Mathematical Methods Unit 4: Topic 15 Unit Plan – Sampling Theory

<table>
<tr><th>Skill</th><th>Text Book Work</th><th>Deadline</th></tr>
<tr>
<td>

POPULATIONS AND SAMPLES.

• Random samples.

• Population proportion $p = \dfrac{\text{Number in population with attribute}}{\text{Population size}}$.

This value is constant.

• Sample proportion $\hat{p} = \dfrac{\text{Number in sample with attribute}}{\text{Sample size}}$.

This value is NOT constant, it varies from sample to sample.

The sample proportions $\hat{p}$ are the values of a random variable $\hat{P}$.

</td>
<td>• Ex 17A - Q1; Q2; Q3; Q5; Q5; Q7; Q8; Q11; Q12; Q13.</td>
<td>Mon 12 Sept</td>
</tr>
<tr>
<td>

THE EXACT DISTRIBUTION OF THE SAMPLE PROPORTION.

• Sampling from a **small population**: The probabilities associated with the possible values of $\hat{p}$ can be calculated either by direct consideration of the sample outcomes or by using counting methods.

Let X be the random variable '*number in sample of size n with attribute*'.

Then $\hat{P} = \dfrac{X}{n}$ and $\hat{p} = \dfrac{x}{n}$.

A 'scaffolded' probability distribution table can be constructed:

<table>
<tr><td>x</td><td>0</td><td>1</td><td>....</td></tr>
<tr><td>$\Pr(X = x)$</td><td></td><td></td><td></td></tr>
<tr><td>$\hat{p} = \dfrac{x}{n}$</td><td>0</td><td>....</td><td></td></tr>
<tr><td>$\Pr(\hat{P} = \hat{p}) = \Pr(X = x)$</td><td></td><td></td><td></td></tr>
</table>

from which a final probability distribution table can be written:

<table>
<tr><td>$\hat{p}$</td><td>0</td><td>....</td><td></td></tr>
<tr><td>$\Pr(\hat{P} = \hat{p})$</td><td></td><td></td><td></td></tr>
</table>

• Sampling from a **large population (or a small population which is known to have a binomial distribution)**: The probabilities associated with the possible values of $\hat{p}$ can be calculated using the binomial distribution.

'Rule of thumb': $n < 0.1$ (population size).

Let X be the random variable '*number in sample of size n with attribute*'.

$X \sim \text{Binomial}(n, p)$.

A 'scaffolded' probability distribution table can then be constructed from which a final probability distribution table can be written.

• The mean and standard deviation of the sample proportion:

$E(\hat{P}) = p$ and $\text{sd}(\hat{P}) = \sqrt{\dfrac{p(1-p)}{n}}$.

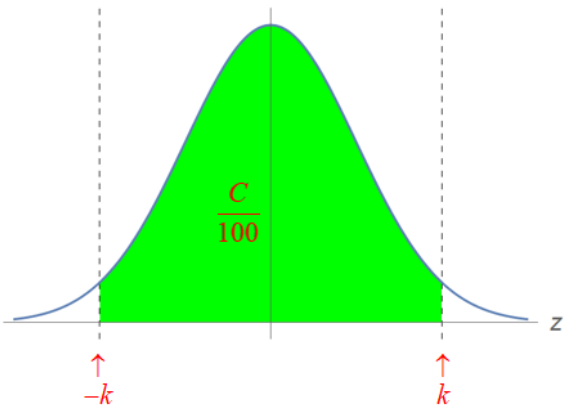
</td>
<td>

• Ex 17B - Q1; Q2; Q3; Q4; Q6; Q7; Q9; Q10; Q11; Q12; Q13; Q14.

• VCAA 2016 Mathematical Methods Sample Exam 1 - Q7.

• VCAA 2016 Mathematical Methods Sample Exam 2 - Extended Response Q3 a. iii., iv., v.

</td>
<td>Mon 12 Sept</td>
</tr>
</table>

APPROXIMATING THE DISTRIBUTION OF THE SAMPLE PROPORTION. • The normal approximation to the binomial distribution. When $np > 5$ and $n(1-p) > 5$ (or even better, when $np > 10$ and $n(1-p) > 10$): $\hat{P} \sim \text{Normal}\left(\mu = p,\ \sigma = \sqrt{\dfrac{p(1-p)}{n}}\right)$. The question will say whether or not you can use the normal approximation.	• Ex 17C - Q1; Q2; Q3; Q6; Q7; Q8; Q9; Q10.	Mon 12 Sept
CONFIDENCE INTERVALS FOR THE POPULATION PROPORTION • The value of the sample proportion $\hat{p}$ can be used to estimate the population proportion *p*. This is called a **point estimate** of *p*. • An interval that we are reasonably sure contains the population proportion *p* is called a **confidence interval** for *p*. • Approximate 95% confidence interval for *p* (when $n\hat{p} > 5$ and $n(1-\hat{p}) > 5$ (or even better, when $n\hat{p} > 10$ and $n(1-\hat{p}) > 10$)): $\left(\hat{p} - 1.96\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 1.96\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right)$. Interpretation: The probability that such an interval contains the unknown population proportion *p* is 0.95. Given a set of *n* 95% confidence intervals, the number that contain the unknown population proportion *p* is a random variable with a distribution $\text{Binomial}(0.95,\ n)$. $1.96\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ is called the **margin of error**. • Approximate *C*% confidence interval for *p* (when when $n\hat{p} > 5$ and $n(1-\hat{p}) > 5$ (or even better, when $n\hat{p} > 10$ and $n(1-\hat{p}) > 10$)): $\left(\hat{p} - k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right)$ where *k* (the quantile) is such that $\Pr(-k < Z < k) = \dfrac{C}{100}$. The **margin of error** is $k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.  • Using *Mathematica* to directly get a *C*% confidence interval. • Finding the minimum sample size for a given margin of error.	• Ex 17D - Q1; Q2; Q3; Q5; Q6; Q7; Q9; Q10; Q12. • Chapter Review - Extended-response questions - Q4. • VCAA 2016 Mathematical Methods Sample Exam 2 - Multiple Choice Q14. • Example 1.	Mon 12 Sept

Mathematica glossary:

Example: A sample of size $n = 100$ is collected and the sample proportion was found to be $\hat{p} = 0.1$.

• Calculating a Confidence Interval for the Population Proportion:

Example: Find a 90% confidence interval.

Method 1: Manual calculation using $\left(\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ and *Mathematica*.

The quantile k must first be calculated:

```
Solve[Probability[-k < x < k, x ≈ NormalDistribution[]] == 0.90, k]

{{k → 1.64485}}
```

```
Quantile[NormalDistribution[], 0.95]

1.64485
```

Note: Quantile is calculated using the probability to the **left**, hence 0.95.

Substitute the value of k into the confidence interval formula:

```
{0.1 - 1.64485 Sqrt[0.1 (1 - 0.1)/100], 0.1 + 1.64485 Sqrt[0.1 (1 - 0.1)/100]}

{0.0506545, 0.149346}
```

Method 2: Use the Hypothesis Testing feature.

Mathematica has a range of statistical functions that require the importing of a package.

Use: Needs[“HypothesisTesting’ ”]. Remember to press shift+enter. Note the single quotation mark at the end.

```
Needs["HypothesisTesting`"]
```

By default *Mathematica* calculates a 95% confidence interval based on the sample mean μ and the population standard deviation σ:

```
NormalCI[0.1, Sqrt[0.1 (1 - 0.1)/100]]

{0.0412011, 0.158799}
```

Use ConfidenceLevel $\rightarrow \frac{C}{100}$ to calculate a C % confidence interval.

Example: Find a 90% confidence interval.

```
NormalCI[0.1, Sqrt[0.1 (1 - 0.1)/100], ConfidenceLevel → 0.90]

{0.0506544, 0.149346}
```

• Calculating the minimum sample size for a required margin of error:

Example: Find the minimum sample size so that the margin of error for a 90% confidence interval is less than 0.05.

The margin of error is $M = k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

The quantile k must first be calculated:

```
Solve[Probability[-k < x < k, x ≈ NormalDistribution[]] == 0.90, k]

{{k → 1.64485}}

Quantile[NormalDistribution[], 0.95]

1.64485
```

Note: Quantile is calculated using the probability to the **left**, hence 0.95.

Substitute the value of k into M and use either Solve or Reduce.

```
In[1]:= Solve[1.64485 Sqrt[(0.1) (1 - 0.1) / n] == 0.05, n]
Out[1]= {{n → 97.3991}}
```

Rounding: Sample size is a discrete quantity therefore rounding to an integer is required.

$n = 97$ (rounding down) and $n = 98$ (rounding up) must be checked to see which satisfies $M < 0.05$.

```
In[2]:= 1.64485 Sqrt[(0.1) (1 - 0.1) / n] /. n → {97, 98}
Out[2]= {0.0501028, 0.0498465}
```

Therefore $n = 97$ is rejected and so $n = 98$ is the answer.

```
In[3]:= Reduce[1.64485 Sqrt[(0.1) (1 - 0.1) / n] < 0.05, n]
   Reduce: Reduce was unable to solve the system with inexact coefficients. The answer was obtained by solving a
   corresponding exact system and numericizing the result.
Out[3]= n > 97.3991
```

From this it can be seen that $n = 98$ is the answer.

VCAA 2016 Mathematical Methods Sample Exam 1 - Q7:

Question 7 (4 marks)

A student performs an experiment in which a computer is used to simulate drawing a random sample of size n from a large population. The proportion of the population with the characteristic of interest to the student is p.

a. Let the random variable $\hat{P}$ represent the sample proportion observed in the experiment.

If $p = \frac{1}{5}$, find the smallest integer value of the sample size such that the standard deviation of $\hat{P}$ is less than or equal to $\frac{1}{100}$. 2 marks

Each of 23 students in a class independently performs the experiment described above and each student calculates an approximate 95% confidence interval for p using the sample proportions for their sample. It is subsequently found that exactly one of the 23 confidence intervals calculated by the class does not contain the value of p.

b. Two of the confidence intervals calculated by the class are selected at random without replacement.

Find the probability that exactly one of the selected confidence intervals does not contain the value of p. 2 marks

Answers:

a. $n = 1600$.

b. $\frac{2}{23}$.

VCAA 2016 Mathematical Methods Sample Exam 2 - Extended Response Q3 a. iii., iv., v:

Question 3 (20 marks)

FullyFit is an international company that owns and operates many fitness centres (gyms) in several countries. It has more than 100 000 members worldwide. At every one of FullyFit's gyms, each member agrees to have their fitness assessed every month by undertaking a set of exercises called S. If someone completes S in less than three minutes, they are considered fit.

a. It has been found that the probability that any member of FullyFit will complete S in less than three minutes is $\frac{5}{8}$. This is independent of any other member. A random sample of 20 FullyFit members is taken. For a sample of 20 members, let X be the random variable that represents the number of members who complete S in less than three minutes.

For samples of 20 members, $\hat{P}$ is the random variable of the distribution of sample proportions of people who complete S in less than three minutes.

iii. Find the expected value and variance of $\hat{P}$. 3 marks

iv. Find the probability that a sample proportion lies within two standard deviations of $\frac{5}{8}$. Give your answer correct to three decimal places. Do not use a normal approximation. 3 marks

v. Find $\Pr(\hat{P} \geq \frac{3}{4} \mid \hat{P} \geq \frac{5}{8})$. Give your answer correct to three decimal places. Do not use a normal approximation. 2 marks

Answers: a. iv. 0.959. **v.** 0.352

VCAA 2016 Mathematical Methods Sample Exam 2 - Multiple Choice Q14:

Question 14

An opinion pollster reported that for a random sample of 574 voters in a town, 76% indicated a preference for retaining the current council.

An approximate 90% confidence interval for the proportion of the total voting population with a preference for retaining the current council can be found by evaluating

A. $\left(0.76-\sqrt{\dfrac{0.76\times 0.24}{574}},\ 0.76+\sqrt{\dfrac{0.76\times 0.24}{574}}\right)$

B. $\left(0.76-1.65\sqrt{\dfrac{0.76\times 0.24}{574}},\ 0.76+1.65\sqrt{\dfrac{0.76\times 0.24}{574}}\right)$

C. $\left(0.76-2.58\sqrt{\dfrac{0.76\times 0.24}{574}},\ 0.76+2.58\sqrt{\dfrac{0.76\times 0.24}{574}}\right)$

D. $\left(436-1.96\sqrt{0.76\times 0.24\times 574},\ 436+1.96\sqrt{0.76\times 0.24\times 574}\right)$

E. $\left(0.76-2\sqrt{0.76\times 0.24\times 574},\ 0.76+2\sqrt{0.76\times 0.24\times 574}\right)$

Answer: B.

Example 1 (4 marks)

A sample of size 80 is used to construct the $C\%$ confidence interval (0.3023, 0.4377) for the population proportion.

a. Find the sample proportion, correct to two decimal places. 2 marks

b. Find the value of C, correct to the nearest whole number. 2 marks

Answers:

a. $\hat{p}=0.37$. **b.** $C=79$.