Math 125—Introductory Statistics

Key Definitions and Verbal Procedures for Fall 2025 Final Exam

Mathematics Department, UMass Boston

Reference text: Numbers in brackets refer to sections of Freedman, Pisani, & Purves, *Statistics,* fourth edition.

Note: Outcomes marked **(Optional)** may appear on the final exam only with the unanimous consent of all instructors.

Descriptive Statistics

The Histogram [Chapter 3].

Histogram: page 56, summary point 1. A hoistogram ia a set of blocks.

> In a histogram, the areas of the blocks represent percentages.

Class Intervals: range over which the data is distributed according to the variable. The base of each block must be proportional to the width of its interval. The horizontal scale must be a true scale, with equal intervals having the same witdths on the graph. Try Problem 3, page 33.

Distribution Table: shows the percentage of data n each class interval.

> To figure out the height of a block over a class interval, divide the percentage by the length of the interval.

The units will be percent per unit length.

Density Scale:

> In a histogram, the height of a block represents crowding—percentage per horizontal unit.

> With the density scale on the vertical axis, the areas of the blocks come out in percent.
> The area uner the histogram over an interval equals the precentage of cases in that interval.
> The total area under the histogram is 100%.

Also presented on page 56, summary points 2 and 3.

Formula for the Block of a Histogram

Area = base × height. (the general formula for a rectangle)

(The percent for the block) = (the length of the interval) × (the height of the block, a density)

For Example: (12%) = (3 in.) × (4% per inch) or (12%) = (3 in.) × (4%/in.)

See Example 1 on page 40 and Problem 5 on page 52. Also try problem 1 on page 41.

Variables: qualitative or quantitative, discrete or continuous: qualitative variable have a verbal description; quntitative variable have numeric description. For a discrete variable, the vaues can only differ by fixed amounts. For a continuous variable, the difference in value between two data points cna be arbitrarily small. See page 56, Point 4; and Section 3.4.

Plotting a Histogram for a Discrete Variable: The convention is to center the class intervals at the possible values. See the extensive discussion on page page 44 (bottom) and the graph on the top of page 44, and try paroblems 2a and 2b on page 44.

The Average and the Standard Deviation [Chapter 4].

The Average:

The average of a set of numbers equals their sum, divided by how many there are.

Look at page 76, summary points 1 to 4; snd the technical note from page 65 to the top of age 66.

Symmetric Histograms: If a histogram is symmetrical around a value, that value equals the average. Furthermore, half the area under such a histogram lies to the left of of that value, and half to the right. So, for such a symmetric histogram, the median equals the average.

The Root-mean-square: Try problem 5 an page 67, and see summary point 5 on page 77.

The r.m.s. size of a list $= \sqrt{\textbf{average of (entries}^2\textbf{)}}$.

The SD: Refer to page 77, summary point 7.

The SD says how far away numbers on a list are from their average. Most entries on the list will be somewhere around one SD away from the average. Very few will be more than two or three SDs away.

Roughly 68% of the entries on a list (two in three) are within one SD of the average, the other 32% are further away. Roughly 95% (19 in 20) are within 2 SDs of the average, the other 5% are further away. This is so for many lists, but not all.

See problems 9 on psge 71, and problem 10 on pages 75 and 76.

Computing the Standard Deviation:

SD = r.m.s. deviation from average.

Procedure to find the SD.

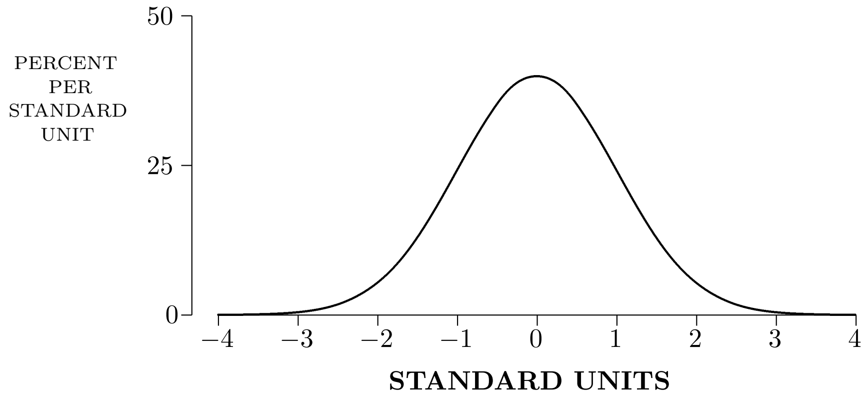
- Find the average of the list.
- Subtract the average from each member of the list. This will give the deviations from average.

  [Now find the root-mean-square size of these deviations.]
- SQUARE all the entries, getting rid of the signs.
- Take the MEAN (average) of the squares.
- Take the square ROOT of the mean.

Example 2 on page 72 shows the application of this procedure.

The Normal Approximation for Data [Chapter 5].

The Normal Curve:



A normal curve with a horizontal scale in standard units is symmetric about 0, has under it area in percent and a total area of 100%, is always above the horizontal axis, and gets and remains very close to the horizontal axis between 3 and 4 and (in the negative direction) between −3 and −4. The vertical axis is a density with scale in percent per standard unit.

Standard Units: page 96, summary point 2; problem 1 on page 82.

A value is converted to standard units by seeing how many SDs it is above or below average.

(This is the most important stated concept in the course.)

Formula. Its formula is: $\text{standard units} = \dfrac{\text{observation} - \text{average}}{\text{SD}}$.

Finding Areas under the Normal Curve: Review exercise 1, page 50; section 5.2, examples 2 to 7; page 84, problem 1; page 85, problems 2 and 3 **(important: problem 3b)**.

An area from minus a value to plus the same value is read off from the Normal Table on page A-104; other areas are found by making a sketch and expressing the desired area in terms of areas that may be found by using the Table.

The Normal Approximation for Data: pages 85 to 87: examples 8 and 9; page 88: problems 1 and 2; page 94: review exercises 3 and 4; page 96: summary point 4.

If a histogram follows the normal curve, approximate areas may be found by converting the endpoints to standard units and finding the appropriate areas under the normal curve by using the Table.

Percentiles: refer to page 95, problem 7.

The nth percentile is a value in the distribution for which n% of the values in the distribution were equal to or smaller than that value. For example, the 23rd percentile of a distribution of family incomes is a family income (say $17,349). That means that 23% of all family incomes were equal to or smaller than $17,349.

A percentile is not a percent, but a number which has the given percentage of the area in the distribution below it. That number puts you at the n$^{\text{th}}$ percentile of the number distribution. (The percentile rank is a percent, the percent associated with that number.)

Percentiles and the Normal Curve: Example 10 on pages 90 and 91; page 92, problems 1 to 3. When calculating percentiles for a histogram that follows the normal curve, remember that the normal table cannot be used directly because it gives areas between $-z$ and z, while a percentile involves the area to the left of z.

Percentiles and the Normal Curve: technical details

A. For a percentile rank above 50%, subtract 50% from the rank and double the result. That will give you the area between $-z$ and z. Use the normal table to detemine the z's. Here the positive answer is the relevant one.

B. Given a positive z, to find the associated percentile rank, take half of the area between $-z$ and z and add to it the area below $z = 0$ (50%).

C. For a percentile rank below 50%, to find the associated z, double the rank and subtract from 100%. From the table get $-z$ to z for that answer. Choose the negative one here.

D. Given a negative z, to find the associated percentile rank, look up $-z$ to z, subtract that middle area from 100%, and then divide by 2 to get the tail area. The tail area of the left-hand tail gives the percentile rank.

Examples: (a) Percentile rank is 91%. $(91\% - 50\%) \times 2 = 82\%$ z is 1.35.

(b) $z = 0.20$. Half of area is $15.85\%/2 \approx 8\%$. $8\% + 50\% = 58\%$.

(c) Percentile rank is 5%. $100\% - (2 \times 5\%) = 100\% - 10\% = 90\%$. z is -1.65.

(d) $z = -0.50$. -0.50 to 0.50 is about 38% on the table. Subtract $100\% - 38\% = 62\%$.
The tail area of the left-hand tail is $62\%/2 = 31\%$. That is the percentile rank.

Change of Scale: section 5.6; page 96, summary point 7.

- Adding the same number to every entry on a list adds that constant to the average; the SD does not change.
- Multiplying every entry on a list by the same positive number multiplies the average and the SD by that constant.
- These changes of scale do not change the standard units.

Correlation and Regression

Correlation [Chapter 8], pages 139 and 140, Points 1 to 7; Section 8.4.

Scatter Diagram: a set of plotted points, each of which represents (by its 2 coordinates) an associated pair of values of the distribution.

Point of Averages: the point showing the average of the x-values and the average of the y-values. (average x, average y)

The Correlation Coefficient:

The correlation coefficient is a measure of linear association, or clustering around a line. The relationship between two variables can be summarized by

- the average of the x-values, the SD of the x-values,
- the average of the y-values, the SD of the y-values,
- the correlation coefficient r.

Correlations are always between -1 and 1, but can take any value in between. A positive correlation means that the cloud slopes up; as one variable increases, so does the other. A negative correlation means that the cloud slopes down; as one variable decreases, the other decraeses.

Computing the Correlation Coefficient: See example 1 on pages 132 to 133; problems 1a and 1b on page 134; problems 9a and 9b on page 137; and summary points 1 to 5 and 7 on page 140. For efficient work, create a table like Table 2 on page 133, and follow Steps 1 to 4, as shown in Example 1 on pages 132 to 133.

Convert each variable to standard units. The average of the products gives the correlation coeficient.

Regression [Chapter 10], page 161, problems 1, 2ab; page 178, summary point 1.

> The regression line for y on x estimates the average value for y corresponding to each value of x.

> Associated with each increase of one SD in x there is an increase of only r SDs in y, on the average.

Formula: A formula for the estimate of the average y for a given x is

$$\text{estimated average} = \text{average } y + (x \text{ in standard units} \times r \times \text{SD of } y)$$

The R. M. S. Error for Regression [Chapter 11], page 201, Points 1 to 4.

The Regression Line [Chapter 12], pages 216 and 217, Points 1 to 6, and 8 and 11.

Probability

What Are the Chances? [Chapter 13], page 236, Points 1 to 10.

More About Chance [Chapter 14], page 254, Points 1 to 3.

The Binomial Formula [Chapter 15], pages 268 and 269, Points 1, 2, and 4.

Chance Variability

The Law of Averages [Chapter 16], page 287, Points 1 to 6.

The Expected Value and Standard Error [Chpater 17], page 307, Points 1 to 6.

The Normal Approximation for Probability Histograms [Chapter 18], pages 329–30, Points 1 to 5.

Sampling

Sample Surveys [Chapter 19], pages 353 to 354, Points 1 to 3, and 9.

Chance Errors in Sampling [Chapter 20], page 373, Points 1 to 3.

The Accuracy of Percentages [Chapter 21], page 394, Points 1 to 5.

The Accuracy of Averages [Chapter 23], pages 436 and 437, Points 1, 2, and 4.

Chance Models

A Model for Measurement Error [Chapter 24], page 457, Points 1 to 5.

Tests of Significance

Tests of Significance [Chapter 26], page 500, Points 1 To 5.

The Chi-Square Test [Chapter 28], page 544, Points 1 to 4.

A Closer Look at Tests of Significance [Chapter 29], page 576, Points 3 and 4.