**1. Normal Distributions (Lecture 5)**

The Normal Distribution

- Bell shaped and symmetrical about the mean
- Location is determined by the mean, mu
- Spread is determined by the standard deviation, sigma
- The random variable X has an infinite theoretical range from –infinity to +infinity

Normal probabilities (using the standardised normal distribution – Z)

- The probability of a continuous variable having a specific value = 0
- Therefore, don't use greater than or equal to (or visa versa)
- Area to the left and right of mu = 0.5, total area =1

$$Z = \frac{X - \mu}{\sigma}$$

**2. Sampling Distributions (Lecture 6)**

Sampling Distribution Properties

- As n increases, standard deviation decreases (curve gets thinner and taller)

2.1 Sampling Distribution of the Sample Mean

Z value for the Sampling Distribution of the Sample Mean:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Standard\ Error = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem (CLT)

- If n is greater than or equal to 30, the sampling distribution of sample means will be approximately normally distributed (Only applies to xbar questions)

2.2 Sampling Distribution of the Sample Proportion

Z value for the Sampling Distribution of the Sample Mean:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$Standard\ Error = \sqrt{\frac{p(1-p)}{n}}$$

## 3. Confidence Intervals (I & II) (Lecture 7 and 8)

Confidence Interval

- Interval estimates, built around a sample statistic
- Draws conclusions from a sample statistic about a population parameter

Conclusion [sample]

- "Based on the sample mean, I have estimated, with 95% confidence, that the true population mean μ will be between some upper and lower calculated value."

Level of Confidence (LOC)

- X% of all confidence intervals, constructed using all possible samples, will contain the population parameter
- How confident you are that the confidence interval will contain the population parameter (will either contain it or not, not a probability)

General Confidence Interval Form

$$CI\ for\ LOC = Point\ Estimate \pm (critical\ value \times standard\ error)$$

<u>3.1 Population mean ($\sigma$ known)</u>

$$\mu: \bar{X} \pm Z_{crit} \times \frac{\sigma}{\sqrt{n}}$$

<u>3.2 Population mean (s known)</u>

$$\mu: \bar{X} \pm t_{crit} \times \frac{s}{\sqrt{n}}$$

<u>3.2 Population proportion</u>

$$p: \hat{p} \pm Z_{crit} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

<u>3.3 Determining the sample size for estimating the mean proportion ($\sigma$ must be known)</u>

$$n \geq \left(\frac{Zcrit \times \sigma}{E}\right)^2 \quad [round\ up]$$

Sampling Error (E) observations

- As sample size (n) increases, E decreases
- As spread of data ($\sigma$) decreases, E decreases
- As LOC increases (Z), E increases

<u>3.4 Determining the sample size to estimate the true proportion</u>

$$n \geq \frac{{Z_{crit}}^2 \times \hat{p}(1-\hat{p})}{E^2}$$

- $\hat{p}$ = 0.5 if there is no information at all given about the value of p or $\hat{p}$
  - produces the maximum number of trials – never underestimates