Midterm exam Statistical Methods – Solutions

22 November 2016

1.

a) Incorrect. About 95% of the data points in a sample originating from a normal distribution will fall within 2 standard deviations of the population mean.

b) Incorrect. The sample mean cannot be computed for ordinal data.

c) Correct. In the formula for sample standard deviation we divide by $n-1$, so for $n=1$ the sample standard deviation cannot be computed.

d) Correct. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and $P(A \cap B)$ is nonnegative, so $P(A \cup B) \leq P(A) + P(B)$.

NB. *Other examples/arguments are of course also possible.*

2.

a) In the table of measurements, 13 of them exceed 30 seconds. Therefore, an estimate based on the relative frequency is $\frac{13}{25} = 0.52$.

b) The amount of time spent on the main page is assumed to be normally distributed with mean $\mu = 30.00$ and standard deviation $\sigma = 10.00$. Therefore, the z-score of 28 seconds is $z = \frac{28.00-30.00}{10.00} = -0.20$. We use Table 2 and see that $P(Z \leq -0.20) = 0.4207$, where Z denotes the standard normal distribution. Hence, the probability that a single randomly selected visitor will spend less than 28 seconds on the main page of the portal is 0.4207 (or 42.07%).

c) The mean amount of time spent on the main page by $n = 36$ randomly selected visitors is normally distributed with mean $\mu = 30.00$ and standard deviation $\sigma = 10.00/\sqrt{36} = \frac{5}{3}$, by the Central Limit Theorem. Therefore, the z-score of 28 seconds is $z = \frac{28.00-30.00}{5/3} = -1.20$. We use Table 2 again and see that $P(Z \leq -1.20) = 0.1151$. Hence, the probability that the mean time spent on the main page by 36 randomly selected visitors is less than 28 seconds is 0.1151 (or 11.51%).

3.

a) Recall that $A = \{\text{Verge crashed at least once}\}$ and $B = \{\text{Marble is installed}\}$. We have $P(A) = 0.80$, and $P(\bar{A}) = 1 - P(A) = 0.20$, where $\bar{A}$ is the complement of the event A. We also have $P(B|A) = 0.92$ and $P(B|\bar{A}) = 0.46$. By the law of total probability

$$P(B) = P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A}) = 0.92 \cdot 0.80 + 0.46 \cdot 0.20 = 0.828.$$

b) We are interested in $P(A|B)$. By Bayes' theorem

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}.$$

Therefore

$$P(A|B) = \frac{0.92 \cdot 0.80}{0.92 \cdot 0.80 + 0.46 \cdot 0.20} = \frac{2 \cdot 4 \cdot 0.46 \cdot 0.20}{2 \cdot 4 \cdot 0.46 \cdot 0.20 + 0.46 \cdot 0.20} = \frac{8}{9} \approx 0.889.$$

c) Crashes of Verge increase the probability of installing Marbles, since $P(B|A) = 0.920 > 0.828 = P(B)$. It also implies that A and B are dependent, because if they were independent, $P(B|A) = P(B)$.

4.

a) $\Omega = \{(1,H),(2,H),(3,H),(4,H),(5,H),(6,H),(1,T),(2,T),(3,T),(4,T),(5,T),(6,T)\}$. Both the die and the coin are fair, so all outcomes are equally likely and $P(\omega) = 1/12$ for all ω in Ω.

b) The roll and the toss are independent by definition, so $P\big((i,H)\big) = \frac{1}{6} \cdot P(H)$ and $P\big((i,T)\big) = \frac{1}{6} \cdot P(T)$, where $i = 1, 2, \ldots, 6$. Therefore all six outcomes with *Heads* have probability $\frac{1}{6} \cdot 0.6 = \frac{1}{10}$ and all six outcomes with *Tails* have probability $\frac{1}{6} \cdot 0.4 = \frac{1}{15}$.

c) We find what Bridget earns for each outcome, and get the following table:

x	$\{X = x\}$	$P(\{X = x\})$
3	$\{(1,H),(2,H),(3,H),(4,H),(5,H),(6,H)\}$	$6 \cdot 1/12 = 1/2$
-1	$\{(1,T)\}$	$1/12$
-2	$\{(2,T)\}$	$1/12$
-3	$\{(3,T)\}$	$1/12$
-4	$\{(4,T)\}$	$1/12$
-5	$\{(5,T)\}$	$1/12$
-6	$\{(6,T)\}$	$1/12$

d) By definition of expectation we get

$$E(X) = \sum_x x \cdot P(X = x) = 3 \cdot \frac{1}{2} - 1 \cdot \frac{1}{12} - 2 \cdot \frac{1}{12} - 3 \cdot \frac{1}{12} - 4 \cdot \frac{1}{12} - 5 \cdot \frac{1}{12} - 6 \cdot \frac{1}{12} = -0.25.$$

5.

a) The histogram of Sample A resembles the bell-shaped curve. On the other hand, the other histogram indicates that Sample B comes from a population distribution that concentrates a lot of mass in the center (and looks quite uniform). Therefore, both tails of the population distribution of Sample A will be heavier than both tails of the population distribution of Sample B.

b) The bottom line is the first quartile, the middle line is the second quartile (in other words, the median), and the top line is the third quartile.

c) Sample A – normal QQ-plot 2, Sample B – normal QQ-plot 1. Both samples seem quite symmetric, and this automatically rules out the third normal QQ-plot. Sample A seems quite normal, and Sample B has lighter tails than normal distribution.