A ***sample*** is a small, representative portion of a larger ***population***
***Mean*** = Average ***Median*** = Middle Number in Set (or Average of two middle numbers)
$Mean (written \overbar{x} or μ)=\frac{\sum\_{i=1}^{n}Xi}{n}$ $Variance s^{2}$= $\frac{\sum\_{}^{}\left(x\_{i}- \overbar{x}\right)^{2}}{n-1}$ ***Standard deviation*** s = $\sqrt{s^{2}}$

***Unitized/Standardized value*** z = $\frac{x\_{i}- \overbar{x}  }{s}$ ***Linear Regression*** Y= A+Bx where A=intercept, B=slope

***Least Squares Regression Formulas***: $B=r \left(\frac{s\_{y}}{s\_{x}}\right) A= \overbar{y} -B\overbar{x}$

***Covariance*** $S\_{xy}$ = $\frac{\sum\_{}^{}(x\_{i}-\overbar{x})(y\_{i}-\overbar{y})}{(n-1)}$ ***Pearson correlation*** $ r= \frac{S\_{xy}}{S\_{x}S\_{y}}$

***Mutual Exclusivity***  is when two events cannot both occur together, such as a coin simultaneously H & T
***Sample Space:*** total number of possible permutations of simple events. Sample space of 7 coin flips = $2^{7}$. Also see Extended *mn* rule, where sample space = n1 \* n2 \* n3 … \* nk .

Also consider ***sampling without replacement,*** where each trial has one less option than the previous trial. Sample space for n stages in sampling without replacement, equals n!. 5! = 5\*4\*3\*2\*1.

***Permutation*** $\frac{n!}{\left(n-r\right)!}$ ***Combinations*** $\frac{n!}{r!\left(n-r\right)!}$

If all simple events are equally likely, probability of event A = number of combinations of simple events that result in A, divided by the total number of events N. i.e. probably of exactly 1 head in 2 coin flips = 2/4 = ½, because it could be TH or HT out of {HH, TH, HT, TT}.

Two events A and B are ***independent*** if A does not affect B and B does not affect A.
***Conditional Probability.*** P(A | B) = Probability of A, Given that we know that B occurred.
***General Multiplication Rule.*** $P\left(A∩B\right)=P\left(A\right)P(B|A)$ ***.*** $∩ $means “AND”.
But when A and B are independent, $P\left(A∩B\right)=P\left(A\right)P(B)$**.**
***Conditional Probability Formula P(B|A) =*** $\frac{P(A∩B)}{P(A)}$ ***Bayes Rule*** $P\left(B\right)= \frac{P\left(A\right)P(A)}{P(B)}$

***Law of Total Probability.*** Given set of events S1, S2, S3…Sk that are mutually exclusive and exhaustive,
P(A) = P(S1)P(A|S1) + P(S2)P(A|S2) + P(S3)P(A|S3) + … P(Sk)P(A|Sk)

***Extended Form of Bayes Rule.*** $P(S\_{i}|A$***) =*** $\frac{P\left(S\_{i}\right)P\left(A \right|S\_{i})}{\sum\_{j=1}^{k}P\left(S\_{j}\right)P(A|S\_{j})}$

***Expected Value (Mean) of Discrete Random Variable: E(x) =*** $\sum\_{}^{}x\*p(x)$ ***SD:*** $\sqrt{\sum\_{}^{}\left(x-μ\right)^{2}p(x)}$

***Mean of Binomial Distribution:*** $μ=np$ ***SD of Binomial Distribution:*** $σ=\sqrt{npq}$ ***Prob. for Binomial Distribution*** $P\left(x=k\right)=(\frac{n!}{k!\left(n-k\right)!})(p^{k}q^{(n-k)})$ ***Cumulative:P(X<=x)=***$\sum\_{i=0}^{x}P(i)$ ***Binomial Distribution:*** n independent trials, same probability each trial for exactly two possible results. Calculate number of successes x during the n trials.

***Flat Distribution:*** F(X)=0 if X<i or X>j F(X)=1 if X>i and X<j Where i<j
***Continuous probability distribution***: probability of value between a & b = area under curve betw. a & b

***Controlled Experiment/RCT:*** Random assignment to control (“business as usual”) or other condition;
assignment allowing valid inference, i.e. random or stratified; identical experiences except treatment
***Quasi-Experiment:*** Same as experiment except assignment is not random or stratified
***Stratified Sample:*** Sample randomly within each group, to condition(s)

***Central Limit Theorem:*** For some non-normal populations, if you take large random sample, the sampling distribution of the mean will be approximately normally distributed

Normal Distribution: $SE= \frac{σ}{\sqrt{n}}$ Binomial Distribution: SE = $\sqrt{\frac{pq}{n}}$ Maximum margin of error: p=q=0.5

***CI (Z):*** $\overbar{X} \pm Z\_{∝/2}SE$ ***CI (t):*** $\overbar{X} \pm t(df)\_{∝/2}SE$$ SE \left(Z, two group\right)= \sqrt{\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}}}$

$SE (Z, two proportions)= \sqrt{\frac{p\_{1}q\_{1}}{n\_{1}}+\frac{p\_{2}q\_{2}}{n\_{2}}}$ $Z\left(two group\right)= \frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{SE (Z, two group)}$

**Type I Error: False Positive:** Rejecting the Null Hypothesis when the Null Hypothesis is True: 
**Type II Error: False Negative:** Accepting the Null Hypothesis when the Null Hypothesis is False: 

***Power:*** 1-P(reject H0 when Ha is true) $\overbar{x}\_{1.96}= $ + (1.96)($\frac{σ}{\sqrt{n}}$) $\overbar{x}\_{-1.96}= $ - (1.96)($\frac{σ}{\sqrt{n}}$)

***Power 95% CI bounds:*** $\frac{\overbar{x}\_{-1.96}-\overbar{x}}{\frac{σ}{\sqrt{n}}} \frac{\overbar{x}\_{1.96}-\overbar{x}}{\frac{σ}{\sqrt{n}}}$$β=P(Z\_{right}>Z> Z\_{left})$

***One-group t-test:*** $t= \frac{\overbar{x}-v}{\frac{s}{\sqrt{n}}} $***, df = n-1 Two-group t-test(unpooled) t =*** $\frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{\sqrt{\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}}}}$

***Use non-pooled variance when*** $\frac{:Larger s^{2}}{Smaller s^{2}}>3$ ***df =*** $\frac{(\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}})^{2}}{\frac{(\frac{s\_{1}^{2}}{n\_{1}})^{2}}{(n\_{1}-1)}+\frac{(\frac{s\_{2}^{2}}{n\_{2}})^{2}}{(n\_{2}-1)}}, $ ***round down***

***Two-group t-test (pooled) t =*** $\frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{\sqrt{s\_{pooled}^{2}(\frac{1}{n\_{1}}+\frac{1}{n\_{2}})}}$$s\_{pooled}^{2}$ ***=*** $\frac{\left(n\_{1}-1\right)s\_{1}^{2}+(n\_{2}-1)s\_{2}^{2}}{n\_{1}+n\_{2}-2}$ ***df =***$n\_{1}+n\_{2}-2$

***Paired t-test:*** $t= \frac{\overbar{d}}{\frac{s\_{d}}{\sqrt{n}}}$$s\_{d}=\sqrt{\frac{\sum\_{}^{}(d\_{i}- \overbar{d})^{2}}{n-1}}$ ************$= \sum\_{}^{}\frac{(observed-expected)^{2}}{expected}$$ \hat{E}\_{ij}=\frac{r\_{i}c\_{j}}{n}$

Total SS = $\sum\_{}^{}(x\_{ij} - \overbar{x})^{2}$ Total SS = SST + SSE SST = $\sum\_{}^{}n\_{i}(\overbar{x}\_{i}-\overbar{x})^{2}$ df(SST)= (k-1) df(SSE)= n-k
$df $= (r-1)(c-1) df(TSS) = (n-1) $MSS= \frac{TSS}{df (TSS)} MST= \frac{SST}{df (SST)}$ $MSE= \frac{SSE}{df (SSE)}$ $F= \frac{MST}{MSE}$