

# ST3009: Statistical Methods for Computer Science

---

## Week 8 Assignment - Senán d'Art - 17329580

### Question 1

#### (a)

- Students who answer are likely to be more involved in the module and therefore less likely to be "studying to pass". This could seriously skew the results.
- The poll is only sampling 3<sup>rd</sup> Year students taking the ST3009 module. It does not consider other years or students not taking the module. This cannot therefore be used to accurately estimate what fraction of TCD CS students are "studying to pass".

#### (b)

- improve: sample more than just students taking the ST3009 module
- improve: make the responses mandatory

### Question 2

#### (a)

Independent -  $P(A \cap B) = P(A)P(B)$  ie. one event occurring has no effect on the probability of the other or in this case: the value of one variable has no effect on the value of the other

Identical Distribution - the probability distribution of both is the same

#### (b)

No.  $\bar{Y}$  is the mean of the random variables.

**(c)**

Chebychev:

$$\mu - \frac{\sigma}{\sqrt{0.05N}} \leq Y \leq \mu + \frac{\sigma}{\sqrt{0.05N}}$$

$$\mu = 0.1, \sigma = \sqrt{\mu(1-\mu)}$$

$$0.1 - \frac{\sqrt{0.1(1-0.1)}}{\sqrt{0.05N}} \leq Y \leq 0.1 + \frac{\sqrt{0.1(1-0.1)}}{\sqrt{0.05N}}$$

$$N = 100$$

$$0.1 - \frac{\sqrt{0.1(1-0.1)}}{\sqrt{0.05 * 100}} \leq Y \leq 0.1 + \frac{\sqrt{0.1(1-0.1)}}{\sqrt{0.05 * 100}}$$
$$-0.03416 \leq Y \leq 0.23416$$

**(d)**

CLT:

$$\frac{\frac{(X_1+X_2+\dots+X_n)}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\frac{(X_1+X_2+\dots+X_n)}{n} = Y$$

$$n = 100, \sqrt{n} = 10$$

$$\frac{Y - \mu}{\frac{\sigma}{10}}$$

$$\sigma = \sqrt{0.1(1-0.1)}$$

$$-1.96 * \frac{0.3}{10} \leq Y - 0.1 \leq 1.96 * \frac{0.3}{10}$$

$$-1.96 * \frac{0.3}{10} + 0.1 \leq Y \leq 1.96 * \frac{0.3}{10} + 0.1$$

$$0.0412 \leq Y \leq 0.1588$$

Chebychev Pros:

- Provides an actual bound as opposed to an approximation
- Works for all N

CLT Pros:

- Only requires mean and variance to fully describe

(e)

```
N = 100;  
Size = 10000;  
means = zeros(1,Size);  
for t = 1:Size  
    Sample = randi([0 1], 1, N);  
    tmpMean = 0;  
    for i = 1:N  
        tmpMean = tmpMean + Sample(i);  
    end  
    means(t) = tmpMean / N;  
end  
  
Mean = mean(means)  
PD = std(means)  
  
histogram(means)
```

