf{A} = \{$ **odd** numbers $\}$, $\mathbf{B} = \{$ **even** numbers $\}$, and $\mathbf{C} = \{$ numbers **greater than** 7 $\}$.

(3 pt) (a) Using set notation, express the **sample space**.

(3 pt) (b) Using set notation, express the event A and calculate $P(A)$.

(3 pt) (c) Using set notation, express the event C^c and calculate $P(C^c)$.

(3 pt) (d) Using set notation, express the event $A \mid C^c$ and calculate $P(A \mid C^c)$.

Problem 7: 19 pts

You decide to go to an online casino instead of Vegas and play a virtual slot machine. Assume each play is independent and the probability of winning on any play is 0.25.

(3 pt) (a) What is the probability of winning three times in a row?

For parts (b)-(e), suppose the slot machine is played 20 times in a row.

(3 pt) (b) What is the probability of winning **exactly** 13 times?

(3 pt) (c) What is the probability of winning **at least** 15 times?

(3 pt) (d) What is the probability of winning **less than** 6 times?

(3 pt) (e) Calculate the **mean** and **standard deviation**.

(4 pt) (f) Are 10 wins out of 20 **significantly high**? Justify your answer with statistical reasoning and give your answer in a complete sentence ($M \to E$).

Formula Sheet

- $\boxed{P(A \textbf{ or } B) = P(A) + P(B) - P(A \textbf{ and } B)}$
- $\boxed{P(A \textbf{ and } B) = P(A) * P(B)}$
- $\boxed{P(A \textbf{ and } B) = P(A) * P(B|A)}$
- $\boxed{P(B|A) = \dfrac{P(A \textbf{ and } B)}{P(A)}}$
- $\boxed{P(\bar{A}) = 1 - P(A)}$
- $\boxed{\text{binompdf}(n, p, x)}$
- $\boxed{\mu = \sum[x \cdot P(x)]}$
- $\boxed{\sigma^2 = \sum(x-\mu)^2 \cdot P(x)]}$
- $\boxed{\sigma = \sqrt{\sum[x^2 \cdot P(x)] - \mu^2}}$
- $\boxed{E(X) = \sum[x \cdot P(x)]}$
- $\boxed{\text{binomcdf}(n, p, x)}$

Table II

Critical Values for Correlation Coefficient

n	
3	0.997
4	0.950
5	0.878
6	0.811
7	0.754
8	0.707
9	0.666
10	0.632
11	0.602
12	0.576
13	0.553
14	0.532
15	0.514
16	0.497
17	0.482
18	0.468
19	0.456
20	0.444
21	0.433
22	0.423
23	0.413
24	0.404
25	0.396
26	0.388
27	0.381
28	0.374
29	0.367
30	0.361

Problem 5: 8 pts

Let X be a random variable for a game determined by chance and x be the amount won or lost.

(2 pt) (a) The table below gives the probability distribution for X. What does p have to be to guarantee X is a **probability distribution**? Show your mathematical reasoning.

x	$P(x)$
-30	0.65
10	0.20
25	p
50	0.05

(4 pt) (b) Find the **mean** and the **standard deviation** of the probability distribution. Show your mathematical reasoning.

(2 pt) (c) Would you play this game? *Justify your answer with statistical reasoning*

(a) $0.65 + 0.2 + p + 0.05 = 1 \quad \left(\sum P(x) = 1\right)$

$0.9 + p = 1$

$p = 1 - 0.9$

$\boxed{p = 0.1}$ +2

(b) $\mu = \sum[x \cdot P(x)] = (-30)(0.65) + (10)(0.2) + (25)(0.1) + (50)(0.05)$

$= \boxed{-12.5 \text{ \$}}$ or \$/game +2

$\sigma = \sqrt{\underbrace{\sum[x^2 \cdot P(x)]}_{\text{Do on calc}} - \mu^2} = \sqrt{792.5 - (-12.5)^2} = 25.22399\ldots$

$= \boxed{25.2 \text{ \$}}$ +2

(c) Don't play! No, I would not play this game! If I played it alot, I would expect to leave having lost an average of \$12.50. +2

+8

Problem 6: 12 pts

A fair, **12-sided die** is rolled. Let $\mathbf{A} = \{$ **odd** numbers $\}$, $\mathbf{B} = \{$ **even** numbers $\}$, and $\mathbf{C} = \{$ numbers **greater than** 7 $\}$.

(3 pt) (a) Using set notation, express the **sample space**.

(3 pt) (b) Using set notation, express the event A and calculate $P(A)$.

(3 pt) (c) Using set notation, express the event C^c and calculate $P(C^c)$.

(3 pt) (d) Using set notation, express the event $A \mid C^c$ and calculate $P(A \mid C^c)$.

(a) $\boxed{S = \{1,2,3,4,5,6,7,8,9,10,11,12\}}$ +3 or $S = \{1,2,3,\dots,11,12\}$.

(b) $A = \{1,3,5,7,9,11\} \rightarrow P(A) = \frac{\#A}{\#S} = \frac{6}{12} = \boxed{0.5}$ +3

(c) $C = \{8,9,10,11,12\}$

$C^c = \{1,2,3,4,5,6,7\} \rightarrow P(C^c) = \frac{\#C^c}{\#S} = \frac{7}{12} = \boxed{0.583}$ +3

(d) $A \mid C^c = A \text{ and } C^c = \{1,3,5,7\} \rightarrow P(A \mid C^c) = \frac{\#A \text{ and } C^c}{\# C^c} = \frac{4}{7}$

+3 $= \boxed{0.571}$

+12

Problem 7: 19 pts

You decide to go to an online casino instead of Vegas and play a virtual slot machine. Assume each play is independent and the probability of winning on any play is 0.25. win or lose → this is a binomial probability problem

(3 pt) (a) What is the probability of winning three times in a row?

For parts (b)-(e), suppose the slot machine is played 20 times in a row.

(3 pt) (b) What is the probability of winning **exactly** 13 times? $x=13$

(3 pt) (c) What is the probability of winning **at least** 15 times? $x \geq 15$

(3 pt) (d) What is the probability of winning **less than** 6 times? $x < 6$

(3 pt) (e) Calculate the **mean** and **standard deviation**.

(4 pt) (f) Are 10 wins out of 20 **significantly high**? Justify your answer with statistical reasoning and give your answer in a complete sentence ($M \to E$).

(a) $n=3$ trials

$P(W \text{ and } W \text{ and } W) = P(W)^3 = 0.25^3 = \boxed{0.0156}$ +3

b/c independent

Note can also do it w/ $P(x=3) = \text{binompdf}(\underset{n}{3}, \underset{p}{0.25}, \underset{x}{3})$

Now ASSUME $n=20$

let X = # of wins in 20 plays, x = # of X

(b) $P(x=13) = \text{binompdf}(\underset{n}{20}, \underset{p}{0.25}, \underset{x}{13}) = \boxed{1.54\text{E}-4}$ or $\boxed{0.000154}$ +3

(c) at least 15 → $\underbrace{0\,1\,2 \cdots 14}_{\text{cummulative}}\,|\,15\,16 \cdots 20$

$P(x \geq 15) = 1 - P(x \leq 14) = 1 - \text{binomcdf}(20, 0.25, 14) = \boxed{3.81\text{E}-6 = 0.000\,003\,81}$ +3

↳ use cdf

(d) $P(x<6) = \text{binomcdf}(20, 0.25, 5) = \boxed{0.617}$ +3

less than 6 → $0\,1 \cdots 5\,|\,6\,7 \cdots 20$

↳ cummulative

(e) Binomial Distributions have simple formulas for μ & σ!

$\mu = np$ & $\sigma = \sqrt{npq}$

$\mu = 20 * 0.25$ $\quad$ $\sigma = \sqrt{20 * 0.25 * 0.75}$

$\boxed{\mu = 5 \text{ wins}}$ $\quad$ $\boxed{\sigma = 1.9 \text{ wins}}$ +3

(f) $P(x=10) = \text{binompdf}(20, 0.25, 10)$
$= 0.00992$.

+4 Yes, since the probability of 10 wins is less than 1% it is unusually high # of wins.