MATH10282 Introduction to Statistics
Class Test - feedback and 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.]

Summary

I was pleased to see a large number of students doing very well; around 48.9% scored 14 out of 20 or higher. Well done if you did this! However, there was also a long tail of low marks; around 27.5% of students got less than 10 out of 20. I would say that those in the second group need to do rather more work to ensure a good chance of passing the final exam.

I am always very happy to take questions to help people improve, either in person at office hours, at other times (just email me to check I'll be in), or over email.

Detailed feedback on the questions

1. Consider the following data:

   3.83, 5.63, 6.29, 6.31, 6.63, 8.99

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

   (a) **Correct answer: 6.28** (b) 6.30 (c) 7.54 (d) 30.19 (e) 34.49

   The sample mean is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(3.83 + 5.63 + 6.29 + 6.31 + 6.63 + 8.99) = 6.28\,.$$

   No rounding was needed. This was answered correctly by 99% of students.

   If you got one of the other answers, you may have made one of the following errors:

   - incorrectly using $n-1$ rather than n in the denominator, which gives 7.54
   - incorrect use of brackets, which can give 30.19 or 34.49
   - 6.30 is the sample median, not the sample mean.

   [1 mark]

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

   3.83, 5.63, 6.29, 6.31, 6.63, 8.99

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

   (a) 2.32 (b) **Correct answer: 2.78** (c) 5.01 (d) 10.67 (e) 35.18

The sample variance is

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = (1/5)\times(250.523 - 6\times 6.28^2) = 2.77852.$$

Rounded to 2 d.p., the answer is 2.78. This was only answered correctly by 45.3% of students.

If you chose one of the other options you may have made one of the following mistakes:

- incorrectly using the denominator n instead of $n-1$, which gives $(1/\mathbf{6})(250.523-6\times 6.28^2) = 2.32$. As explained in Chapter 5, to obtain an *unbiased* estimator of σ^2 you need to use denominator $n-1$. This option was chosen 48.2% of students – slightly more popular than the correct answer!
- incorrectly using $n-1$ twice, i.e. calculating $(1/5)\times(250.523 - \mathbf{5}\times 6.28^2) = 10.67$.
- incorrectly used denominator n rather than $n-1$ and made an error with brackets, giving $(1/\mathbf{6})(250.523 - 6.28^2) = 35.18$.

Certain formulae need to be memorised to do well on the exam for this course: the formula for s^2 is one of them. The sample variance is an important quantity and it is used in several places in the course (e.g. when calculating confidence intervals if σ^2 is unknown, and also later in hypothesis testing).

[3 marks]

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

3.83, 5.63, 6.29, 6.31, 6.63, 8.99

To 2 d.p., what is $\hat{Q}(0.33)$? Calculate your answer using the formula given in lectures for the Type 6 quantile. Do not use R.

(a) 1.98 (b) 2.31 (c) 5.59 (d) 5.63 (e) **Correct answer: 5.83**

Let $r = p\times(n+1) = 0.33\times 7 = 2.31$, and $r' = 2$ be the integer part of r. Using linear interpolation, $\hat{Q}(p) = \hat{Q}(0.33)$ is given in terms of order statistics by

$$x_{(r')} + (r'-r)(x_{(r'+1)} - x_{(r')}) = 5.63 + 0.31\times(6.29 - 5.63) = 5.8346 \approx 5.83\,.$$

This was answered correctly by 64.4% of students. If you chose a different answer you may have made one of the following mistakes:

- incorrectly using $r = p\times n$ rather than $r = p\times(n+1)$, giving $\hat{Q}(0.33) = 5.59$.
- incorrectly mixing up the r value and the quantile, giving 2.31 (or 1.98 if you incorrectly thought that $r = p\times n$)
- incorrectly thinking that $p = \frac{1}{3}$ and $r = \frac{1}{3}\times 6 = 2$, giving $\hat{Q} = 5.63$.

Again this is mostly a case of making sure to memorise the correct formulae for the exam - make sure you do it, it well help a lot! [2 marks]

4. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(12, 7^2)$ distribution. What is the sampling distribution of $\bar{X}$?

(a) $N(1.2, 4.9)$ (b) **correct answer:** $N(12, 4.9)$ (c) $N(120, 70)$
(d) $N(12.0.7)$ (e) $N(120, 490)$

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 $\bar{X} \sim N(12, \frac{49}{10})$ i.e. $N(12, 4.9)$. Pleasingly, this was answered correctly by 89.0% of students.

[1 mark]

5. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 10$ from a $N(12, 7^2)$ distribution. To 3 d.p., what is the probablity that $\bar{X} \leq 13$?

(a) **Correct answer: 0.674** (b) 0.629 (c) 0.685 (d) 0.714 (e) 0.742

$$\Pr(\bar{X} \leq 13) = \Pr\left(\frac{\bar{X} - 12}{7/\sqrt{10}} \leq \frac{13 - 12}{7/\sqrt{10}}\right) = \Phi(0.4518)\,, \quad \text{since } \frac{\bar{X} - 12}{7/\sqrt{10}} \sim N(0, 1)\,.$$

From the table given in the test, $\Phi(0.45) = 0.674$ and $\Phi(0.46) = 0.677$ and so $\Phi(0.4518)$ must lie between 0.674 and 0.677 (inclusive). The only one of the available options that satisfies this is (a) 0.674. This was answered correctly by 82.2% of students. [3 marks]

6. Suppose that $X_1, \ldots, X_{13}$ denote a random sample of size $n = 13$ from a $N(10, 4^2)$ distribution. Let S^2 denote the sample variance. Which of the following statements is true?

(a) **Correct answer:** $0.75\, S^2 \sim \chi^2(12)$ (b) $\frac{4}{3S^2} \sim \chi^2(12)$ (c) $\frac{1.23}{S^2} \sim \chi^2(13)$
(d) $0.81\, S^2 \sim \chi^2(13)$ (e) $S^2 \sim \chi^2(12)$

Theorem 4.4 states 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 $12 \times S^2/16 = 0.75\, S^2 \sim \chi^2(12)$, so the correct answer is (a).

This was answered correctly by 59.6% of students. This is a very important result about sampling distributions, and it needs to be learnt in order to understand confidence intervals for σ^2, or confidence intervals and hypothesis tests involving the t-distribution (which are really a bread and butter examination topic at this level).

[2 marks]

7. Suppose that $X_1, \ldots, X_{10}$ denote a random sample of size $n = 13$ from a $N(10, 4^2)$ distribution. To 3 d.p., what is the probability that $S \leq 6.143$?

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
12	21.026	23.337	26.217	28.300	32.909
13	22.362	24.736	27.688	29.819	34.528

(a) 0.950 (b) 0.975 (c) 0.990 (d) **Correct answer:** 0.995 (e) 0.999

Use the fact that $0.75\,S^2 \sim \chi^2(12)$. Hence,

$$\begin{aligned}\Pr(S \le 6.143) &= \Pr(0.75\,S^2 \le 0.75 \times 6.143^2)\\ &= \Pr(\chi^2(12) \le 28.302)\\ &\approx \Pr(\chi^2(12) \le 28.300) = 0.995\,, \text{ from the table.}\end{aligned}$$

This was answered correctly by 57.6% of students, so it seems that most who knew the distributional fact from the previous question could apply it to find the probability.

[2 marks]

8. Suppose that $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ independently with $n \ge 5$. Consider the following estimators of μ. Which estimator is unbiased for μ?

   (a) $\hat{\mu}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3$
   (b) $\hat{\mu}_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2$ **correct answer**
   (c) $\hat{\mu}_3 = \frac{1}{2}X_2 + \frac{1}{8}X_4$
   (d) $\hat{\mu}_4 = \frac{3}{8}X_3 + \frac{2}{8}X_5$

   An estimator $\hat{\mu}$ of the parameter μ is unbiased if $\mathrm{E}(\hat{\mu}) = \mu$. Note that

$$\mathrm{E}(\hat{\mu}_2) = \mathrm{E}\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = \frac{1}{4}\,\mathrm{E}\,X_1 + \frac{3}{4}\,\mathrm{E}\,X_2 = \frac{1}{4}\mu + \frac{3}{4}\mu = \mu$$

   hence $\hat{\mu}_2$ is unbiased. All the other estimators are biased:

$$\begin{aligned}\mathrm{E}(\hat{\mu}_1) &= \mathrm{E}\left(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3\right) = \frac{1}{3}\,\mathrm{E}(X_1) + \frac{1}{3}\,\mathrm{E}(X_2) + \frac{1}{4}\,\mathrm{E}(X_3)\\ &= \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{4}\mu = \frac{11}{12}\mu \ne \mu\end{aligned}$$

$$\mathrm{E}(\hat{\mu}_3) = \mathrm{E}\left(\frac{1}{2}X_2 + \frac{1}{8}X_4\right) = \frac{1}{2}\,\mathrm{E}\,X_2 + \frac{1}{8}\,\mathrm{E}\,X_4 = \frac{1}{2}\mu + \frac{1}{8}\mu = \frac{5}{8}\mu \ne \mu$$

$$\mathrm{E}(\hat{\mu}_4) = \mathrm{E}\left(\frac{3}{8}X_3 + \frac{2}{8}X_5\right) = \frac{3}{8}\,\mathrm{E}\,X_3 + \frac{2}{8}\,\mathrm{E}\,X_5 = \frac{3}{8}\mu + \frac{2}{8}\mu = \frac{5}{8}\mu \ne \mu$$

   This was answered correctly by 78.3% of students. [2 marks]

9. Suppose that $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ independently with $n \ge 5$. Consider the following estimators of μ. Which estimator has variance $\frac{5\sigma^2}{8}$?

(a) $\hat{\mu}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3$

(b) $\hat{\mu}_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2$ **correct answer**

(c) $\hat{\mu}_3 = \frac{1}{2}X_2 + \frac{1}{8}X_4$

(d) $\hat{\mu}_4 = \frac{3}{8}X_3 + \frac{2}{8}X_5$

[2 marks]

We have that

$$\begin{aligned}\operatorname{Var}(\hat{\mu}_2) &= \operatorname{Var}\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = \operatorname{Var}\left(\frac{1}{4}X_1\right) + \operatorname{Var}\left(\frac{3}{4}X_2\right) \text{ by independence} \\ &= \left(\frac{1}{4}\right)^2 \operatorname{Var}(X_1) + \left(\frac{3}{4}\right)^2 \operatorname{Var}(X_2) \\ &= \frac{1}{16}\sigma^2 + \frac{9}{16}\sigma^2 = \frac{10}{16}\sigma^2 = \frac{5}{8}\sigma^2\end{aligned}$$

so the correct answer is $\hat{\mu}_2$. The other estimators have different variances:

$$\operatorname{Var}(\hat{\mu}_1) = \frac{41}{144}\sigma^2, \quad \operatorname{Var}(\hat{\mu}_3) = \frac{17}{64}\sigma^2, \quad \operatorname{Var}(\hat{\mu}_4) = \frac{13}{64}\sigma^2.$$

This question was answered correctly by 59.9% of students.

If you chose a different answer you may have made the following mistake:

- incorrectly argued that $\operatorname{Var}(\hat{\mu}_3) = \operatorname{Var}\left(\frac{1}{2}X_2 + \frac{1}{8}X_4\right) = \frac{1}{2}\operatorname{Var} X_2 + \frac{1}{8}\operatorname{Var}(X_4) = \frac{5}{8}\sigma^2$, or $\operatorname{Var}(\hat{\mu}_4) = \operatorname{Var}\left(\frac{3}{8}X_2 + \frac{2}{8}X_4\right) = \frac{3}{8}\sigma^2 + \frac{2}{8}\sigma^2$. These are both **wrong** because the variance is NOT linear, you need to square the fractions when taking them outside the variance! Around 34% of students selected one of these two wrong answers.

10. Suppose that $X_1, \ldots, X_n \sim N(\mu, 3)$ independently, with μ unknown, and that a data set is obtained with $n = 10$ and $\bar{x} = 17.2$. Which of the following is a 95% confidence interval for μ?

(a) **correct answer:** (16.13, 18.27) (b) (15.34, 19.06)

(c) (16.61, 17.79) (d) (16.29, 18.10)

The end points of the $100(1-\alpha)\%$ confidence interval are given by

$$\bar{x} \pm \frac{z_{\alpha/2}\sigma}{\sqrt{n}}.$$

Here $100(1-\alpha) = 95$, so $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025}$. Recall that here $z_{\alpha/2}$ denotes the *upper* $\alpha/2$ point, i.e. $P(Z > z_{\alpha/2}) = \alpha/2$, so that $P(Z \le z_{\alpha/2}) = 1 - \alpha/2$. Thus $z_{0.025}$ corresponds to the $1 - 0.025 = 0.975$ percentage point, and so using the table with $q = 0.975$ we find that $z_{0.025} = 1.96$.

Thus the end points are

$$17.2 \pm 1.96 \times \frac{\sqrt{3}}{\sqrt{10}}$$

and the 95% CI is $(16.13, 18.27)$. This was answered correctly by only 52.8% of students, but the question is straightforward if you have learnt the formulae.

If you chose a different answer you may have made one of the following mistakes:-

- incorrectly using $\sigma = 3$ rather than $\sigma = \sqrt{3}$, giving (15.34, 19.06)
- incorrectly omitting the square root, giving (16.61, 17.79)
- incorrectly using z_α rather than $z_{\alpha/2}$, giving (16.29, 18.10).

[2 marks]