Problem Set — Week 1, Session 1 · Solutions

Mathematical Statistics I · Wackerly §§1.1–1.2 · ADA University

> This is one randomly generated version of the set. Every student receives different numbers, so the values here will not match your own copy — follow the method, not the arithmetic.

Problem 1

seed 20260909

A bank has 4550 retail deposit accounts at one branch. To estimate the average balance across all of them, an analyst interviews the holders of 145 accounts chosen at random.

Give the proportion and percentage to at least four significant figures.

(a) How many of the branch's accounts are **not** in the sample? ____________

(b) What proportion of the population was sampled? ____________

(c) Express that as a percentage. ____________ %

Solution

(a) $4550 - 145 = 4405$ accounts are outside the sample.

(b) $145/4550 = 0.0318681318681319$.

(c) $100 \times 0.0318681318681319 = 3.18681318681319\%$.

The vocabulary matters more than the arithmetic here. The **population** is the 4550 balances the analyst would like to know about; the **sample** is the 145 she actually measured. The average of the sample is a **statistic** she can compute; the average of the population is a **parameter** she cannot, and the whole point of the course is to say something defensible about the second using only the first — together with a measure of how good that statement is.

Note that this population is real and finite: all 4550 balances exist today. Contrast it with the population of **next** year's balances, which is conceptual — those measurements do not exist yet, and a sample of today's accounts is treated as representative of them.

Problem 2

seed 20260909

A commercial bank classifies its corporate loans by days past due.

Frequency distribution of the loan book

Days past due	Loans
0 (current)	20

1 to 30	22
31 to 60	31
61 to 90	16
more than 90	29

Give the proportions to at least four significant figures.

(a) How many loans are in the book altogether? _______________

(b) The relative frequency of the 31-to-60-day class is _______________

(c) The cumulative relative frequency of the first three classes is _______________

(d) The relative frequency of loans more than 60 days past due is _______________

Solution

(a) $n = 20 + 22 + 31 + 16 + 29 = 118$.

(b) A relative frequency is the class count divided by n: $31/118 = 0.26271186440678$.

(c) Add the first three counts and divide once: $(20 + 22 + 31)/118 = 0.61864406779661$.

(d) $(16 + 29)/118 = 0.38135593220339$.

A relative frequency distribution is what a histogram draws: the height of each rectangle is proportional to the fraction of measurements in that class, so the five heights must sum to 1. Parts (c) and (d) are complements of one another for exactly that reason — check that your two answers add to 1, and if they do not, a class has been dropped.

Problem 3

seed 20260909

A fund's monthly returns over several years are summarised in the relative frequency distribution below. The five classes have equal width and one relative frequency has been lost.

Relative frequency distribution of monthly fund returns

Monthly return (%)	Relative frequency
-6 up to -3	0.1
-3 up to 0	0.18
0 up to 3	0.29
3 up to 6	0.16
6 up to 9	?

Give each answer to at least four significant figures.

(a) The missing relative frequency for the top class is _______________

(b) A month is selected at random from this record. The probability that its return falls between -3% and 6% is ____________

(c) The probability that it falls **outside** that range is ____________

(d) The probability that the return is above 3% is ____________

Solution

(a) The five classes are exhaustive and non-overlapping, so their relative frequencies sum to 1: $1 - 0.1 - 0.18 - 0.29 - 0.16 = 0.27$.

(b) The range -3% to 6% is exactly the union of the second, third and fourth classes. Because the classes do not overlap, add their relative frequencies:

$$0.18 + 0.29 + 0.16 = 0.63$$

(c) $1 - 0.63 = 0.37$.

(d) The top two classes: $0.16 + 0.27 = 0.43$.

This is the step that makes a histogram more than a picture. If a measurement is drawn at random from the data set, the probability that it lands in an interval is **proportional to the area** of the histogram over that interval — and since the total area is 1, the probability simply **is** that area. Every answer above is an area, computed by adding the rectangles that sit over the interval in question.

Problem 4

seed 20260909

An analyst has the credit spreads of a portfolio of bonds, in basis points. The smallest is 24 and the largest is 142. She decides on classes of equal width 16 basis points, and — following the book's advice that no measurement should fall on a point of subdivision — starts the first class at 23.5.

(a) What is the range of the data? ____________

(b) How many classes of width 16 are needed to span the range? ____________

(c) What is the upper boundary of the first class? ____________

(d) What is the upper boundary of the last class? ____________

Solution

(a) $\text{range} = 142 - 24 = 118$ basis points.

(b) Each class covers 16 basis points, so the number needed is the range divided by the width, **rounded up** — a fractional class is still a whole class:

$$\left\lceil \frac{118}{16} \right\rceil = 8$$

(c) $23.5 + 16 = 39.5$.

(d) $23.5 + 8 \times 16 = 151.5$.

Two points of method. The boundary at 23.5 is deliberately placed half a basis point below the smallest observation: Wackerly's guideline is that the points of subdivision be chosen so that **no measurement can fall on one**, which removes any question of which class an observation belongs to. And the book suggests spanning the data with 5 to 20 classes, using more of them when there is more data — too many classes with little data barely summarises anything, and too few hides the shape the histogram exists to show.

Problem 5

seed 20260909

A manufacturer has completed 360 custom contracts. On 93 of them the profit margin exceeded 8%. The firm treats these as a sample from a **conceptual** population: all contracts it might complete, now and in future, at similar settings of contract size and competition.

Give each answer to at least four significant figures; the count in (c) need not be a whole number.

(a) What is the relative frequency of contracts with a margin above 8%? ______________

(b) Taking that relative frequency as the probability, what is the probability that the **next** contract has a margin of 8% or less? ______________

(c) The firm expects 75 contracts next quarter. If they behave like the historical record, about how many will have a margin above 8%? ______________

Solution

(a) $93/360 = 0.258333333333333$.

(b) The two outcomes are exhaustive, so $1 - 0.258333333333333 = 0.741666666666667$.

(c) $75 \times 0.258333333333333 = 19.375$ contracts.

The move made in part (b) is the whole of Chapter 1 in miniature. A relative frequency is a fact about 360 contracts that have already happened; a probability is a statement about one that has not. They are connected by the observation that relative frequencies in a long series of trials are remarkably stable — which is why we are willing to read the first as an estimate of the second.

What is missing, and what the rest of the course supplies, is the **measure of goodness**: how close is 0.258333333333333 likely to be to the true long-run proportion? That question cannot even be posed until Chapter 2 gives probability a rigorous definition.

Problem 6

seed 20260909

Two branches are audited. At branch A, 28 of 74 sampled files contain a documentation error; at branch B, 18 of 79.

Give each proportion to at least four significant figures.

(a) The sample proportion with an error at branch A is ________________

(b) At branch B it is ________________

(c) Treating the two samples as one sample from the combined population of both branches, the pooled proportion is ________________

Solution

(a) $28/74 = 0.378378378378378$. **(b)** $18/79 = 0.227848101265823$.

(c) Pool the counts, not the proportions:

$$\frac{28+18}{74+79} = \frac{28+18}{74+79} = 0.300653594771242$$

The pooled figure is **not** in general the average of 0.378378378378378 and 0.227848101265823 — it is a weighted average, with each branch weighted by its own sample size. Averaging the two proportions directly would silently give the smaller sample equal say, and that is a mistake worth learning to recognise now: it reappears as Simpson's paradox, and in any report that combines desks or regions of unequal size.

Problem 7

seed 20260909

A histogram of daily trade counts at a brokerage uses six classes of equal width, with these frequencies:

class 1: 17 class 2: 19 class 3: 24 class 4: 15 class 5: 23 class 6: 29

The analyst decides the picture is too choppy and merges classes into pairs, so classes 3 and 4 become one class of double width.

Give the proportions to at least four significant figures.

(a) The total number of days recorded is ________________

(b) The frequency of the merged class (old classes 3 and 4) is ________________

(c) Its relative frequency is ________________

(d) The relative frequency of the top two classes, 5 and 6, is ________________

Solution

(a) $n = 17 + 19 + 24 + 15 + 23 + 29 = 127$.

(b) Merging adjacent classes simply adds their counts, because no day is in two classes: $24 + 15 = 39$.

(c) $39/127 = 0.307086614173228$.

(d) $(23 + 29)/127 = 0.409448818897638$.

One caution the merge exposes. With equal-width classes, a taller rectangle always means more data — so comparing heights is safe. Once a class is twice as wide, that stops being true: the merged class holds the **sum** of two classes' data but covers twice the axis, and if it were drawn at the height of its relative frequency it would appear to hold twice the mass it does. This is why the honest statement is that **area**, not height, is proportional to the fraction of measurements. With equal widths the distinction does not bite; with unequal widths it does.

Problem 8

seed 20260909

Two funds are compared over their whole histories. Fund X has 330 daily returns on record, of which 31 fall in the two worst classes of its histogram. Fund Y has 340 daily returns, of which 18 fall in the two worst classes of its own histogram. The two histograms use the same class boundaries.

Give each answer to at least four significant figures.

(a) The relative frequency in the lower tail for fund X is ______________

(b) For fund Y it is ______________

(c) The difference, X minus Y, is ______________

(d) The ratio of X's tail proportion to Y's is ______________

Solution

(a) $31/330 = 0.0939393939393939$. **(b)** $18/340 = 0.0529411764705882$.

(c) $0.0939393939393939 - 0.0529411764705882 = 0.0409982174688057$.

(d) $0.0939393939393939/0.0529411764705882 = 1.77441077441077$.

Notice that the raw counts 31 and 18 are not comparable on their own, because the two records are of different lengths. Only after dividing by n does the comparison mean anything — which is the reason a **relative** frequency distribution, rather than a frequency distribution, is the object the book characterises a population by.

A caution about what has and has not been shown. The difference in (c) is a difference between two samples, and two samples drawn from identical populations would still differ by something. Whether 0.0409982174688057 is large enough to conclude that fund X genuinely has the heavier tail is a question of inference, and answering it needs the measure of goodness that Chapters 8 and 10 supply.

Problem 9

seed 20260909

Five equal-width classes of client portfolio values have relative frequencies 9%, 15%, 40%, 13% and one unknown, in order from lowest to highest.

Give each answer as a proportion to at least four significant figures.

(a) The relative frequency of the top class is ____________

(b) The cumulative relative frequency through class 2 is ____________

(c) The cumulative relative frequency through class 3 is ____________

(d) The median lies in class 3. Having passed through class 2, what further proportion of the data must be crossed to reach the median? ____________

Solution

(a) The five relative frequencies sum to 1, so the top class has $1-(9+15+40+13)/100=0.23$.

(b) $(9+15)/100=0.24$.

(c) $(9+15+40)/100=0.64$.

(d) The median is the value with half the data at or below it. The cumulative proportion is 0.24 at the top of class 2 and 0.64 at the top of class 3, and since $0.24<0.5<0.64$ the median sits inside class 3. The proportion still to be crossed on entering that class is

$$0.5-0.24=0.26$$

Cumulative relative frequency is how any quantile is **located** from grouped data: find the first class whose cumulative proportion reaches the quantile you want. Pinning the median down to a number inside that class needs an extra assumption — usually that the data are spread evenly across the class — and that assumption, not the arithmetic, is where grouped-data quantiles go wrong.

Problem 10

seed 20260909

Daily settlement volumes at a clearing house run from a low of 37 to a high of 293 thousand instructions.

Give each answer to at least four significant figures; the count in (c) is a whole number.

(a) The range of the data is ____________ thousand.

(b) An analyst wants exactly 13 classes of equal width spanning that range. What width must each class have? ____________ thousand.

(c) A colleague prefers a round width of 18 thousand instead. How many classes of that width are needed to span the range? ____________

Solution

(a) $293 - 37 = 256$ thousand.

(b) $256/13 = 19.6923076923077$ thousand — the width is forced once the number of classes is fixed.

(c) Now the width is fixed and the count is forced, rounded up:

$$\left\lceil \frac{256}{18} \right\rceil = 15$$

The two parts are the same decision taken from opposite ends, and Wackerly is explicit that neither the number nor the width of the classes is prescribed — the choice is at the discretion of whoever builds the histogram. The guideline is to span the data with 5 to 20 classes, taking more when there is more data.

The reason it is a guideline rather than a rule is that both failure modes are real. Too many classes with little data barely summarises anything and returns a picture close to the raw list; too few smooths away the shape the histogram was drawn to reveal. Judge your answer to (c) against the 5-to-20 band: if it falls outside, the round width is the thing to reconsider.

Problem 11

seed 20260909

An insurer has 10800 policies in force. An auditor examines 250 of them chosen at random and finds 47 to be mispriced.

Give each answer to at least four significant figures; the estimated counts need not be whole numbers.

(a) The sample proportion mispriced is ______________

(b) Estimate how many of the 10800 policies in force are mispriced. ______________

(c) Estimate how many are correctly priced. ______________

Solution

(a) $47/250 = 0.188$.

(b) Scale the sample proportion up to the population: $10800 \times 0.188 = 2030.4$ policies.

(c) $10800 - 2030.4 = 8769.6$, or equivalently $10800 \times (1 - 0.188)$.

Part (b) is an **inference**, and it is worth being precise about what has and has not been claimed. The sample proportion 0.188 is a statistic: it is known exactly, and a different sample of 250 policies would have produced a different value. The population proportion is a parameter: fixed, unknown, and the thing actually of interest. Multiplying by 10800 gives a point estimate of the number mispriced — a single number with no statement of how far off it might be.

A regulator reading this would immediately ask how far off. Answering that requires attaching a measure of goodness to the estimate, and that is what the machinery of the next several chapters is built to supply.

Problem 12

seed 20260909

Two analysts, working independently, each draw a random sample from the **same** population of outstanding invoices and record how many are more than 30 days old. The first samples 130 invoices and finds 56; the second samples 140 and finds 34.

Give each answer to at least four significant figures.

(a) The first analyst's sample proportion is ____________

(b) The second analyst's sample proportion is ____________

(c) The absolute difference between the two is ____________

(d) Pooling both samples into one gives a proportion of ____________

Solution

(a) $56/130 = 0.430769230769231$. **(b)** $34/140 = 0.242857142857143$.

(c) $|0.430769230769231 - 0.242857142857143| = 0.187912087912088$.

(d) Pool the counts over the pooled sample size: $(56 + 34)/(130 + 140) = 0.333333333333333$.

The point of this problem is part (c). Both analysts sampled the **same** population, so there is a single true proportion, and yet they report different numbers. Neither has made a mistake: a statistic varies from sample to sample because the sample does.

This is why an estimate without a measure of goodness is close to useless. A report saying "the proportion over 30 days is 0.430769230769231" invites a reader to treat that figure as the truth, when a second, equally valid sample produced a figure 0.187912087912088 away from it. The whole apparatus of this course exists to turn "here is a number" into "here is a number, and here is how far from the truth it is likely to be" — and the pooled estimate in (d), resting on more data, is the better of the three precisely because more information buys a tighter statement.

Problem 13

seed 20260909

A histogram of 1000 settled trades groups them into three bands by settlement delay. 16% fall in the same-day band and 35% in the one-to-two-day band.

The counts below are whole numbers; give the proportion to at least four significant figures.

(a) How many trades are in the same-day band? ____________

(b) What percentage fall in the band of three days or more? ____________ %

(c) How many trades is that? ________________

(d) Express the one-to-two-day percentage as a relative frequency. ________________

Solution

(a) $1000 \times 16/100 = 160$ trades.

(b) The three bands are exhaustive, so their percentages sum to 100: $100 - 16 - 35 = 49$%.

(c) $1000 \times 49/100 = 490$ trades.

(d) A percentage divided by 100: $35/100 = 0.35$.

Three ways of saying one thing, and it is worth keeping them straight because they are not interchangeable in a sentence. A **frequency** is a count and depends on n; a **relative frequency** is a proportion between 0 and 1 and does not; a **percentage** is a relative frequency multiplied by 100. The histogram in the book is drawn from the middle one, because that is the version whose rectangles have total area 1 and can therefore be read as probabilities.

A practical check: your three counts should add back to 1000. If they do not, a rounding step has gone astray.

Problem 14

seed 20260909

A histogram of loan sizes uses **five classes of equal width**, with relative frequencies 6%, 17%, 42%, 15% and one unknown, in order from smallest loans to largest.

Consider the interval covered by the two lowest classes. Give each answer as a proportion to at least four significant figures; (c) may be negative.

(a) The relative frequency of the top class is ________________

(b) What fraction of the **measurement axis** spanned by the histogram do the two lowest classes cover?

____

(c) What fraction of the **area** under the histogram lies over them? ________________

(d) By how much does the area share exceed the axis share? ________________

Solution

(a) $1 - (6 + 17 + 42 + 15)/100 = 0.2$.

(b) The classes have equal width, so two of five cover $2/5 = 0.4$ of the axis. This is a fact about the horizontal scale alone and has nothing to do with the data.

(c) Area over an interval is the sum of the relative frequencies of the classes in it: $(6 + 17)/100 = 0.23$.

(d) $0.23 - 0.4 = -0.17$.

The two numbers in (b) and (c) are answers to different questions, and reading one as the other is the most common way a histogram is misread. “Two classes out of five” describes how much of the **range** you are looking at; only the relative frequencies tell you how much of the **data** sits there.

The probability statement follows the area, not the axis: a loan drawn at random from this book falls in the two lowest classes with probability 0.23, not 0.4. A negative answer in (d) simply means those classes are sparser than an even spread would give — which is what a mound-shaped distribution does to its own tails.

Problem 15

seed 20260909

An insurer records claim sizes in six classes of equal width, from smallest to largest:

class 1: 56 class 2: 69 class 3: 49 class 4: 15 class 5: 9 class 6: 8

Give the proportions to at least four significant figures.

(a) How many claims are recorded? ______________

(b) The relative frequency in the two lowest classes is ______________

(c) The relative frequency in the two highest classes is ______________

(d) The lower two exceed the upper two by ______________

Solution

(a) $n = 206$.

(b) $(56 + 69)/206 = 0.606796116504854$.

(c) $(9 + 8)/206 = 0.0825242718446602$.

(d) $0.606796116504854 - 0.0825242718446602 = 0.524271844660194$.

The two tails cover equal stretches of the axis — both are two classes wide — and yet they hold very different fractions of the data. That asymmetry is what is meant by saying the distribution is **skewed to the right**: a long thin tail of large claims, with the bulk of the mass packed against the left.

This matters for what comes next. Claim sizes, loan losses and equity returns are all commonly skewed or heavy-tailed, and the empirical rule — which assumes a mound-shaped, roughly symmetric distribution — has no licence here. Tchebysheff’s theorem does, because it assumes nothing about shape at all. When you meet both rules in the next session, this histogram is the case that tells you which one you are allowed to use.

Problem 16

seed 20260909

A histogram of 580 daily trading volumes has one class running from 35 to 49 lots, containing 64 days. A further 48 days fall in classes entirely above 49.

The desk wants to know how often volume exceeded the midpoint of that class. Assume the 64 days are spread evenly across the class.

Give the proportion to at least four significant figures.

(a) What is the midpoint of the class? ________________ lots.

(b) Under the even-spread assumption, how many of the 64 days lie above the midpoint? ________________

(c) How many days in total had volume above the midpoint? ________________

(d) What proportion of all 580 days is that? ________________

Solution

(a) $(35 + 49)/2 = 42$ lots.

(b) The midpoint splits the class into halves of equal width. Spreading 64 days evenly puts half in each: $64/2 = 32$ days.

(c) Add the days in the classes entirely above: $32 + 48 = 80$ days.

(d) $80/580 = 0.137931034482759$.

The honest part of this answer is (c) and (d) **given** (b), and the assumption in (b) is doing real work. Grouping has destroyed the information about where inside the class those 64 days actually fell; the even spread is a guess, reasonable for a narrow class and poor for a wide one, and worse still if the class sits in a steep part of the distribution where far more of its data will lie towards one edge.

This is the standing cost of grouped data: anything asked at a point that is not a class boundary can only be answered by assuming something about the inside of a class. Ask the question at a boundary and no assumption is needed.

Problem 17

seed 20260909

A histogram of corporate bond maturities has several classes. The three shortest classes contain 56, 59 and 56 bonds. The analyst reports that the remaining classes together hold 28% of the portfolio.

Give each answer to at least four significant figures; the counts need not be whole numbers.

(a) How many bonds are in the portfolio altogether? ________________

(b) How many bonds are in the remaining classes? ________________

(c) What is the relative frequency of the shortest class? ________________

Solution

(a) The three known classes hold $56 + 59 + 56 = 171$ bonds, and they are the part of the portfolio **not** in the remaining classes — that is, $100 - 28 = 72\%$ of it. So

$$n = \frac{171}{1 - 28/100} = 237.5$$

(b) $237.5 - 171 = 66.5$ bonds, which you can check against $237.5 \times 28/100$.

(c) $56/237.5 = 0.235789473684211$.

The step worth noticing is that the 28% had to be converted into its complement before it could be used. A count and a percentage cannot be added; the count of known classes corresponds to the complementary percentage, and dividing by that complement is what recovers n. Getting this backwards — dividing by $28/100$ instead — is the standard error here, and it produces an n smaller than the classes you already counted, which is the check that catches it.

Problem 18

seed 20260909

Five equal-width classes of customer waiting times at a bank branch, from shortest to longest, contain

class 1: 41 class 2: 59 class 3: 65 class 4: 22 class 5: 15

Give the proportions to at least four significant figures.

(a) How many customers were timed? ________________

(b) What proportion waited long enough to fall in class 4 or class 5? ________________

(c) What proportion fell below class 4? ________________

(d) What proportion fell in class 3 or above? ________________

Solution

(a) $n = 202$.

(b) $(22 + 15)/202 = 0.183168316831683$.

(c) Everything not in (b): $1 - 0.183168316831683 = 0.816831683168317$. You could also add the first three classes and divide — the two routes must agree, and disagreeing means a class was double-counted or dropped.

(d) $(65 + 22 + 15)/202 = 0.504950495049505$.

Accumulating from the top rather than the bottom is often the natural direction for a service-level question — "what fraction of customers waited too long" is a statement about the upper tail. Both directions carry the same information, since each is one minus the other at every boundary, and choosing the direction that makes the arithmetic shorter is the only reason to prefer one.

Problem 19

seed 20260909

The daily returns of an equity fund are grouped into five classes of equal width:

Frequency distribution of daily returns

Daily return (%)	Days
-4.0 up to -2.0	18
-2.0 up to 0.0	42
0.0 up to 2.0	76
2.0 up to 4.0	36
4.0 up to 6.0	20

Give the proportions to at least four significant figures.

(a) How many trading days are in the record? ________________

(b) The relative frequency of the class 0.0 up to 2.0 is ________________

(c) A day is drawn at random from the record. The probability its return lies between -2.0% and 4.0% is

____

(d) The probability it lies outside that interval is ________________

(e) The probability the return is at least 2.0% is ________________

Solution

(a) $n = 18 + 42 + 76 + 36 + 20 = 192$.

(b) $76/192 = 0.395833333333333$.

(c) The interval is the union of classes 2, 3 and 4, which do not overlap, so the areas add: $(42 + 76 + 36)/192 = 0.802083333333333$.

(d) $1 - 0.802083333333333 = 0.197916666666667$, which is also $(18 + 20)/192$ — the two tails.

(e) Classes 4 and 5: $(36 + 20)/192 = 0.291666666666667$.

Every one of parts (b) to (e) is an area under the histogram, and that is the single idea worth carrying out of this session. Dividing the counts by n turns the frequency distribution into a relative frequency distribution whose rectangles have total area 1; the probability that a randomly drawn measurement falls in an interval is then just the area sitting over it.

Two habits follow. Check that your class proportions sum to 1, and check that complementary answers — (c) and (d) here — sum to 1. Both checks cost seconds and catch the dropped class, which is the error that actually happens.

Problem 20

seed 20260909

A machining firm has completed 240 custom contracts. Of these, 19 lost money and a further 56 returned a margin below 5%; the rest returned 5% or more. The firm regards this record as a sample from the conceptual population of all contracts it might take on at similar settings of contract size and competition.

Give each answer to at least four significant figures; the count in (d) need not be a whole number.

(a) The relative frequency of loss-making contracts is ______________

(b) The relative frequency of contracts returning 5% or more is ______________

(c) Taking the relative frequency as the probability, the probability that the **next** contract loses money is

____

(d) The firm plans 55 contracts next year. About how many would you expect to lose money? ______________

Solution

(a) $19/240 = 0.0791666666666667$.

(b) The three categories are exhaustive: $1 - 0.0791666666666667 - 0.233333333333333 = 0.6875$.

(c) The same number as (a) — and that is the point of the question, not an oversight.

(d) $55 \times 0.0791666666666667 = 4.35416666666667$ contracts.

Parts (a) and (c) are numerically identical and logically different. (a) is a **description**: a fact about 240 contracts that are finished and counted. (c) is an **inference**: a statement about a contract that does not exist yet, drawn from a population that does not exist yet either. The bridge between them is the empirical observation that relative frequencies in long series are stable, which is exactly why the book is willing to read a histogram as a probability distribution.

What the firm has **not** been given is how much to trust 4.35416666666667. A different 240 contracts would have produced a different proportion, so the forecast inherits that variability, and a single number conceals it. Supplying the missing measure of goodness — an interval, and a statement of how often such an interval contains the truth — is the work of the rest of the course.

Answer key

Problem	Answers

1	(a) 4405, (b) 0.0318681, (c) 3.18681
2	(a) 118, (b) 0.262712, (c) 0.618644, (d) 0.381356
3	(a) 0.27, (b) 0.63, (c) 0.37, (d) 0.43
4	(a) 118, (b) 8, (c) 39.5, (d) 151.5
5	(a) 0.258333, (b) 0.741667, (c) 19.375
6	(a) 0.378378, (b) 0.227848, (c) 0.300654
7	(a) 127, (b) 39, (c) 0.307087, (d) 0.409449
8	(a) 0.0939394, (b) 0.0529412, (c) 0.0409982, (d) 1.77441
9	(a) 0.23, (b) 0.24, (c) 0.64, (d) 0.26
10	(a) 256, (b) 19.6923, (c) 15
11	(a) 0.188, (b) 2030.4, (c) 8769.6
12	(a) 0.430769, (b) 0.242857, (c) 0.187912, (d) 0.333333
13	(a) 160, (b) 49, (c) 490, (d) 0.35
14	(a) 0.2, (b) 0.4, (c) 0.23, (d) -0.17
15	(a) 206, (b) 0.606796, (c) 0.0825243, (d) 0.524272
16	(a) 42, (b) 32, (c) 80, (d) 0.137931
17	(a) 237.5, (b) 66.5, (c) 0.235789
18	(a) 202, (b) 0.183168, (c) 0.816832, (d) 0.50495
19	(a) 192, (b) 0.395833, (c) 0.802083, (d) 0.197917, (e) 0.291667
20	(a) 0.0791667, (b) 0.6875, (c) 0.0791667, (d) 4.35417