**Chap6**

Discrete: Number of outcomes can be counted

E.g.Roll of die

Continuous: Outcome occurs over a range

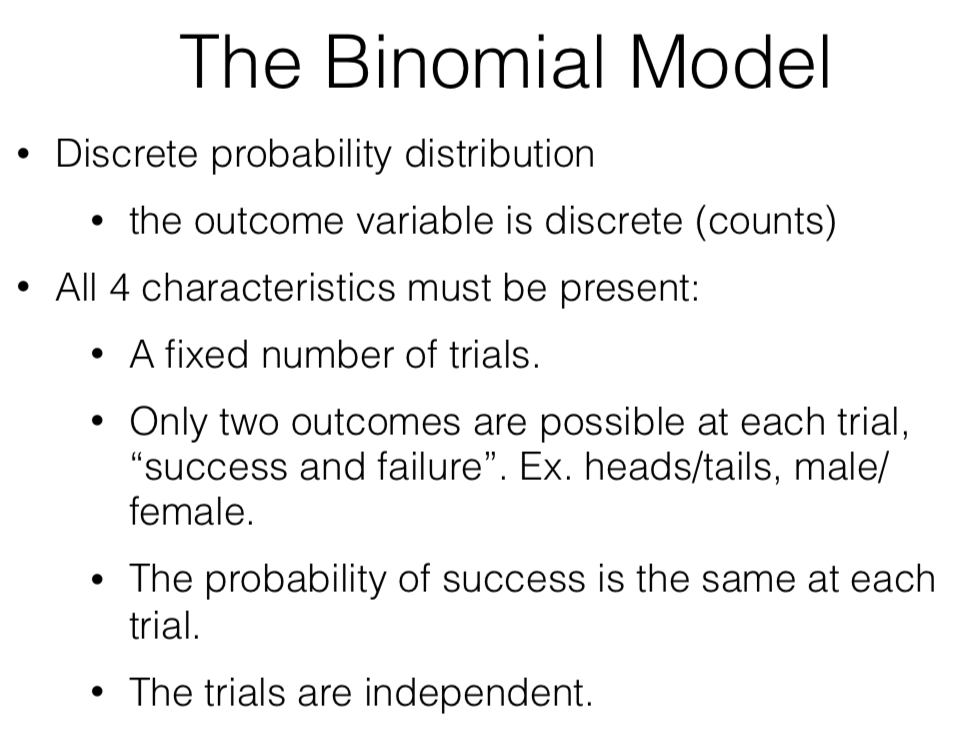
E.g.Exact Time to finish exam

Residual: observed value-predicted value

Normal Model: continuous

Binomial Model: discrete

**b(n,p,x)**: n->fixed number of trials, p->probability of success, x->number of success interested



µ=np ; Stdev=√(np(1-p))

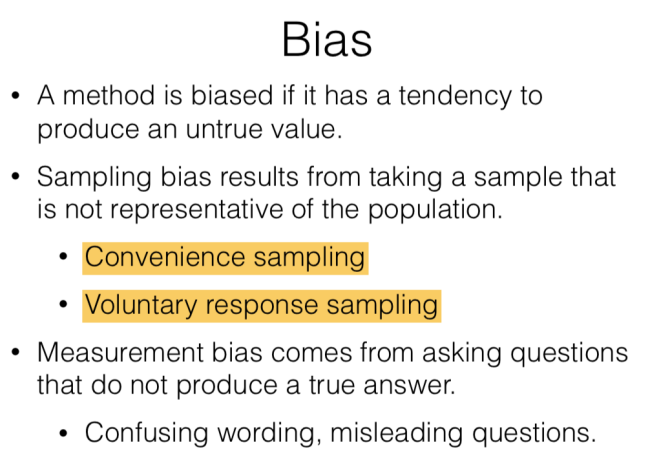
**Chap7**

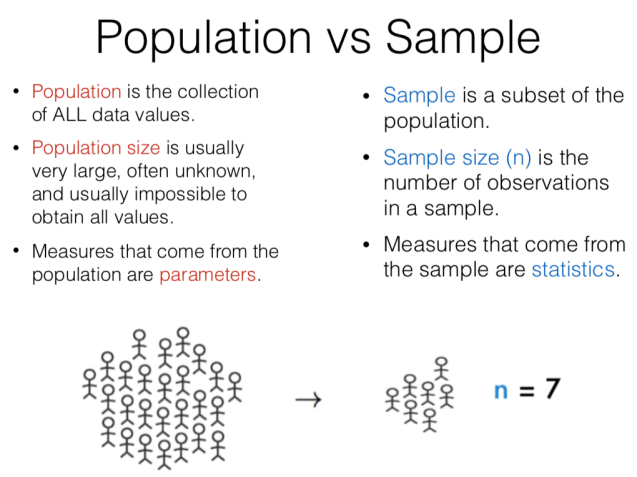
Sample entire population: census

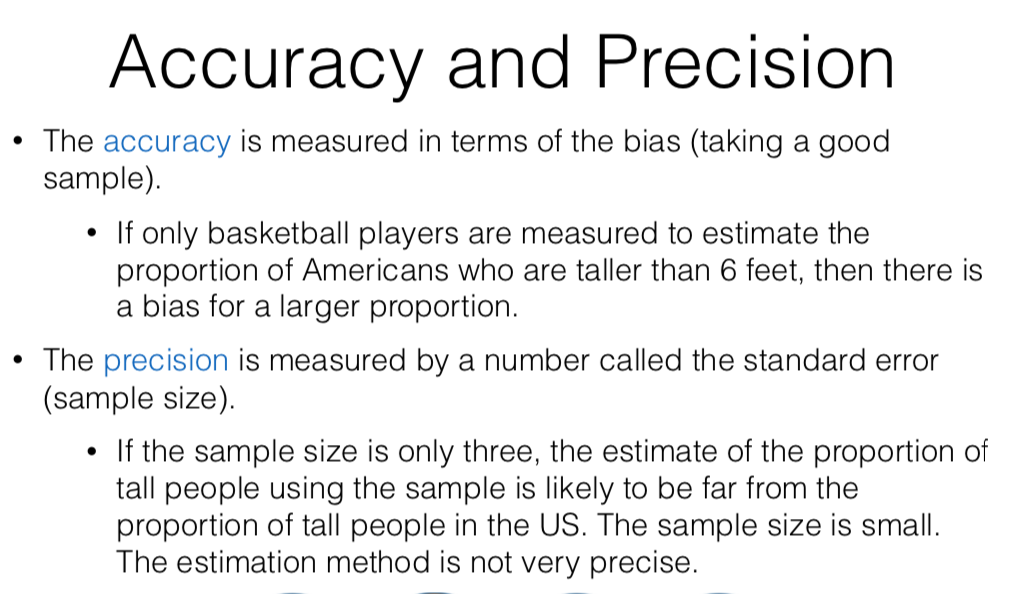
Voluntary response samples are almost always biased, not representative

Non-response error occurs when people respond differ from people don’t

Simple random Sampling(SRS): Randomly draw people from population without replacement

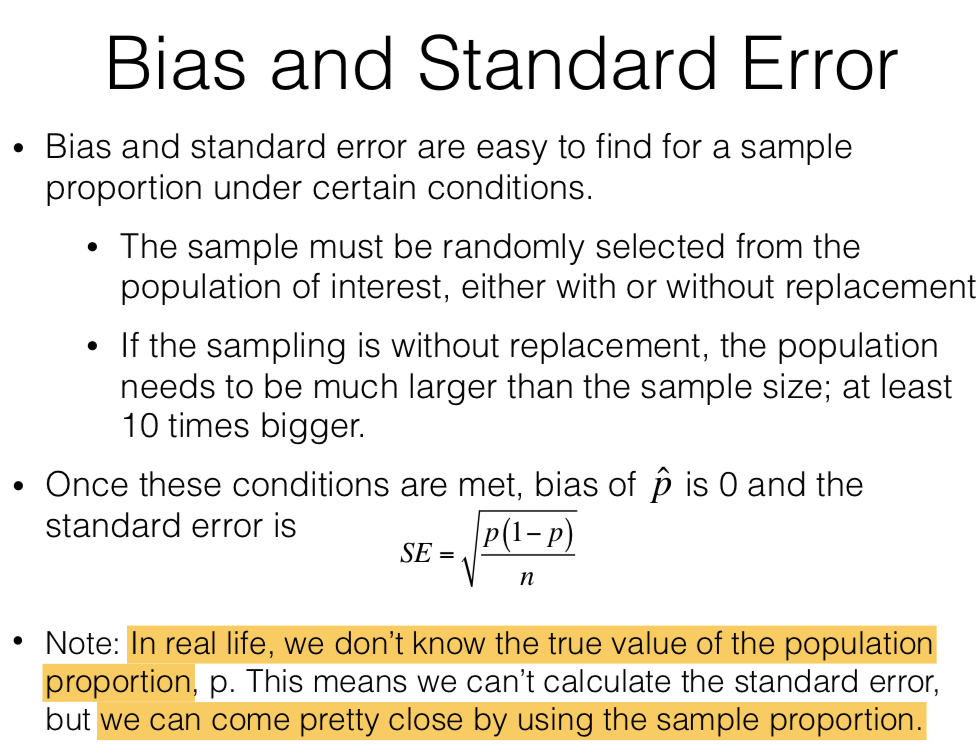




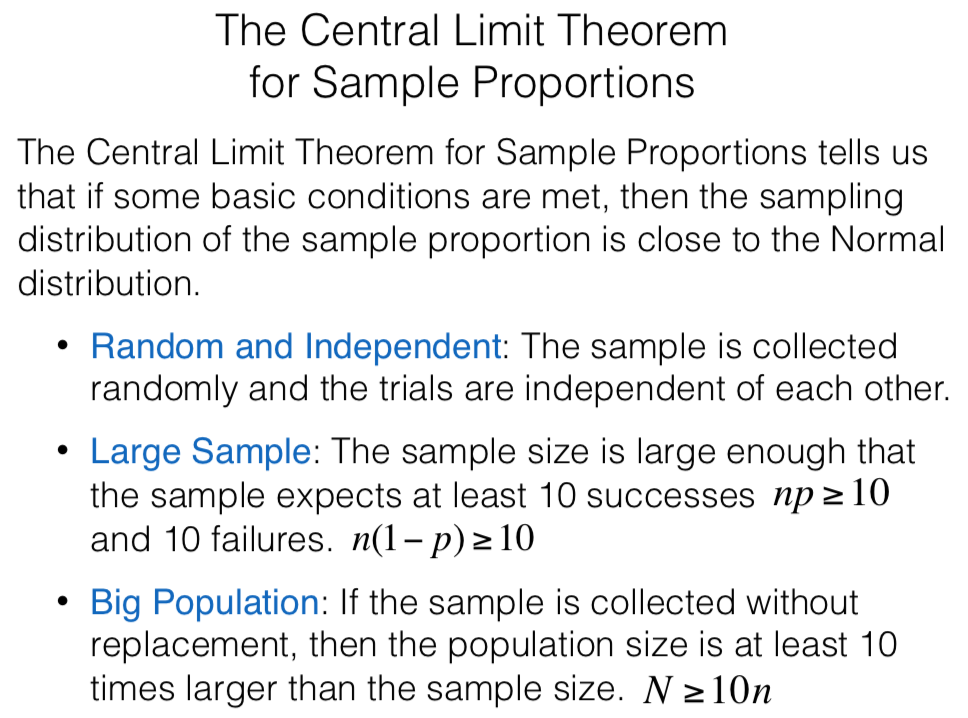


Precision: spread; Accuracy: mean

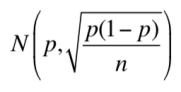
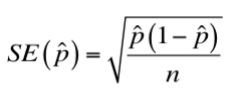
Precision and bias are independent of population size as long as population size is at least 10 times larger than the sample

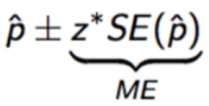


n: number of people in sample

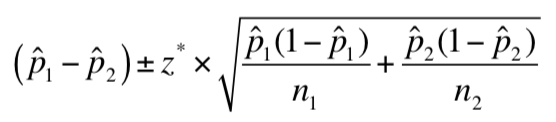


CLT for Sample Proportion:

Confidence Interval 

Difference in population mean



Random sampling is used in observational studies, allowing us to make inferences from the sample to the population

Random assignment is used in experiments, allowing us to possibly conclude causation

**Chap8**

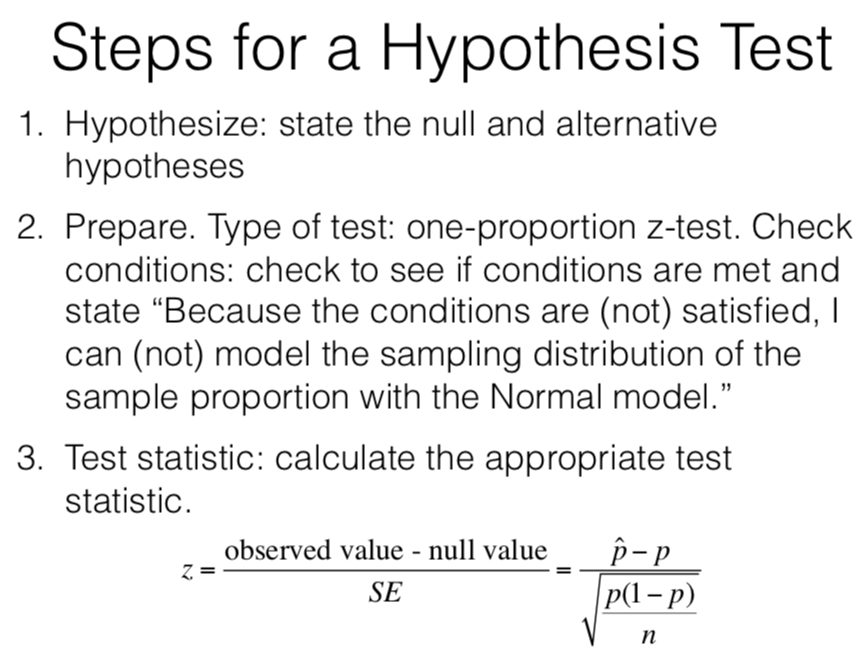
Significance level

The probability of making the mistake of rejecting H0 when H0 is true(Type I error)

We never "accept" but only "fail to reject” H0

If H0 correct, p-value should be close to 0

Hypothesis test for one proportion

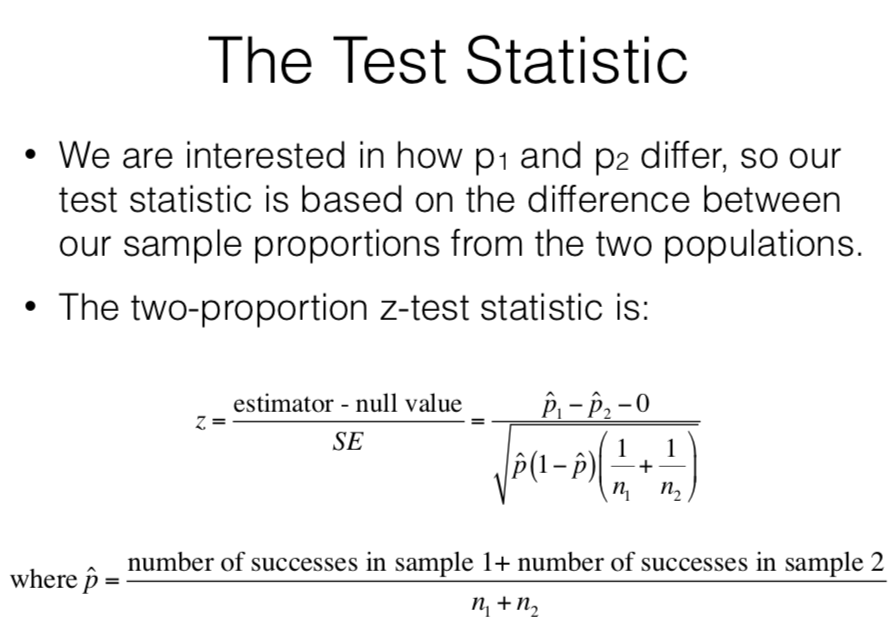


If p-value<a, reject H0, sufficient evidence that Ha is plausible

If p-value>a, fail to reject H0

Smaller p-value means easier to reject H0(less possible that H0 is true)

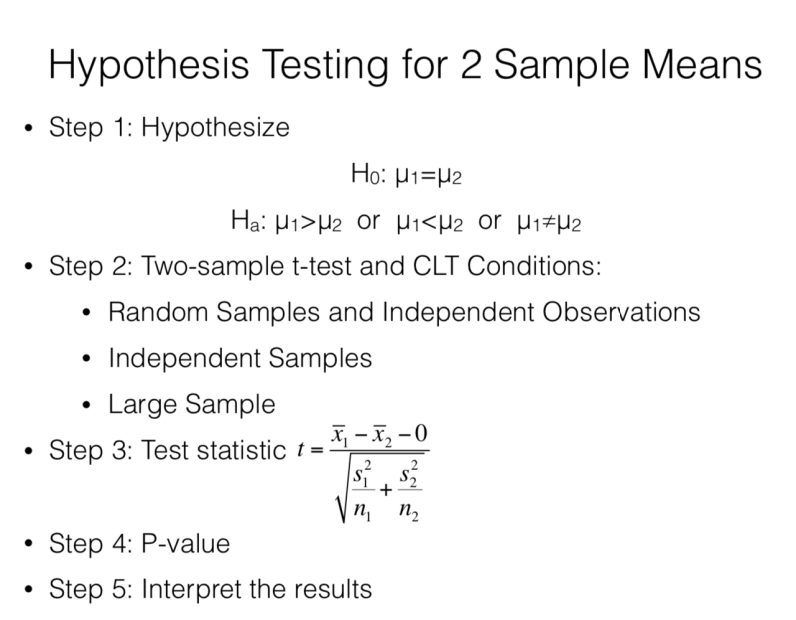
Hypothesis testing for two proportion

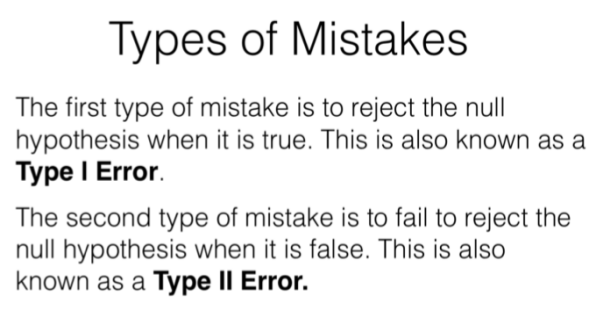


**Chap9**

Central Limit Theorem

Must assume that the population is normally distributed



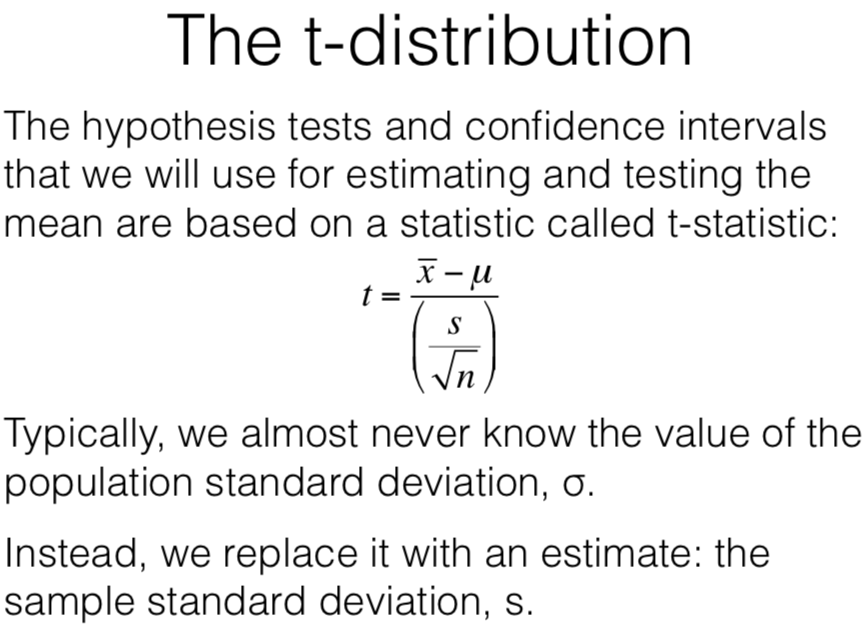


Type I: false positive, the truth is positive

Type II: false negative, the truth is negative

Use z-test for proportion->categorical;

Use t-test for mean->numerical



The larger the df, the thinner the tails and distribution is more similar to N(0,1)