Statistics 1

Daniel Renschler

June 16, 2023

Contents

Lecture 1: intro

02-06-2023

Chapter 1

1.1 – Key Words and Definitions

1.1 Key words

Data	Any observations that hvae been collected.
Statistics	Collect, analyze, summarize, interpret and draw conclustions from there.
Population	The complete set of elements being studied.
Samples	Some subset of the population.
Census	Collection from every member of a population.

Table 1.1: Statistics Vocabulary

→ If you take a sample, it must be collected **randomly**.

1.2 Types pf Data

P-P	**Parameter**	A characteristic of a population.
S-S	**Statistic**	A characteristic of a sample.

Table 1.2: Statistics Vocabulary

1.3 Two Types of Data

Qualitative (Categorical)	Data that is non-numerical e.g. color, gender, race, zip-codes... Mathematical operations are **meaningless**.
Quantitative	Numerical e.g. height/weight, wages, temperature, time. Mathematical operations are **meaningful**.

Table 1.3: table

1.3.1 Two types of Quantitative Data

Discrete data	Countabe or finite Numbers of eggs, dice...
Continious Data:	Infinite number of possible values (not countable) Usually a **measurement**, e.g. temperature.

Table 1.4: Quantitative data

1.4 4 Levels of Measurement

Nominal	Categories **not** ordered. e.g. religion
Ordinal	Can be ordered, differences are meaningless Rank, color (spectrum)...
Interval	Ordered, differences are meaningful, no "Natural Zero" e.g. temperature
Ratio	Just like interval, but with a natural zero. e.g. amount of money

Table 1.5: Measurements

1.5 Design of Experiments/Observations

1.5.1 Observation vs. Experiment

An **observation** measures specific traits, but noes **not** modify subjects.

An **experiment** applies a treatment and then measuers the effect on the subjects.

1.5.2 Random

Each member of a population, has an equal chance of being selected in a sample.

Simple random sample

Each group of size 'n' has an equal chance of being selected.

1.5.3 Common techniques to get a sample

Table 1.6: 4 Common techniques to get a sample

Convenience sample	You use the results, which you easily get (not random)
Systematic sampling	Put a population in some order and select every "k^{th}" member.
Stratafied Sample	Breaking population into sub-groups based on some characteristic, and then take a simple random sample out of each sub-groups.
Cluster sample	Divide population into "clusters" (regardless of characteristic), randomly select a certain number of clusters, and then collect data from the entire cluster.

Chapter 2

Frequency Distribution

A frequency distribution is a list of values with corresponding frequencies.

Class width	Difference between two "lower class limits"
Lower class limit	Smallest value belonging to a class
Upper class limit	Highest value belonging to a class

Table 2.1: Frequency Distribution Terms

Steps:

1. Determine number of classes: 8
2. class width:

$$\frac{\text{Max Value - Min value}}{\text{number of classes}} \rightsquigarrow \frac{44-18}{8} \rightsquigarrow \frac{26}{8} \rightsquigarrow 3.25$$

   Round **up**. $\rightsquigarrow 4$
3. Start with the minimum value: 18
4. Creat classes with class width (4)
5. Find the class midpoint:

$$\frac{\text{upper class limit} + \text{lower class limit}}{2} \rightsquigarrow .$$

6. Class boundaries: used to seperate classes without gaps.

class width: 4

Lower class limit: 18, 22, 26, . . . 46

upper class limit: 21,25 . . . 49

class midpoint:

$$\frac{\text{upper class limit} + \text{lower class limit}}{2}$$

$\rightsquigarrow 19.5, 23.5, 27.5, 31.5, 35.5, 39.5, 43.5, 47.5$

class-width inbetween

class boundaries: Used to seperate classes without gaps. 17.5, 21.5, 25.5, 29.5, 33.5, 37.5, 41.5, 49.5

Relative frequency distribution: Percentage

$$\frac{\text{class } f.}{\sum f.(n)}$$

Cumulative Frequency Distribution Adds sequential classes together.

Age	Freq.	Rel. Freq.	Cum. Freq.
18-21	25	58.1%	25
22-25	10	23.3%	35
26-29	4	9.3%	39
30-33	2	4.7%	41
34-37	1	2.3%	42
38-41	0	0%	42
42-43	1	2.3%	43
46-49	0	0%	43
	n=43 $\sum f \uparrow$	100%	

Table 2.2: Frequency Distribution

2.1 Touching Bar Chart

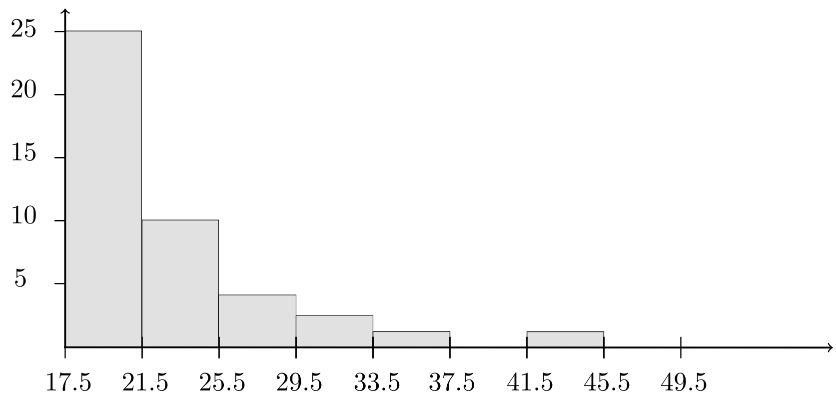


Figure 2.1: figures/stats-1

A cumulative chart would look exactly the same, but instead of having boundaries numered it'd be in the middle of the bars with teh cumulative frequency from class 1-8. And also the y-axis would be the percentage.

There is also a last one, where one takes the cumulative stuff, so that the graph colums are getting bigger and bigger...

Horizontal: Class midpoints or boundaries.

Vertical: Frequency.

Chapter 3

Describing Data

3.1 5 Caracteristics

1. Center
2. Variation
3. Distribution
4. Outliers
5. Changes over time.

Center The "middle: of the data set. 3 ways:

1. **mean:** Arthimetric Average, add all the values and divide by the numbers you added.

$$\begin{aligned} \text{Mean} &= \frac{\sum x}{\text{Number of values}} \\ \sum &= \text{sum} \\ x &= \text{data value} \\ n &= \text{number of items in a sample} \\ N &= \text{Number of items in a population} \\ \overline{x} &= \text{sample mean} \\ \mu &= \text{population mean} \end{aligned}$$

We can write the sameple mean then as:

$$\overline{x} = \frac{\sum x}{n}.$$

And the population mean as

$$\mu = \frac{\sum x}{N}.$$

Example

Sample data: 5.40, 1.10, 0.42, 0.73, 0.48, 1.10
$\overline{x} = \frac{\sum x}{n}$ is the formula we have to use, because it's a sample, then we get:

$$\overline{x} = \frac{5.40 + 1.10 + 0.42 + 0.73 + 0.48 + 1.10}{6} = \frac{9.23}{6} = 1.54.$$

2. **Median:** The middle value of the dataset.
   - Must be in order.
   - Find middle value.
     - If odd number of values, the median is the middle number.
     - If even number of values, the median is the **mean** of the two middle values.

   Example

   8, 3, 5, 11, 13, 4, 6
   To find the median we first need to order em, so:
   3, 4, 5, 6, 8, 11, 13.
   We have seven values so we can just take the middle one which is 6.
   If we'd then add 412, so our numbers are:
   3, 4, 5, 6, 8, 11, 13, 412.
   Then our median is: $M = \frac{6+8}{2} = 7$

   And it's obviously the same with decimals.

   The Median is **not** affected by outliers, the mean is.

3. **Mode:** The most commonly ocurring data value.

Example

(a) 5.40, 1.10, 0.42, 0.73, 0.48, 1.10
Here the mode is 1.10 becuase it's ocuring most often.
(b) 27, 27, 27, 55, 55, 55, 88, 88, 89
Modes: 27, 55
(c) 1, 2, 4, 7, 9, 10, 12
Mode: $\emptyset$

One rounds always to one more value than the beginning values, so one more decibel, and rounded is not before the most final step.

3.2 Mean of a frequency distribution

Age	freq.	x (midpoint)	$freq. \cdot x$
21-30	28	25.5	714
31-40	30	35.5	1065
41-50	12	45.5	546
51-60	2	55.5	111
61-70	2	65.5	131
71-80	2	75.5	151
	n=76		$\sum f \cdot x = 2718$

Table 3.1: Another age distribution

So now we can get the sample mean:

$$\overline{x} = \frac{\sum f \cdot x}{n} = \frac{2718}{76} = 35.76.$$

And here another table, this time about a grade's distribution.

	w	points	$x \cdot w$
Hw	15%	70	10.5
T_1	20%	90	18.0
T_2	20%	68	13.6
T_3	20%	85	17.0
F	25%	95	23.75
			$\sum x \cdot w = 82.85$

Table 3.2: Grade example

$$\overline{x} = \frac{\sum x \cdot w}{\sum \cdot w} \to \frac{82.85}{100} = .8285 \to 82.85\%$$

We can also do the same just half way in the class with the following table:

	w	points	$x \cdot w$
Hw	15%	70	10.5
T_1	20%	90	18.0
T_2	20%	68	13.6
	55%		$\sum x \cdot w = 42.10$

Table 3.3: Grade example

$$\overline{x} = \frac{\sum x \cdot w}{\sum \cdot w} \to \frac{42.10}{55} = .765 \to 76.50\%$$

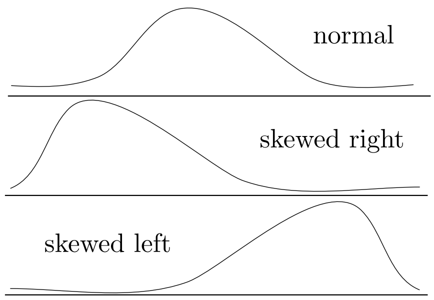


Figure 3.1: figures/stats-2

3.3 Variation

- How the data is spread.

no				$\overline{x}$
no. 1	6	6	6	6
no. 2	4	7	7	6
no. 3	1	3	14	6

Table 3.4: Bank lines (Waiting times, different strategies.

Ways to measure Variation

1. Range: Max Value - Min Value
   - easy to find
   - Does not consider all values
2. Standard deviation: Measures the average distance your data values are from the mean.
   - Never negative and never 0 unless all entries are the same.
   - Greatly affected by outliers.

Sample standard deviation is denoted "s". The formula is:

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

or

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

Here you dont need the mean.

Example

Find the standard deviation of: 1, 3, 14.

x	$x-\overline{x}$	$(x-\overline{x})^2$
1	$1-6=-5$	25
3	$3-6=-3$	9
14	$14-6=8$	64
		$\sum(x-\overline{x})^2=98$

Table 3.5: table to make it easier

So we get the standard deviation as:

$$s=\sqrt{\frac{98}{3-1}} \to s=\sqrt{\frac{98}{2}} \to s=\sqrt{49}=7.$$

Now we take the other formula:

x	x^2
1	1
3	9
14	196
$\sum x=18$	$\sum(x^2)=206$

Table 3.6: another standard deviation

$$s=\sqrt{\frac{3\cdot 206-(18)^2}{3(3-1)}}=\sqrt{\frac{618-324}{3\cdot 6}}=\sqrt{\frac{294}{6}}=7$$

Example

Do standard deviation on 4, 7, 7.

One first needs the superior formula to solve it:

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

Then it's easier when making a table to solve for some things:

x	x^2
4	16
7	49
7	49
$\sum x=18$	$\sum(x^2)=114$

So now we just do it inside the function to get the solution:

$$s=\sqrt{\frac{3\cdot 114-(18)^2}{3(3-1)}}=\sqrt{3}\approx 1.73.$$

3.3.1 Standard deviation for a population

The differences to a sample deviation are firstly the following notations, $n \rightsquigarrow N$ and $\overline{x} \rightsquigarrow \mu$.

To get sigma (σ) we do the following:

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}.$$

3.3.2 Variance

The number you have before you take the squareroot of the deviation formula.

Sample Variance: s^2

Population Variance: σ^2

So as an example, when s is $\sqrt{49}$ it is 49. Or when s is 7 then you got to squre it so you get 49.

some general properties

- closely grouped data will have a small standard deviation.
- spread-out data will have a large standard deviation.

3.3.3 empiric rules

If a data set is normally distributed, we can use the **empirical rule**.

- 68% of the data will fall within 1 standard deviation of the mean.
- 95% of the data will fall within 2 standard deviations of the mean.
- 99.7% of the data will fall within 3 standard deviations of the mean.

If a data value lies within 2 standard deviations of the mean, it's considered "normal". A data values outside of 3 standard deviations is very rare ($\frac{3}{1000}$).

sample Heights are normally distributed with a mean of 65 and a standard deviation of 3.

Example

mean is 34kg, standard deviation 8kg. What percent of data will fall between 10kg and 58kg? to the left one s.d. is 26, two are 18 and three are 10. to the right three are 58.
So with the some empiric rules we get 99.7% of the sample within our three standard deviations left and right of the mean.

	$\overline{x}$	s
Height	65	3
Weight	175	7

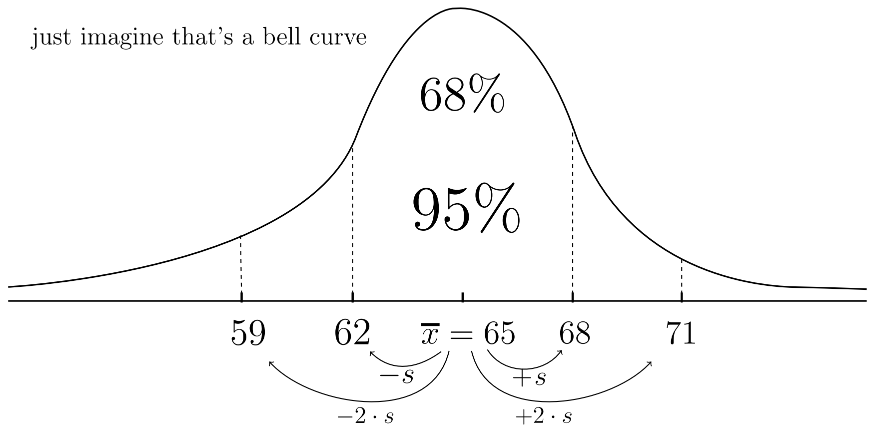


Figure 3.2: figures/stats-3

Coefficient of Variation:

$$C.V. = \frac{s}{x} \cdot 100\%.$$

So in the top example that would be:

$$\frac{3}{65} \cdot 100\% = 4.6\%$$

$$\frac{4}{175} \cdot 100\% = 23\%.$$

3.4 Measures of relative standing

(Comparing measures between or within data sets)

Z-Score The number of standard deviations a data value (x) is away from the mean ($\overline{x}$) for sample:

$$Z = \frac{x - \overline{x}}{s}.$$

for population:

$$Z = \frac{x - \mu}{\sigma}.$$

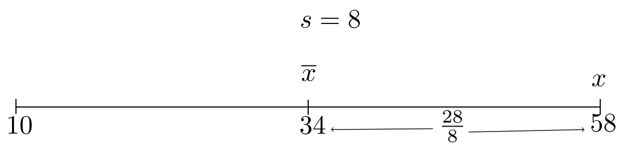


Figure 3.3: figures/stats-4

> Allows comparison of the variation in two different sample/population.

Example

Who is relatively taller? (use z-score for populations)

1. LBJ's height → 76"
   Mean for presidents → 71.5"
   Std. Dev. → 2.1"

2. Shaq's height → 86"
   Mean for Heat → 80.0"
   Std. Dev. → 3.3"

First the president:

$$Z = \frac{x - \mu}{\sigma} = \frac{76 - 71.5}{2.1} = 2.14.$$

And the basketball players:

$$Z = \frac{x - \mu}{\sigma} = \frac{86 - 80.0}{3.3} = 1.82.$$

The Z-Score shows how many standard deviations the individual is away from the mean, saying how "rare" it it. Given from the empiric principles we know 68% will fall within 1, 95% will fall within 2, and 99.7% will fall within a Z-score of 3.
That means a Z-score between -2 and 2 is considered "usual". If it isn't within, it's unusual.
So we can compare if LBJ's height is rarer within presidents than shaq's within the Heats (team). So we can just read the Z-Score, LBJ is 2.14 away from the mean and Shaq "just" 1.82, so we can say shaq's height is less rare within his team than LBJ's within presidents.

The larger the Z-Score, in terms of absolute value, the rarer the data value.

3.4.1 Quartiles

1st quartile: Q_1 → bottom 25% of **sorted** data.

2nd quartile: Q_2 (Median M) → bottom 50% of **sorted** data.

3rd quartile: Q_3 → bottom 75% of **sorted** data.

Example

1, 3, 6, 10, 15, 21, 28, 36

1. first, find median (Q_2).
   We have eight values, so we have to find what's between value 4&5 to find it.
   Value four is 10, and five is 15, so we take what's in between (12.5).
2. next we find the other Quartiles, Q_1 is between 3&6, so 4.5, to find Q_3 we need to find the mid of 21&28, which is 24.5.

If we'd add a number, so 1, 3, 6, 10, 15, 21, 28, 36, 39:

1. median can be picked in the mid (15).
2. if we now get quartiles, we act like the median 15 isn't there, so it's just there to split our data.
   Further the quartiles are then $Q_1 = 4.5$ and $Q_2 = 32$.

3.4.2 Percentiles

Percentiles: Seperates data into 100 parts.
Therefore, there are 99-Percentiles.

Percentile of x:

$$\frac{\text{Number of values less than x}}{\text{total number of values}} \cdot 100$$

Example

You scored $\frac{87}{100}$ on a test, 39 people scored lower than you, there are 54 People in the class. What percentile did you score in?

$$\frac{39}{54} \cdot 100 = 72^{nd} \text{ percentile.}$$

normally written P_{72}

$P_{25} = Q_1$, $P_{50} = M$, $P_{75} = Q_3$,

3.4.3 IQR

IQR: Is $Q_3 - Q_1$, middlt 50% of the data.

3.4.4 Box Plot

A boxplot is like a summary, it shows the following five:

1. Minimum
2. Q_1
3. Median
4. Q_3
5. Maximum

With the figure it should be obvious.

Figure 3.4: figures/stats-5

Example

1, 4, 5, 5, 7, 9, 12, 13, 13, 15, 21.

1. Minimum = 1
2. $Q_1 = 5$
3. Median = 9
4. $Q_3 = 13$
5. Maximum = 21

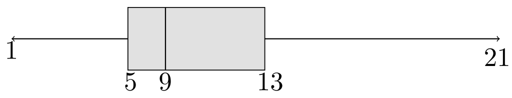


Figure 3.5: figures/stats-6

There are obviously ways to find an outlier:

Find IQR $\rightarrow$ 13-5=8

$IQR \cdot 1.5 \rightarrow 1.5 \cdot 8 = 12$

And lastly subtract it from Q_1 and add it to Q_3:

$Q_1 - 1.5 \cdot IQR = -7$

$Q_3 + 1.5 \cdot IQR = 34$

So everything that is outside of those “borders” is mathematically considered an outlier.

Chapter 4

Probability

- Low probability = Rare/Unusual occurances

4.1 Vocabulary

Event: A Collection of outcomes of a procedure.

Simple Event: A single outcome.

Sample Space: All simple events that can happen. (Every possible outcome)

Procedure	Event	Sample space
Flip coin 1 time	Head	{Head Tail}
Flip coin 3 times	1 Head, 2 Tails 3 Heads	{ H-H-H, H-H-T, H-T-H, H-T-T, T-T-T, T-H-H, T-T-H T-H-T T-H-H} *Here the event is three times. once there.*

Table 4.1: ex 1

4.2 Probability

The likelyhood of an event occuring.

Denoted P

Event: $A, B, C \dots$

$P(A)$= The Probability of event "A" happening.

4.2.1 3 types

1. **Observed probability:** Probability that is estimated based on observations.

$$P(A) = \frac{\text{Number of times "A" occured}}{\text{Number of times the procedure was repeated}}.$$

(What did happen)

2. **Classical probability:** Probability based on the **chance** of an event occuring. (Each event must have an equal chance of occuring)

$$P(A) = \frac{\text{Number of ways "A" could occur}}{\text{Number of simple evens (outcomes)}}.$$

(What should happen.)

3. **Subjective probability:** Educated guess.

4.2.2 Examples

Example

The probability of selecting a ♡ from a standard deck of cards

$$P(\heartsuit) = \frac{13}{52} = 0.25\,(\text{classical} \rightsquigarrow \text{should happen})$$

Example

Flip coin 100 times. You get 64 Tails.

$$P(T) = \frac{64}{100} = .64\,(\text{Observed} \rightsquigarrow \text{did happen})$$

Example

Peyton completed 385 out of his first 525 passes.
Find probability that Peyton will complete a pass.

$$P(\text{Completing a pass}) = \frac{385}{525} \approx 0.729 \text{ or } 72.9\%.$$

This is observed.

Example

Random deck of cards, probability of selecting a two.

$$P(2) = \frac{4}{52} = 0.76 \to 7.6\%.$$

This is classical.

Example

poll about some cloning:
91 people ... cloning good
901 people ... cloning bad
20 people ...
find probability of randomly selecting a person on the street that says cloning is a good idea.

$$p(good) = \frac{91}{91 + 901 + 20} = \frac{91}{1012} = 0.09 = 9\%.$$

the probability to find a person that says cloning good it 9%.

Example

find the probability that a bird will poop on your car today. **subjective**

Example

find probability that if a couple ha three kids, two will be boys. (assuming equal chance of boy/girl)
procedure: having 3 children.
event: 2 boys, 1 girl.
(classical theory.)
sample space: {bbb, **bgg, bgb,** bgg, ggg, ggb, gbg, **gbb**}

$$P(2B, 1G) = \frac{3}{8} = 37.5\%$$

- Probabilities are always between 0 and 1.
  - $P = 0 \rightarrow$ Impossible event.
  - $P = 1 \rightarrow$ Certain event.
- The more a procedure is repeated. The closer observed probability will get to classical prob.

4.2.3 python example

You can nicely test it with the following python script.

```
import random

def roll_dice():
    return random.randint(1,6)

def main():
    n = int(input("How many times should the dice be rolled? "))
    rolls = []
    for i in range(n):
        rolls.append(roll_dice())
    for i in range(1,7):
        print("The probability of the number", i, "is", rolls.count(i)/n)

main()
```

4.2.4 Complementary Events

Events which are mutually exclusive (can't happen at the same time).

Complement: The complement of event "A" is denoted $\overline{A}$ and is all the outcomes when event "A" does not occur.

Example

Event: Rolling a 5
Complement: Not Rolling a 5 (1,2,3,4,6)

$$P(5) = \frac{1}{6} P(\overline{5}) = \frac{5}{6} \quad P + \overline{P} = 1.$$

The probability of an event plus the probability of the complement must equal 1.

4.3 Addition Rule

Compound event: An event which joins two more more simple events.

Example

Probability of rolling a ⚀ or a ⚄
"or" $\rightarrow$ one or the other, or both.
⚅ + ⚀ $\rightarrow$ 7
⚅ + ⚄ $\rightarrow$ 11

$P(A\ or\ B)$ = Prob. of "A" occuring or "B" occuring or both "A" and "B" occuring in a single trail.
P(Blonde or Female)
$\uparrow$ not mutually exclusive.

	Didn't do it	Did it
Guilty	11, false positive	72, true positive
Not guilty	85, true negative	9, false negative

Table 4.2: Table for the following example.

Example

How many people are "guilty" or "did it"?
Not mutually exclusive.

$$11 + 72 + 9 = 92$$

$$P(\text{Guilty or did it}) = \frac{92}{177} = .520 = 52\%.$$

4.3.1 Addition Rule

P(A or B) Requires elimination of any "double count".

$$P(\text{A or B}) = P(A) + P(B) - P(\text{“}A\text{” and “}B\text{”}) \quad (\textit{in a single try.}).$$

Disjoint Events

Events which are **mutually exclusive**, means $\rightarrow$ can't happen at the sam time.

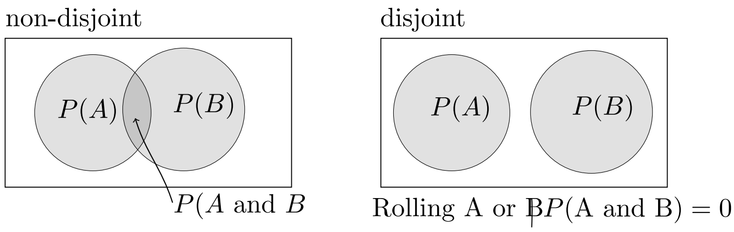


Figure 4.1: figures/stats-7

Example

Probability of selecting a "heart" or a "spade".

$$
\begin{aligned}
P(\heartsuit \text{ } or \text{ } \spadesuit) &= P(\heartsuit) + P(\spadesuit) - P(\heartsuit \text{ } and \text{ } \spadesuit) \\
&= \frac{13}{52} + \frac{13}{52} - 0 \\
&= \frac{26}{52} \\
&= \frac{1}{2} = .50
\end{aligned}
$$

Example

Probability of "blond" or female:
Blonde: .18
Female: .60
Blonde and Female: .12

$$
\begin{aligned}
P(\text{Blond or Female}) : .12 &= .18 + .66 - .12 \\
&= 0.66
\end{aligned}
$$

Example

$$
\begin{aligned}
P(\diamondsuit \text{ } or \text{ } K) &= P(\diamondsuit) + P(K) - P(\diamondsuit \text{ and } K) \\
&= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} \\
&= \frac{16}{52} \approx 0.310
\end{aligned}
$$

4.3.2 Complementary Events

$$P(A \text{ } or \text{ } \overline{A}) = P(A) + P(\overline{A}) - P(A \text{ } and \text{ } \overline{A}) = 1.$$

Example

$P(Girls) = 0.512$, what's then the probability to have a boy?

$$P(Boy) = P(\overline{g}) = 1 - P(Girl).$$

023

Lecture 2: 4.4

$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B) \leftarrow \text{(Both at them same time)}.$$

4.4 Multiplication rule

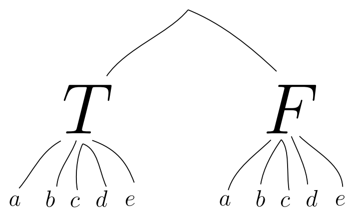
$P(A\ and\ B) =$ Probability of "A" occuring, **and then** "B" occuring in **successive trails**.

Example

Two questions

1. (T/F): Mr. Leonard drives an audi.
2. (Multiple choice): My favourite color is:
   - red
   - blue
   - green
   - black
   - yellow

If you guess randomly, what's the probability that you will get the right answer on both problems?



So the probability is $\frac{1}{10}$.

Prob. of selecting guily and then not guity as in an earlier table (4.2). (without replacment)

$$P(Guilty) = \frac{83}{177} \text{ and then } P(Not\ guilty) = \frac{94}{176}.$$

4.4.1 Conditional probability

The probability of an event occurring given that some other event has already occured.

$P(A|B) =$ The probability of event "B" occurring **given that** "A" has already occurred.

Independent events

The occurrance of one event does **not** affect the occurrance of another event. (Non-independent events are Dependend)

If A&B are independent,

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

Example

1. Roll a die: **Independent**

$$P(2|3) = P(2) = \frac{1}{6}.$$

2. Drawing cards:

$$\text{with replacment:} \quad P(Q|9)\frac{4}{52} = \frac{1}{13} \approx 0.077.$$

$$\text{without replacment:} \quad P(Q|9) = \frac{4}{51} \approx 0.078.$$

3. Another card one:

$$\text{without replacment:} \quad P(Q|Q) = \frac{3}{51} \approx 0.059.$$

$$\text{with replacment:} \quad P(Q|Q)\frac{4}{52} \approx 0.77.$$

4. with replacment:

$$P(\heartsuit|J\diamondsuit) = \frac{13}{52}.$$

5. without

$$P(\heartsuit|J\diamondsuit) = \frac{13}{51}.$$

4.4.2 Multiplication Rule

$$P(A \textit{ and } B) = P(A) \cdot P(B|A).$$

Recall: if independent $P(B|A) = P(B)$ so $P(A \textit{ and } B) = P(A) \cdot P(B)$.

Example

Bag of marbles
in the bag are 9 marbles:

- 3 red
- 2 blue
- 4 green

with replacement:

$$P(G \text{ } and \text{ } B) = P(G) \cdot P(B|G) = \frac{4}{9} \cdot \frac{2}{9} = \frac{8}{81}.$$

without replacement:

$$P(G \text{ } and \text{ } B) = P(G) \cdot P(B|G) = \frac{4}{9} \cdot \frac{2}{8} = \frac{1}{9}.$$

without replacement:

$$P(R \text{ } and \text{ } R) = \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{12}.$$

without replacement:

$$P(B \text{ } and \text{ } B \text{ } and \text{ } B) = \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} = \frac{1}{21}.$$

Dice: $P(1 \text{ } and \text{ } 2 \text{ } and \text{ } 3 \text{ } and \text{ } 4) = \frac{1}{6}^4 = \frac{1}{1296} = 0.00077$

Cards: without replacement

$P(A \text{ } and \text{ } K \text{ } and \text{ } Q \text{ } and \text{ } J \text{ } and \text{ } 10) = \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50} \cdot \frac{4}{49} \cdot \frac{4}{48}$

023

Lecture 3: 4.5

4.5 Complementary events, at least one

"At least one" Means one or more.

- the **complement** of "at least one" is none

$P(\text{“}at\ least\ one\text{”}) = 1 - P(\text{“}none\text{”})$

Example

Flip coin 3 times. Prob of getting at least one head?

$$P(\textit{at least one head}) = \frac{7}{8}.$$

Sample space: {HHH, HHT, HTH, HTT, TTT, TTH, THT, THH}

$$\begin{aligned}P(\textit{at least one head}) &= 1 - P(\textit{no heads})\\ &= 1 - P(T \textit{ and } T \textit{ and } T) = 1 - \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\\ &= 1 - \frac{1}{8} = \frac{7}{8}\end{aligned}$$

Flip coin 20 times, prob of getting at least one head:

$$P = 1 - \frac{1}{2}^{20} = 1 - \frac{1}{1,0485,276} = \frac{1,048,575}{1,048,576} = 0.999999046.$$

Less than $P = 0.05$ means something is "rare".

$P(B|A)$ = probability that B happens given that A already happened. $P(B \textit{ and } A)$ = probability that B and then A.

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

$$P(A \textit{ and } B) = P(A)\cdot P(B|A).$$

Again refering to guilty table 4.2.

$$P(\textit{Guilty}|\textit{did it}) = \frac{72}{81}.$$

$$P(\textit{did it}|\textit{guilty}) = \frac{72}{83}.$$

Lecture 4: 4.7 03-06-2023

4.6 Counting

Coin & Die: Outcomes: two for coin, six for die. Thoes together have 12 outcomes.

4.6.1 Fundamental counting rule

If you have an event that can occur "m" ways and another event that can occur "n" ways, then together they can occur $m \cdot n$ ways.

Example

Four number security code

- 1^{st} digit can't be zero.

$$9 \cdot 10 \cdot 10 \cdot 10 = 9000.$$

probability of guessing: $\frac{1}{9000}$

If you have different arangements, which are all unique it's like the following. Given you have 5 unique items, you can have $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ different arrangements. In conclusion, if you have 5 unique things, there are 5! constelations, if you have 10 unique things there are 10! constelations.

- Remember: $0! = 1$
- for any set of "n" **different** items, there are $n!$ different arrangements possible.

Example

There are 7 cool rides at D.L., how many different ways could you write all seven?
$7! = 5040$

023

Lecture 5: 5.0

Chapter 5

Discrete Probabilities

→ Countable of finite.

5.1 Vocabulary

Random variable	A variable, x, that has a value for each outcome of a procedure, that is determined by chance.
Probability distribution	Table that gives the probability for each value of a random variable.
Discrete random variable	A variable with a countable or finite number of values. (Chapter 5)
Continous random variable	A variable with an infinit amount of values. (usually measurements) (Chapter 6)

Table 5.1: Vocabulary

Example

Probability distribution:

x	$P(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

5.2 Histogram from Probability Distribution.

Horizontal	Values of random variable.
Vertical	Probability.

Example

Weighted die:

x	$P(x)$
1	.05
2	.15
3	.35
4	.30
5	.10
6	.05

Notitz:

$$0 \leq P(x) \leq 1.$$
$$\sum P(x) = 1.$$

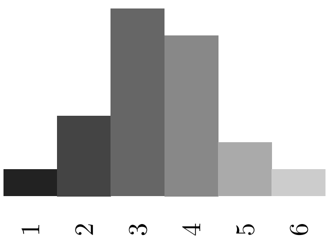


Figure 5.1: Prob. distribution of the weighted die

5.2.1 Mean, Variance, Standard Deviation

Mean:

$$\mu = \frac{\sum(x \cdot f)}{N}$$
$$\mu = \sum \left[\frac{x \cdot f}{N}\right]$$
$$= \sum \left[x \cdot \frac{f}{n}\right]$$
$$\mu = \sum [x \cdot P(x)] \leftarrow \text{mean, expected value}$$

x	$P(x)$	$x \cdot P(x)$	x^2	$x^2 \cdot P(x)$
1	.05	.05	1	.05
2	.15	.30	4	.60
3	.35	1.05	9	3.15
4	.30	1.20	16	4.8
5	.10	.50	25	2.5
6	.05	.30	36	1.8
	1.00	$\sum[x \cdot P(x)] = 3.4$		$\sum[x^2 \cdot P(x)] = 12.9$

So the mean is: $\mu = 3.4$

Variance: $\sigma^2 = \sum \left[x^2 \cdot P(x)\right] - \mu^2$

with a little solving of the table we then get:

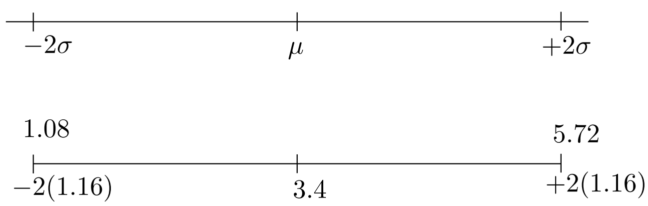
$$\begin{aligned}
\sigma^2 &= 12.9 - (3.4)^2 \\
\sigma^2 &= 12.9 - 11.56 \\
\sigma^2 &= 1.34 \\
\sigma &= \sqrt{1.34} \\
\sigma &= 1.16
\end{aligned}$$

remember standard deviation is just the sqareroot of the variance, as 1.16 above.

5.2.2 Usual vs. Unusual

Values are unusual if they lie outside of:

- $\mu + 2\sigma$
- $\mu - 2\sigma$



Here we took the dice example, and 1.08 as 5.72 mark our range of $\pm$ two standard deviations, whatever's outside is considered unusual. So we get 1 & 6 to be unusual.

If $P(A) \leq$ .05 "A" is considered unusual.

Example

Flip a coin 1000 times

$$P(\text{exactly 501 heads}) = 0.0252 \leq .05 \Rightarrow \textit{ unusual}$$

$$P(\text{501 or more heads}) = 0.487 \Rightarrow \textit{ usual}$$