MATH10282 Introduction to Statistics
Class Test - solutions

[N.B. the options may appear in a different order to in your online test, as Blackboard randomizes the order of the options. Blackboard takes this into account when marking.]

1. Consider the following data:

   0.01, 2.53, 2.64, 3.22, 3.73, 3.80, 4.30, 4.71, 4.85

   To 2 d.p., what is the sample median?

   (a) 3.72 (b) 3.73 (c) 3.74 (d) 3.75 (e) 3.76

   The sample median is given by the rth order statistic, $x_{(r)}$, with $r = 0.5(n+1) = 5$, i.e. (b) 3.73. This was answered correctly by 98.1% of students.

2. Consider the following data, which is the same as in Question 1:

   0.01, 2.53, 2.64, 3.22, 3.73, 3.80, 4.30, 4.71, 4.85

   To 3 d.p., what is the sample lower quartile?

   Use the main method discussed in lectures to calculate the 'Type 6' lower quartile. Do not use R.

   (a) 2.584 (b) 2.585 (c) 2.586 (d) 2.587 (e) 2.588

   The lower quartile is the $r = 0.25 \times (n+1) = 2.5$th order statistic as computed by linear interpolation, i.e.

   $$x_{(2)} + 0.5(x_{(3)} - x_{(2)}) = 2.53 + 0.5 \times (2.64 - 2.53) = 2.585\,.$$

   This was answered correctly by 93.34% of students.

3. Consider the following data set:

   2.93, 6.21, 7.99, 10.81, 11.06, 12.35, 14.34, 14.58, 14.83, 15.84

   To 2 d.p., what is the sample mean?

   (a) 11.05 (b) 11.06 (c) 11.07 (d) 11.08 (e) 11.09

   The sample mean is $(2.93 + 6.21 + \ldots + 15.84)/10 = 11.094$. Rounded to 2 d.p. the answer is 11.09. This was answered correctly by 98.42% of students.

4. Consider the following data set, which is the same as in Question 3:

$$2.93,\ 6.21,\ 7.99,\ 10.81,\ 11.06,\ 12.35,\ 14.34,\ 14.58,\ 14.83,\ 15.84$$

To 2 d.p., what is the sample standard deviation?

(a) 4.21 (b) 4.22 (c) 4.23 (d) 4.24 (e) 4.25

The sample variance is

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = (1/9) \times (1391.738 - 10 \times 11.094^2) = 17.88552.$$

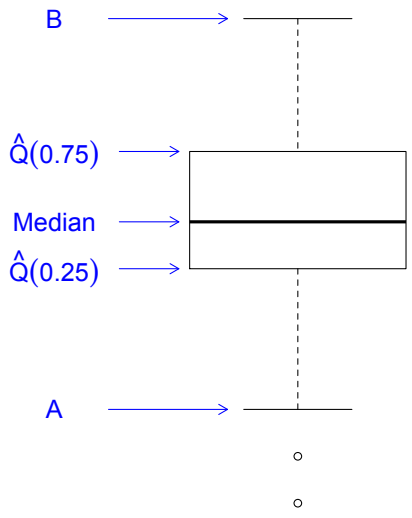
Thus the standard deviation is

$$s = \sqrt{17.88552} = 4.229128\,.$$

Rounded to 2 d.p., the answer is 4.23. A number of people were unsure whether to use the rounded value of $\bar{x}$; this would lead to an answer of 4.24. Part of the question was to realise that it is more accurate to leave any rounding until the very last step.

For this question, 64.13% of students gave the correct answer.

5. Study the box plot below.



Which of the following statements about the box plot is true?

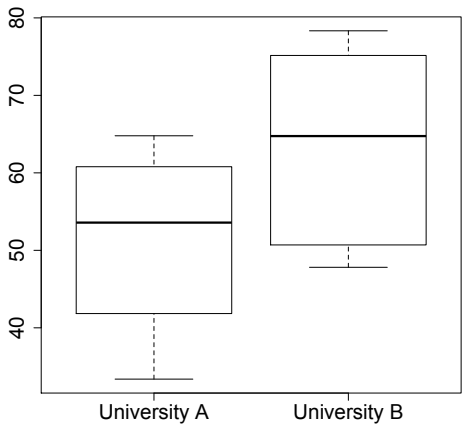
(i) A corresponds to the sample minimum

(ii) B corresponds to the sample maximum

(a) Both (i) and (ii), (b) Only (i), (c) Only (ii), (d) Neither

In the figure, B corresponds to the sample maximum, while A corresponds to the lower adjacent value. As there are two low-valued outliers in the sample (indicated by the circles on the plot), A is not the sample minimum. Thus the correct answer is (c), chosen by 53.02% of students. A common mistake, made by 35.56% of students, was to select (a).

6. Students from two universities were asked to sit the same exam. The box plot below shows the distribution of marks from the two groups. Study the box plot and answer the question below.



Which of the following can be concluded from the figure above?

i The lowest mark in University A is smaller than the lowest mark in University B

ii The highest mark in University A is greater than the lowest mark in University B

iii The median mark in University A is greater than the lower quartile in University B

(a) All of the above (b) Only (iii) (c) None of the above (d) Only (i) and (ii)

The correct answer is (a), selected by 87.62% of students.

This is easy to check provided you know that (i) the whiskers extend to the maximum and minimum of the sample (since here there are clearly no outliers), (ii) the line across the centre of the box gives the sample median, and (iii) the edges of the box give the lower and upper quartiles.

7. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(10, 5^2)$ distribution. What is the sampling distribution of $\bar{X}$?

(a) $N(1, 2.5)$ (b) $N(10, 2.5)$ (c) $N(100, 50)$ (d) $N(10, 0.5)$ (e) $N(100, 250)$

In general, if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, the sample mean satisfies $\bar{X} \sim N(\mu, \sigma^2/n)$. In this case, that gives $\bar{X} \sim N(10, \frac{25}{10})$ i.e. $N(10, 2.5)$. This was answered successfully by 91.43% of students.

8. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(10, 5^2)$ distribution. To 3 d.p., what is the probablity that $\bar{X} \leq 12$?

   Hint: you may use the following table of values for the standard normal c.d.f., $\Phi(z)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944

   (a) 0.897 (b) 0.913 (c) 0.916 (d) 0.929 (e) 0.941

$$\Pr(\bar{X} \leq 12) = \Pr\left(\frac{\bar{X} - 10}{5/\sqrt{10}} \leq \frac{12 - 10}{5/\sqrt{10}}\right) = \Phi(1.2649)\,, \quad \text{since } \frac{\bar{X} - 10}{5/\sqrt{10}} \sim N(0,1)\,.$$

From the table, $\Phi(1.26) = 0.896$ and $\Phi(1.27) = 0.898$ and so $\Phi(1.2649)$ must lie between 0.896 and 0.898. The only one of the available options that satisfies this is (a).

79.37% of students did this correctly.

9. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(10, 5^2)$ distribution. Which of the following statements is true?

   (a) $0.36\,S^2 \sim \chi^2(9)$ (b) $\frac{2.78}{S^2} \sim \chi^2(9)$ (c) $\frac{2.5}{S^2} \sim \chi^2(10)$

   (d) $0.4\,S^2 \sim \chi^2(10)$ (e) $S^2 \sim \chi^2(9)$

In general, it is an important fact that if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ independently, then the sample variance S^2 satisfies

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)\,.$$

In this particular example, this means that $9 \times S^2/25 = 0.36\,S^2 \sim \chi^2(9)$, so the correct answer is (a).

This was answered correctly by 56.51% of students, perhaps surprisingly low considering the online mock test contained a question about this distributional fact.

10. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(10, 5^2)$ distribution. To 3 d.p., what is the probability that $S \leq 7.269$?

Hint: You may use the following table, which lists some important quantiles of the $\chi^2(\nu)$ distribution, i.e. the value q such that $P(Y \leq q) = p$, where $Y \sim \chi^2(\nu)$.

			p		
ν	0.950	0.975	0.990	0.995	0.999
9	16.919	19.023	21.666	23.589	27.877
10	18.307	20.483	23.209	25.188	29.588

(a) 0.950 (b) 0.975 (c) 0.990 (d) 0.995 (e) 0.999

Use the fact that $0.36\,S^2 \sim \chi^2(9)$. Hence,

$$\begin{aligned}\Pr(S \leq 7.269) &= \Pr(0.36\,S^2 \leq 0.36 \times 7.269^2) \\ &= \Pr(\chi^2(9) \leq 19.02181) \\ &\approx \Pr(\chi^2(9) \leq 19.023) = 0.975\,, \text{ from the table.}\end{aligned}$$

Thus the correct answer is (b), selected by 59.69% of students.

11. Which of the following is true of an unbiased estimator?

(a) it is equal to the true value of the parameter

(b) its variance is small

(c) it is representative

(d) its mean is equal to the true value of the parameter

An estimator $\hat{\theta}$ of a parameter θ is said to be unbiased if $E(\hat{\theta}) = \theta$, which corresponds to statement (d). The correct answer was selected by 63.81% of students.

A common mistake was to select (c) (17.47% of students), but representativeness is a property of unbiased *sampling* from a population rather than an unbiased *estimator*.

12. Before the experiment, which of the following is true of a 95% confidence interval?

(a) it will always contain the true value of the parameter

(b) the confidence interval will be wide

(c) it will contain the true value of the parameter with 95% probability

(d) the confidence interval will be symmetric

The correct answer is (c), by definition of a 95% confidence interval. This was selected by 76.2% of students. A common mistake was to select (d), but for example the confidence intervals for σ^2 in Chapter 7 are not symmetric.