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

Overall, most of the questions were quite straightforward provided you had amply studied the material. I think that someone taking the course seriously should be aiming for at least 15/20, especially given that it is possible to benefit from guessing. Well done to the good number of students who obtained scores of 17+/20.

In the group overall, there were a few areas where knowledge was weaker than it should have been. One was the calculation of the sample variance (Q2), where the formula seemed to have been misremembered by the majority of students. Another was the sampling distribution of S^2, and the χ^2 distribution (Q6/7). This is an important topic that underpins much of the later work on confidence intervals and hypothesis tests (e.g. the t-distribution), and it needs to be learnt.

I was concerned about the large number of students who scored 12 or less: more work is definitely needed here to give yourself the best chance of passing the final exam. Remember, the module is compulsory and you do have to pass it (or at the very least obtain a compensatable mark $\geq$30% while not failing too many credits overall). Last year those who scored 12 or less were more likely to fail the final exam than to pass.

Detailed feedback on the questions

1. Consider the following data:

$$2.92, 4.20, 5.26, 7.98, 8.23, 8.84, 9.79, 10.94, 15.95, 16.05$$

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

(a) 9.01 (b) 9.02 (c) 9.03 (d) 9.04 (e) 9.05

The sample mean is $\frac{1}{10}(2.92+4.20+5.26+7.98+8.23+8.84+9.79+10.94+15.95+16.05) = 9.016$. Rounding to 2 d.p., we obtain 9.02. This was answered correctly by 94.5% of students. Somewhat concerningly, 4.86% incorrectly selected 9.01. I am not sure why, perhaps due to a rounding error or a slip somewhere. It is always worth double checking numerical calculations like this if you have time.

[1 mark]

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

$$2.92, 4.20, 5.26, 7.98, 8.23, 8.84, 9.79, 10.94, 15.95, 16.05$$

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

(a) 17.80 (b) 19.78 (c) 23.92 (d) 25.71 (e) 28.81

The sample variance is

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = (1/9) \times (990.9256 - 10 \times 9.016^2) = 19.78256.$$

Rounded to 2 d.p., the answer is 19.78. I was surprised that this was answered correctly by just 36.9% of students.

A very common mistake, made by more than half (54.7%) of students, was to select the incorrect answer (a) 17.80. This is what you would obtain if you incorrectly used n as the denominator in the formula for s^2 instead of $n-1$. Recall that you need to use denominator $n-1$ to obtain an unbiased estimator of σ^2.

Certain formulae need to be memorised to do well on the exam for this course: this is one of them. The sample variance s^2 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:

   2.92, 4.20, 5.26, 7.98, 8.23, 8.84, 9.79, 10.94, 15.95, 16.05

   To 3 d.p., what is the sample lower quartile? Use the main method discussed in lectures to calculate the 'Type 6' lower quartile.

   (a) 4.465 (b) 4.601 (c) 4.730 (d) 4.995 (e) 5.101

   The lower quartile is the $r = 0.25 \times (n+1) = 2.75$th order statistic as computed by linear interpolation, i.e.

   $$x_{(2)} + 0.75(x_{(3)} - x_{(2)}) = 4.20 + 0.75 \times (5.26 - 4.20) = 4.995\,.$$

   This was answered correctly by 67.97% of students.

   A common mistake, made by 18.45% of students, was to select (c) 4.73. This is the answer you would obtain if you incorrectly used $r = 0.25 \times n$ rather than $r = 0.25 \times (n+1)$. 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 $N(11, 6^2)$ distribution. What is the sampling distribution of $\bar{X}$?

   (a) $N(1.1, 3.6)$ (b) $N(11, 3.6)$ (c) $N(110, 60)$ (d) $N(11, 0.6)$ (e) $N(110, 360)$

   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 $\bar{X} \sim N(11, \frac{36}{10})$ i.e. $N(11, 3.6)$. I was pleased to see that this was answered successfully by 89.65% of students.

   [1 mark]

5. Suppose that $X_1, \dots, X_{10}$ denote a random sample of size $n = 10$ from a $N(11, 6^2)$ distribution. To 3 d.p., what is the probablity that $\bar{X} \le 13.5$?

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) 0.901 (c) 0.906 (d) 0.911 (e) 0.925

$$\Pr(\bar{X} \le 13.5) = \Pr\left(\frac{\bar{X} - 11}{6/\sqrt{10}} \le \frac{13.5 - 11}{6/\sqrt{10}}\right) = \Phi(1.3176)\,, \quad \text{since } \frac{\bar{X} - 11}{6/\sqrt{10}} \sim N(0,1)\,.$$

From the table, $\Phi(1.31) = 0.905$ and $\Phi(1.32) = 0.907$ and so $\Phi(1.31)$ must lie between 0.905 and 0.907. The only one of the available options that satisfies this is (c) 0.906.

Pleasingly, 80.59% of students did this correctly.

[3 marks]

6. Suppose that $X_1, \dots, X_{10}$ denote a random sample of size $n = 10$ from a $N(11, 6^2)$ distribution. Let S^2 denote the sample variance. Which of the following statements is true?

(a) $0.25\, S^2 \sim \chi^2(9)$ (b) $\frac{4}{S^2} \sim \chi^2(9)$ (c) $\frac{3.6}{S^2} \sim \chi^2(10)$

(d) $0.28\, S^2 \sim \chi^2(10)$ (e) $S^2 \sim \chi^2(9)$

In general, it is an important fact that if $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ 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 $9 \times S^2/36 = 0.25\, S^2 \sim \chi^2(9)$, so the correct answer is (a). This was answered correctly by 54.37% 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 $X_1, \dots, X_{10}$ denote a random sample of size $n = 10$ from a $N(11, 6^2)$ distribution. To 3 d.p., what is the probability that $S \le 9.713$?

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 $P(Y \le q) = p$, where $Y \sim \chi^2(\nu)$.

ν	p 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 $0.25\,S^2 \sim \chi^2(9)$. Hence,

$$\begin{aligned}\Pr(S \le 9.713) &= \Pr(0.25\,S^2 \le 0.25 \times 9.713^2) \\ &= \Pr(\chi^2(9) \le 23.58559) \\ &\approx \Pr(\chi^2(9) \le 23.589) = 0.995\,, \text{ from the table.}\end{aligned}$$

Thus the correct answer is (d), selected by 51.78% of students. It seems that most of those who knew the distributional fact from question 6 were able to apply it to solve this question - well done.

[2 marks]

8. Suppose that $X_1, \ldots, X_n \sim \text{Po}(\lambda)$ independently with $n \ge 5$, and define the following estimators of λ:

$$\hat{\lambda}_1 = \bar{X}\,, \qquad \hat{\lambda}_2 = \frac{1}{6}(X_1 + 4X_2 + X_3)\,.$$

Which of the following statements is true?

(i) The estimator $\hat{\lambda}_1$ is unbiased for λ.

(ii) The estimator $\hat{\lambda}_2$ is unbiased for λ.

(iii) $\text{Var}(\hat{\lambda}_1) < \text{Var}(\hat{\lambda}_2)$

(iv) cannot say whether $\text{Var}(\hat{\lambda}_1) < \text{Var}(\hat{\lambda}_2)$

(a) none of the above (b) (i) only (c) (i) and (iv) only
(d) (i), (ii) and (iv) only (e) (i), (ii) and (iii) only

An estimator $\hat{\lambda}$ of the parameter λ is said to be unbiased if $\text{E}(\hat{\lambda}) = \lambda$. Note that

$$\text{E}\,\hat{\lambda}_1 = \text{E}(\bar{X}) = \text{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\text{E}\,X_i = \frac{1}{n}(n\lambda) = \lambda$$

$$\text{E}\,\hat{\lambda}_2 = \text{E}\left(\frac{1}{6}(X_1 + 4X_2 + X_3)\right) = \frac{1}{6}(\text{E}(X_1) + 4\,\text{E}(X_2) + \text{E}(X_3)) = \frac{1}{6}(\lambda + 4\lambda + \lambda) = \frac{6\lambda}{6} = \lambda$$

Hence both estimators are unbiased. Moreover, $\text{Var}\,\bar{X} = \lambda/n$ and

$$\text{Var}\,\hat{\lambda}_2 = \frac{1}{6^2}(\text{Var}\,X_1 + 4^2\,\text{Var}\,X_2 + \text{Var}\,X_3) = \frac{\lambda + 16\lambda + \lambda}{36} = \lambda/2 > \lambda/n = \text{Var}\,\bar{X}$$

since $n \ge 5$. Hence statement (iii) is true. Thus the correct answer is (e). This was answered correctly by only 35.6% of students. Well done if you managed this.

Almost all students seemed to realise that $\hat{\lambda}_1 = \bar{X}$ is unbiased for λ, with only 2.59% selecting option (a) 'none of the above'. Pleasingly, a good majority (64.08%) also seemed to realise that $\hat{\lambda}_2$ is unbiased, as they selected options (d) or (e) – though I was perhaps surprised the figure was not even higher. However, fewer than I expected realised that it can be shown that $\text{Var}(\hat{\lambda}_1) < \text{Var}(\hat{\lambda}_2)$.

Nonetheless, lots of students seemed to have the technical skills to calculate the variances, as evidenced by the large number of successful responses to the next question. It seems mainly to be an issue of having the confidence or skill to string it all together, which will come if you put in sufficient practice.

[2 marks]

9. Suppose again that $X_1, \ldots, X_n \sim \text{Po}(\lambda)$ independently with $n \geq 5$.

   Which of the estimators below has variance $\frac{7\lambda}{18}$?

   (a) $\bar{X}$ (b) $\frac{1}{3}X_1 + \frac{2}{3}X_2$ (c) $\frac{1}{6}(X_1 + 4X_2 + X_3)$ (d) $\frac{1}{6}(3X_2 + 2X_4 + X_5)$

   Note that

$$\begin{aligned}
\text{Var}(\bar{X}) &= \text{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}\, X_i = n\lambda/n^2 = \lambda/n \\
\text{Var}\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) &= \frac{1}{9}(\lambda + 4\lambda) = 5\lambda/9 \\
\text{Var}\left(\frac{1}{6}(X_1 + 4X_2 + X_3)\right) &= \frac{1}{36}(\lambda + 4^2\lambda + \lambda) = \lambda/2 \\
\text{Var}\left(\frac{1}{6}(3X_1 + 2X_4 + X_5)\right) &= \frac{1}{6^2}(3^2\,\text{Var}\, X_1 + 2^2\,\text{Var}\, X_2 + \text{Var}\, X_5) \\
&= \frac{1}{36}(9 + 4 + 1)\lambda = \frac{14\lambda}{36} = \frac{7\lambda}{18}
\end{aligned}$$

   Hence the correct answer is (d). Pleasingly, 74.44% of students answered this correctly.

[2 marks]

10. Suppose that $X_1, \ldots, X_n \sim N(\mu, 2)$ independently, with μ unknown, and that a data set is obtained with $n = 20$ and $\bar{x} = 15.1$. Which of the following is a 95% confidence interval for μ?

    (a) (14.48, 15.72) (b) (14.29, 15.91) (c) (14.36, 15.75) (d) (13.86, 16.34)

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

$$\bar{x} \pm \frac{z_{\alpha/2}\sigma}{\sqrt{n}} = 15.1 \pm 1.96 \times \sqrt{2}/\sqrt{20}$$

    thus the answer is (a) (14.48, 15.72), since $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$. This was answered correctly by 58.9% of students. The three incorrect answers were roughly equally popular.

A possible stumbling block was not remembering that $z_{0.025} = 1.96$ — in the final exam you would have a table to check this. Alternatively, perhaps some simply did not remember the formula as the topic had only been covered recently. Nonetheless confidence intervals are an important topic and it is difficult to do well in the final exam if you don't know how to calculate them (and hypothesis tests also need to be understood).

If you didn't know that $z_{0.025} = 1.96$ but you did know the formulae for the end-points, you could have always tried to work backwards. By rearranging, it is true that

$$z_{\alpha/2} = \frac{\text{upper end point} - \bar{x}}{\sigma/\sqrt{n}}$$

Thus answers (a), (b), (c), or (d) would thus correspond to a value for $z_{\alpha/2}$ of 1.96, 2.56, 2.45, or 3.92. I would hope at this point your memory might be jogged and that you might be able to select the correct value from this list.

[2 marks]