CENTRAL LIMIT THEOREM AND HYPOTHESIS TESTING

There are three key points to remember about Central Limit Theorem. If you can remember these, then you will be well on your way to understanding contextual problems and interpreting questions. These three rules are as follows;

- The mean of the population (μ) is approximately equal to the mean of sample means ($\mu_{\bar{x}} = \mu$)
- Sample standard deviation (a measure of spread) is always equal to the population standard deviation (σ) divided by square root n (where n is the sample size). *To be technical this generally works so long as no more than 5% of the population represents the sample size.*
- As the sample size (n) increases, the population will begin increasingly approximate a Gaussian Distribution (that is the fancy name for a Normal Distribution)

1. Below represents another point based second dotpoint regarding Central Limit Theorem. What does this show and why may it be of significance?

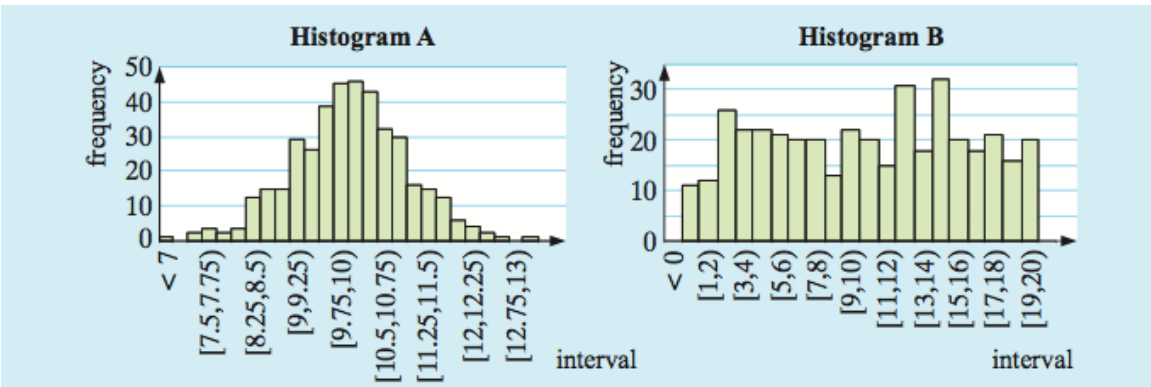
$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\textit{If } \sigma \textit{ is fixed},$$

$$as\ n->+\infty, \frac{\sigma}{\sqrt{n}}->0$$

2. Sampling error is denoted in the same way sample standard deviation is. This is a measure of how precise our distribution is and whether there is a large or small amount of variability. Calculate the sampling error associated with a population standard deviation of 9 and a sample size of 129.

3. Below is a series of two histograms. Each has a sample size of 650. One is a uniform distribution with standard deviation of 5 and a mean of 10. The other is a distribution $\overline{X}_{42}$ of the sample means of size 42 selected from the distribution X.



I. Which of the histograms represents $\overline{X}_{42}$? Give you reasons as to why this distribution is so?

II. From the histogram estimate Pr ($X_{42} < 9.5$) by calculating the bins

III. Find the mean and standard deviation of $\overline{X}_{42}$

IV. Use the histogram to estimate the probability X_{36} is one standard deviation from the mean.

4. The mean time for a bus to travel from Adelaide to Noarlunga is 51 minutes with a standard deviation of 2.5 minutes. A bus conductor travels on seven trips between Adelaide and Noarlunga Centre. Let X_7 be the mean of this sample of seven and hence find the mean and standard deviation of this sample.

5. A small packet of potato crisps has a mean energy amount of 600kJ. The manufacturer states that the standard deviation is 4.75KJ. Find the probability that from a sample of 31 packets, that the energy amount will be greater than 605kJ.

6. A sample of 189 year 12 students is used to estimate the number of lessons that stage 2 students skip. Last year, the standard deviation of hours lost by year 12 students was 5.8 and it is assumed that this is the same amount this year. What is the probability that the estimate produced for the mean number of lessons skipped will be out by less than 4 hours?

7. A manufacturer claims that the mean battery life of one of their new mobile phones is 787 minutes of 4G LTE internet browsing. A group of mobile phone enthusiasts tested this and found that the mean battery life was 769 minutes. Assuming that there were 35 people who tested the claim, that the data could be represented as a normal distribution and that the test was conducted appropriately, determine whether the null hypothesis should be rejected at the 0.05 significance level if standard deviation is 6.1 minutes.

a) What is the null hypothesis?

b) What is the alternate hypothesis?

c) What is the sample error?

d) What is the null distribution?

e) Calculate the test statistic

f) Based on the test statistic should the null hypothesis be rejected? Explain why (without calculating a p value)

g) Prove your statement made in (f) by calculating a p value

h) For what values of the sample mean should the alternate hypothesis be supported?

H) Continued

I) What is the smallest standard deviation needed to ensure that the manufacturer's claim is accurate?

8. The cost of a taxi fare in Sydney depends on the amount of traffic on the road at any given time of the day. It is believed that the mean time at which the taxi fare is most expensive is at 8.20am. A sample of 30 commuters found the mean time at which the taxi fare is most expensive to be at 8.45am. Assuming a standard deviation of 20 minutes, determine whether the null hypothesis should be rejected at the 0.05 significance level.

8) continued

9) The mean number of people in an elevator at any given time is 6. To test this claim, eleven elevator trips were surveyed and it was found that the mean number of people in an elevator at any given time was 8. Assuming a standard deviation of 1 person, determine whether the null hypothesis should be rejected at the 0.05 significance level.

10) What assumptions and limitations are there to Central Limit Theorem?

FOR THOSE WHO ARE INTERESTED…

11) Describe how order and chaos can be seen in the Central Limit Theorem. Is CLT chaotic or ordered?