Last updated: November 1, 2026

🗺️ Four-year roadmap · Year 3, Spring

About this course

This free course has 10 units and 6 labs. Those are the real counts, not a longer syllabus cut short.

A check is one number or code result on a course page that a script recomputes before the page is published. The printed number and the recomputed number have to match.

> This is a free self-study course outline written by Saving Planets. It is not a university course, it carries no academic credit, and no school is involved.
>
> Its scope follows a one-semester junior probability and statistics course for engineers. It is modeled on the topics and depth of MIT 18.05 Introduction to Probability and Statistics and MIT 6.041 Probabilistic Systems Analysis (free on MIT OpenCourseWare), using the free OpenStax Introductory Statistics text and the NIST/SEMATECH e-Handbook of Statistical Methods, which we link where they help. None of those authors, schools or agencies endorses this page.
>
> What we add: every unit ties to real work with a cited public source (USGS flood and earthquake probabilities, NOAA hurricane outlooks and forecast cones, the Mauna Loa CO_2 record and NASA's temperature analysis), every number on the page is re-computed by a checking script before it goes live, and every lab has a free simulation or a spreadsheet.
>
> Work through it in order. Each unit has goals, key ideas, equations, step-by-step examples, practice with full answers, common mistakes, flashcards and a quiz. Your progress is saved only in this browser.

How this course helps save the planet

Protecting people and the planet is a question of odds: how likely a flood, earthquake or hurricane is, how sure we are that CO_2 is rising faster, and how reliable the systems that warn us are.

Statistics turns data into those odds and tells us how much to trust them. Each unit ends with a short, sourced section. Values that are not from a source are labeled as sample or textbook values.

For instructors

You can teach this course, or let a student work through it alone. It is free. Saving Planets does not give a grade or school credit.

- This page is the whole course: 10 units and 6 labs (MATH 350).
- Plan on 15 weeks, about 6 to 9 hours a week, including the lab weeks.
- Follow the 15-week schedule. Read the unit, do the practice on paper, then open the answer. In a lab week, do the lab or the home version.
- A check is one number or code result on the page that a script recomputes before we publish. The number you see and the recomputed number have to match.
- Use Print this page, or download the PDF study guide. The formula sheet is near the end.
- Ask: how does this unit help people and the planet? Each unit has a green box with a sourced fact.
- Copy the teacher guide template and write in your class name, the time you have, and the questions you will ask. Open the teacher guide template

Before you start

Prerequisites

- Calculus I and II: derivatives, integrals and series.
- Linear Algebra: vectors and least squares (helpful for regression).
- Intro Programming or spreadsheet skills: formulas, ranges and charts.
- Numerical Methods (helpful, not required).

How to study

- Write down the random variable and its units before computing anything.
- Draw the distribution or a tree diagram first.
- Check that probabilities lie between 0 and 1 and add up to 1.
- State assumptions such as independence out loud; most errors start there.
- Simulate when in doubt: a spreadsheet with RAND() can check almost any probability.
- Do every practice problem on paper before opening the answer, and retake the quizzes a few days later.

What you need

- A computer or tablet for the free PhET Center and Variability, Plinko Probability and Least-Squares Regression simulations.
- Spreadsheet software (any free one works) with RAND(), NORM.DIST and LINEST, or Python.
- NOAA's Mauna Loa CO_2 data file in a web browser.
- A calculator; a pair of dice is optional.

Suggested 15-week schedule

About 6 to 9 hours a week: reading, practice problems and, in lab weeks, the lab.

Week	Unit	Topics	Reading
1	Describing Data: Center, Spread and Plots	Describing data; Lab A	OpenStax Intro Stats
2	Probability Rules, Counting and Independence	Probability rules and counting	MIT 18.05
3	Conditional Probability and Bayes' Theorem	Conditional probability and Bayes	MIT 6.041
4	Discrete Random Variables: Binomial and Poisson	Discrete variables and the binomial; Lab B	MIT 18.05
5	Discrete Random Variables: Binomial and Poisson	Poisson and geometric models	MIT 18.05
6	Continuous Distributions: Normal, Exponential and Extremes	Continuous distributions; Lab C	USGS 100-year flood
7	Sampling, Simulation and the Central Limit Theorem	Sampling and the CLT; Lab D; midterm practice	MIT 18.05
8	Confidence Intervals	Confidence intervals for means	OpenStax Intro Stats
9	Confidence Intervals	Intervals for proportions and sample size	NOAA CPC
10	Hypothesis Tests and p-Values	Hypothesis tests and p-values	MIT 18.05
11	Correlation and Linear Regression	Correlation and regression	OpenStax Intro Stats
12	Correlation and Linear Regression	Residuals and model checking; Lab E	NOAA GML
13	Reliability and Risk	Reliability and MTBF	NIST e-Handbook
14	Reliability and Risk	System reliability; Lab F	NIST e-Handbook
15	Reliability and Risk	Risk and review; final practice	NOAA CPC; USGS

Your progress

Tick units and labs as you finish them, and take the quizzes. Progress and quiz scores are saved only in this browser, never sent to us.

Units

Describing Data: Center, Spread and Plots

Types of data, histograms and box plots, mean and median, range, variance and standard deviation, quartiles and outliers, and why the sample standard deviation divides by n − 1.

Learning objectives

By the end of this unit you can:

- Compute mean, median, range and sample standard deviation.
- Choose between mean and median.
- Read a histogram and box plot.
- Flag outliers with the 1.5 IQR rule.

Key concepts

- The mean x̄ = Σx/n is the balance point; the median is the middle value and resists outliers.
- The sample variance s² = Σ(x − x̄)²/(n − 1) measures spread; s, its square root, has the data's units. Dividing by n − 1 corrects for using x̄ instead of the true mean.
- The range is max − min; the interquartile range IQR = Q3 − Q1 holds the middle half of the data.
- Points more than 1.5 IQR beyond the quartiles are flagged as possible outliers: check them, do not just delete them.
- Always plot first: a histogram or box plot shows skew, gaps and outliers that a single number hides.

Key equations

$$\bar{x} = \frac{1}{n}\sum x_i$$

`x=1n x_i`

Sample mean

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$

`s^2=(1)/(n-1)(x_i- x)^2`

Sample variance

$$IQR = Q_3 - Q_1$$

`IQR=Q_3-Q_1`

Interquartile range

$$z = \frac{x - \bar{x}}{s}$$

z=(x- x)/(s)

Standard score

Worked examples, step by step

Example 1.1 NOAA's Mauna Loa annual means give yearly CO_2 increases for 2021–2025 of 2.20, 2.12, 2.55, 3.53 and 2.74 ppm. Find the mean and median.

1. Sum = 13.14; divide by 5.
2. Sort: 2.12, 2.20, 2.55, 2.74, 3.53.

Answer: Mean 2.628 ppm/yr; median 2.55 ppm/yr.

Example 1.2 Find the range and sample standard deviation.

1. Range = 3.53 − 2.12.
2. s = √(Σ(x − 2.628)²/4).

Answer: Range 1.41 ppm/yr; s = 0.564 ppm/yr.

Example 1.3 Find the z-score of the 3.53 ppm increase (2024).

1. z = (3.53 − 2.628)/0.564.

Answer: z = 1.60.

Practice problems

Work each on paper first, then check.

1. Which is pulled more by the 3.53 value, mean or median?

   The mean (2.628 vs 2.55).

2. Why divide by n − 1?

   The deviations from x̄ are slightly too small; n − 1 corrects the bias.

3. Units of s for CO_2 increases?

ppm per year.

4. What does a long right tail do to mean vs median?

   Mean > median.

Common misconceptions

Myth: The mean is always typical.
Actually: Skew or outliers can pull it away.

Myth: Outliers are mistakes.
Actually: They may be real events; investigate.

Myth: Standard deviation can be negative.
Actually: It is a square root, never negative.

Flashcards

Mean

$\Sigma x/n$.

Median

Middle value.

Standard deviation

$\sqrt{(\Sigma(x - \bar{x})^2/(n - 1))}$.

IQR

Q3 − Q1.

Outlier rule

1.5 IQR beyond quartiles.

z-score

$(x - \bar{x})/s$.

Self-check quiz

1. Resists outliers...

median

mean

range

variance

median. Middle value.

2. Sample variance divides by...

n − 1

n

n + 1

2n

n − 1. Bias correction.

3. IQR is...

Q3 − Q1

max − min

mean − median

s^2

Q3 − Q1. Middle half.

4. Plot data...

first

never

last

only if asked

first. See shape.

Try this quiz again

How this helps save the planet

NOAA reports that the Mauna Loa CO_2 record, begun by C. David Keeling of Scripps in March 1958, is the longest record of direct measurements of CO_2 in the atmosphere. NOAA's annual means rose from 414.21 ppm in 2020 to 427.35 ppm in 2025.

This unit's summary statistics turn that record into a rate: the five yearly increases average 2.628 ppm per year, with a standard deviation of 0.564 ppm per year from year-to-year variation.

On Saving Planets: Solutions · Earth Observation

Sources: NOAA Global Monitoring Laboratory: Trends in atmospheric CO_2 (Mauna Loa) (https://gml.noaa.gov/ccgg/trends/) · NOAA GML: Mauna Loa annual mean CO_2 data (https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_annmean_mlo.txt)

Free learning resources

- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.
- PhET: Center and Variability (https://phet.colorado.edu/en/simulations/center-and-variability): Mean, median and spread interactively.

☐ **I finished this unit**

Probability Rules, Counting and Independence

Sample spaces and events, the axioms, complements, addition and multiplication rules, independence, permutations and combinations, and “at least one” problems.

Learning objectives

By the end of this unit you can:

- Apply the complement, addition and multiplication rules.
- Test independence.
- Count with permutations and combinations.
- Solve “at least one” problems.

Key concepts

- Probabilities lie between 0 and 1 and the whole sample space has probability 1. P(not A) = 1 − P(A).
- Addition rule: P(A or B) = P(A) + P(B) − P(A and B).
- A and B are independent if P(A and B) = P(A)P(B); then knowing one tells you nothing about the other.
- For “at least one” in n independent trials, use the complement: 1 − (1 − p)^n.
- Combinations C(n, k) = n!/(k!(n − k)!) count unordered selections; permutations count ordered ones.

Key equations

$$P(A^c) = 1 - P(A)$$

`P(A^c)=1-P(A)`

Complement

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

`P(A B)=P(A)+P(B)-P(A B)`

Addition rule

$$P(A \cap B) = P(A)P(B)$$

`P(A B)=P(A)P(B)`

Independence

$$P(\geq 1) = 1 - (1 - p)^n$$

`P(1)=1-(1-p)^n`

At least one

Worked examples, step by step

Example 2.1 A 100-year flood has a 1% chance each year. Find the chance of at least one in a 30-year mortgage.

1. P(none) = 0.99^{30} = 0.740.
2. P(at least one) = 1 − 0.740.

Answer: 26.0%.

Example 2.2 Find the chance of at least one 100-year flood in 100 years.

1. 1 − 0.99^{100}.

Answer: 63.4%, not 100%.

Example 2.3 Find the chance of at least one 50-year flood (2% a year) in 30 years.

1. 1 − 0.98^{30}.

Answer: 45.5%.

Practice problems

Work each on paper first, then check.

1. At least one 250-year flood (0.4% a year) in 30 years?

   1 − 0.996^{30} = 11.3%.

2. How many ways to choose 6 of 9 items?

   C(9, 6) = 84.

3. Can two 100-year floods happen in back-to-back years?

   Yes: each year starts with the same 1% chance.

4. Are yearly floods treated as independent here?

Yes; the USGS notes a rare flood does not reduce the chance of another soon after.

Common misconceptions

Myth: A 100-year flood happens once every 100 years.

Actually: It has a 1% chance every year.

Myth: After a flood, the next is less likely.

Actually: Each year is independent.

Myth: P(A or B) = P(A) + P(B) always.

Actually: Only if A and B cannot both happen.

Flashcards

Complement

1 − P(A).

Addition rule

Subtract the overlap.

Independence

$P(A \cap B) = P(A)P(B)$.

At least one

$1 - (1 - p)^n$.

Combination

$n!/(k!(n - k)!)$.

AEP

Annual exceedance probability.

Self-check quiz

1. P(at least one in n)...

$1 - (1 - p)^n$

np

p^n

$1 - p^n$

$1 - (1 - p)^n$. Complement.

2. Independent means...

$P(A \cap B) = P(A)P(B)$

disjoint

equal

$P = 0$

$P(A \cap B) = P(A)P(B)$. Definition.

3. 100-year flood annual chance...

1%

100%

10%

0.1%

1%. USGS.

4. C(n, k) counts...

unordered choices

orders

outcomes always

pairs only

unordered choices. Combinations.

Try this quiz again

How this helps save the planet

The USGS explains that a “100-year flood” is a flood with a 1 percent chance of happening in any year, and that 100-year floods can happen two years in a row. Its table pairs each recurrence interval

with an annual exceedance probability: 2 years with 50%, 10 years with 10%, 100 years with 1% and 500 years with 0.2%.

The USGS notes that in the 1960s the United States chose the 1-percent annual exceedance probability flood as the basis for the National Flood Insurance Program. This unit's complement rule shows that a 30-year home in that zone has a 26% chance of seeing at least one such flood.

On Saving Planets: Oceans · Solutions

Sources: USGS Water Science School: The 100-Year Flood (https://www.usgs.gov/special-topics/water-science-school/science/100-year-flood)

Free learning resources

- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.
- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.

☐ **I finished this unit**

Conditional Probability and Bayes' Theorem

Conditional probability, the multiplication rule, tree diagrams, the law of total probability, Bayes' theorem, base rates, and turning a 30-year probability into a yearly rate.

Learning objectives

By the end of this unit you can:

- Compute conditional probabilities from a table or tree.
- Apply the law of total probability.
- Use Bayes' theorem and explain base-rate effects.
- Convert multi-year probabilities to annual rates.

Key concepts

- P(A | B) = P(A and B)/P(B) is the probability of A once we know B happened.
- The law of total probability adds the branches of a tree: P(A) = Σ P(A | B_i)P(B_i).
- Bayes' theorem reverses a condition: P(B | A) = P(A | B)P(B)/P(A).
- Rare events give many false alarms: even a good test has a low positive predictive value when the base rate is small.
- If events arrive at a constant rate λ, the chance of at least one in t years is 1 − e^(−λt), so λ = −ln(1 − P)/t.

Key equations

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

`P(A|B)=(P(A B))/(P(B))`

Conditional probability

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

`P(A)=_i P(A|B_i)P(B_i)`

Total probability

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

`P(B|A)=(P(A|B)P(B))/(P(A))`

Bayes' theorem

$$\lambda = -\frac{\ln(1-P)}{t}$$

```
=-((1-P))/(t)
```

Rate from a t-year probability

Worked examples, step by step

Example 3.1 A sensor flags a fault 95% of the time when one exists and falsely 5% of the time otherwise; 2% of units have the fault (sample values). Find P(fault | flag).

1. P(flag) = 0.95 × 0.02 + 0.05 × 0.98 = 0.068.
2. Bayes: 0.019/0.068.

Answer: 27.9%: most flags are false alarms.

Example 3.2 USGS gives a 72% chance of at least one M ≥ 6.7 earthquake in the San Francisco Bay region in 30 years. Find the yearly rate, assuming a constant rate.

1. λ = −ln(0.28)/30.

Answer: λ = 0.0424 per year.

Example 3.3 Find the chance of at least one such earthquake in a single year.

1. 1 − e^(−0.0424).

Answer: 4.15% per year.

Practice problems

Work each on paper first, then check.

1. Chance of one in the next 10 years at that rate?

   1 − e^(−0.424) = 34.6%.

2. Yearly rate for the Los Angeles region's 60% in 30 years?

   −ln(0.4)/30 = 0.0305 per year.

3. Why is P(fault | flag) so low?

The base rate (2%) is small.

4. Is P(A | B) equal to P(B | A)?

   Generally no.

Common misconceptions

Myth: P(A | B) = P(B | A).
Actually: Bayes' theorem relates them through base rates.

Myth: A 95% accurate test is right 95% of the time it fires.
Actually: Not when the condition is rare.

Myth: A 30-year probability divides evenly by 30.
Actually: Use the rate formula.

Flashcards

Conditional

$P(A \cap B)/P(B)$.

Total probability

Sum over branches.

Bayes

$P(A|B)P(B)/P(A)$.

Base rate

Prior frequency.

Rate

$\lambda = -\ln(1 - P)/t$.

Tree diagram

Branches multiply, paths add.

Self-check quiz

1. Bayes reverses...

a condition

a sum

a mean

a sample

a condition. P(B|A) from P(A|B).

2. Low base rates cause...

false alarms

more accuracy

nothing

bias in s

false alarms. Few true positives.

3. P(A|B) is...

$P(A \cap B)/P(B)$

P(A)P(B)

P(A) + P(B)

P(B)/P(A)

$P(A \cap B)/P(B)$. Definition.

4. Bay Area 30-year M ≥ 6.7 chance...

72%

7%

99%

20%

72%. USGS.

Try this quiz again

How this helps save the planet

The USGS, using the Uniform California Earthquake Rupture Forecast 3, gives 30-year probabilities of 72% for a magnitude 6.7 earthquake in the San Francisco Bay region and 60% in the Los Angeles region. It also notes that around the world, the chance an earthquake is followed by a larger one nearby within a week is about 5%.

Turning those 30-year numbers into a yearly rate of about 4% is how engineers and emergency planners compare earthquake risk with other hazards.

On Saving Planets: Solutions · Earth Observation

Sources: USGS FAQ: What is the probability that an earthquake will occur in the Los Angeles area? In the San Francisco Bay area? (https://www.usgs.gov/faqs/what-probability-earthquake-will-occur-los-angeles-area-san-francisco-bay-area)

Free learning resources

- MIT OpenCourseWare 6.041 Probabilistic Systems Analysis and Applied Probability (Fall 2010) (https://ocw.mit.edu/courses/6-041-probabilistic-systems-analysis-and-applied-probability-fall-2010/): Free MIT course.
- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.

Watch the lecture

Lecture 2: Conditioning and Bayes' Rule2 (https://ocw.mit.edu/courses/6-041sc-probabilistic-systems-analysis-and-applied-probability-fall-2013/resources/lecture-2-video-2/) (in English)

MIT 6.041SC Probabilistic Systems Analysis and Applied Probability (Fall 2013), MIT OpenCourseWare. License: CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/) · Terms of use (https://ocw.mit.edu/pages/privacy-and-terms-of-use/). Linked for free study only; Saving Planets is not affiliated with or endorsed by this university.

☐ **I finished this unit**

Discrete Random Variables: Binomial and Poisson

Random variables, probability mass functions, expected value and variance, the Bernoulli and binomial distributions, the Poisson distribution for rare events, and the geometric distribution.

Learning objectives

By the end of this unit you can:

- Compute expected value and variance.
- Apply the binomial distribution.
- Apply the Poisson distribution to rare events.
- Choose the right discrete model.

Key concepts

- A random variable assigns a number to each outcome. Its expected value E[X] = Σx p(x) is the long-run average.
- Var(X) = E[(X − μ)²]; the standard deviation is its square root.
- Binomial: the number of successes in n independent trials with probability p. P(X = k) = C(n, k)p^k(1 − p)$^{n-k}$, mean np, variance np(1 − p).
- Poisson: counts of rare events at rate λ per interval. P(X = k) = e^(−λ)$λ^k$/k!, mean and variance both λ.
- Geometric: the number of trials until the first success, mean 1/p, which is where “recurrence interval” comes from.

Key equations

$$E[X] = \sum x\, p(x)$$

E[X]= x p(x)

Expected value

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

P(X=k)= nk p^k(1-p)^n-k

Binomial

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

P(X=k)=^-^kk!

Poisson

$$E[T] = 1/p$$

E[T]=1/p

Geometric mean wait

Worked examples, step by step

Example 4.1 NOAA sizes its hurricane forecast cone so two-thirds of historical track errors fall inside. Over 9 independent forecasts, find the expected number inside and P(all 9 inside).

1. E = 9 × 2/3.
2. P = $(2/3)^9$.

Answer: E = 6; P(all 9) = 2.60%.

Example 4.2 Find P(exactly 6 of 9 inside).

1. $C(9, 6)(2/3)^6(1/3)^3 = 84 \times 0.0878 \times 0.0370$.

Answer: 27.3%.

Example 4.3 Model 100-year floods over 100 years as Poisson with λ = 1. Find P(none) and P(2 or more).

1. $P(0) = e^{-1}$.
2. $P(\geq 2) = 1 - 2e^{-1}$.

Answer: P(0) = 36.8%; P(≥2) = 26.4%.

Practice problems

Work each on paper first, then check.

1. Standard deviation of the number inside the cone, n = 9?

   √(9 × 2/3 × 1/3) = 1.414.

2. Mean wait for a 1%-per-year event (geometric)?

1/0.01 = 100 years.

3. Poisson P(at least one) with $\lambda = 1$?

   $1 - e^{-1} = 63.2\%$, close to the binomial 63.4%.

4. When does Poisson approximate binomial well?

   Large n, small p.

Common misconceptions

Myth: Expected value is the most likely value.
Actually: It is the long-run average.

Myth: Cone = where the storm's impacts stay.
Actually: It shows only the probable track of the center.

Myth: Poisson needs a fixed number of trials.
Actually: It counts events in an interval.

Flashcards

Random variable

A number per outcome.

E[X]

Long-run average.

Binomial

Successes in n trials.

Poisson

Rare-event counts.

Geometric

Trials to first success.

Variance

$E[(X - \mu)^2]$.

Self-check quiz

1. Binomial mean...

np

n/p

p

n

np. Sum of Bernoullis.

2. Poisson mean and variance...

both λ

λ and λ^2

1 and λ

0

both λ. Property.

3. NHC cone circles contain...

two-thirds of errors

all errors

half

95%

two-thirds of errors. NOAA definition.

4. Geometric mean wait...

1/p

p

np

p^2

1/p. Recurrence interval.

Try this quiz again

How this helps save the planet

NOAA's National Hurricane Center builds its track forecast cone from circles sized so that two-thirds of historical official forecast errors over a 5-year sample fall inside. For 2026 in the Atlantic, based on 2021–2025 errors, the circle radius grows from 25 nautical miles at 12 hours to 95 at 72 hours and 200 at 120 hours.

So the center of a storm will be outside the cone about one time in three, a binomial fact worth knowing before deciding the cone means "safe" or "not safe."

On Saving Planets: GOES-19 · Oceans

Sources: NOAA National Hurricane Center: Definition of the NHC Track Forecast Cone (https://www.nhc.noaa.gov/aboutcone.shtml)

Free learning resources

- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.
- PhET: Plinko Probability (https://phet.colorado.edu/en/simulations/plinko-probability): Binomial interactively.

Watch the lecture

Lecture 5: Discrete Random Variables; Probability Mass Functions; Expectations Video (https://ocw.mit.edu/courses/6-041sc-probabilistic-systems-analysis-and-applied-probability-fall-2013/resources/lecture-5-video/) (in English)

MIT 6.041SC Probabilistic Systems Analysis and Applied Probability (Fall 2013), MIT OpenCourseWare. License: CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/) · Terms of use (https://ocw.mit.edu/pages/privacy-and-terms-of-use/). Linked for free study only; Saving Planets is not affiliated with or endorsed by this university.

☐ I finished this unit

Continuous Distributions: Normal, Exponential and Extremes

Probability density and cumulative distribution functions, the uniform, normal and exponential distributions, standardizing with z, the memoryless property, and return periods for extremes.

Learning objectives

By the end of this unit you can:

- Use a PDF and CDF.
- Find normal probabilities with z-scores.
- Apply the exponential distribution to waiting times.
- Relate return period and annual exceedance probability.

Key concepts

- For a continuous variable, probabilities are areas under the density f(x); the CDF F(x) = P(X ≤ x).
- The normal distribution is bell-shaped with mean μ and standard deviation σ. About 68%, 95% and 99.7% of values fall within 1, 2 and 3σ.
- Standardize with z = (x − μ)/σ and read Φ(z) from a table or NORM.S.DIST.
- Waiting times between events at constant rate λ are exponential: P(T > t) = e^(−λt), mean 1/λ. The process has no memory: past waiting does not change future odds.
- Return period T and annual exceedance probability are reciprocals: a 0.2 AEP event has a 5-year return period.

Key equations

$$P(a < X < b) = \int_a^b f(x)\,dx$$

`P(a<X<b)=_a^b f(x) dx`

Probability as area

$$z = \frac{x - \mu}{\sigma}$$

`z=(x-μ)/(σ)`

Standardizing

$$P(T > t) = e^{-\lambda t}$$

`P(T>t)=e^- t`

Exponential survival

$$T = \frac{1}{AEP}$$

T=1/AEP

Return period

Worked examples, step by step

Example 5.1 Find P(Z < 1) and the share within ±1σ of a normal.

1. Φ(1) = 0.8413.
2. Φ(1) − Φ(−1).

Answer: 0.8413; 68.3% within ±1σ.

Example 5.2 With the Bay Area rate λ = 0.0424 per year from Unit 3, find the mean and median wait for an M ≥ 6.7 earthquake.

1. Mean = 1/λ.
2. Median = ln 2/λ.

Answer: Mean 23.6 years; median 16.3 years.

Example 5.3 What return period matches a 0.2 annual exceedance probability, and a 0.002 one?

1. T = 1/AEP.

Answer: 5 years and 500 years, matching the USGS table.

Practice problems

Work each on paper first, then check.

1. P(Z < −1.5)?

   0.0668.

2. Share within ±2σ?

   95.4%.

3. Φ(1.96)?

0.975.

4. A quake has not happened for 20 years. Does the exponential model raise the odds next year?

   No: it is memoryless (real faults may differ).

Common misconceptions

Myth: P(X = x) > 0 for continuous X.

Actually: Single points have probability 0.

Myth: All data are normal.

Actually: Check with a plot; extremes are often skewed.

Myth: Memoryless means events are scheduled.

Actually: It means the opposite.

Flashcards

PDF	**CDF**
Density; area = probability.	$P(X \le x)$.
Normal	**z**
Bell curve, μ and σ.	$(x - \mu)/\sigma$.
Exponential	**Return period**
Memoryless waiting time.	1/AEP.

Self-check quiz

1. Within ±2σ of a normal...

about 95%

50%

68%

99.7%

about 95%. Empirical rule.

2. Exponential mean...

1/λ

λ

λ^2

ln λ

1/λ. Property.

3. Return period of 1% AEP...

100 years

1 year

10 years

1000 years

100 years. Reciprocal.

4. Φ(1) is...

0.8413

0.5

0.68

1

0.8413. Table.

Try this quiz again

How this helps save the planet

The USGS says recurrence intervals come from frequency analysis of historical data, that ten or more years of data are needed, and that an analysis based on 30 years of record earns more confidence than one based on 10. It defines a 1,000-year flood as one with a 0.1% chance in any given year.

Because flood levels are estimated from past data, the USGS notes the 100-year flood level changes as new floods are recorded or as dams and development alter a basin, so the distribution behind a flood map is an estimate, not a fixed fact.

On Saving Planets: Oceans · Calculus II

Sources: USGS Water Science School: The 100-Year Flood (https://www.usgs.gov/special-topics/water-science-school/science/100-year-flood)

Free learning resources

- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.
- NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/): Free handbook.

Watch the lecture

Lecture 10: Continuous Bayes' Rule Derived Distributions (https://ocw.mit.edu/courses/6-041sc-probabilistic-systems-analysis-and-applied-probability-fall-2013/resources/lecture-10-video-1/) (in English)

MIT 6.041SC Probabilistic Systems Analysis and Applied Probability (Fall 2013), MIT OpenCourseWare. License: CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/) · Terms of use (https://ocw.mit.edu/pages/privacy-and-terms-of-use/). Linked for free study only; Saving Planets is not affiliated with or endorsed by this university.

☐ **I finished this unit**

Sampling, Simulation and the Central Limit Theorem

Populations and samples, bias and random sampling, the sampling distribution of the mean, standard error, the central limit theorem, the law of large numbers, and Monte Carlo simulation.

Learning objectives

By the end of this unit you can:

- Distinguish population and sample quantities.
- Compute a standard error.
- State and apply the central limit theorem.
- Estimate a probability by Monte Carlo simulation.

Key concepts

- A statistic computed from a sample varies from sample to sample; its distribution is the sampling distribution.
- The standard error of the mean is σ/√n: four times the data halves it.
- Central limit theorem: for large n, the sample mean is approximately normal whatever the shape of the data.
- Law of large numbers: sample averages and frequencies settle toward their true values as n grows.
- Monte Carlo simulation repeats a random experiment many times on a computer; its error also falls like 1/√n.

Key equations

$$SE = \frac{\sigma}{\sqrt{n}}$$

`SE=(σ)/( n)`

Standard error

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Central limit theorem

$$\hat{p} = \frac{\#\text{hits}}{N}$$

`p=\#hitsN`

Monte Carlo estimate

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{N}}$$

`SE_ p=(p(1-p))/(N)`

Proportion standard error

Worked examples, step by step

Example 6.1 Measurements have σ = 10 (sample value). Find the standard error of the mean of 25.

1. 10/√25.

Answer: SE = 2.

Example 6.2 A fair die has mean 3.5 and σ = 1.708. Find the standard error of the average of 25 rolls.

1. 1.708/5.

Answer: SE = 0.342.

Example 6.3 How many Monte Carlo runs estimate a probability near 0.5 to ±0.05 (95%)?

1. n = 1.96² × 0.25/0.05².

Answer: About 384 runs.

Practice problems

Work each on paper first, then check.

1. SE of a die average over 100 rolls?

   0.171.

2. Double the sample size. SE?

   Divided by √2.

3. Does the CLT make the data normal?

   No, only the sample mean.

4. 95% margin for the σ = 10, n = 25 mean?

   1.96 × 2 = 3.92.

Common misconceptions

Myth: A big sample fixes bias.

Actually: It reduces random error only.

Myth: n = 30 always suffices.

Actually: Skewed data may need more.

Myth: Simulation is exact.

Actually: It has its own sampling error.

Flashcards

Population

The whole group.

Sample

The part measured.

Standard error

σ/√n.

CLT

Means become normal.

Law of large numbers

Averages settle.

Monte Carlo

Repeat random trials.

Self-check quiz

1. SE of the mean...

σ/√n

σ/n

σ

nσ

σ/√n. Formula.

2. CLT is about…

the sample mean

every data point

the median only

bias

the sample mean. Sampling distribution.

3. Monte Carlo error falls like…

1/√n

1/n

n

constant

1/√n. SE.

4. Random sampling prevents…

selection bias

variance

all error

nothing

selection bias. Design.

Try this quiz again

How this helps save the planet

NASA's GISS Surface Temperature Analysis (GISTEMP v4) estimates global surface temperature change from NOAA station and ocean data and is updated about the 10th of every month. NASA also publishes a 2024 uncertainty ensemble for it.

An uncertainty ensemble is this unit's Monte Carlo idea at planetary scale: many plausible versions of the record, whose spread shows how sure we are of each year's value.

On Saving Planets: Earth Observation · Solutions

Sources: NASA GISS: GISS Surface Temperature Analysis (GISTEMP v4) (https://data.giss.nasa.gov/gistemp/)

Free learning resources

- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.
- PhET: Plinko Probability (https://phet.colorado.edu/en/simulations/plinko-probability): Watch sums become normal.

Watch the lecture

Lecture 20: Central Limit Theorem (https://ocw.mit.edu/courses/6-041sc-probabilistic-systems-analysis-and-applied-probability-fall-2013/resources/lecture-20-video-2/) (in English)

MIT 6.041SC Probabilistic Systems Analysis and Applied Probability (Fall 2013), MIT OpenCourseWare. License: CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/) · Terms of use (https://ocw.mit.edu/pages/privacy-and-terms-of-use/). Linked for free study only; Saving Planets is not affiliated with or endorsed by this university.

☐ **I finished this unit**

Confidence Intervals

Point and interval estimates, z and t intervals for a mean, intervals for a proportion, margin of error and sample size, and what a confidence level does and does not mean.

Learning objectives

By the end of this unit you can:

- Build a t interval for a mean.
- Build an interval for a proportion.
- Choose a sample size for a target margin.
- Interpret a confidence level correctly.

Key concepts

- A confidence interval is estimate ± critical value × standard error.
- With σ unknown and small n, use Student's t with n − 1 degrees of freedom; t has fatter tails than z.
- For a proportion, p̂ ± z√(p̂(1 − p̂)/n).
- A 95% level means 95% of intervals built this way would contain the true value; it is not a 95% chance for one computed interval.
- Halving the margin needs four times the sample size.

Key equations

$$\bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

t interval

$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Proportion interval

$$n = \frac{z^2 p(1-p)}{E^2}$$

`n=(z^2p(1-p))/(E^2)`

Sample size

$$E = z\frac{\sigma}{\sqrt{n}}$$

$$E = z(\sigma)/(\sqrt{n})$$

Margin of error

Worked examples, step by step

Example 7.1 Build a 95% interval for the mean yearly CO_2 increase from Unit 1 ($\bar{x}$ = 2.628, s = 0.564, n = 5, t = 2.776).

1. Margin = 2.776 × 0.564/√5.

Answer: 2.628 ± 0.700: from 1.93 to 3.33 ppm/yr.

Example 7.2 NOAA says 22 of the 31 Atlantic seasons since 1995 were above normal. Give a 95% interval for that proportion.

1. $\hat{p}$ = 0.710; SE = √(0.710 × 0.290/31).

Answer: 0.710 ± 0.160: from 0.55 to 0.87.

Example 7.3 How large a sample gives a ±0.05 margin for a proportion near 0.5?

1. $1.96^2 \times 0.25/0.05^2$.

Answer: About 385 (384.2, rounded up).

Practice problems

Work each on paper first, then check.

1. Why t instead of z for n = 5?

   σ is estimated from few points.

2. Effect of 99% instead of 95%?

   Wider interval.

3. Does NOAA's 70% range for named storms mean 70% of storms?

   No: the season total falls in the range in 7 of 10 similar seasons.

4. Margin for the proportion with n = 124?

About half (0.080).

Common misconceptions

Myth: 95% chance the true value is in this interval.

Actually: The procedure captures it 95% of the time.

Myth: Wider is worse.

Actually: Wider is honest when data are few.

Myth: Intervals fix bias.

Actually: They describe random error only.

Flashcards

Confidence interval

Estimate ± margin.

t distribution

Fatter tails, $n - 1$ df.

Margin of error

Critical value × SE.

Confidence level

Long-run capture rate.

Proportion interval

$\hat{p} \pm z\sqrt{(\hat{p}(1-\hat{p})/n}$.

Sample size

$z^2p(1-p)/E^2$.

Self-check quiz

1. Small n, σ unknown: use...

t

z

Poisson

none

t. Student.

2. Halve the margin…

4× the data

2× the data

same data

½ the data

4× the data. √n.

3. NOAA likely ranges hold in…

7 of 10 similar seasons

every season

half

95%

7 of 10 similar seasons. CPC.

4. CI = estimate ± …

critical value × SE

mean

range

σ^2

critical value × SE. Structure.

Try this quiz again

How this helps save the planet

NOAA's Climate Prediction Center outlook for 2026, updated August 6, gives a 75% chance of a below-normal Atlantic season, 20% near-normal and 5% above-normal, and a 70% probability for

each range, including 7–13 named storms and 2–6 hurricanes.

NOAA explains that the season is expected to fall within these ranges in 7 out of 10 seasons with similar conditions, which is exactly how a 70% interval should be read. It also stresses that it only takes one hurricane to cause a disaster.

On Saving Planets: GOES-19 · Oceans

Sources: NOAA Climate Prediction Center: Atlantic Hurricane Outlook (updated Aug. 6, 2026) (https://cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml)

Free learning resources

- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.
- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.

☐ **I finished this unit**

Hypothesis Tests and p-Values

Null and alternative hypotheses, test statistics, p-values, one- and two-sided tests, t tests and proportion tests, Type I and Type II errors, power, and statistical versus practical significance.

Learning objectives

By the end of this unit you can:

- Set up null and alternative hypotheses.
- Run a one-sample t test and a proportion test.
- Interpret a p-value.
- Explain Type I and Type II errors and power.

Key concepts

- A test asks whether data are surprising if the null hypothesis H_0 were true.
- The p-value is the probability, under H_0, of a result at least as extreme as the one seen. Small p is evidence against H_0; it is not the probability H_0 is true.
- A Type I error rejects a true H_0 (rate α); a Type II error misses a false H_0 (rate β). Power = 1 − β.
- More data raise power; a tiny, unimportant effect can be “significant” with enough data.
- Decide α and the test before looking at the data.

Key equations

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

```
t=( x-μ_0)/(s/ n)
```

t statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

```
z= p-p_0_0(1-p_0)/n
```

Proportion z

$$\text{power} = 1 - \beta$$

```
power=1-β
```

Power

$$p = P(\text{as extreme} \mid H_0)$$

p=P(as extreme H_0)

p-value

Worked examples, step by step

Example 8.1 Test H_0: μ = 2 ppm/yr against μ > 2 for the five CO_2 increases ($\bar{x}$ = 2.628, s = 0.564).

1. t = (2.628 − 2)/(0.564/√5).

Answer: t = 2.49 with 4 df, above the one-sided 5% cutoff of 2.132: reject H_0.

Example 8.2 If each Atlantic season had a 1/3 chance of being above normal, how many of 31 would we expect?

1. 31/3.

Answer: About 10.3; NOAA reports 22.

Example 8.3 Find the exact binomial P(22 or more of 31) under p = 1/3.

1. Sum $C(31, k)(1/3)^k(2/3)^{31-k}$ for k ≥ 22.

Answer: p ≈ 2.06 × 10^{-5}: strong evidence the high-activity era is real.

Practice problems

Work each on paper first, then check.

1. Proportion z for 22 of 31 vs p_0 = 1/3?

   z = 4.45.

2. Is p = 0.03 the chance H_0 is true?

   No.

3. Type I error in a flood-warning test?

   A false alarm.

4. How do you raise power?

   More data, larger effect, or larger α.

Common misconceptions

Myth: p > 0.05 proves H_0.
Actually: It only fails to reject.

Myth: Significant means important.
Actually: Check the effect size.

Myth: Test many things, report the winner.
Actually: That inflates false positives.

Flashcards

H_0

Null hypothesis.

p-value

Surprise under H_0.

α

Type I rate.

β

Type II rate.

Power

1 − β.

Effect size

How big, not just whether.

Self-check quiz

1. p-value is...

P(data this extreme | H_0)

$P(H_0)$

α

power

P(data this extreme | H_0). Definition.

2. Type I error...

reject true H_0

miss false H_0

bad data

none

reject true H_0. False alarm.

3. Power is...

$1 - \beta$

α

p

$1 - \alpha$

$1 - \beta$. Detect a real effect.

4. More data...

raise power

lower power

no change

raise α

raise power. Smaller SE.

Try this quiz again

How this helps save the planet

NOAA's Climate Prediction Center reports that since the current Atlantic high-activity era began in 1995, 22 of 31 seasons (71%) were above normal, 5 (16%) near normal and 4 (13%) below normal.

NOAA also notes that every hurricane season coinciding with a strong El Niño since 1950 has had below-normal activity. Testing claims like these against chance is how forecasters decide which signals belong in an outlook.

On Saving Planets: GOES-19 · Solutions

Sources: NOAA Climate Prediction Center: Atlantic Hurricane Outlook (updated Aug. 6, 2026) (https://cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml)

Free learning resources

- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures.
- NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/): Free handbook.

- [ ] **I finished this unit**

Correlation and Linear Regression

Scatter plots, correlation r, least-squares regression, slope interpretation, R^2, residuals, prediction versus extrapolation, and correlation versus causation.

Learning objectives

By the end of this unit you can:

- Compute and interpret r and R^2.
- Fit and interpret a least-squares line.
- Use residual plots to check a model.
- Separate correlation from causation.

Key concepts

- Correlation r measures linear association from −1 to 1; it has no units.
- The least-squares line $\hat{y} = b_0 + b_1 x$ has slope $b_1 = Sxy/Sxx$ and passes through $(\bar{x}, \bar{y})$.
- $R^2 = r^2$ is the share of variation in y explained by the line.
- Residual plots reveal curvature or changing spread that r hides.
- Correlation is not causation; physical mechanisms and designed studies are needed for that. Extrapolation beyond the data is risky.

Key equations

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

```
b_1=_xyS_xx
```

Slope

$$b_0 = \bar{y} - b_1 \bar{x}$$

```
b_0= y-b_1 x
```

Intercept

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

```
r=_xy_xxS_yy
```

Correlation

$$R^2 = r^2$$

R^2=r^2

Explained share

Worked examples, step by step

Example 9.1 Fit a line to NOAA's Mauna Loa decade means 1960–2020 (316.91, 325.68, 338.76, 354.45, 369.71, 390.10, 414.21 ppm), with x in years since 1960.

1. $\bar{x} = 30$, $\bar{y} = 358.55$.
2. $b_1 = Sxy/Sxx$.

Answer: $\hat{y} = 310.15 + 1.613x$ ppm.

Example 9.2 Find r and R^2.

1. $r = Sxy/\sqrt{(SxxSyy)}$.

Answer: $r = 0.990$; $R^2 = 0.979$.

Example 9.3 Do the residuals support a straight line?

1. Residuals are positive at both ends and negative in the middle.

Answer: No: the curve bends upward, so growth is accelerating despite $R^2 = 0.979$.

Practice problems

Work each on paper first, then check.

1. Units of the slope?

   ppm per year.

2. Line prediction for 2025 (x = 65)?

   415.0 ppm vs NOAA's 427.35.

3. Does r = 0.99 prove the line is right?

No; check the residuals.

4. Does a time trend alone prove its cause?

   No; the cause comes from physics and other evidence.

Common misconceptions

Myth: High r means a good model.

Actually: Curvature can hide behind high r.

Myth: r has units.

Actually: It is unitless.

Myth: Extrapolating is safe if R^2 is high.

Actually: Not outside the data range.

Flashcards

Correlation

−1 to 1, unitless.

Slope

Sxy/Sxx.

R^2

Share explained.

Residual

$y - \hat{y}$.

Extrapolation

Predicting outside the data.

Causation

Needs mechanism or design.

Self-check quiz

1. R^2 is...

share of variation explained

slope

p-value

residual

share of variation explained. Definition.

2. The line passes through...

$(\bar{x}, \bar{y})$

origin

$(0, \bar{y})$

max point

$(\bar{x}, \bar{y})$. Property.

3. Curved residuals mean...

wrong model form

perfect fit

no data

bias in x

wrong model form. Pattern.

4. Correlation shows...

association

causation

nothing

error

association. Caution.

Try this quiz again

How this helps save the planet

NOAA notes that the Mauna Loa data are taken at an altitude of 3400 m and may differ from the global average, and that during the 2022 Mauna Loa eruption measurements moved to Maunakea, about 21 miles north, until July 2023.

Knowing where and how data were collected is part of regression too: a fit is only as good as the record behind it, and NOAA documents its gaps and site changes openly.

On Saving Planets: Solutions · Linear Algebra

Sources: NOAA Global Monitoring Laboratory: Trends in atmospheric CO_2 (Mauna Loa) (https://gml.noaa.gov/ccgg/trends/)

Free learning resources

- PhET: Least-Squares Regression (https://phet.colorado.edu/en/simulations/least-squares-regression): Fit lines interactively.
- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.

☐ **I finished this unit**

Reliability and Risk

Reliability functions, failure rates and MTBF, series and parallel systems, redundancy, the exponential and Weibull life models, and combining probability with consequences into risk.

Learning objectives

By the end of this unit you can:

- Compute reliability of series and parallel systems.
- Use the exponential life model with MTBF.
- Explain the bathtub curve and Weibull shapes.
- Combine likelihood and consequence into risk.

Key concepts

- Reliability R(t) is the probability a part survives to time t. With constant failure rate λ, R(t) = e^(−λt) and MTBF = 1/λ.
- Series systems need every part: R = $R_1R_2…R_n$, always worse than the weakest part.
- Parallel (redundant) systems need only one: R = 1 − (1 − R_1)(1 − R_2)….
- The Weibull model covers early failures (shape < 1), random failures (shape = 1, exponential) and wear-out (shape > 1), the three parts of the bathtub curve.
- Risk combines probability and consequence; redundancy and warnings reduce it, but only if failures are independent.

Key equations

$$R(t) = e^{-\lambda t},\ MTBF = 1/\lambda$$

`R(t)=e^- t,\ MTBF=1/`

Exponential reliability

$$R_s = \prod R_i$$

`R_s= R_i`

Series system

$$R_p = 1 - \prod (1 - R_i)$$

`R_p=1-(1-R_i)`

Parallel system

$$R(t) = e^{-(t/\eta)^{\beta}}$$

R(t)=e^-(t/)^β

Weibull

Worked examples, step by step

Example 10.1 Three parts in series each have R = 0.99 (sample values). Find the system reliability.

1. 0.99^3.

Answer: R = 0.9703.

Example 10.2 Two redundant pumps each have R = 0.9. Find the system reliability.

1. $1 - 0.1^2$.

Answer: R = 0.99.

Example 10.3 A sensor has MTBF 50,000 h. Find its reliability over a 1000-hour mission.

1. e^(−1000/50,000).

Answer: R = 0.980.

Practice problems

Work each on paper first, then check.

1. Three redundant units of 0.9?

   $1 - 0.1^3 = 0.999$.

2. A 0.9 part in series with a redundant pair of 0.9?

   $0.9 \times 0.99 = 0.891$.

3. Reliability over a year (8760 h) at MTBF 100,000 h?

   e^(−0.0876) = 0.916.

4. Two 0.95 units in parallel?

0.9975.

Common misconceptions

Myth: Redundancy always helps.

Actually: Not against common-cause failures.

Myth: MTBF is how long every unit lasts.

Actually: It is an average; many fail sooner.

Myth: A series system is as reliable as its best part.

Actually: It is worse than its weakest.

Flashcards

Reliability

P(survive to t).

MTBF

1/λ.

Series

Multiply R.

Parallel

1 − Π(1 − R).

Weibull

Bathtub shapes.

Risk

Probability × consequence.

Self-check quiz

1. Series reliability...

product

sum

max

min

product. All must work.

2. Exponential MTBF...

1/λ

λ

e^λ

ln λ

1/λ. Mean life.

3. Weibull shape > 1 means...

wear-out

infant mortality

random

none

wear-out. Rising hazard.

4. Redundancy fails against...

common causes

random failures

nothing

small λ

common causes. Shared failure.

Try this quiz again

How this helps save the planet

NOAA's 2026 hurricane outlook reminds residents, businesses and governments that hurricane-related disasters can occur in any season, even one with low activity, and that coastal communities

should prepare every year regardless of the outlook.

That is reliability thinking: a low probability is not zero, so warning systems, backups and plans are designed to work in the rare bad year, not just the average one.

On Saving Planets: GOES-19 · Solutions

Sources: NOAA Climate Prediction Center: Atlantic Hurricane Outlook (updated Aug. 6, 2026) (https://cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml)

Free learning resources

- NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/): Free handbook (reliability chapter).
- MIT OpenCourseWare 6.041 Probabilistic Systems Analysis and Applied Probability (Fall 2010) (https://ocw.mit.edu/courses/6-041-probabilistic-systems-analysis-and-applied-probability-fall-2010/): Free MIT course.

☐ **I finished this unit**

Problem Sets

These practice problems live inside each unit. Work on paper first, then open the answer.

1. Describing Data: Center, Spread and Plots
   4 problems
2. Probability Rules, Counting and Independence
   4 problems
3. Conditional Probability and Bayes' Theorem
   4 problems
4. Discrete Random Variables: Binomial and Poisson
   4 problems
5. Continuous Distributions: Normal, Exponential and Extremes
   4 problems
6. Sampling, Simulation and the Central Limit Theorem
   4 problems
7. Confidence Intervals
   4 problems
8. Hypothesis Tests and p-Values
   4 problems
9. Correlation and Linear Regression
   4 problems
10. Reliability and Risk
    4 problems

The lab (Statistics lab)

The lab is the hands-on part of the course. Here it is free and for practice only, with no credit.

Every lab runs in a free simulation or a spreadsheet, so there are no physical hazards. Take screen breaks, sit with good posture, and keep cables out of walkways.

Safety reference: OSHA: Walking-working surfaces (slip, trip and fall hazards) (https://www.osha.gov/walking-working-surfaces)

Lab A: Center and spread of the CO_2 record

WEEK 1

Goal: Explore mean, median and spread in PhET Center and Variability, then summarize NOAA's recent yearly CO_2 increases.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

```
s=((x- x)^2)/(n-1)
```

Sample standard deviation

Procedure

1. In PhET Center and Variability, drag values and watch the mean and median move.
2. Open NOAA's Mauna Loa annual means and compute the yearly increases for 2021–2025.
3. Compute mean, median, range and s.
4. Make a dot plot and describe the 2024 value.

Data table

Statistic	Value (ppm/yr)
Mean	2.628
Median	2.55
Range	1.41
Standard deviation	0.564

Data from NOAA GML; statistics computed.

Sample calculation

1. Increases: 2.20, 2.12, 2.55, 3.53, 2.74.
2. s = √(1.2735/4) = 0.564.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Any spreadsheet works.

PhET: Center and Variability (https://phet.colorado.edu/en/simulations/center-and-variability) · NOAA GML: Mauna Loa annual mean CO_2 data (https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_annmean_mlo.txt)

Planet link: NOAA's Mauna Loa monthly mean was 426.07 ppm in September 2026. NOAA Global Monitoring Laboratory: Trends in atmospheric CO_2 (Mauna Loa) (https://gml.noaa.gov/ccgg/trends/)

☐ **I completed this lab (wet or virtual)**

Lab B: Plinko and the binomial distribution

WEEK 4

Goal: Drop balls in PhET Plinko Probability and compare the histogram with the binomial model.

$$\mu = np,\ \sigma = \sqrt{np(1-p)}$$

$\mu=np,\ \sigma=(1-p)$

Binomial mean and spread

Procedure

1. Set 12 rows and p = 0.5.
2. Drop 1000 balls and record the mean and standard deviation.
3. Compute the binomial values.
4. Change p to 0.7 and predict the new mean before dropping.

Data table

Quantity	Binomial value
Mean bin	6
Standard deviation	1.732
P(bin 6)	0.2256

Computed values; your simulation should come close.

Sample calculation

1. σ = √(12 × 0.25) = 1.732.
2. P(6) = C(12, 6)/2^{12} = 924/4096 = 0.2256.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Any browser works.

PhET: Plinko Probability (https://phet.colorado.edu/en/simulations/plinko-probability)

Planet link: NOAA sizes its hurricane cone so two-thirds of past track errors fall inside. NOAA National Hurricane Center: Definition of the NHC Track Forecast Cone (https://www.nhc.noaa.gov/aboutcon

e.shtml)

- [ ] **I completed this lab (wet or virtual)**

Lab C: Simulating 100-year floods

WEEK 6

Goal: Use a spreadsheet Monte Carlo to estimate the chance of at least one 1%-per-year flood in 30 years.

$$P = 1 - 0.99^{30}$$

```
P=1-0.99^30
```

Exact answer

Procedure

1. Make 30 columns of =RAND()<0.01 (one per year).
2. In each row, test whether any year flooded.
3. Copy to 1000 rows and average the result.
4. Compare with the exact answer and the expected Monte Carlo error.

Data table

Quantity	Value
Exact probability	0.260
Monte Carlo SE, N = 1000	0.0139
Expected range (±2 SE)	0.232 to 0.288

Computed values; your estimate will vary each recalculation.

Sample calculation

1. SE = √(0.26 × 0.74/1000) = 0.0139.
2. 0.260 ± 0.0278.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Any spreadsheet works.

MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/)

Planet link: The USGS stresses that 100-year floods can happen two years in a row. USGS Water Science School: The 100-Year Flood (https://www.usgs.gov/special-topics/water-science-school/science/100-year-flood)

- [ ] **I completed this lab (wet or virtual)**

Lab D: Dice and the central limit theorem

WEEK 7

Goal: Average simulated (or real) dice rolls and watch the distribution of averages become normal.

$$SE = \frac{\sigma}{\sqrt{n}}$$

SE=(σ)/(n)

Standard error

Procedure

1. Simulate rolls with =RANDBETWEEN(1,6).
2. Average groups of 1, 5, 25 and 100 rolls, 500 times each.
3. Plot histograms and compute their standard deviations.
4. Compare with σ/√n.

Data table

Rolls per average	Predicted SE
1	1.708
25	0.342
100	0.171

Computed values; a die has mean 3.5 and σ = √(35/12).

Sample calculation

1. σ = √(35/12) = 1.708.
2. SE(25) = 1.708/5 = 0.342.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Real dice work too.

PhET: Plinko Probability (https://phet.colorado.edu/en/simulations/plinko-probability)

Planet link: NASA publishes an uncertainty ensemble for its GISTEMP temperature record. NASA GISS: GISS Surface Temperature Analysis (GISTEMP v4) (https://data.giss.nasa.gov/gistemp/)

- [ ] **I completed this lab (wet or virtual)**

Lab E: Regression and residuals

WEEK 12

Goal: Fit a line in PhET Least-Squares Regression, then fit NOAA's Mauna Loa decade means and study the residuals.

$$\hat{y} = b_0 + b_1 x$$

y=b_0+b_1x

Least-squares line

Procedure

1. In PhET, place points and compare your line with the best fit.
2. Enter the 1960–2020 decade means with x = years since 1960.
3. Use LINEST or a trendline to get b_0, b_1 and r.
4. Plot the residuals and describe the pattern.

Data table

Quantity	Value
Slope	1.613 ppm/yr
Intercept	310.15 ppm
r	0.990
Residuals 1960 / 2000 / 2020	+6.76 / −4.97 / +7.27 ppm

Data from NOAA GML; fit computed. The U-shaped residuals show upward curvature.

Sample calculation

1. b_1 = Sxy/Sxx = 1.613.
2. Residual 2020 = 414.21 − (310.15 + 1.613 × 60) = 7.27.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Any spreadsheet works.

PhET: Least-Squares Regression (https://phet.colorado.edu/en/simulations/least-squares-regression) · NOAA GML: Mauna Loa annual mean CO_2 data (https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_annmean_mlo.txt)

Planet link: NOAA notes Mauna Loa samples air at 3400 m, which may differ from the global average. NOAA Global Monitoring Laboratory: Trends in atmospheric CO_2 (Mauna Loa) (https://gml.noaa.gov/ccgg/trends/)

☐ **I completed this lab (wet or virtual)**

Lab F: Reliability of a warning system

WEEK 14

Goal: Build a spreadsheet for series and parallel reliability and test how redundancy changes a sample flood-warning chain.

$$R_s = \prod R_i,\ R_p = 1 - \prod (1 - R_i)$$

```
R_s= R_i,\ R_p=1-(1-R_i)
```

System reliability

Procedure

1. Model a sample chain: sensor, radio link and siren in series.
2. Compute R with each at 0.99, then with a redundant siren pair.
3. Use e^(−t/MTBF) for a 1000 h mission at 50,000 h MTBF.
4. Find which single upgrade helps most.

Data table

Configuration	Reliability
Three parts in series (0.99 each)	0.9703
Two redundant units (0.9 each)	0.99
0.9 part + redundant 0.9 pair	0.891
1000 h at MTBF 50,000 h	0.980

Sample values; computed.

Sample calculation

1. $0.99^3 = 0.9703$.
2. $0.9 \times (1 - 0.1^2) = 0.891$.

> **Safety**
>
> - Computer only.

Home or virtual alternative

- Any spreadsheet works.

NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/)

Planet link: NOAA says disasters can occur even in low-activity hurricane seasons, so prepare every year. NOAA Climate Prediction Center: Atlantic Hurricane Outlook (updated Aug. 6, 2026) (https://cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml)

☐ **I completed this lab (wet or virtual)**

Exam prep practice sets

Time yourself: about 10 minutes per problem. Then compare with the full solutions.

Midterm practice set (Units 1–5)

1. Find the mean, median and sample standard deviation of 4, 8, 6, 10, 12.

    1. Mean = 40/5.
    2. s = √(40/4).

    Answer: Mean 8, median 8, s = 3.16.

2. Find the chance of at least one 25-year flood (4% a year) in 10 years.

    1. $1 - 0.96^{10}$.

    Answer: 33.5%.

3. A test detects a condition 90% of the time and false-alarms 10%; the base rate is 1%. Find P(condition | positive).

    1. 0.009/(0.009 + 0.099).

    Answer: 8.33%.

4. Find P(X = 0) for a binomial with n = 5, p = 0.2.

    1. 0.8^5.

    Answer: 0.328.

5. Find P(X = 0) for a Poisson with λ = 2.

    1. e^{-2}.

    Answer: 0.135.

Final practice set (all units)

1. Find P(Z > 1.96).

1. 1 − 0.975.

Answer: 0.025.

2. Find the standard error for σ = 6, n = 36.

    1. 6/6.

Answer: 1.

3. Build a 95% interval for $\bar{x}$ = 50, s = 4, n = 16 (t = 2.131).

    1. 2.131 × 4/4.

Answer: 50 ± 2.13.

4. Test H_0: μ = 50 with $\bar{x}$ = 52, s = 4, n = 16.

    1. t = 2/(4/4).

Answer: t = 2.0, just below 2.131: do not reject at 5% (two-sided).

5. Find the reliability of four 0.95 parts in series.

    1. 0.95^4.

Answer: 0.815.

6. A part has MTBF 2000 h. Find P(it survives 500 h).

    1. e^(−0.25).

Answer: 0.779.

Formula sheet

$$\bar{x} = \sum x/n,\ s^2 = \sum (x-\bar{x})^2/(n-1)$$

$$P(\geq 1) = 1-(1-p)^n,\ P(A \cup B) = P(A) + P(B$$

$$P(B|A) = P(A|B)P(B)/P(A),\ \lambda = -\ln(1-P)/$$

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k},\ P(X=k) = e^{-\lambda}\lambda$$

$$z = (x-\mu)/\sigma,\ P(T>t) = e^{-\lambda t},\ T = 1/AEP$$

$$SE = \sigma/\sqrt{n}$$

$$\bar{x} \pm t\, s/\sqrt{n},\ \hat{p} \pm z\sqrt{\hat{p}(1-\hat{p})/n}$$

$$t = (\bar{x}-\mu_0)/(s/\sqrt{n})$$

$$b_1 = S_{xy}/S_{xx},\ r = S_{xy}/\sqrt{S_{xx}S_{yy}}$$

$$R_s = \prod R_i,\ R_p = 1 - \prod(1 - R_i),\ MTBF = 1/$$

Constants

Constant	Value
Standard normal (textbook values)	Φ(1) = 0.8413, Φ(1.96) = 0.975, Φ(2) − Φ(−2) = 0.9545
Student t, 4 degrees of freedom (textbook value)	$t_{0.975}$ = 2.776
e	2.718281828
68–95–99.7 rule	share of a normal within 1, 2 and 3 standard deviations
1 ppm	one part per million by mole fraction
Annual exceedance probability	1/(recurrence interval)

Source: NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/)

Free texts and simulations

Use these free lectures, textbooks, handbooks and simulations alongside the course.

- MIT OpenCourseWare 18.05 Introduction to Probability and Statistics (Spring 2022) (https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/): Free MIT lectures, readings and problem sets.
- MIT OpenCourseWare 6.041 Probabilistic Systems Analysis and Applied Probability (Fall 2010) (https://ocw.mit.edu/courses/6-041-probabilistic-systems-analysis-and-applied-probability-fall-2010/): Free MIT applied probability course.
- OpenStax Introductory Statistics 2e (https://openstax.org/books/introductory-statistics-2e/pages/1-introduction): Free textbook.
- NIST/SEMATECH e-Handbook of Statistical Methods (https://www.itl.nist.gov/div898/handbook/): Free engineering statistics handbook.
- PhET: Plinko Probability (https://phet.colorado.edu/en/simulations/plinko-probability): Binomial and normal distributions interactively.
- PhET: Center and Variability (https://phet.colorado.edu/en/simulations/center-and-variability): Mean, median and spread interactively.

Glossary

Annual exceedance probability
Chance an event is equalled or exceeded in a year.

Bayes' theorem
Reverses a conditional probability.

Binomial distribution
Successes in n independent trials.

Central limit theorem
Sample means become normal.

Conditional probability
Probability given that another event occurred.

Confidence interval
Range from a procedure that captures the truth at a set rate.

Correlation
Strength of linear association.

Expected value
Long-run average.

Exponential distribution
Memoryless waiting time.

Independence
One event does not change another's probability.

MTBF
Mean time between failures.

Median
Middle value.

Monte Carlo
Estimating by repeated random simulation.

Normal distribution
Bell-shaped distribution.

Poisson distribution
Counts of rare events.

Recurrence interval
Average years between events of a given size.

Reliability
Probability of surviving to a given time.

Standard error
Standard deviation of a statistic.

Type I error
Rejecting a true null hypothesis.

p-value
Chance of data this extreme if H_0 holds.

Self-assessment rubric

No grade or credit is given. Use this to judge honestly where you are in each unit.

Level	What it looks like
4 · Mastery	Names the random variable, states assumptions, picks the right model, reports uncertainty and interprets results in plain words.
3 · Proficient	Solves the practice problems with occasional slips in assumptions or units.
2 · Developing	Follows worked examples but misreads p-values or confidence levels.
1 · Beginning	Plugs numbers into formulas without checking assumptions; needs to revisit the unit.

Suggested weighting if you want a score

Part	Weight
Practice problems	20%
Labs	25%
Midterm practice set	25%
Final practice set	30%

Aim for level 3 or better in every unit before moving to the next course.

Related courses

- Numerical Methods MATH 340
- Propulsion AERO 410
- Aerodynamics I AERO 310
- All courses