**Question 4 cont.**

(b)    Determine the cumulative distribution function (CDF) for $1 \leq t \leq 3$.

$$\checkmark \int_1^t \frac{1}{x \ln 3} dx = \left[ \frac{\ln(x)}{\ln 3} + c \right]_1^t$$

$$= \frac{\ln(t)}{\ln(3)} \checkmark$$

**Question 5**                                                    **[3 marks: 2, 1]**

A political party is keen to know the proportion of voters in Australia who support a policy. A census is not possible. A far more efficient approach would be to sample voters and then form a reliable estimate for the proportion $p$.

(a)    List two things that you can do to make the policy support proportion estimate more reliable.

$> 30$ per sample $\checkmark$

Many samples $\checkmark$
of voters!

(b)    What statistical effect are you trying to reduce?

Bias. $\checkmark$

4

## Question 5

[13 marks: 2, 2, 1, 3, 2, 3]

Surveys have show that 31% of the Australians went on an overseas trip during the year.

(a) Determine the standard deviation of proportion $p$. Assume 100 were surveyed.

$$\sqrt{\frac{0.31(1-0.31)}{100}} =$$

(2)

$$0.0462$$

(b) Determine a 95% confidence interval for the proportion of the Australian population that had taken an overseas trip.

Casio $\rightarrow$ Interval
One-Prop Z Int
C-Level 0.95
$x$   31
$n$   100

(2)

$\downarrow$

$$0.2194 \le p \le 0.4006$$

(c) Assume the 31% sample proportion applies to the whole population.

A sample of 350 was taken and X = the number of people who took and overseas trip in 2012 was recorded. Give a range, using the 95% confidence interval, within which you would expect X to lie.

$$350 \times 0.2194 = 77$$
$$350 \times 0.4006 = 140$$

(1)

$$77 \text{ to } 140$$

5

**Question 9 cont.**

(b)     Assume the 31% sample proportion applies to the whole population.

A sample of 350 was taken and X = the number of people who took and overseas trip in 2012 was recorded.  Give a range, using the 95% confidence interval, within which you would expect X to lie.

$$350 \times 0.2194 = 77 \quad \checkmark$$
$$350 \times 0.4006 = 140$$

(c)     Determine the probability that in a random sample of 150 people, the number who had taken an overseas trip was greater than 50

$$\text{Bino CD} \quad \text{Lower 51} \qquad \checkmark\checkmark$$
$$\text{Upper 150} \qquad 0.2383$$
$$\text{Num trial 150}$$
$$\text{Pos 0.31}$$

(d)     If 25 surveys were taken and for each a 95% confidence interval for $p$ was calculated, determine the probability that at least 10 of the intervals included the true value of $p$.

$$\text{Bino}\left(25, 0.95\right)$$
$$P(10 \le X \le 25) = 1 \quad \checkmark$$

(e)     A follow-up survey is to be conducted to confirm the results of the original survey.  Working with a confidence interval of 95%, estimate the sample size necessary to ensure a margin of error at most 5%.

$$0.05 = 1.96 \sqrt{\frac{0.31(1-0.31)}{n}}$$
$$n = 328.7 \rightarrow 329 \quad \checkmark\checkmark$$

**Question 10**                                                    **[4 marks: 2, 1, 1]**

The Education Deparmentt of WA is keen to know the proportion of students who are not visual learners. A census can be done but it would be labour and time intensive, and costly to interview every public student in the state. A far more efficient approach would be to sample students from across the state and from this sample, form a reliable estimate for the proportion $p$.

(a)   List two things that will make the non-visual learner proportion estimate more reliable.

Move sampling ✓
>30 per sample ✓

(b)   In statistical terms, student samples must be unbiased. What is the other aspect to consider?

Randomised ✓

(c)   Stratified sampling is the best sampling method in terms of bias reduction. State what could wrong with this method

Various strata need to be clearly and correctly ✓ identified.

OT Lee p205
§18 Sampling.

4

**Question 11**

$\hat{p} = 0.687$ **[2 marks]**

An initial survey found a proportion confidence interval of   $0.649 \le p \le 0.725$ .

A second survey is conducted and it found a confidence interval of
$0.319 \le p \le 0.502$ .   $\hat{p} = 0.4105$

Comment on the conduct of the second survey.   Not done ✓✓
correctly.

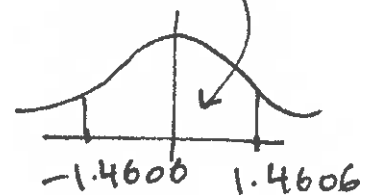**Question 12**                                                                 **[3 marks]**

Determine the confidence interval, as a %, given the following information.

$0.2 \le p \le 0.3$   $\hat{p} = 0.25$
Sample size = 40.

$$MoE = 0.1 = z\sqrt{\frac{0.25 \times 0.75}{40}} \; \checkmark$$

$$z = 1.4606 \rightarrow 86\% \; \checkmark$$
$\checkmark$

-1.4606    1.4606

5

## Question 11

$\hat{p} = 0.68$ **[2 marks]**

An initial survey found a proportion confidence interval of $0.649 \leq p \leq 0.725$ .

A second survey is conducted and it found a confidence interval of
$0.319 \leq p \leq 0.502$ . $\hat{p} = 0.4105$

Comment on the conduct of the second survey. Not done ✓✓
correctly.

## Question 12

**[3 marks]**

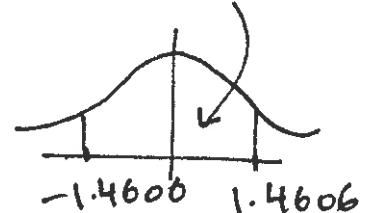Determine the confidence interval, as a %, given the following information.

$0.2 \leq p \leq 0.3$ $\hat{p} = 0.25$
Sample size = 40.

$$MoE = 0.1 = z\sqrt{\frac{0.25 \times 0.75}{40}} \checkmark$$

$$z = 1.4606 \rightarrow 86\% \checkmark$$



$-1.4606 \quad 1.4606$

5

## Question 12

It is known that 3% of a batch of canned beef stew is contaminated with horse meat. A sample of $n$ cans was randomly selected from this batch. Find the maximum value of $n$ so that the probability that there is at least one contaminated can is no more than 50%.

$$P(X \geq 1) \leq 0.5$$
$$1 - P(X=0) \leq 0.5$$
$$P(X=0) \geq 0.5$$
$$0.97^n \geq 0.5$$
$$n \leq 22.75$$
$$\boxed{22}$$

## Question 13
[3 marks]

X is a normal variable with a mean of 120 and a standard deviation of 20.

Find $k$ if $P(115 \leq X \leq k) = 0.5$

Method 1
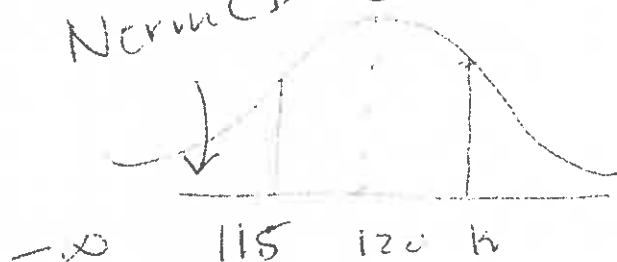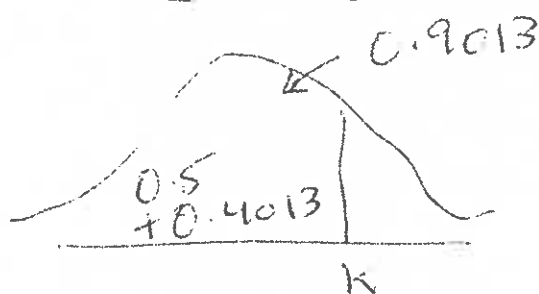Solve (normCDF(115, k, 20, 120) = 0.5, k)
$$k = 145.78 \checkmark\checkmark$$

Method 2
NormCD = 0.4013

$-\infty$  115  120  k

Inv Norm CD
0.9013
0.5 + 0.4013
k
$$\boxed{145.78}$$

6

## Question 14

[5 marks: ⌋ ⌋

14

Let the proportion of people in a city that are able to "roll their tongue" be $p$. A sample of 400 residents in this city yielded a confidence interval for $p$ as $0.23 \leq p \leq 0.29$

(a)    Explain why the proportion able to "roll their tongues" is 0.26.

Mean or centre of
0.23 + 0.29.

(b)    If 50 samples of 400 residents each were selected, and the associated confidence intervals for $p$ calculated in the same manner. How many confidence intervals would actually contain $p$ ?

Error = 0.29 - 0.26

= 0.03

$$Z \sqrt{\frac{0.26(1-0.26)}{400}} = 0.03$$

$$Z = 1.3679 ✓✓$$

$$P(-1.3679 \leq Z \leq 1.3679)$$

$$= 0.8287$$

82.9% CI
✓✓

5