Review questions for the first exam

The solutions

Nikos Apostolakis

1. For the following data
$$47 \quad 49 \quad 50 \quad 51 \quad 52 \quad 53 \quad 56 \quad 56 \quad 57 \quad 59.$$
   (a) Find the sample mean x.
   (b) Find the sample standard deviation s.

   Solution. We compute:

	x	x^2
	47	2209
	49	2401
	50	2500
	51	2601
	52	2704
	53	2809
	56	3136
	56	3136
	57	3249
	59	3481
$\sum$	530	28226

   We have $n = 10$, so
$$\bar{x} = \frac{\sum x}{n} = \frac{530}{10} = 53.$$
   For the variance of a sample we have the formula:
$$\begin{aligned} s^2 &= \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \\ &= \frac{28226 - \frac{(530^2)}{10}}{9} \\ &= \frac{28226 - \frac{280900}{10}}{9} \\ &= \frac{28226 - 28090}{9} \\ &= \frac{136}{9} \\ &= 15.1111. \end{aligned}$$
   And so
$$s = \sqrt{15.1111} \approx 3.887.$$
$\square$

2. Find the sample standard deviation for the data

x	1	2	3	4	5	6	7	8	9	10
f	384	208	98	56	28	12	8	2	3	1

   Solution. We will use the computational formula for frequency tables:
$$s^2 = \frac{\sum x^2\, f - \frac{(\sum x\, f)^2}{\sum f}}{\sum f - 1}.$$
   We compute, new rows for $x\,f$, x^2, and $x^2\,f$ and the sums.

											Σ
x	1	2	3	4	5	6	7	8	9	10	
f	384	208	98	56	28	12	8	2	3	1	800
$x\,f$	384	416	294	224	140	72	56	16	27	10	1639
x^2	1	4	9	16	25	36	49	64	81	100	
$x^2\,f$	384	832	882	896	700	432	392	128	243	100	4989

We see then that

$$\sum f = 800$$
$$\sum x^2\,f = 4989$$
$$\sum x\,f = 1639$$

and so

$$\left(\sum x\,f\right)^2 = (1639)^2 = 2686321, \qquad \frac{(\sum x\,f)^2}{\sum f} = \frac{2686321}{800} = 3357.90125.$$

Substituting into the formula we find

$$s^2 = \frac{4989 - 3357.90125}{800 - 1} = \frac{1631.09875}{799} \approx 2.04,$$

and so

$$s = \sqrt{s^2} \approx 1.43.$$

□

3. The following data represent the duration (in days) of U.S. space shuttle voyages for the years 1992-94.

$$8 \quad 9 \quad 9 \quad 14 \quad 8 \quad 8 \quad 10 \quad 7 \quad 6 \quad 9 \quad 7 \quad 8 \quad 10 \quad 14 \quad 11 \quad 8 \quad 14 \quad 11$$

(a) Find the mode.

(b) Find the the median.

(c) Find the first and the third quartile.

(d) What is the percentile rank of 7?

Solution. We first put the data in order. We have

$$6 \quad 7 \quad 7 \quad 8 \quad 8 \quad 8 \quad 8 \quad 8 \quad 9 \quad 9 \quad 9 \quad 10 \quad 10 \quad 11 \quad 11 \quad 14 \quad 14 \quad 14.$$

(a) The most frequent value is 8 that occurs 5 times. Therefore the mode is 8.

(b) We have a sample of size 18, an even number. So the median is the average of the two middle values at positions 9 and 10. Both of numbers are 9, and therefore the median is $q_2 = 9$.

(c) The first quartile is the median of the lower half of the data, that is the first 9 values. That is the fifth number, and the first quartile is $q_1 = 8$.
The third quartile is the the median of the upper half of the data, that is the first 9 values. That is the fourteenth number, and the first quartile is $q_3 = 11$.

(d) There are 3 numbers equal to 7 or less. We calculate $3/18 \approx 0.17$. So 17% of numbers are less or equal to 7. This means that 7 is at the 17th percentile.

□

4. The following table summarizes the exam scores of 100 students.

Score	Frequency
49.5 – 59.5	5
59.5 – 69.5	10
69.5 – 79.5	30
79.5 – 89.5	40
89.5 – 100	15

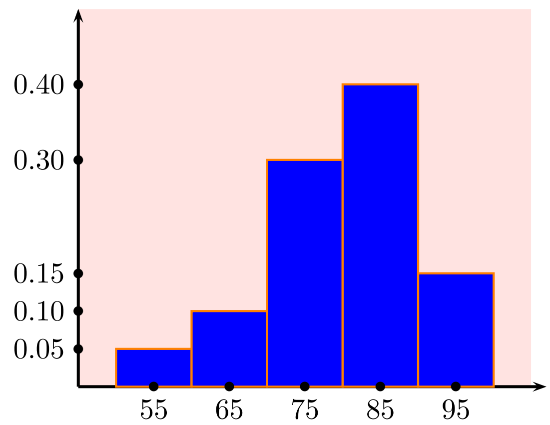
(a) Find the relative frequency of each grade range.

(b) Construct a relative frequency histogram.

Solution. (a) To find the relative frequency we divide the frequency by the size. There are 100 students so we will divide by 100. We get

Score	Frequency	Relative Frequency
$49.5 - 59.5$	5	0.05
$59.5 - 69.5$	10	0.1
$69.5 - 79.5$	30	0.3
$79.5 - 89.5$	40	0.4
$89.5 - 100$	15	0.15

(b) We have the following relative frequency histogram. We put the name each class by its midpoint.



□

5. Anna and Benjamin took the same Statistics course, Anna in the fall, Benjamin o in the spring. Anna's score on the final exam was79, and on that exam the mean was 72 and the standard deviation 7. Benjamin's score on the final exam that he took was 82, and on that exam the mean was 74 and the standard deviation 10.

(a) Who did relatively better, Anna or Benjamin?

Solution. We compare the z–scores. Anna's z–score is

$$z = \frac{79 - 72}{7} = 1,$$

while Benjamin's is

$$z = \frac{82 - 74}{10} = 0.8.$$

Since Anna's z–score was higher, she did relatively better. □

(b) What score in Anna's version of the exam, does Benjamin score correspond to?

Solution. For Anna's version of the exam a z–score of 0.8 corresponds to the raw score

$$x = 72 + 0.8 \cdot 10 = 80.$$

□

6. The distribution of the scores of the MTH 23 final exam over the last 10 years is roughly bell shaped and has a mean of 72 and the standard deviation 6.

(a) Approximately what percentage of students scores above 84 in the MTH 23 final?

Solution. The raw score of 84 corresponds to the z–score of

$$z = \frac{84 - 72}{6} = 2.$$

Since the distribution of scores is bell shaped we can apply the Empirical Rule and conclude that about 95% of z–scores are between -2 and 2, and therefore about 5% is not in that interval. Since bell shaped curves are symmetrical about the mean we know that this 5% is evenly distributed between the two ends. So about 2.5% of students have z–scores above 2. So about 2.5% of students scores above 84. □

(b) If 200 students take the final exam this semester, about how many will score below 66?

Solution. The z–score of 66 is

$$z = \frac{66 - 72}{6} = -1.$$

From the Empirical Rule we conclude that about 68% of students have z–scores between -1 and 1, and so abut 34% lie outside that interval. Again by symmetry we conclude that half of those z–scores are below -1. So about 17% of students scores below 66.
Since there are 200 this means that about

$$0.17 \cdot 200 = 34$$

students will score below 66. □

7. In a class of 80 students the professor announced that the mean numerical grade was 74 with a standard deviation of 6. In the chat group of the class 10 different students claimed that they got an A+, which means numerical grade between 93 and 97.

   (a) Explain why some of them are lying about their grade.

   Solution. By Chebyshev's Theorem we have that *at least* 88.89% of students have z–scores between -3 and 3. In our case $\mu = 74$ and $\sigma = 6$, so $3\sigma = 18$ so $z = -3$ corresponds to $x = 74 - 18 = 66$, and $z = 3$ corresponds to $x = 74 + 18 = 92$.

   So we know that at least 88.89% of students scored between 66 and 82. Since there were 80 students, 88.89% means $0.8889 \cdot 80 = 71.112$ students scored between 66 and 82. Of course the numbers of students is a whole number and it has to be more than 71.111. So we know that at least 72 students scored between 66 and 92. So *at most* 8 students scored above 92. It's impossible to have 10 students with scores in the range of 93 to 97. Therefore some of those 10 students are lying. ☐

   (b) At least how many are lying?

   Solution. Since at most 8 students scored above 92, we know that at lease 2 of those 10 students are not telling the truth. ☐

8. An animal shelter has a 60% adoption rate for puppies. Of all puppies in the shelter, 75% live to be 6 years or older. Of the puppies who are adopted, 87% live to be 6 years or older.

   (a) What is the probability that a randomly selected puppy in the shelter will get adopted **and** live 6 or more years?

   Solution. Let A be the event "the puppy gets adopted" and L the event "the puppy lives six or more years". We are given:

   $$p(A) = 0.6, \quad p(L) = 0.75, \quad p(L \textbf{ given } A) = 0.78.$$

   We use the formula

   $$p(A \textbf{ and } L) = p(A)\, p(L \textbf{ given } A).$$

   We get

   $$p(A \textbf{ and } L) = 0.78 \cdot 0.87 = 0.522.$$

   ☐

   (b) What is the probability that a randomly selected puppy in the shelter will get adopted **or** live 6 or more years?

   Solution. We will use the formula

   $$p(A \textbf{ or } L) = p(A) + p(L) - p(A \textbf{ and } L).$$

   We have

   $$p(A \textbf{ or } B) = 0.6 + 0.75 - 0.522 = 0.828.$$

   ☐

9. Maria is applying for a job. The application consists of two steps. In the first step she has to submit a written application, and then to be interviewed by a hiring committee. 65% of the written applications are approved, and 60% of the applicants pass the hiring committee interview. We also know that 80% of those whose written application has been approved, pass the interview by the hiring committee.

   (a) What is the probability that Maria's written application is approved **and** she passes the hiring committee interview?

   Solution. Let A be the event that Maria's written application is approved, and I that she passes the hiring committee interview. We are given:

   $$p(A) = 0.65, \quad p(I) = 0.60, \quad p(I \textbf{ given } A) = 0.8.$$

   So as in the previous exercise we compute:

   $$p(A \textbf{ and } I) = 0.65 \cdot 0.80 = 0.52.$$

   ☐

   (b) What is the probability that Maria's written application is approved **or** she passes the hiring committee interview?

Solution. Again as in the previous exercise we have:

$$p(A \textbf{ or } I) = 0.64 + 0.6 - 0.52 = 0.72.$$

□

10. A tourist who speaks English and German but no other language visits a region of Slovenia. If 35% of the residents speak English, 15% speak German, and 3% speak both English and German, what is the probability that the tourist will be able to talk with a randomly encountered resident of the region?

Solution. The tourist will be able to talk with a randomly encountered resident of the region if that person speaks English or German. If E stands for the event that the randomly encountered resident speaks English, and G for the event that the randomly encountered resident speaks English, we are given

$$P(E) = 0.35, \quad P(G) = 0.15, \quad P(E \textbf{ and } G) = 0.03.$$

Therefore

$$P(E \textbf{ or } G) = 0.35 + 0.15 - 0.03 = 0.47.$$

□

11. The following table relates the weights and heights of a group of individuals participating in an observational study.

	Tall	**Medium**	**Short**	**TOTAL**
Obese	18	28	14	60
Normal	20	51	28	99
Underweight	12	25	9	46
TOTAL	50	104	51	205

(a) Find the total for each row and column.

Solution. Done above. □

(b) Find the probability that a randomly chosen individual from this group is Tall.

Solution. Out of a total of 205 individuals 50 are tall. Therefore, if T stands for the event that the selected individual is tall, we have

$$p(T) = \frac{50}{205} \approx 0.244.$$

□

(c) Find the probability that a randomly chosen individual from this group is Tall **given** that the individual is Obese.

Solution. Out of a total of 60 obese individuals 18 are tall. Therefore, if O stands for the event that the selected individual is obese, we have,

$$p(T \textbf{ given } O) = \frac{18}{60} = 0.3.$$

□

(d) Are the events Obese and Tall independent?

Solution. Since the probabilities computed in parts (a) and (b) are different the events are *not* independent. For independent events the probability of t, and T **given** O should be the same. □

(e) Find the probability that a randomly chosen individual from this group is Obese **or** Tall.

Solution. To compute $p(O \textbf{ or } T)$ we need $p(T)$, $p(O)$, and $p(O \textbf{ and } T)$. We have already computed $p(T)$. We have 60 obese individuals out of 205, so

$$p(O) = \frac{18}{205} \approx 0.29.$$

From the table we see that out of 205 individuals 18 are both obese and tall. Therefore,

$$p(O \textbf{ and } T) = \frac{18}{205} \approx 0.088.$$

Therefore,

$$p(O \textbf{ or } T) \approx 0.446.$$

□

(f) Find the probability that a randomly chosen individual from this group is **not** Obese.

Solution. The probabilities of complementary events add up to 1. We know $p(O) = 0.29$. Therefore,

$$p(\textbf{not } O) = 1 - 0.29 = 0.71.$$

□

12. The sample space of equally likely outcomes for the experiment of rolling two fair dice is

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

Let N be the event "the sum is at least nine", T be the event "at least one of the dice is a two", and F be the event "at least one of the dice is a five".

(a) List the outcomes that comprise each of the events N, T, and F.

Solution. The largest possible sum we can get is 12 and so the event N consists of those outcomes that have sum 9, 10, 11, or 12. Therefore,

$$N = \{36, 45, 54, 63, 46, 55, 64, 56, 65, 66\}.$$

The event T consists of the outcomes in the second row or in the second column in the table above. Therefore,

$$T = \{21, 22, 23, 24, 25, 26, 12, 32, 42, 52, 62\}.$$

The event F consists of the outcomes in the second row or in the second column in the table above. Therefore,

$$F = \{51, 52, 53, 54, 55, 56, 15, 25, 35, 45, 65\}.$$

□

(b) Find $P(N)$.

Solution. Since all the outcomes are equally likely, the probability of N is the number of outcomes in N divided by the total number of outcomes. Counting we see that there are 10 outcomes in N while the total number of outcomes is 36. Thus

$$P(N) = \frac{10}{36} = \frac{5}{18} \approx 0.28.$$

□

(c) Find $P(N \textbf{ given } F)$.

Solution. Out of the 11 outcomes in F only 5 outcomes, namely 45, 54, 55, 56, and 65. Therefore

$$P(N \textbf{ given } F) = \frac{5}{11} \approx 0.45.$$

□

(d) Find $P(N \textbf{ given } T)$.

Solution. No outcomes from T are in N. Therefore

$$P(N \textbf{ given } T) = 0.$$

□

(e) Determine from the previous answers whether or not the events N and F are independent; whether or not N and T are.

Solution. From the calculations in parts (b) and (c) we see that

$$P(N) \neq P(N \textbf{ given } F).$$

Thus N and F are not independent events.
From the calculations in parts (b) and (d) we see that

$$P(N) \neq P(N \textbf{ given } T).$$

Thus N and T are not independent events. □

13. A manufacturer examines its records over the last year on a component part received from outside suppliers. The breakdown on source (supplier A, supplier B) and quality, (High, Usable, or Defective) is shown in the two-way contingency table.

	High	**Usable**	**Defective**	**TOTAL**
Supplier A	0.6937	0.0049	0.0014	0.7
Supplier B	0.2982	0.0009	0.0009	0.3
TOTAL	0.9919	0.0058	0.0023	1

The record of a part is selected at random. Find the probability of each of the following events.

(a) The part was defective.

Solution. We have

$$P(D) = 0.0023.$$

□

(b) The part was either of high quality or was at least usable.

Solution. The events H and U are mutually exclusive, so

$$\begin{aligned} P(H \textbf{ or } U) &= P(H) + P(U) \\ &= 0.9919 + 0.0058 \\ &= 0.9977. \end{aligned}$$

□

(c) The part was defective and came from supplier B.

Solution. From the table we have

$$P(D \textbf{ and } B) = 0.0009.$$

□

(d) The part was defective or came from supplier B.

Solution. We have

$$\begin{aligned} P(D \textbf{ or } B) &= P(D) + P(B) - P(D \textbf{ and } B) \\ &= 0.0023 + 0.3 - 0.0009 \\ &= 0.3014. \end{aligned}$$

□

14. A survey of the customers of an ice cream shop found that 70% like chocolate flavor, 45% like vanilla flavor, and 30% like both. If we randomly select a customer what is the probability that

(a) they like either chocolate or vanilla flavored ice cream.

(b) They don't like either flavor.

(c) They like vanilla flavor given that they like chocolate flavor.

Solution. Let C be the event "they like chocolate flavor" and V be the event "they like vanilla flavor". Then we are given that

$$P(C) = 0.7, \quad P(V) = 0.45, \quad P(C \textbf{ and } V) = 0.3.$$

(a) We have

$$\begin{aligned} P(C \textbf{ or } V) &= P(C) + P(V) - P(C \textbf{ and } V) \\ &= 0.7 + 0.45 - 0.3 \\ &= 0.85. \end{aligned}$$

(b) The event "they don't like either flavor" is the complement of the event C **or** V. Thus,

$$P\left(\textbf{not } (C \textbf{ or } V)\right) = 1 - 0.85 = 0.15.$$

(c) We have

$$\begin{aligned} P(V \textbf{ given } C) &= \frac{P(V \textbf{ and } C)}{P(C)} \\ &= \frac{0.3}{0.7} \\ &\approx 0.43. \end{aligned}$$

□

15. The sample space that describes all three-child families according to the genders of the children with respect to birth order is

$$S = \{\text{BBB}, \text{BBG}, \text{BGB}, \text{BGG}, \text{GBB}, \text{GBG}, \text{GGB}, \text{GGG}\}$$

In the experiment of selecting a three-child family at random, compute each of the following probabilities, assuming all outcomes are equally likely.

(a) The probability that the family has at least two boys.

Solution. Since all the outcomes are equally likely the probability of an event is the proportion of the outcomes in the event out of all possible outcomes. By counting we see that there are 8 possible outcomes.
Let A be the event "family has at least two boys". Then

$$A = \{\text{BBB}, \text{BBG}, \text{BGB}, \text{GBB}\},$$

and so A is comprised of 4 outcomes. Therefore,

$$P(A) = \frac{4}{8} = 0.5.$$

□

(b) The probability that the family has at least two boys, given that not all of the children are girls.

Solution. Let B be the event "not all of the children are girls". Then

$$B = \{\text{BBB}, \text{BBG}, \text{BGB}, \text{BGG}, \text{GBB}, \text{GBG}, \text{GGB}\},$$

and so B is comprised of 7 outcomes. All the outcomes of A are also in B, so

$$p(A \textbf{ given } B) = \frac{4}{7} \approx 0.57.$$

□

(c) The probability that at least one child is a boy.

Solution. The event "at least one child is a boy" is the the event B. Therefore

$$P(B) = \frac{7}{8} = 0.875.$$

□

(d) The probability that at least one child is a boy, given that the first born is a girl.

Solution. Let C be the event "the first born is a girl". Then

$$C = \{\text{GBB}, \text{GBG}, \text{GGB}, \text{GGG}\},$$

and is comprised of 4 outcomes. Of those outcomes 3 are in B and therefore,

$$p(B \textbf{ given } C) = \frac{3}{4} \approx 0.75.$$

□

16. Suppose for events A and B in a random experiment $P(A) = 0.70$ and $P(B) = 0.30$. Compute the indicated probability, or explain why there is not enough information to do so.

(a) $P(A \textbf{ and } B)$.

Solution. We don't have enough information. □

(b) $P(A \textbf{ and } B)$, with the extra information that A and B are independent.

Solution. Since the events A and B are independent we can use the formula

$$P(A \textbf{ and } B) = P(A)P(B).$$

We find

$$P(A \textbf{ and } B) = 0.70 \cdot 0.30 = 0.21.$$

□

(c) $P(A \textbf{ and } B)$, with the extra information that A and B are mutually exclusive.

Solution. If A and B are mutually exclusive we have

$$P(A \textbf{ and } B) = 0.$$

□