

Remote Proctoring Instructions.

- Gradescope assignment with the PDF **entire exam** will be available on the “Final assignment” (on either the regular or Alternate Gradescope).
- **Be sure to download** the *PDF* from the Final Gradescope assignment.
- There will be no clarifications made directly to individuals. We will listen to issues, but if a problem is identified to be in error, we may choose to address it during the midterm or after the midterm. Please keep moving through the exam.
- **Remote: You have 180 minutes to do the exam and then an extra twenty minutes to scan your answer sheet to the Final assignment.**
- **Remote: Clarification Request form:** <https://forms.gle/GYVTRkoZeTScgt7t9>
- **Clarification Doc:** <https://docs.google.com/document/d/1240jC66ZIkHlyDQA00JaAnadz111zKpyUPKeuCTQow/>
- For emergencies, email fa21@eecs70.org or use the disruption form at: <https://forms.gle/fufi8sKN3m5p5jet8>. Again, keep working as best as possible, as we cannot respond in real time.

Advice.

- The questions vary in difficulty. In particular, some of the proof questions at the end are quite accessible, and even those are in not necessarily in order of difficulty. Unless stated otherwise, All short answers and true false questions are worth 3 points and each written problem is worth 15. No negative points on true/false. **So do really scan over the exam a bit.**
- The question statement is your friend. Reading it carefully is a tool to get to your “rational place”.
- You may consult *three sheet of notes on both sides*. Apart from that, you may not look at books, notes, etc. Calculators, phones, computers, and other electronic devices are NOT permitted.
- **You may, without proof, use theorems and lemmas that were proven in the notes and/or in lecture, unless otherwise stated. That is, if we ask you to prove a statement, prove it from basic definitions, e.g., “ $d \mid x$ means $x = i(d)$ for some integer i ” is a definition.**

Major Gradescope Issues. If there is a global issue and it is not affecting you, please continue. If you are experiencing difficulties with Gradescope or Zoom, you may check your email, and we will post a global message on Piazza and bypass email preferences to inform you of what to do.

1. Pledge.

Berkeley Honor Code: As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.

In particular, I acknowledge that:

- I alone am taking this exam. Other than with the instructor and GSIs, I will not have any verbal, written, or electronic communication about the exam with anyone else while I am taking the exam or while others are taking the exam.
- I will not have any other browsers open while taking the exam.
- I will not refer to any books, notes, or online sources of information while taking the exam, other than what the instructor has allowed.
- I will not take screenshots, photos, or otherwise make copies of exam questions to share with others.
- I will make an effort to be polite to people from Stanford. [This is optional.]

Signed: _____

2. Long Ago.

1. $\forall x, y \in \mathbb{N}, x - y \leq \min(x, y)$. True False
2. $(\neg(\exists x \in S, (\neg P(x) \vee Q(x)))) \equiv (\forall x \in S, (P(x) \wedge (\neg Q(x))))$. True False
3. $\forall p \in \mathbb{N}, ((\forall n \in \mathbb{N}, (n < 2 \vee n = p \vee \neg(n|p))) \implies (\forall a \in \mathbb{N}, a^p - a = 0 \pmod{p}))$. True False
4. If $x|y$ and $x|z$, then $x|\gcd(y, z)$. True False

5. Let A and B be finite sets, and let $f : A \rightarrow B$ and $g : B \rightarrow A$ be two functions such that

$$\forall x \in B, f(g(x)) = x.$$

Fill in the blank with one of $>, \geq, <, \leq, =$ or “Incomparable” in the statement

$$|A| \text{ ___ } |B|.$$

Choose the most specific answer that is always true.

- \geq $>$ $=$ $<$ \leq Incomparable.

6. For natural numbers k and n , any solution to $x^k = n$ must be an integer or irrational. True False
7. For a natural number n which is *not* a perfect square, then $x^3 = n$ has no rational solutions. True False
8. For a natural number n which is *not* a perfect square, then $x^4 = n$ has no rational solutions. True False
9. Consider job optimal and candidate optimal stable pairings, P and P' , in a stable matching instance, consider a graph G where vertices consist of jobs and candidates and the pairs in P and P' form edges.
- (a) The resulting graph is simple. True False
- (b) If the graph consists of a single simple cycle, then there are at most two stable pairings. True False
10. If no one gets rejected in the job propose stable matching algorithm, then:
- (a) Every job gets the first candidate on its list. True False

(b) The resulting pairing is *both* job optimal and candidate optimal.

True False

11. There are always at least two distinct stable pairings in any stable matching instance; the job optimal and the candidate optimal pairing.

True False

12. What is the number of edges in an n -vertex simple graph where every vertex has degree 2?

13. A simple graph with n vertices and c connected components, where each has degree exactly 2, has _____ simple cycles.

14. Give as tight an upper bound as possible on the number of edges in a simple planar graph with n vertices and c connected components.

15. How many paths of length d are there from the all 1's vertex to the all 0's vertex in a d -dimensional hypercube?

3. More and more pi.

Prove *using induction* that the sum of the interior angles of a convex polygon with n sides (and n vertices) is $(n-2)\pi$. You may assume this is true for a triangle. (For a convex polygon, you may assume a line segment between “non-adjacent” vertices u and v splits the polygon into two convex polygons each containing u and v as vertices and are otherwise disjoint.)

4. Three, what's up with thee?

1. What are all the possible values of perfect squares modulo 3?

-
2. Prove that any integer solution to the equation $x^2 + y^2 = 3xy$ must have $x \equiv y \equiv 0 \pmod{3}$.

3. Prove that the equation $x^2 + y^2 = 3xy$ has no solutions for positive integers x and y . (Hint: Suppose for contradiction that there was a solution (a, b) , with a and b positive integers, where a was as small as possible. Find a positive integer solution (a', b') where $a' < a$. You may find part (b) helpful.)

5. Polynomials.

1. Suppose $P(x)$, $Q(x)$ are distinct degree d polynomials. Then they have at most ____ intersections. Recall an intersection is a value x , where $P(x) = Q(x)$. (Give the tightest upper bound possible.)

2. Suppose polynomial $P(x)$ is of degree $2d$ and $Q(x)$ is of degree d . Then they have at most ____ intersections. Recall an intersection is a value x , where $P(x) = Q(x)$. (Give the tightest upper bound possible.)

3. Give a polynomial with roots at r_1, r_2 that has value v at r_3 modulo a prime p .

4. Every polynomial modulo a prime p is equivalent to a polynomial of degree at most _____. (Your answer should be as small as possible.)

5. The Berlekamp-Welch algorithm is used to send a message of size $n = 3$ sent over a noisy channel, which possibly corrupts at most 2 packets.

(a) How many packets does one send in this situation? (Looking for a number.)

(b) If exactly 2 errors occur somewhere in the set of packets used, how many possible polynomials $Q(x)$ (working modulo a prime p) could one possibly reconstruct?

-
- (c) If the error polynomial is $E(x) = x^2 + 4x + 2 \pmod{7}$, where are the errors? (Give the x value(s) for your answer.)

6. Some counting.

1. Sylvia has found seven different NFTs (non-fungible tokens or digitally stamped pictures), and plans to acquire them for herself by taking screenshots. She wishes to end up with ten screenshots in total. She does not necessarily need a screenshot of every one.

- (a) How many different sets of screenshots can she obtain? The order in which she takes the screenshots does not matter, and screenshots of the same NFT are indistinguishable.

- (b) (5 points) Sylvia decides she does not want to have more than four screenshots of the same NFT. How many ways could she take screenshots now?

2. (5 points) How many ways are there to express 10 as the sum of positive integers, where the order of the sum matters? For example, $10 = 10$, $10 = 4 + 3 + 3$ and $10 = 3 + 3 + 4$ count as three different ways.

3. (5 points) How many positive factors of 1008 are also factors of 840? (Hint: Think of factors of $\gcd(840, 1008)$.)

7. Countability

1. The set of indices (or locations) of the 1's in the decimal representation of π . (For example, $\pi = 3.14159\dots$ so the set contains the indices 1 and 3 as there is a 1 at these locations in the decimal representation of π .)

Countable

Uncountable

2. The set of subsets of a countably infinite set.

Countable

Uncountable

3. The set of finite sized subsets of a countably infinite set.

Countable Uncountable

4. The set of all pairs of elements of two countably infinite sets.

Countable Uncountable

8. Computability

1. Given a program P , an input x , and a number n , does the program P halt on x in n steps?

Computable Uncomputable

2. Given a program P , an input x , and a number n , does the program P touch the memory location numbered n ?

Computable Uncomputable

3. Given a program P , an input x , and a number n , does the program P touch more than n memory locations?

Computable Uncomputable

9. Probability: Based!

Given a probability space (Ω, \mathbb{P}) and events A, B and C , fill in $\leq, <, =, \geq, >$, or “Incomparable” for questions 1-7 such that the statement always holds. If no inequality always holds use “Incomparable.”

1. $\mathbb{P}[A \cup B \cup C]$ _____ $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C]$.

\geq $>$ $=$ $<$ \leq Incomparable.

2. $\mathbb{P}[A \cap B \cap C]$ _____ $\mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]$

\geq $>$ $=$ $<$ \leq Incomparable.

3. $\mathbb{P}[A \cap B]$ _____ $\mathbb{P}[A]$.

\geq $>$ $=$ $<$ \leq Incomparable.

4. $\mathbb{P}[A|B]$ _____ $\mathbb{P}[A]$.

\geq $>$ $=$ $<$ \leq Incomparable.

5. $\mathbb{P}[A \cup B \cup C]$ _____ $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] - \mathbb{P}[A \cap B] - \mathbb{P}[A \cap C] - \mathbb{P}[B \cap C]$.

\geq $>$ $=$ $<$ \leq Incomparable.

For the following, let X_A, X_B , and X_C be indicator random variables for events A, B, C .

6. If $\text{Cov}(X_A, X_B) > 0$, then $\mathbb{P}[A \cap B]$ _____ $\mathbb{P}[A]\mathbb{P}[B]$.

\geq $>$ $=$ $<$ \leq Incomparable.

7. If $\text{Cov}(X_A, X_B) > 0$ and $\text{Cov}(X_B, X_C) > 0$, then $\text{Cov}(X_A, X_C)$ _____ 0.

\geq $>$ $=$ $<$ \leq Incomparable.

8. If $\mathbb{P}[A|B] = .6$ and $\mathbb{P}[A] = .5$ and $\mathbb{P}[B] = .5$, what is the linear function $f(X_B)$ that minimizes $\mathbb{E}[(X_A - f(X_B))^2]$?

10. More Probability: Cringe?

1. For random variables X and Y , the linear regression line of Y given X goes through the origin if and only _____.

2. For a random variable X , $\mathbb{E}[X^2] =$ _____. (In terms of $\text{Var}(X)$ and $\mathbb{E}[X]$.)

3. For random variables X and Y , $\text{Cov}(XY) = \mathbb{E}[XY]$ if $\mathbb{E}[X] = \mathbb{E}[Y] =$ _____.

4. For independent random variables X and Y , what is the best linear estimator for Y given X , i.e., $\hat{y}(X)$? (Fully simplify the answer for this setup.)

5. Consider the process of sampling n people who tested for flu last year to determine the fraction, p , of the population would test positive. To set this up, let X_i be the random indicator variable that i person in the sample got the flu for $i \in \{1, \dots, n\}$.

(a) What is $\mathbb{E}[X_i]$ in terms of p ?

(b) What is the tightest possible upper bound on $\text{Var}(X_i)$, independent of the value of p ?

(c) Using Chebyshev, give a 95% confidence interval for p , given that 50 people in your sample of 100 people tested positive for having flu.

11. The die is cast.

A bag contains a 4-sided die and a 6-sided die. Your friend Lucas pulls a die out of the bag uniformly at random, rolls it, and gets a 1. Conditional on this event, what is the probability they pulled the 4-sided die out of the bag? Show your work.

12. Fixed Points.

Recall that the number of fixed points X in a permutation π of the n elements $\{1, 2, \dots, n\}$ is the number of points such that $\pi(i) = i$.

Use the union bound to show that the probability that there are at least k fixed points in a random permutation of $\{1, 2, \dots, n\}$ is at most $\frac{1}{k!}$.

(Hint: What is the probability $1, 2, \dots, k$ are all fixed points?)

13. Happier apart.

There are 6 students (Alice, Bob, Catherine, Dustin, Emily, and Frank) in a classroom in which 6 seats are arranged in a circle. When the semester begins, the teacher assigns a seat to each student randomly; that is, all possible seating assignments are equally likely.

Alice and Bob do *not* want to be seated next to each other. If they're assigned to adjacent seats, they will complain of unhappiness. Determine the probability that the teacher's seating assignment does not elicit a complaint.

14. Go big! Or not!

You and a friend play marbles. Each player starts with 10 marbles, and to win you must get all 20 marbles. After deciding on a game, you are given two options of game format. For the game in each part, select the option maximizing your chance of getting all 20 marbles.

- Option A: Play one round; the winner gets all 20 marbles
- Option B: Play a 20-round series (in the event of a 10-to-10 tie, replay until the tie is broken): whoever wins more individual rounds gets all 20 marbles
- Option C: Doesn't matter between A and B. This is the only option that should be selected if A and B yield equal probability of winning.

(a) Odds and Evens, where you have a 35% chance of winning each round.

- (A) (B) (C)

(b) Rock-Paper-Scissors, where you have a 50% chance of winning each round.

(A) (B) (C)

(c) Mancala, where you have a 90% chance of winning each round.

(A) (B) (C)

(d) Nim, where you have a 100% chance of winning each round.

(A) (B) (C)

15. Spam Filter

Jon Byrnu Lee receives many emails each day. Each email is as likely to be *Spam* (S) as it is to be *Not Spam* (S^c)—that is, $\mathbb{P}[S] = \mathbb{P}[S^c] = 1/2$.

To increase his daily productivity, Jon decides to implement a content-based statistical filtering algorithm. By parsing the words in the subject line—perhaps in addition to analyzing other aspects of each email as well—his spam filter directs each email either to his Inbox or to his Spam Folder. Jon *never* reads an email that the filter places in his Spam Folder.

The table below depicts a listing of the probabilities Jon uses to design his spam filter; for ease of notation and grading, the *Symbol* column indicates the symbol you should use to denote each word in your mathematical expressions.

Word	Symbol	$\mathbb{P}[\text{Word} S]$	$\mathbb{P}[\text{Word} S^c]$
Bonds	B	$\mathbb{P}[B S] = 0.1$	$\mathbb{P}[B S^c] = 0.15$
Reward	R	$\mathbb{P}[R S] = 0.15$	$\mathbb{P}[R S^c] = 0.25$
Winner	W	$\mathbb{P}[W S] = 0.08$	$\mathbb{P}[W S^c] = 0.02$

By way of example, the table specifies that the probability that the word "Winner" appears in the subject line, given that it is *Spam*, is $\mathbb{P}[W|S] = 0.08$. The probability that the same word appears in the subject line, given that it is *Not Spam*, is $\mathbb{P}[W|S^c] = 0.02$.

(a) Determine $\mathbb{P}[R]$, the probability of the occurrence of the word "Reward" in the subject line of his emails.

(b) Given that an incoming email has the word "Reward," what is the probability that Jon will read it? Note that Jon will read it if it is *Not Spam*. You may leave your answer here in terms of $\mathbb{P}[R]$.

-
- (c) Jon receives an email at noon containing the words "Bonds" and "Winner" in the subject line. Determine the probability that this email is *Spam*.

You may *not* assume that the words appearing in the subject line are independent. However, you may assume that the words "Bonds" and "Winner" are independent, *conditioned on Spam* (S). You may also assume that the words "Bonds" and "Winner" are independent, *conditioned on Not Spam* (S^c).

16. Going to a party!

Nate is waiting for the bus to go to Professor Rao's social. The time in between buses arriving at Nate's bus stop is an exponential distribution with an average arrival time of 20 minutes. Nate decides that if the next bus arrives within m minutes, he will take the bus to Professor's Rao social, and the bus ride takes 10 minutes. Otherwise, Nate will walk to the social, which takes 30 minutes. Prove that no matter the choice of m , the expected amount of time it will take for Nate to get to the social is 30 minutes.

17. Positively Gaussian.

Consider a *zero-mean* Gaussian random variable X whose probability density function (PDF) is given by

$$\forall x \in \mathbb{R}, \quad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

where $\sigma > 0$ is the standard deviation of X .

Define another random variable $Y = |X|$.

- (a) Let K denote an indicator random variable for the event $Y \leq \sigma$. Evaluate—that is, determine a numerical value for— $\mathbb{E}[K]$, the mean of K . (Use the standard normal table attached to the final page of the exam.)

- (b) (5 points) Determine a reasonably simple expression for $\mathbb{E}[Y]$. Show your setup.

Note: You do *not* need the PDF $f_Y(y)$ to determine the mean $\mathbb{E}[Y]$.

- (c) Determine a reasonably simple expression for the variance of Y .

Note: You do *not* need the PDF $f_Y(y)$ to determine the variance σ_Y^2 . You may leave your answer in terms of $\mathbb{E}[Y]$ if you wish.

- (d) Determine a reasonably simple expression for, and provide a well-labeled plot of, $f_Y(y)$, the PDF of Y . Your plot must be as close to hand-sketched as possible. It must *not* be generated by a computing device or software.

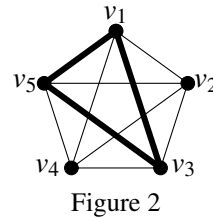
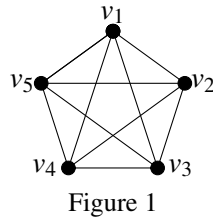
Please note that you must account for the *entire* y axis—that is, your expression must specify what the PDF $f_Y(y)$ is for *every* real y .

Expression:

Well-labeled plot of the $f_Y(Y)$ with y -intercept and units in terms of σ on the x -axis.

18. You complete me.

A *complete graph* K_n is a collection of n nodes where every pair of nodes is connected by an edge. For example, the complete graph K_5 is shown in Figure 1. Throughout this problem, assume that $n \geq 4$.



- (a) A **triangle** is a set of three nodes, each two of which are connected by an edge. Determine t_n , the number of triangles in the complete graph K_n .

Now assume that we randomly thicken some of the edges in the complete graph K_n . Specifically, each edge is thickened with probability p , or it's left intact (thin) with probability $1 - p$, independently of all other edges. Figure 2 depicts such a random thickening of edges as applied to the complete graph K_5 of Figure 1.

A **thick triangle** is a triangle that has three thick edges. For example, in Figure 2, the nodes v_1 , v_3 , and v_5 form the only thick triangle in the graph.

- (b) Determine the probability q that the nodes v_1 , v_2 and v_3 form a **thick triangle**.

- (c) Let X_n be the number of thick triangles in K_n . Determine $\mathbb{E}[X_n]$ in terms of possibly t_n , p , q , n .

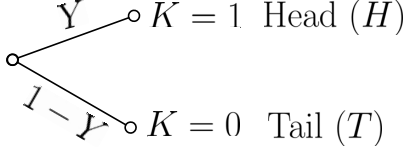
Hint: Appropriately-chosen indicator random variables may be useful to you.

- (d) (10 points.) Determine the variance of X_n , the number of thick triangles in K_n . Express your answer in terms of t_n , p , q , n , or a combination of these. Show your work.

Hint: The number of pairs of triangles in K_n that share exactly one edge is $6 \binom{n}{4}$.

19. What is the bias?

Consider a coin whose bias (probability) for a "Head" is determined by a continuous random variable Y uniformly distributed in the interval $(0, 1)$. This is shown in the figure below, with random variable K acting as the indicator variable for the outcome "Head."

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$


Let H_i denote the event that the outcome of the i^{th} toss is a Head. Similarly, T_i denotes the event that the outcome of the i^{th} toss is a Tail.

Given that $Y = y$, then we have that

$$\mathbb{P}[H_i | Y = y] = \mathbb{P}[K_i = 1 | Y = y] = \mathbb{E}[K_i | Y = y] = y,$$

and

$$\mathbb{P}[T_i | Y = y] = \mathbb{P}[K_i = 0 | Y = y] = 1 - y,$$

Moreover, we have the coin flips are conditionally independent, *given* $Y = y$. For example,

$$\mathbb{P}[H_1 \cap H_2 | Y = y] = \mathbb{P}[H_1 | Y = y] \mathbb{P}[H_2 | Y = y]$$

and

$$\mathbb{P}\left[\bigcap_{i=1}^n H_i \mid Y = y\right] = \prod_{i=1}^n \mathbb{P}[H_i | Y = y],$$

and so on.

However, suppose you have no observation on Y . In this problem, you'll show that the coin flips are *not* independent.

In what follows, it may be useful to know the Law of Total Probability for an event A and a continuous random variable Y :

$$\mathbb{P}[A] = \int_{-\infty}^{+\infty} \mathbb{P}[A | Y = y] f_Y(y) dy.$$

(a) (5 points) Determine a reasonably simple closed-form expression for

$$\mathbb{P} \left[\bigcap_{i=1}^n H_i \right] = \mathbb{P}[H_1 \cap H_2 \cap \cdots \cap H_n],$$

in terms of n . Show your setup.

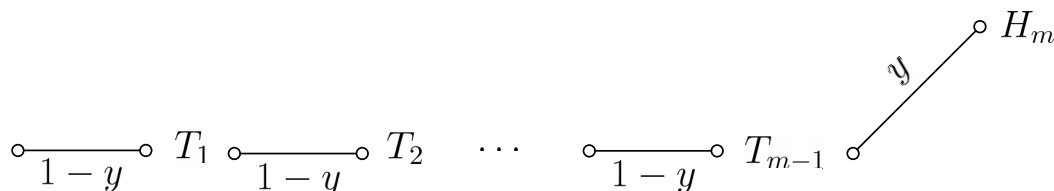
Hint: The events H_i , conditioned on $Y = y$, are independent.

(b) Determine a numerical value for $\mathbb{P}[H_j | H_i]$ for any pair of distinct positive integers i and j ($i \neq j$). Your result should indicate that H_i and H_j are not independent.

Hint: Recall the events H_i and H_j , conditioned on $Y = y$, are independent.

(c) Let M be the random variable denoting the number of tosses up to, and including, the first Head. Determine a reasonably simple expression for the probability mass function (PMF) $p_M(m)$.

A probability tree diagram that might be helpful to you is shown below:



$M = m$ tosses up to, and including, the first Head.

You may find it useful to know the following facts:

- Law of Total Probability involving a discrete and a continuous random variable:

$$p_M(m) = \int_{-\infty}^{+\infty} p_{M|Y}(m | y) f_Y(y) dy.$$

- For $0 < y < 1$ and nonnegative integers α and β ,

$$\int_0^1 y^\alpha (1-y)^\beta dy = \frac{\alpha! \beta!}{(\alpha + \beta + 1)!}.$$

(d) Let L denote the number of Heads in a set of n tosses of the coin. In particular,

$$L = K_1 + \cdots + K_n = \sum_{i=1}^n K_i,$$

where K_i denotes the value of the indicator random variable K for the i^{th} toss. Our primary goal in this part and the next is estimate the bias Y based on the observed Head count L (pun intended!).

For this part, you may find it useful to know two facts.

- For $0 < y < 1$ and nonnegative integers α and β ,

$$\int_0^1 y^\alpha (1-y)^\beta dy = \frac{\alpha! \beta!}{(\alpha + \beta + 1)!}.$$

- The conditional PDF of Y , given L , is of the form

$$f_{Y|L}(y|\ell) = \begin{cases} c(n, \ell) y^\ell (1-y)^{n-\ell} & \text{if } \ell \in \{0, 1, 2, \dots, n\} \text{ and } 0 < y < 1. \\ 0 & \text{otherwise,} \end{cases}$$

where $c(n, \ell)$ is the normalization factor, dependent on n and ℓ .

- (i) What is the normalization factor $c(n, \ell)$?

- (ii) The minimum mean squared-error estimator of Y , based on L , is defined by

$$\hat{Y}_{\text{MMSE}}(L = \ell) = \mathbb{E}[Y | L = \ell] = \int_{-\infty}^{+\infty} y f_{Y|L}(y|\ell) dy.$$

Determine a reasonably simply expression for $\hat{Y}_{\text{MMSE}}(L)$ in terms of L and n . (If you are unsure of your expression for $c(n, \ell)$, you may leave your here in terms of $c(n, \ell)$.)

- (e) Using a formula below, or through some other means, determine a reasonably simple expression for $\hat{Y}_{\text{LLSE}}(L)$, the *linear* least squares estimator (LLSE) for Y , based on L .

Recall that the LLSE is given by:

$$\hat{Y}_{\text{LLSE}}(L) = \mathbb{E}[Y] + \frac{\sigma_{LY}}{\sigma_L^2} [L - \mathbb{E}[L]]$$

Note: Please pause and contemplate before you dive into complicated mathematical manipulations. This part can be answered *without* resorting to the LLSE formulas above!