

MATH 203: PRINCIPLES OF STATISTICS 1
FALL 2024

ASSIGNMENT 1: SOLUTIONS

1. (a) Popn: All children undergoing elective surgery requiring BMV and intubation. 1
(b) Exp. units: A child undergoing elective surgery requiring BMV and intubation. 1
(c) Sample: Children, ages of 0 and 8 years, that underwent elective surgery requiring bag-mask ventilation and intubation at the Montreal Children's Hospital. 1
(d) Variables: DBMV status (presence or absence), and Age of the patient. 1

This is an observational study, as the condition is observed and not induced. For ethical reasons, it is not possible to provoke difficult mask ventilation on a patient while undergoing surgery. This is something that the investigators can only observe when it happens. 1
2. (a) Quantitative 1
(b) Qualitative 1
(c) Qualitative 1
(d) Qualitative 1
(e) Quantitative 1
3. (a) Variable: Cause of spill 1
(b) Qualitative 1
(c) Example solution: groundings seems to be the major cause of spills, followed by Collisions. 2
4. For example, the use of headgear is not very popular compared to not using it. This seems to be the case regardless of position played. 2
5. (a) Pilot (A/B), Time of flight (Day/Night) and Landing status (On time/Not on time). 1
(b) Pilot A landed on time 100 times out of 120, i.e. **83.33%** were on time; Pilot B landed on time 94 times out of 120, i.e. 78.33% were on time. 2
(c) For day flights, 2
 - Pilot A landed on time 90 times out of 100, i.e. 90% were on time.
 - Pilot B landed on time 19 times out of 20, i.e. **95%** were on time.For night flights,
 - Pilot A landed on time 10 times out of 20, i.e. 50% were on time.
 - Pilot B landed on time 75 times out of 100, i.e. **75%** were on time.
- (d) This is a case of Simpson's paradox: We can say that the relationship between Pilot and landing Status is confounded by time of the flight. 2

6. (a) The sum of the frequencies between the intervals from 115 to 120, around 20, and the interval from 120 to 125, about 60. Thus, approximately 80. 2
- (b) The sum of the frequency of the interval from 120 to 125 is approximately 60. If we make the assumption that between 125 and 127.5 (which is half of the interval between 125 and 130) we have half the frequency in the interval from 125 to 130 (approximately $120/2=60$). Thus, in total approximately 120. 2
- (c) This is due to skewness; the sample mean is affected by (pulled towards) the extreme observations on the left (lower grades), but the median is not affected. 2
7. (a) We add the relative frequencies (percentages) of the 3 last intervals. It shows, approximately, $8\% + 3\% + 0.05\% = 11.05\%$. The exact value is 11.90% 2
- (b) No: we do not have the individual data 1
- (c) Yes, because the histogram looks roughly symmetric and mound-shaped. 1
- (d) No: we do not have the individual data 1
- (e) Yes, approximately 160. 1
8. (a) There are 35 observations in the honey dosage group. The median is the middle number, the 18th observation, which is 11. 2
- (b) There are 33 observations in the DM dosage group. The median is the 17th observation which is 9. 2
- (c) There are 37 observations in the control group. The median is the 19th observation which is 7. 2
- (d) Since the median of the honey dosage group is the highest, the median of the DM group is the next highest and the median of the control group is the smallest, we may conclude that the honey dosage group is the most effective, followed by the DM dosage and then the control group. 2
9. (a) The range is the difference between the largest and the smallest observations: $71 - 37 = 34$ years. 1
- (b) The variance is

$$\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) = \frac{1}{49} \left(125541 - \frac{(2487)^2}{49} \right) = 37.50245 \text{ years}^2$$
2

although computing using software is allowed.

- (c) The standard deviation is $s = \sqrt{s^2} = \sqrt{37.50245} = 6.123$ years. 2
- (d) The age data for Europe were more variable 1
- (e) The variance will decrease to 28.675 years², and the s.d. to 5.355 years. 1