



Full guide + ask Sia
about LGT5101

0 • Exam Map

SYLLABUS · LEC 0

15 25 10 FINAL 50%

Assignment 1 15% · Assignment 2 25% · participation 10% · final 50%.

"Finally, there is an open-book final exam (50%)."

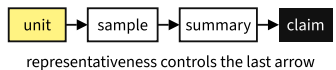
(LGT5101 Statistics for Management syllabus, Semester 1 2026/27) · "Final Exam 50% Lectures 1-13 TBD (Open-book)" (LGT5101 Lecture 0 course overview, Semester 1 2026/27)

The captured materials do not state duration, sections, question count or a more detailed materials rule. **Show formula → substitution → result → interpretation:** Assignment 1 explicitly asks for formulas and calculation steps wherever possible.

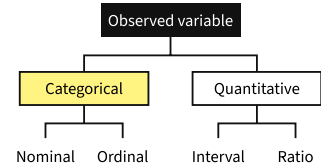
1 • Start With Scope

C1

★ **Five-line opening:** experimental unit (one row) · target population · observed sample · variables · intended inference. Then record source, filters and missing values. A complete branch roster is a population for that branch but a sample for a company-wide claim.



Source scope logic, redrawn: a correct calculation can still overreach its sample.



Source decision tree, redrawn: meaning—not digits—sets the scale.

SCALE	PERMITTED MEANING	SAFE SUMMARIES
Nominal	labels only	count, mode
Ordinal	order; gaps unproved	rank, median
Interval	equal differences	mean, SD
Ratio	real zero; ratios	all arithmetic

OPENING AUDIT

One row = one level. Do not turn a delivery row into a customer or depot claim. Description organises observed units; inference extends beyond them and needs a sampling argument. Record any grouping because it loses detail.

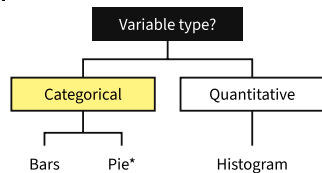
CHECK → If "code 4 is twice code 2" is meaningless, the digits are labels. Generalisation beyond observed rows needs representativeness and stable conditions.

2 • Frequencies & Displays

C2

FREQUENCY LEDGER

f_i = count in class i
relative frequency $r_i = f_i / n$
checks: $\sum f_i = \text{usable } n \cdot \sum r_i = 1$ (rounding aside)



Source chart map, redrawn. *Pie only for a small partition of one whole.

NEED	USE	GEOMETRY / CHECK
category mix	bar	separated; common baseline
small whole	pie	slices sum to 100%
distribution	histogram	touching intervals
unequal bins	density	area = frequency

Histogram boundary rule: intervals are adjacent, non-overlapping and collectively cover the values; state which endpoint is included. The lecture guidance normally suggests **5–15 classes**, depending on the number of values. Compare reasonable widths: broad bins can hide a cluster; narrow bins can manufacture noise.

BUILD THE DISPLAY IN ORDER

- Name the usable denominator after filters and missing values.
- List mutually exclusive, collectively exhaustive classes.
- Count f_i ; divide by n when relative frequency is required.
- Reconcile totals before drawing.
- Label title, axes, units and count/relative-frequency scale.

MISLEADING MOVE	REPAIR
truncated bar axis	use an honest common baseline
unordered ordinal bands	restore the natural order
ambiguous bin edge	state one inclusion rule
percent with missing rows	show the usable denominator

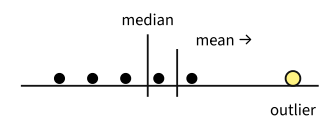
2b • Read, Then Question

C2
WORKFLOW

Describe quantitative pattern in the fixed order **centre → spread → shape → unusual features**. Add title, labelled axes, frequency scale, denominator, observation window and missing-value rule. Preserve natural order for ordinal categories.

3 • Centre, Spread, Shape

C3



Source dot plot, redrawn: a high tail pulls mean and expands spread.

SAMPLE
 $\bar{x} = \sum x_i / n$
 $s^2 = \sum (x_i - \bar{x})^2 / (n - 1) \cdot s = \sqrt{s^2}$
POPULATION
 $\mu = \sum x_i / N$
 $\sigma^2 = \sum (x_i - \mu)^2 / N \cdot \sigma = \sqrt{\sigma^2}$

POSITION
range = max - min
 $z = (x - \text{mean}) / \text{standard deviation}$

MEASURE	BEST FOR	MAIN WEAKNESS
mean	all magnitudes; total impact	pulled by tails
median	ordinary case under skew	ignores distances
mode	most common; categorical	may be multiple/none
range	quick span	only extremes
SD	typical distance; original units	tail-sensitive

Arithmetic checks: order before median; $\sum (x_i - \bar{x}) = 0$ apart from rounding; squared deviations ≥ 0 ; variance uses squared units, SD original units. Keep precision until the final line. **Scope, not dataset size, chooses n-1 versus N.**

CHOOSE THE LEAD STATISTIC

Use the **median** when a skewed distribution or distant values make the mean unrepresentative of an ordinary case. Retain the **mean** when total labour, cost or customer impact matters because every magnitude contributes. Mode answers "most common", including for categories. Pair every centre with spread: equal means can conceal very different consistency.

EVIDENCE	REPORT MOVE
symmetric, one cluster	mean + SD; still inspect extremes
skew / distant values	median leads; note tail and range
several clusters	split by a defensible process factor
extreme z	verify record, measure and process

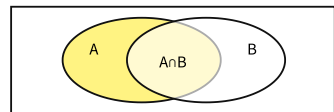
3b • Distribution Rules

C3

Right skew: long high tail, typically mean > median.
Left skew: long low tail, typically mean < median.
Multiple clusters can make any single centre incomplete.

4 • Event Probability

C4



"given B" → denominator becomes B

Source event lens, redrawn: overlap is numerator; given event is denominator.

EVENT LEDGER

$P(A^c) = 1 - P(A)$
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 $P(A|B) = P(A \cap B) / P(B)$
 $P(A \cap B) = P(B)P(A|B)$

INDEPENDENT ONLY

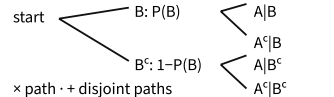
$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

WORDING	TRANSLATE	SANITY CHECK
or / at least one	union / complement of none	$\geq$ each event
and / both	intersection	$\leq$ each event
given	restrict to stated group	denominator = given
independent	conditioning changes nothing	product equality
mutually exclusive	intersection = 0	non-empty events dependent

Basic outcome probabilities lie in $[0, 1]$ and total 1. Label the source: equally likely theoretical outcomes, empirical relative frequency or informed judgement. A conditional difference is association, not proof of causation.

REGION-FIRST AUDIT

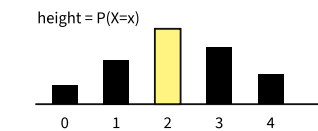
- Name experiment, sample space and events.
- Translate the wording; draw a region, table or tree.
- Apply the rule; check bounds and subsets.
- Interpret using the correct reference group.



Chapter tree rule: second-stage branches are conditional on the first.

5 • Discrete Variables

C5



Source pmf bars, redrawn: exact values carry mass; total mass is 1.

pmf lists $P(X=x)$: each mass $\in [0, 1]$, $\sum P(X=x) = 1$. **cdf** $F(k) = P(X \leq k)$ accumulates left to right, never decreases and ends at 1; each jump recovers the mass.

DISCRETE BOUNDARIES

$P(X=k) = F(k) - F(k-1)$
 $P(X > k) = 1 - F(k)$
 $P(X \geq k) = 1 - F(k-1)$

MOMENTS

$E(X) = \sum xP(X=x)$
 $\text{Var}(X) = \sum (x - \mu)^2 P(X=x)$
 $= E(X^2) - [E(X)]^2 \cdot \text{SD} = \sqrt{\text{Var}(X)}$

Expectation is a long-run weighted centre; it need not be an attainable single outcome. Use a four-column ledger **x, P, xP, x²P**; check mass before computing moments.

OBJECT	QUESTION ANSWERED	VALIDITY CHECK
pmf p(x)	exactly x	$0 \leq p \leq 1$; $\sum p = 1$
cdf F(k)	at or below k	non-decreasing; ends at 1
E(X)	long-run centre	inside possible-value range
Var(X)	weighted squared spread	non-negative

For integer X , strict and inclusive thresholds cannot be exchanged: a single point can carry positive mass. Recover $p(k)$ from the jump $F(k) - F(k-1)$. If listed masses total 0.90, the model is incomplete even though each entry looks plausible.

5b • Binomial Model

C5

Earn the shortcut: fixed n : two outcomes after grouping · constant p · independent trials. Sampling without replacement from a small population can violate independence.

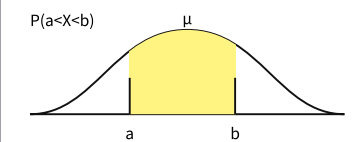
X ~ BINOMIAL(n, P)

$P(X=k) = C(n, k) p^k (1-p)^{n-k}$
 $E(X) = np \cdot \text{Var}(X) = np(1-p)$
 $\text{SD} = \sqrt{np(1-p)}$

PHRASE	CALCULATION
exactly k	mass $P(X=k)$
at most k	$F(k)$
more than k	$1 - F(k)$
at least k	$1 - F(k-1)$

6 • Continuous & Normal

C6



Source normal-area diagram, redrawn: standardise boundaries, preserve area.

For a continuous variable, probability is **area**. Total area = 1; one exact point has zero area, so open/closed endpoints agree. Density height alone is not probability.

X ~ NORMAL(mu, sigma)

$z = (x - \mu) / \sigma$
left: $P(X < x) = \Phi(z)$
right: $P(X > x) = 1 - \Phi(z)$
between: $\Phi(z_u) - \Phi(z_l)$

Forward: boundary → area. Draw, shade, standardise, select left/right/between, interpret in original units.

Inverse: area → boundary. Convert wording to a left-tail p , then $x_0 = \text{NORM.INV}(p, \mu, \sigma)$. "Exceeded on 5%" means $p = 0.95$.

DIRECTION CHECK	MUST BE TRUE
$x > \mu$	left area > 0.5
$p = 0.5$ quantile	$x_0 = \mu$
between	upper cdf - lower cdf
right tail	complement of cdf

FUNCTION TRANSLATION

EVENT	EXCEL ROUTE
$P(X \leq x)$	$\text{NORM.DIST}(x, \mu, \sigma, \text{TRUE})$
$P(X = x)$	$1 - \text{NORM.DIST}(x, \mu, \sigma, \text{TRUE})$
$P(a < X < b)$	$\text{NORM.DIST}(b, \mu, \sigma, \text{TRUE}) - \text{NORM.DIST}(a, \mu, \sigma, \text{TRUE})$
left area $p \rightarrow x$	$\text{NORM.INV}(p, \mu, \sigma)$

The normal model is symmetric and determined by μ and σ ; mean, median and mode coincide. Keep unrounded z -values in the calculation and attach the original unit to a returned boundary. An empirical-rule check is a plausibility screen, not a replacement for the cdf.

6b • Normal Approximation

C6

For a binomial count, the captured course rule requires **$np > 10$ and $n(1-p) > 10$** . Then use $\mu = np$ and $\sigma^2 = np(1-p)$. A large n alone is insufficient when p is near 0 or 1. The smooth curve approximates integer bars; apply the continuity correction to the event boundary.

Continuity direction: an inclusive integer boundary moves half a unit outward when mapped to continuous area; sketch the bars and curve so the included integers determine the corrected boundary. State that the answer is an approximation and, when feasible, compare it with the exact binomial result.



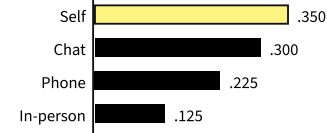
7 · Worked: Display Choice

C2

EXAMPLE

Q. A service desk records 80 tickets: self-service 28, chat 24, phone 18, in-person 10, plus quantitative resolution time. Choose displays and show the channel shares.

CHANNEL	F	F/80
Self-service	28	0.350
Chat	24	0.300
Phone	18	0.225
In-person	10	0.125
TOTAL	80	1.000



New comparison graphic from the guide's 80-ticket example.

Answer. Use separated bars for channel categories; a pie is possible because the four classes partition one whole, but bars support closer comparison. Use equal, non-overlapping, touching histogram bins for resolution time. State boundaries and read centre, spread, shape and unusual features.

MARKS ARE IN THE CHECKS

The frequency checksum is 28+24+18+10=80. The relative-frequency checksum is .350+.300+.225+.125=1.000. State that the bar length encodes a category's share of the **usable 80 tickets**. For the time histogram, exact observations are deliberately sacrificed to reveal concentration, spread, skew, gaps and possible clusters.

CHART STATEMENT	WHAT IT EARNS
channel = categorical	correct separated geometry
time = quantitative	correct adjacent intervals
one boundary rule	each value enters once
pattern, not cause	defensible interpretation

TRAP → Do not use a histogram for coded categories or separated bars for adjacent quantitative intervals.

7b · Set-up Mini-Drill

C1

EXAMPLE

A delivery file records service tier, minutes late, complaint status and depot code. **Unit**: one completed delivery. **Variables**: tier + depot = nominal; complaint = binary categorical; minutes late = quantitative ratio. It describes observed deliveries; inference to future deliveries needs representative operating conditions.

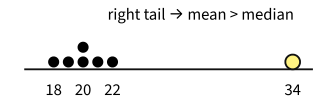
Boundary: describe sampled deliveries; inference to future work requires comparable conditions.

8 · Worked: Centre & Spread

C3

EXAMPLE

Q. Packing times (min): 18, 19, 20, 20, 21, 22, 34. Calculate centre, sample spread and the standard score of 34.



Source outlier logic applied to the seven packing times.

- Centres**: $\Sigma x=154$, so $\bar{x}=154/7=\mathbf{22\ min}$. Ordered middle (4th)=20; median=20; mode=20. Range=34–18=16 min.
- Spread**: $\Sigma(x-22)^2=178$. Sample $s^2\approx 178/(7-1)=29.667\ \text{min}^2$; $s=\sqrt{29.667}\approx\mathbf{5.447\ min}$.
- Standardise**: $z=(34-22)/5.447\approx\mathbf{2.203}$. The job is 2.203 sample SD above the mean, unusual but not beyond the course's three-SD extreme rule.
- Interpret**: the high time creates a right tail, so the mean exceeds the median. Median better represents an ordinary job; mean still matters for total resources.

X	X–22	(X–22) ²
18	–4	16
19	–3	9
20	–2	4
20	–2	4
21	–1	1
22	0	0
34	12	144
TOTAL	0	178

CHECK	OBSERVED
signed deviations	sum to 0
s ² denominator	n–1=6
units	min ² then min
shape	22 > 20; high tail

Decision split: “ordinary packing job” points to the median because it resists the high value; “total staffing minutes” keeps the mean because all minutes contribute. The statistics answer different management questions rather than competing to be universally correct.

RULE DRILLS

An invoice taking 57 minutes under mean 42 and sample SD 6 has $z=(57-42)/6=\mathbf{2.5}$; it is 2.5 sample SD above the mean. A z-score locates; it is not automatically a tail probability. For an approximately bell-shaped distribution, compare with the 68–95–99.7 rule. If shape is unknown, use the weaker Chebyshev guarantee rather than the empirical rule.

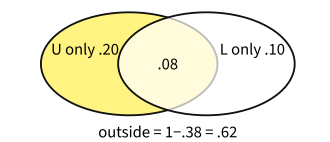
For a population, replace $\bar{x}$, n and s with μ , N and σ and divide the squared-deviation total by N. Scope—not the number of rows—selects the denominator.

9 · Worked: Union + Given

C4

EXAMPLE

Q. P(Urgent)=0.28, P(Late)=0.18 and P(Urgent∩Late)=0.08. Find the union, lateness given urgency and test independence.



Source event lens applied: add both circles, subtract overlap once.

- Union**: $P(U\cup L)=0.28+0.18-0.08=\mathbf{0.38}$.
- Condition**: “given urgent” makes U the denominator: $P(L|U)=0.08/0.28=\mathbf{0.2857}$.
- Independence**: compare conditional with marginal: $0.2857\neq 0.18$, so U and L are **not independent**.
- Meaning**: urgency and lateness are associated in this model; this does not establish that urgency causes delay.

CHECK	RESULT
intersection ≤ each	.08 ≤ .18, .28 ✓
union ≥ each	.38 ≥ .18, .28 ✓
conditional bounds	.2857 ∈ [0,1] ✓
product test	.28 × .18 = .0504 ≠ .08

	LATE	NOT LATE	TOTAL
URGENT	.08	.20	.28
NOT URGENT	.10	.62	.72
TOTAL	.18	.82	1.00

The completed table exposes every denominator: inside the urgent row, late occupies .08/.28; inside the late column, urgent occupies .08/.18. These are different conditional questions. It also makes “neither”= .62 and “exactly one”= .20+.10=.30 visible without a new formula.

Tree alternative: begin with U (.28) and U^c (.72); second-stage late probabilities are conditional on those branches. Multiply along each full path; add only paths that cannot both occur. If the late branch probability were identical after U and U^c, that would visually express independence.

9b · Worked: Count in Stages

C4

DRILL

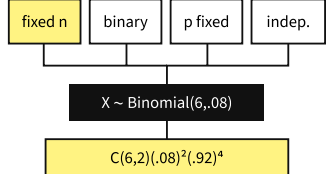
Select 3 from 8 analysts, then name one selected analyst chair. Group first: C(8,3)=56. Role second: ×3. Total= **168 outcomes**. Treating the whole selection as one permutation would count orders that do not create distinct committees.

10 · Worked: Binomial Damage

C5

EXAMPLE

Q. Six parcels have independent damage probability 0.08. Find P(exactly 2 damaged) and summarise the count X.



New audit flow: verify all four conditions before using the mass formula.

- Define**: success=damaged parcel; X counts damaged parcels.
- Model**: fixed 6, damaged/not, constant .08, independent ⇒ X ~ Binomial(6, .08).
- Exact mass**: $P(X=2)=C(6,2)(.08)^2(.92)^4=\mathbf{0.0687737}$.
- Centre**: $E(X)=np=6(.08)=\mathbf{0.48}$ parcel.
- Spread**: $\text{Var}(X)=6(.08)(.92)=\mathbf{0.4416}$; $SD=\sqrt{.4416}\approx\mathbf{0.6645}$ parcel.

The non-integer expectation is a long-run average across repeated six-parcel audits, not a fractional parcel prediction for one audit.

WORDING	EXCEL IDEA
exactly 2	BINOM.DIST(2,6,.08,FALSE)
at most 2	BINOM.DIST(2,6,.08,TRUE)
more than 2	1–BINOM.DIST(2,6,.08,TRUE)

CUMULATIVE BOUNDARY DRILL

If X counts failed inspections in 10 independent items with failure probability .04, “more than one” is $X>1$, so write

$P(X>1)=1-P(X\leq 1)=1-F(1)$ before software. “At least one” would instead be $1-F(0)$. The one-step shift is the discrete-boundary mark.

AUDIT	PARCEL EXAMPLE
mass in [0,1]	.0687737 ✓
mean in [0,6]	.48 ✓
variance ≥ 0	.4416 ✓
SD units	parcel count

Sampling without replacement from a small batch would make each damage probability depend on earlier draws. If p varies by parcel route or packing line, a single constant-p binomial model also fails.

Distribution reconstruction: a full pmf table adds columns $xP(X=x)$ and $x^2P(X=x)$. Sum them for $E(X)$ and $E(X^2)$, then use $\text{Var}(X)=E(X^2)-[E(X)]^2$. Convert pmf to cdf by running sums; recover any mass from the successive jump. These longer routes should agree with np and np(1–p) when the model is binomial.

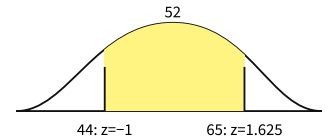
TRAP → TRUE accumulates through k; FALSE returns the exact mass at k.

11 · Worked: Normal Interval

C6

EXAMPLE

Q. Delivery time X is normal with $\mu=52$ min, $\sigma=8$ min. Find $P(44<X<65)$.



Source normal-area diagram applied: between-area = upper cdf – lower cdf.

- Sketch**: shade between 44 and 65; both boundaries are known ⇒ forward calculation.
- Standardise**: $z_1=(44-52)/8=\mathbf{-1}$; $z_2=(65-52)/8=\mathbf{1.625}$.
- Areas**: $\Phi(1.625)=0.9479$; $\Phi(-1)=0.1587$.
- Subtract**: $0.9479-0.1587=\mathbf{0.7892}$.
- Interpret**: the model assigns about **78.92%** probability to delivery times between 44 and 65 minutes.

CHECK	WHY IT PASSES
lower boundary	44<52 ⇒ z<0
upper boundary	65>52 ⇒ z>0
area	0.7892 in [0,1]
operation	upper left tail – lower

SOFTWARE ROUTE

NORM.DIST(65,52,8,TRUE)
– NORM.DIST(44,52,8,TRUE)
= 0.9479 – 0.1587 = 0.7892

Keep 1.625 unrounded while obtaining the cdf. The answer is a model probability, so the interpretation begins “the model assigns”, names the event 44–65 and returns to minutes. A left-tail result of .9479 alone would answer $X<65$, not the requested interval.

11b · Inverse Drill

C6

DRILL

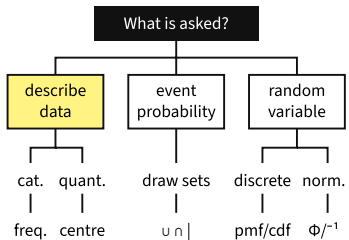
Daily output is normal with $\mu=1,200$ units and $\sigma=90$. A level exceeded on only 5% of days leaves **95% to its left**: solve $P(X< x_0)=0.95$ with NORM.INV(0.95,1200,90). Direction check: $0.95>0.5$, so x_0 must exceed 1,200.

KNOWN	UNKNOWN	FUNCTION
boundary x	left area	NORM.DIST
left area p	boundary x_0	NORM.INV
right-tail q	boundary	NORM.INV(1–q,...)

Model gate: check a histogram for approximate symmetry, one peak and process stability before treating the cdf digits as meaningful. For continuous X, $P(X=\text{boundary})=0$, so < and ≤ have the same area.

12 · Method Chooser

C1–C6



then: conditions → formula → check → context
New retrieval map combining the six supplied topic routes.

QUESTION CUE	METHOD	MUST STATE
code / label / rank	scale + categorical summary	meaning of arithmetic
typical + consistency	centre + spread + shape	sample/population
or / given / both	union / conditional / product	named events
count successes	binomial if 4 conditions hold	n,p,k; success
area from boundary	forward normal cdf	shaded tail
boundary from area	inverse normal	left-tail p

13 · High-Frequency Traps

FAST

SCAN

- Scope first**: n–1 versus N; observed description versus inference.
- Bar ≠ histogram**: categories separated; measurement intervals touch.
- Mean alone is incomplete**: report spread and shape; do not delete a high z automatically.
- Given flips the denominator**: $P(A|B)$ uses P(B), not P(A).
- Exclusive ≠ independent**: non-empty exclusive events cannot co-occur.
- Discrete boundaries matter**: $\geq k$ uses $1-F(k-1)$; $> k$ uses $1-F(k)$.
- Binomial needs all four conditions**: n, two outcomes, constant p, independence.
- Cdf is left-tail**: right tail is 1–cdf; between is a difference.
- Inverse takes left area**: upper 5% ⇒ input .95.
- Interpret**: result + event + unit + management meaning; association is not causation.

FINAL 20-SECOND SCAN	ASK
object	unit / X / events defined?
scope	sample or population?
conditions	model earned?
arithmetic	bounds, direction, units?