

MATH 203: PRINCIPLES OF STATISTICS 1
FALL 2024

ASSIGNMENT 1

Due date: 11.59pm Sunday 22nd September.

Please upload your solutions as a single pdf to myCourses.

1. Incidence of difficult bag-mask ventilation in children

(see *Pædiatr Anaesth.* 2013 Oct; 23(10):920-6).

Difficult airway (DA), including difficult bag-mask ventilation (DBMV), is an important challenge for paediatric anesthesiologists. While expected DBMV can be successfully managed with appropriate equipment and personnel, unexpected DBMV relies on the resources available and the experience of the anesthesiologist at the time of the emergency. The incidence and risk factors of unexpected DA in otherwise healthy children, including DBMV among paediatric patients are not known.

The aim of the study was to expand the scientific knowledge of unexpected DBMV among paediatric patients. Patients age 0-8 years, undergoing elective surgery requiring bag-mask ventilation and intubation at the Montreal Children's Hospital were recruited in this study. Data on DBMV status (presence or absence), age of the patient and other risk factors were collected.

For this study, identify

- (a) the population of interest 1 MARK
- (b) the experimental units 1 MARK
- (c) the sample 1 MARK
- (d) the variables recorded or measured for this study that are mentioned in the above text. (Hint: From the text we do not know what the 'other risk factors' are). 1 MARK

Is this an *Observational* or an *Experimental* Design? Explain your answer in no more than two sentences. 1 MARK

2. Do Mountain Bikers Know When They Have Had a Concussion and, Do They Know to Stop Riding? (*Clin J Sport Med.* 2019 Dec 31).

The objective of this study was to establish the prevalence of concussions in mountain bikers and to determine factors that increase their risk of concussion. The variables recorded in the study included

- (a) age
- (b) gender
- (c) nationality
- (d) style of riding (cross-country, downhill, or free-ride)
- (e) number of years doing mountain bike racing

For each variable, say whether it provides *qualitative* or *quantitative* data.

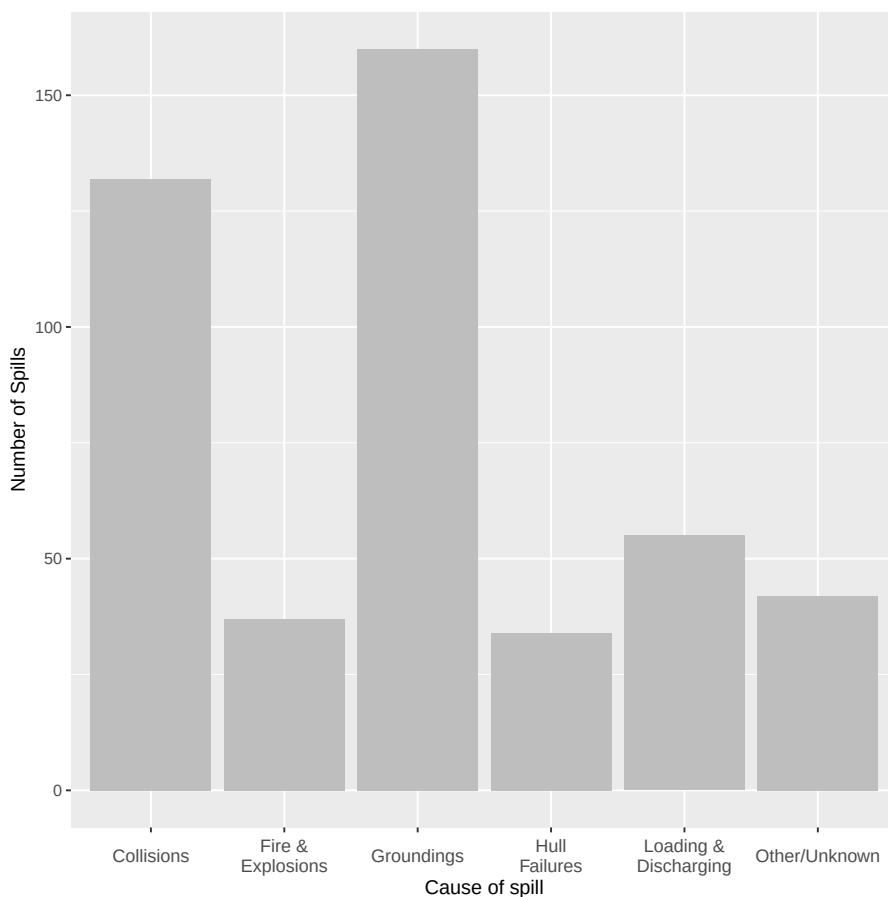
1 MARK each

3. Oil spills as of 2010.

Data from the International Tanker Owners Pollution Federation Limited give the cause of spillage for 460 oil tanker accidents resulting in spills of more than 700 tons from 1970-2010 (www.itopf.com). Results are summarized in the table below.

Cause of spill	Number of spills
Collisions	132
Groundings	160
Hull Failures	34
Fire & Explosions	37
Loading/Discharging	55
Other/Unknown	42

- (a) What is the recorded variable in this study? 1 MARK
- (b) Is the recorded variable qualitative or quantitative? 1 MARK
- (c) The following is a bar plot (or bar graph) of the results:



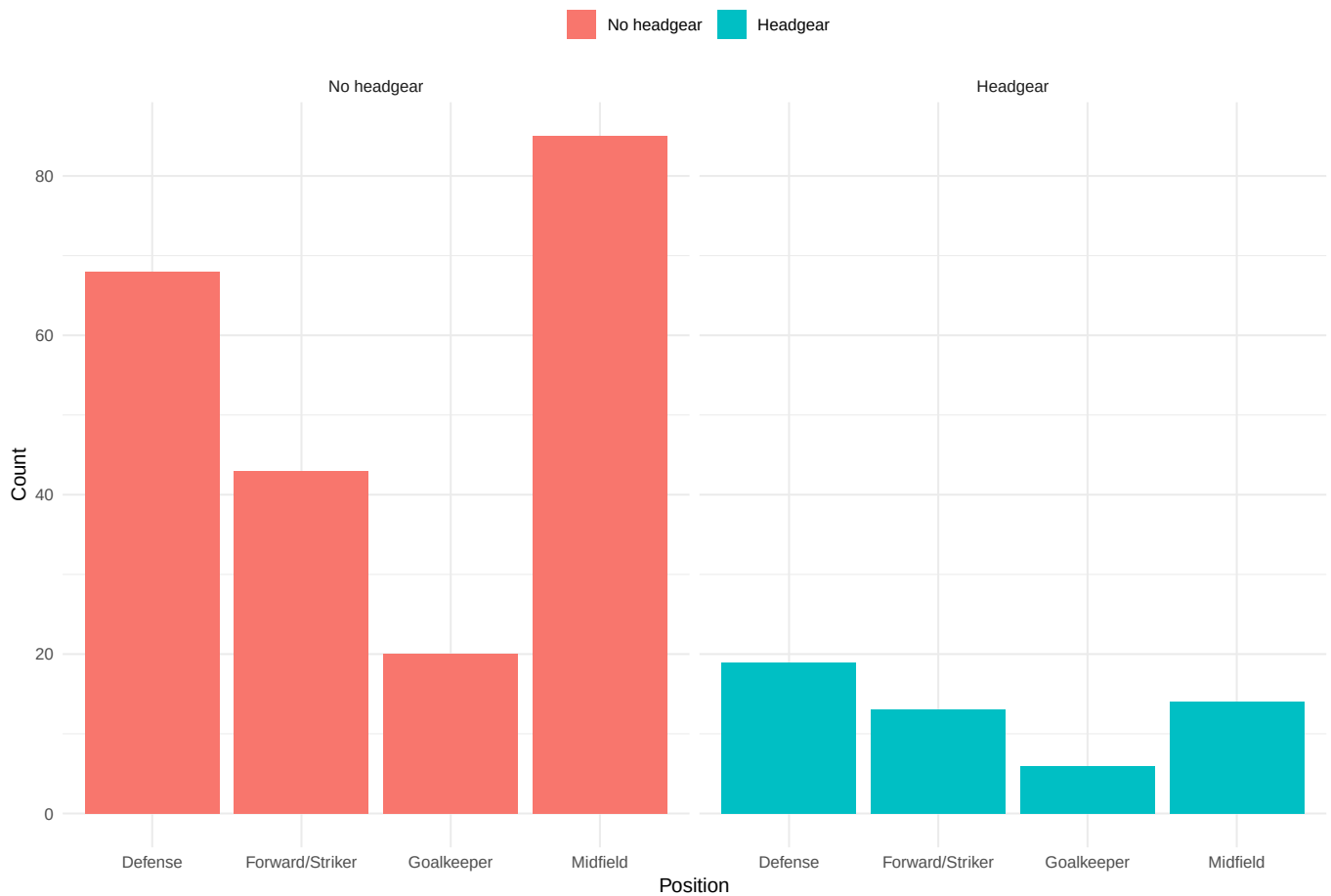
In no more than two sentences, explain what general conclusions may be drawn from this bar plot. 2 MARKS

4. **The effect of protective headgear on head injuries and concussions in adolescent football (soccer) players** (*Br J Sports Med.* 2008 Feb;42(2):110-5)

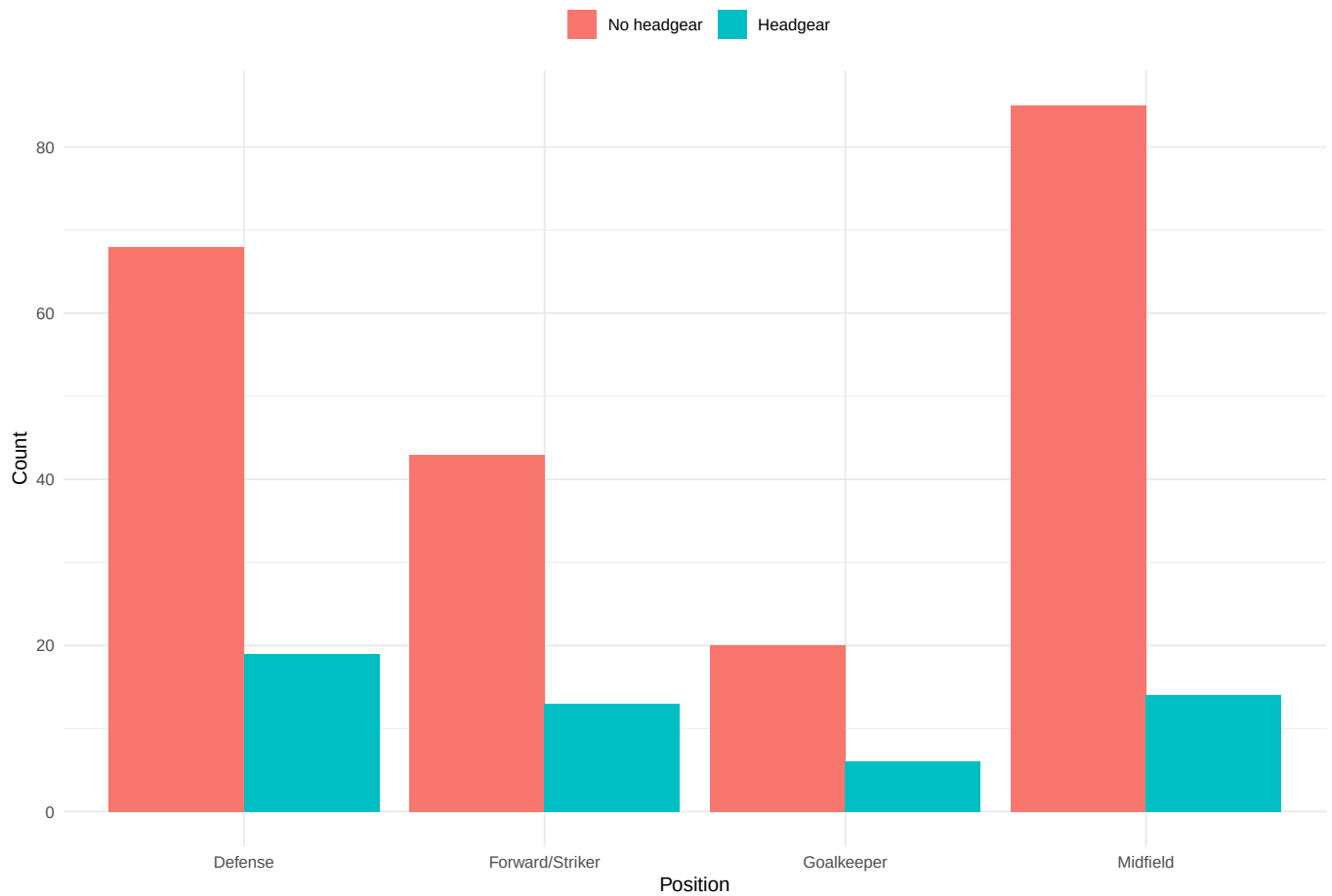
Investigators examined the effects of protective headgear in adolescent soccer players. Participants were soccer players aged 12-17 years from clubs in Ontario. A questionnaire examining the 2006 soccer season was used to investigate self-reported symptoms. There were 278 participants, of which 216 reported not using headgear while playing and 52 reported using headgear while playing. Ten athletes could not reliably be classified as wearing or not wearing headgear.

The following table provides the distribution of players cross-classified by position and whether or not they use headgear.

Position	No headgear	Headgear
Defense	68	19
Forward/Striker	43	13
Goalkeeper	20	6
Midfield	85	14



An alternative way to draw the equivalent bar plot that contains the same information is given below.



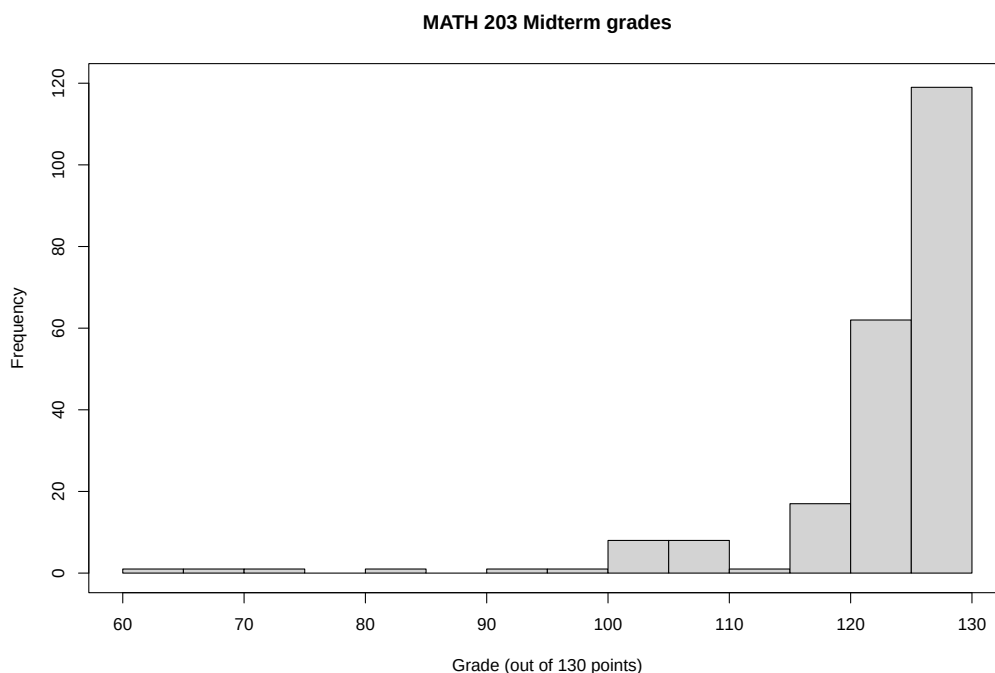
In no more than two sentences, explain what conclusions may be drawn from the graphs about the use of headgear.

2 MARKS

5. Because you are taking Math 203, you have been called as an expert to give your opinion on the following discussion: Pilot A argues that she is the better pilot of the two, since on their last 120 flights she managed to land on time more times compared to pilot B. Pilot B argues that based on the time of the flight she is a better pilot. The following table shows the number of flights landed on time for each pilot by time of the flight.

Day flights	Pilot	Landed on time	Total
	A	90	100
	B	19	20
Night flights	Pilot	Landed on time	Total
	A	10	20
	B	75	100

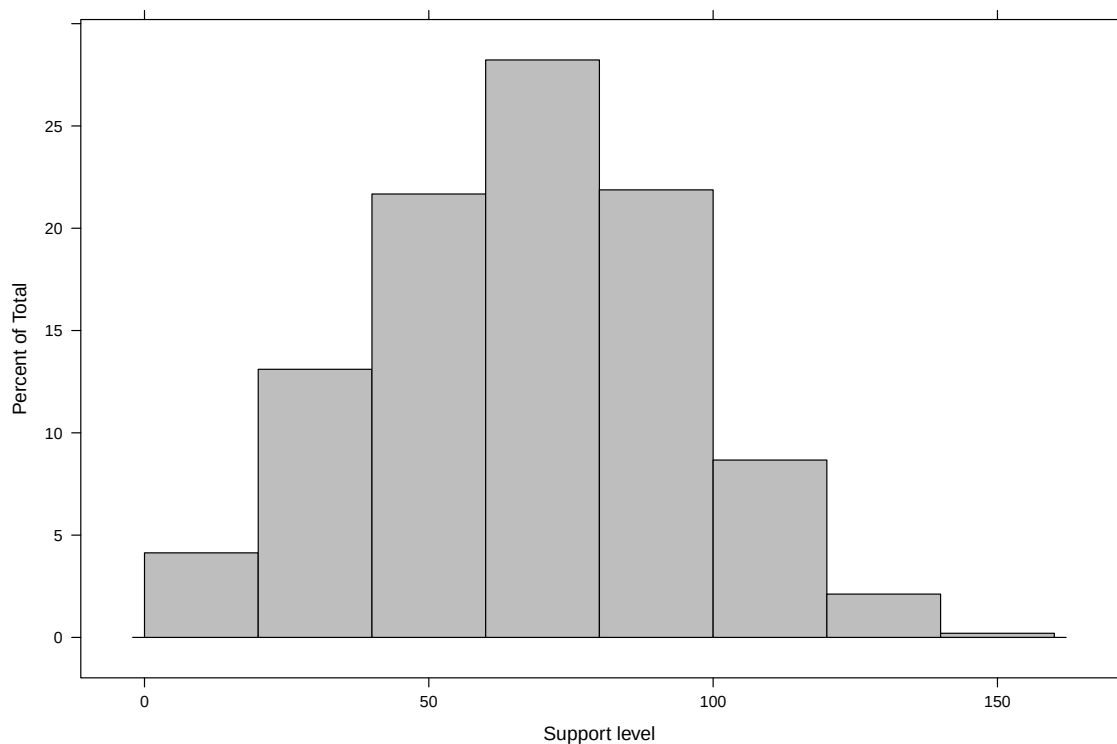
- (a) List the three variables involved in this study. 1 MARK
- (b) Explain why pilot A's argument is correct: overall she has more flights that landed on time. 2 MARKS
- (c) Explain why pilot B's argument is also correct: based on the time of the flight, she has a better record of flights landed on time. 2 MARKS
- (d) Briefly explain how the two pilots are able to reach these contradictory conclusions. 2 MARKS
6. Results of a MATH 203 Midterm exam are summarized in the histogram shown below.



- (a) Approximately how many grades are between 115 and 125 points? 2 MARKS
- (b) Approximately how many grades are between 120 and 127.5 points? 2 MARKS
- (c) It is found from the complete sample that the sample mean is 122.3 points and the sample median is 125 points. In one sentence, say why the median is greater than the mean for these data. 2 MARKS

7. *Corporate sustainability* refers to business practices design around social and environmental considerations. Data on the level of support for corporate sustainability were obtained for 992 senior managers. Level of support was measured quantitatively. Simulation was used to convert the data from the study to a scale ranging from 0 to 160 points, where higher point values indicate a higher level of support for sustainability.

The figure shows a histogram of the distribution of values of level of support for sustainability.



- (a) Report, approximately, what percentage of the 992 senior managers reported a level of support for sustainability of 100 points or greater. 2 MARKS
- (b) Can you provide the exact value of the sample mean ? 1 MARK
- (c) Would you expect the sample mean and sample median to be similar? 1 MARK
- (d) Can you provide the exact value of the sample variance of the data? 1 MARK
- (e) Can you provide an approximate value of the range of the data? 1 MARK

For (b) - (e), justify your answers in one sentence.

8. Does a teaspoon of honey before bed really calm a child's cough? To test the folk remedy, pediatric researchers at Pennsylvania State University carried out a designed study involving a sample of 105 children who were ill with an upper respiratory tract infection (*Archives of Pediatrics and Adolescent Medicine*, December, 2007). On the first night, parents rated their children's cough symptoms on a scale from 0 (no problems at all) to 30 (extremely severe). On the second night, the parents were instructed to give their sick child a dosage of liquid "medicine" prior to bedtime. Unknown to the parents, some were given a dosage of dextromethorphan (DM)—an over-the-counter cough medicine—while others were given a similar dose of honey. Also, a third group of parents (the control group) gave their sick children no dosage at all. Again, the parents rated their children's cough symptoms, and the improvement in total cough symptoms score was determined for each child.

The data (improvement scores) for the study are saved in the file

`https://www.math.mcgill.ca/dstephens/203/Data/HoneyCough.csv`

- (a) Find the median improvement score for the honey dosage group. 2 MARKS
- (b) Find the median improvement score for the DM dosage group. 2 MARKS
- (c) Find the median improvement score for the control group. 2 MARKS
- (d) Based on the results, parts (a)–(c), what conclusions can pediatric researchers draw? (no more than two sentences), 2 MARKS

For anyone using R or RStudio, the data for the three groups can be read in and isolated from the data file using the following code:

```
1 fname<-'https://www.math.mcgill.ca/dstephens/203/Data/HoneyCough.csv'
2 Cough<-read.csv(fname)
3 hg<-Cough$ImproveScore[Cough$Treatment=='H'] #Honey Group
4 dm<-Cough$ImproveScore[Cough$Treatment=='DM'] #DM Group
5 cg<-Cough$ImproveScore[Cough$Treatment=='C'] #Control Group
```

which creates the three R vectors `hg`, `dm` and `cg` that contain the honey dosage, DM dosage and control groups respectively.

However, any method of calculation (e.g. other software, by-hand calculation) may be used.

9. The magazine *Fortune* (November 14, 2005) published a list of the 50 most powerful women in America. The data on age (in years) of each of these 50 women are stored in the file

`https://www.math.mcgill.ca/dstephens/203/Data/Power.txt`

- (a) Find the range of the ages of these 50 women. 1 MARK
- (b) Find the variance of the ages of these 50 women. 2 MARKS
- (c) Find the standard deviation of the ages of these 50 women. 2 MARKS
- (d) Suppose the standard deviation of the ages of the most powerful women in Europe is 10 years. For which location, the United States or Europe, is the age data more variable? 1 MARK
- (e) If the largest age in the data set is omitted, does the standard deviation increase or decrease? Verify your answer. 1 MARK

For anyone using R or RStudio, the data can be read in using the following code:

```
1 fname<-'https://www.math.mcgill.ca/dstephens/203/Data/Power.txt'  
2 wp<-scan(fname) #Ages
```

which creates the R vector `wp` containing the age data. To create a new data vector `wp1` with the largest value of `wp` deleted, the following code may be used:

```
1 wp1<-wp[-which.max(wp)]
```

However, any method of calculation (e.g. other software, by-hand calculation) may be used.

Marks summary: 50 total marks available

Q	1	2	3	4	5	6	7	8	9
Marks	5	5	4	2	7	6	6	8	7