Question 1 **(25 marks)**

A survey of 50 Leaving Certificate candidates in 2014, randomly selected in the Dublin region, found that they had a mean mark of 374 in a certain subject. The standard deviation of this sample was 45.

(a) Find the 95% confidence interval for the mean mark in the subject, in the Dublin region. Interpret this interval.

The 95% confidence interval is

$$\begin{aligned}\left(\bar{x}-\frac{1.96s}{\sqrt{n}},\bar{x}+\frac{1.96s}{\sqrt{n}}\right) &= \left(374-\frac{1.96(45)}{\sqrt{50}},374+\frac{1.96(45)}{\sqrt{50}}\right)\\ &= (374-12.473,374+12.473)\\ &= (361.527,386.473)\end{aligned}$$

We can say with 95% confidence that the mean mark in the subject, in the Dublin area lies in the interval $(361.527, 386.473)$. Having '95% confidence' means that were we to repeat this procedure (i.e sampling and calculating a confidence interval) many times, the true mean would lie inside the calculated interval 95% of the time.

(b) The mean mark in the subject for all Leaving Certificate candidates, in 2014, was 385 and the standard deviation was 45. John suggests that the mean mark in the Dublin region is not the same as in the whole country. Test this hypothesis using a 5% level of significance. Clearly state your null hypothesis, your alternative hypothesis and your conclusion.

H_0: The mean mark in Dublin is the same as the mean mark in the whole country.
H_1: The mean mark in Dublin is different from the mean mark in the whole country.
The mean mark in the whole country is 385 which is inside the confidence interval from the previous part.
That means that we do <u>not</u> reject the null hypothesis at the 5% level of significance.
In other words we have not found significant evidence to suggest that the mean mark in Dublin is different from the mean mark in the whole country.

Question 2 **(25 marks)**

The principal of a large school claims that the average distance from a student's home to the school is 3.5 km. In order to test this claim, a sample of 60 students from the school was randomly selected. The students were asked how far from the school they lived. The mean distance from these students' homes to the school is 3.7 km with a standard deviation of 0.5 km.

(a) Test the principal's claim using a 5% level of significance. Clearly state your null hypothesis, your alternative hypothesis and your conclusion.

H_0: The mean of the distances from the students' homes to the school is 3.5km.
H_1: The mean of the distances from the students' homes to the school is differnt from 3.5km.
We calculate the z-score corresponding to 3.7 given a population mean of 3.5 and a standard error of $\frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{60}} = 0.0645$ (correct to 3 significant places). So

$$z = \frac{3.7 - 3.5}{0.0645} = 3.1$$

At the 5% level of significance, the rejection region is $z \leq -1.96, z \geq 1.96$ and 3.1 lies in this rejection region (since $3.1 > 1.96$).
Therefore we reject the null hypothesis at the 5% level of significance.

(b) In the above sample of 60 students, 20% of them lived within 2.5 km of the school. Find the 95% confidence interval for the proportion of students from that school who live within 2.5 km of the school.

$\hat{p} = 20\% = 0.2$ and $n = 60$. So the confidence interval is

$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

which is

$$\left(0.2 - 1.96\sqrt{\frac{0.2(0.8)}{60}}, 0.2 + 1.96\sqrt{\frac{0.2(0.8)}{60}}\right)$$

That is $(0.2 - 0.1012, 0.2 + 0.1012)$ or

$$(0.0988, 0.3012).$$

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(c) Data from 10 years ago shows that, at that time, 26% of the student population lived within 2.5 km of the school. Based on your answer to part (b) is it possible to conclude, at the 5% level of significance, that the proportion of students living within 2.5 km of the school has changed since that time? Explain your answer.

H_0: The propotion of students living within 2.5km of the school is 26%.
H_1: The proportion of students living within 2.5km of the school is different from 26%.
In this case the z-score corresponding to 0.2 is

$$\frac{0.2-0.26}{\sqrt{\frac{0.26(1-0.26)}{60}}} = -1.06$$

Now

$$-1.96 \leq -1.06 \leq 1.96$$

so we cannot reject the null hypothesis at the 5% level of significance.
Therefore it is not possible to conclude, at the 5% level of significance that the proportion of students living within 2.5km of the school has changed.

MODEL ANSWER BY
thematstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(d) A statistician wishes to estimate, with 95% confidence, the proportion of students who live within a certain distance of the school. She wishes to be accurate to within 10 percentage points of the true proportion. What is the minimum sample size necessary for the statistician to carry out this analysis?

We need Margin of Error $\leq 10\% = 0.1$ (since the Margin of Error is the maximum radius of a 95% confidence interval). Now

$$\text{Margin of Error} = \frac{1}{\sqrt{n}}$$

So we have

$$\frac{1}{\sqrt{n}} \leq 0.1$$

This is equivalent to $\sqrt{n} \geq \frac{1}{0.1}$ or $\sqrt{n} \geq 10$, or $n \geq 100$. So the minimum sample size is 100.

MODEL ANSWER BY
thematstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

Question 3 **(25 marks)**

(a) The mean lifetime of light bulbs produced by a company has, in the past, been 1500 hours. A sample of 100 bulbs, recently produced by the company, had a mean lifetime of 1475 hours with a standard deviation of 110 hours. Test the hypothesis that the mean lifetime of the bulbs has not changed, using a 0.05 level of significance.

H_0: The mean lifetime of the lightbulbs is 1500 hoursor $\mu = 1500$.
H_1: The mean lifetime of the lightbulbs is different from 1500 hours or $\mu \neq 1500$.

Under the null hypothesis $\mu = 1500$ and the standard error is approximately (for a large sample)

$$\frac{s}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11.$$

So the z-score corresponding to the observed value 1475 is

$$\frac{1475 - 1500}{11} = -2.27$$

correct to 2 decimal places.
Now

$$-2.27 < -1.96$$

so we can reject the null hypothesis at the 5% level of significance.
In other words, there is significant evidence that the mean lifetime of the bulbs has changed.

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(b) Find the p-value of the test you performed in part **(a)** above and explain what this value represents in the context of the question.

As in the solution for part **(a)** above, the z-score corresponding to 1475 is -2.27 correct to 2 decimal places.
Now $P(z \leq -2.27) = P(z \geq 2.27)) = 1 - P(z \leq 2.27) = 1 - 0.9884 = 0.0116$ (using the tables p36).
So the required p-value is $2(0.0116) = 0.0232$ or 2.32%.

This p-value represents the probability of randomly selecting a sample of 100 bulbs with mean lifetime that is less than or equal to 1475hrs or greater than or equal to 1525hrs, given that the mean lifetime of the bulbs is 1500hrs.

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

Question 4 **(50 marks)**

A car rental company has been using Evertread tyres on their fleet of economy cars. All cars in this fleet are identical. The company manages the tyres on each car in such a way that the four tyres all wear out at the same time. The company keeps a record of the lifespan of each set of tyres. The records show that the lifespan of these sets of tyres is normally distributed with mean 45 000 km and standard deviation 8000 km.

(a) A car from the economy fleet is chosen at random. Find the probability that the tyres on this car will last for at least 40 000 km.

The z-score for 40 000 is

$$\frac{40000-45000}{8000} = -0.625$$

We want the probability that a randomly selected tyre will last for at least 40 000km, so we need to calculate $P(z \geq -0.625)$.
Now $P(z \geq -0.625) = P(z \leq 0.625) = 0.734$ from Formula and Tables p.36. So the probability that the tyres will last at least 40 000km is 0.734.

MODEL ANSWER BY thematshtutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(b) Twenty cars from the economy fleet are chosen at random. Find the probability that the tyres on at least eighteen of these cars will last for more than 40 000 km.

We have 20 Bernoulli trials with $p = 0.734$ (from part **(a)**). So

$$\begin{aligned} P(20 \text{ successes}) &= (0.734)^{20} = 0.0021 \\ P(19 \text{ successes}) &= \binom{20}{1}(0.734)^{19}(1-0.734) = 0.0149 \\ P(18 \text{ successes}) &= \binom{20}{2}(0.734)^{18}(1-0.734)^2 = 0.0514 \end{aligned}$$

correct to 4 decimal places. So

$$P(\text{at least 18 successes}) = 0.0021 + 0.0149 + 0.0514 = 0.0684$$

Therefore the answer is $0.0684 = 6.84\%$.

MODEL ANSWER BY thematshtutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(c) The company is considering switching brands from *Evertread* tyres to *SafeRun* tyres, because they are cheaper. The distributors of *SafeRun* tyres claim that these tyres have the same mean lifespan as *Evertread* tyres. The car rental company wants to check this claim before they switch brands. They have enough data on *Evertread* tyres to regard these as a known population. They want to test a sample of *SafeRun* tyres against it.

The company selects 25 cars at random from the economy fleet and fits them with the new tyres. For these cars, it is found that the mean life span of the tyres is 43 850 km.

Test, at the 5% level of significance, the hypothesis that the mean lifespan of *SafeRun* tyres is the same as the mean of *Evertread* tyres. State clearly what the company can conclude about the tyres.

The null hypothesis is H_0: The mean lifespan of *SafeRun* is 45 000.
We calculate the z-score corresponding to 43850 given a population mean of 45 000 and a standard error of $\frac{\sigma}{\sqrt{25}} = \frac{8000}{\sqrt{25}} = 1600$ (*). Thus

$$z_1 = \frac{43850 - 45000}{1600} = -0.71875$$

Now $-1.96 < -0.71875 < 1.96$ so we cannot reject the null hypothesis at the 5% level of significance.

The company can conclude that they do not have significant evidence to suggest that *SafeRun*'s claim is false.

(*): We assume that the standard deviation of the lifespan of the *SafeRun* tyres is the same as that of the *Evertread*. We must make some assumption about the standard deviation in order to solve the problem.