ST2334 Probability and Statistics

AY 25/26 Sem 1 — github/omgeta

1. Counting

Counting Formula: $\binom{n}{r} = \frac{n!}{r!(n-r)!}, P(n,r) = \frac{n!}{(n-r)!}$

DeMorgan's Laws:

i. $(A \cup B)' = A' \cap B'$
ii. $(A \cap B)' = A' \cup B'$

Inclusion/Exclusion Principle for finite sets A, B, C:

i. $|A \cup B| = |A| + |B| - |A \cap B|$
ii. $|A \cup B \cup C| = |A| + |B| + |C| + |A \cap B \cap C|$ $- |A \cap B| - |A \cap C| - |B \cap C|$

Number of ways to:

i. Permute n distinct $= n!$
ii. Permute n with n_1, n_2 identical $= \frac{n!}{n_1! n_2!}$
iii. Choose r of n distinct $= \binom{n}{r}$
iv. Choose r groups of n identical $= \binom{n+r-1}{n} = \binom{n+r-1}{r-1}$ $(x_1 + \cdots + x_r = n)$
v. Permute r of n distinct $= P(n,r)$
vi. Permute r of n distinct (repeat) $= n^r$

Useful results:

i. Choose 2 groups of r, m from n distinct $= \binom{n}{r}\binom{n-r}{m}$
ii. Choose k groups of r from n distinct $= \frac{\binom{n}{r}\binom{n-r}{r}\cdots\binom{r}{r}}{k!}$
iii. Permute n distinct with r together $= (n-r+1)!r!$
iv. Permute n, m distinct but separated $= m!\binom{m+1}{n}n!$
v. Permute n distinct in a circle $= (n-1)!$
vi. Permute n distinct with r together in a circle $= (n-r)!r!$
vii. Permute n, m distinct but separated in a circle $= m!\binom{m}{n}n!$
viii. Permute n distinct in a circle with 2 opposite $= (n-2)!$
ix. Permute n distinct in a circle with r identical $= \frac{(n-1)!}{r!}$

2. Probability

Probability of event E in sample space S, $P(E)$, is:

i. $P(E) = \frac{|E|}{|S|}$, where $0 \leq P(E) \leq 1$
ii. $P(E') = 1 - P(E)$ (Complement)
iii. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Union)

Conditional probability of B given A, $P(B \mid A)$, is:

i. $P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A)}$

Mutually exclusive events A, B have special results:

i. $P(A \cap B) = 0$ (Intersection)
ii. $P(A \cup B) = P(A) + P(B)$ (Union)

Independent events $A \perp B$ have special results:

i. $P(A \cap B) = P(A)P(B)$ (Intersection)
ii. $P(A \mid B) = P(A)$ (Conditional)

Total Probability for event B, partition B_1, sB_n of S:

i. $P(A) = P(A \mid B)P(B) + P(A \mid B')P(B')$
ii. $P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$
iii. $P(A \mid C) = \sum_{i=1}^{n} P(A \cap B_i \mid C) = \sum_{i=1}^{n} P(A \mid B_i \cap C)P(B_i \mid C)$

Baye's Theorem for event B, partition B_1, s, B_n of S:

i. $P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B')P(B')}$
ii. $P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)}$
iii. $P(B_k \mid A \cap C) = \frac{P(A \mid B_k \cap C)P(B_k \cap C)}{P(A \cap C)}$
iv. $\frac{P(B \mid A)}{P(B' \mid A)} = \frac{P(A \mid B)}{P(A \mid B')}\frac{P(B)}{P(B')}$ (Odds)

3. Random Variables

Probability mass function (PMF) of a discrete random variable X is:

i. $f(x) = P(X = x)$
ii. $0 \leq f(x_i) \leq 1, \forall x_i \in R_x$ and $f(x_i) = 0, \forall x_i \notin R_x$
iii. $\sum_{x_i \in R_x} f(x_i) = 1$

Probability density function (PDF) of a continuous random variable X is:

i. $\int_a^b f(x)\,dx = P(a \leq X \leq b)$
ii. $f(x) \geq 0, \forall x \in R_x$ and $f(x) = 0, \forall x \notin R_x$
iii. $\int_a^b f(x)\,dx \geq 0$ but not necessarily ≤ 1
iv. $\int_{R_x} f(x)\,dx = 1$

Cumulative density function (CDF) of any random variable X is:

i. $F(x) = P(X \leq x)$
ii. $F(x) = \int_{-\infty}^{x} f(t)dt$ and $f(x) = F'(x)$
iii. Non-decreasing and right continuous
iv. $0 \leq F(x) \leq 1$

Expectation and Variance

Expectation of random variable X, $E(X)$ or μ_X, is:

i. $E(X) = \sum_{x_i \in R_x} x_i f(x_i)$ or $\int_{-\infty}^{\infty} x f(x)\,dx$
ii. $E[g(X)] = \sum_{x_i \in R_x} g(x_i) f(x_i)$ or $\int_{-\infty}^{\infty} g(x) f(x)\,dx$
iii. $E(aX + b) = aE(X) + b$
iv. $E(X + Y) = E(X) + E(Y)$

Variance of random variable X, $V(X)$ or σ_X^2, is:

i. $V(X) = \sum_{x_i \in R_x} (x_i - \mu_X)^2 f(x_i)$ or $\int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx$ $= E(X - \mu_X)^2 = E(X^2) - [E(X)]^2$
ii. $\forall X, V(X) \geq 0$
iii. $V(aX + b) = a^2 V(X)$
iv. Standard deviation, $SD(X) = \sqrt{V(X)}$
v. $V(X + Y) = V(X) + V(Y) + 2Cov(X,Y)$ and $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$
vi. $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X,Y)$

4. Joint Distributions

Joint PMF of discrete random variables X, Y is:

i. $f_{X,Y}(x,y) = P(X=x, Y=y)$
ii. $0 \le f_{X,Y}(x,y) \le 1, \quad \forall (x,y) \in R_{X,Y}$ and $f_{X,Y}(x,y) = 0, \quad \forall (x,y) \notin R_{X,Y}$
iii. $\sum\sum_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y) = 1$

Joint PDF of continuous random variables X, Y is:

i. $P((X,Y) \in D) = \iint_D f(x,y)\,dx\,dy$
ii. $f(x,y) \ge 0, \quad \forall (x,y) \in R_{X,Y}$ and $f(x,y) = 0, \quad \forall (x,y) \notin R_{X,Y}$
iii. $\iint_{R_{X,Y}} f(x,y)\,dx\,dy = 1$

Marginal distribution and conditional distribution are:

i. $f_X(x) = \sum_y f_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
ii. $f_{Y|X}(y \mid x) = P(Y=y \mid X=x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

Independent random variables X, Y have special results:

i. $f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad \forall (x,y) > 0 \in R_{X,Y}$
ii. $R_{X,Y}$ is a product space, $R_{X,Y} = R_X \times R_Y$

Expectation and Variance

Expectation of random variables X, Y, $E(X,Y)$, is:

i. $E[g(X,Y)] = \sum_{R_X}\sum_{R_Y} g(x,y) f_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
ii. If independent, $E(XY) = E(X)E(Y)$

Covariance of random variables X, Y, $Cov(X,Y)$, is:

i. $Cov(X,Y) = \sum_{R_X}\sum_{R_Y}(x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)\,dxdy$ $= E[(X-\mu_X)(Y-\mu_Y)]$ $= E(XY) - \mu_X\mu_Y$
ii. X, Y are independent $\implies Cov(X,Y) = 0$
iii. $Cov(X,Y) = Cov(Y,X)$ and $Cov(X,X) = V(X)$
iv. $Cov(aX+b, cY+d) = ac \cdot Cov(X,Y)$
v. $Cov(W+X, Y+Z) =$ $Cov(W,Y) + Cov(W,Z) + Cov(X,Y) + Cov(X,Z)$

5. Discrete Probability Distributions

Uniform Distribution: $X \sim \text{Unif}(x_1, \cdots, x_k)$

i. $f_X(x) = \frac{1}{k}, \quad x \in x_1, \ldots, x_k$
ii. $\mu_X = \frac{x_1 + \cdots + x_k}{k}, \quad \sigma_X^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu_X^2$

Bernoulli Trial: $X \sim \text{Bern}(p)$ is the outcome of a single trial with success probability p, fail probability $q = 1-p$

i. $f_X(x) = p^x q^{1-x}, \quad x = 0$ (fail), 1 (success)
ii. $\mu_X = p, \quad \sigma_X^2 = pq$

Binomial Distribution: $X \sim \text{Bin}(n,p) = \sum X_i$ is the successes in n independent Bernoulli trials $X_i \sim \text{Bern}(p)$

i. $f_X(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots n$
ii. $\mu_X = np, \quad \sigma_X^2 = npq$

Negative Binomial Distribution: $X \sim \text{NB}(k,p)$ is the number of independent Bernoulli trials until k^{th} success

i. $f_X(x) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, \ldots$
ii. $\mu_X = \frac{k}{p}, \quad \sigma_X^2 = \frac{qk}{p^2}$

Geometric Distribution: $X \sim \text{Geom}(p)$ is the number of independent Bernoulli trials until the first success

i. $f_X(x) = pq^{x-1}, F_X(x) = 1 - q^x \quad x = 1, 2, \ldots$
ii. $\mu_X = \frac{1}{p}, \quad \sigma_X^2 = \frac{q}{p^2}$

Poisson Distribution: $X \sim \text{Poisson}(\lambda)$ is the number of events occurring in a fixed interval or region where $\lambda > 0$ is expected number of occurences in the interval

i. $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$
ii. $\mu_X = \sigma_X^2 = \lambda$
iii. As $n \to \infty$ and $p \to 0$, $X \sim \text{Bin}(n,p)$ converges to $X \sim Poisson(\lambda = np)$. Good approximation if:
   - $n \ge 20$ and $p \le 0.05$, or if
   - $n \ge 100$ and $np \le 10$
iv. Poisson process counts the number of events within interval of time scaled by rate α, such that:
   - expected occurences in interval T is αT
   - no simultaneous occurences
   - number of occurences in disjoint time intervals are independent

6. Continuous Probability Distributions

Uniform Distribution: $X \sim \text{Unif}(a,b)$

i. $f_X(x) = \frac{1}{b-a}, \quad a \le x \le b$
ii. $\mu_X = \frac{a+b}{2}, \quad \sigma_X^2 = \frac{(b-a)^2}{12}$
iii. CDF, $F_X(x) = \frac{x-a}{b-a}, \quad a \le x \le b$

Exponential Distribution: $X \sim \text{Exp}(\lambda)$ is the waiting time for first success in continuous time

i. $f_X(x) = \lambda e^{-\lambda x}, \quad x \ge 0$
ii. $\mu_X = \frac{1}{\lambda}, \quad \sigma_X^2 = \frac{1}{\lambda^2}$
iii. CDF, $F_X(x) = 1 - e^{-\lambda x}, \quad x \ge 0$
iv. $P(X > s+t \mid X > s) = P(X > t)$ (Memoryless)

Normal Distribution: $X \sim \text{N}(\mu, \sigma^2)$ is symmetric about μ and flattens out as σ increases

i. $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad x \in \mathbb{R}$
ii. $\mu_X = \mu, \quad \sigma_X^2 = \sigma^2$
iii. Standard normal: $Z \sim N(0,1) = \frac{X-\mu}{\sigma}$
   - $\phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$
   - $\Phi(z) = \int_{-\infty}^{z}\phi(t)dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp(-\frac{t^2}{2})dt$
   - $P(x_1 < X < x_2) = \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$
   - $P(Z \ge 0) = P(Z \le 0) = \Phi(0) = 0.5$
   - $\Phi(z) = P(Z \le z) = P(Z \ge -z) = 1 - \Phi(-z)$
   - $\sigma Z + \mu \sim N(\mu, \sigma^2)$
iv. Upper α quantile x_α satisfies:
   - $P(X \ge x_\alpha) = \alpha$
   - $P(Z \ge z_\alpha) = P(Z \le -z_\alpha) = \alpha$
v. As $n \to \infty$ but p remains constant, $X \sim \text{Bin}(n,p)$ approximates $Z = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$.
   Good approximation if: $np > 5$ and $n(1-p) > 5$
vi. Apply the continuity corrections for approximating:

Discrete Probability	Normal Approx.
$P(X = k)$	$P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right)$
$P(a \le X \le b)$	$P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right)$
$P(a < X < b)$	$P\left(a + \frac{1}{2} < X < b - \frac{1}{2}\right)$
$P(X \le c)$	$P\left(0 \le X \le c\right)$
$P(X > c)$	$P\left(c < X \le n\right)$

7. Sampling

Population is the totality of all possible outcomes in an experiment. They can be finite or infinite.
Population parameter is a population's numerical fact.
Sample is any subset of a population.

Probability sampling:

i. Simple Random Sampling (finite pop.): every subset of n observations of the population has the same probability of being selected.
ii. Simple Random Sampling (infinite pop.): random sample $X_1, \dots, X_n$ are n independent random variables with same distribution $f_X(x)$ as X s.t. $f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)\dots f_{X_n}(x_n)$.

Statistic is random variable functions of sample data:

i. Sampling Mean, $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$
   - $\mu_{\overline{X}} = \mu_X, \quad \sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}$
   - Standard error, $\sigma_{\overline{X}}$ describes how much $\overline{x}$ tends to vary from sample to sample of size n
ii. Sampling Variance, $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
   - $\mu_{S^2} = \sigma^2, \quad \sigma^2_{S^2} = \frac{1}{n}(\mu_4 - \frac{n-3}{n-1}\sigma^4)$

Law of Large Numbers:

i. As sample size $n \to \infty$, $\frac{\sigma^2}{n} \to 0$ and $\overline{X} \to \mu_X$, $P(|\overline{X} - \mu_X| > \epsilon) \to 0$

Central Limit Theorem (for means):

i. Sampling distribution of sample mean $\overline{X}$ is approximately normal if n is sufficiently large
ii. $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to Z \sim N(0,1)$ as $n \to \infty$
iii. Conditions for population of sample:
   - Symmetric with no outliers, needs 15-20 samples
   - Moderately skewed (exponential or χ^2), can take 30-50 samples
   - Extreme skewness may not be appropriate for CLT even with 1000 samples

8. Sampling Distribution

Diff. of Sample Means: $\overline{X_1} - \overline{X_2} = \frac{\overline{X_1}-\overline{X_2}-\mu_{\overline{X_1}-\overline{X_2}}}{\sigma_{\overline{X_1}-\overline{X_2}}}$ approx. $N(0,1)$ for independent random variables $\overline{X_1} \sim \mathrm{N}(\mu_1, \sigma_1^2/n_1), \overline{X_2} \sim \mathrm{N}(\mu_2, \sigma_2^2/n_2)$

i. $\mu_{\overline{X_1}-\overline{X_2}} = \mu_1 - \mu_2, \quad \sigma^2_{\overline{X_1}-\overline{X_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Chi-Squared Distribution: $Y \sim \chi^2(n) = \sum_{i=1}^{n} Z_i^2$ is the sum of n independent and identically distributed standard normal random variables, with long right tail and n degrees of freedom

i. $\mu_Y = n, \quad \sigma_Y^2 = 2n$
ii. For large n, $\chi^2(n)$ is approximately $N(n, 2n)$
iii. $Y_1 \sim \chi^2(m), Y_2 \sim \chi^2(n)$ are independent $\implies Y_1 + Y_2 \sim \chi^2(m+n)$
iv. Define $\chi^2(n;\alpha)$: $P(Y > \chi^2(n;\alpha)) = \alpha$
v. If S^2 is sample variance of size n from normal population of variance σ^2, then $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$

t-Distribution: $T \sim t(n) = \frac{Z}{\sqrt{U/n}}$ for independent $Z \sim N(0,1)$ and $U \sim \chi^2(n)$, with n degrees of freedom is symmetric vertical and resembles standard normal graph

i. $\mu_T = 0, \quad \sigma_T^2 = \frac{n}{n-2}$ for $n > 2$
ii. For $n \geq 30$, can be replaced by $N(0,1)$
iii. Define $t_{n;\alpha}$: $P(T > t_{n;\alpha}) = \alpha$
iv. If $X_1, \dots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2, $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

F-Distribution: $F \sim F(m,n) = \frac{U/m}{V/n}$ for independent $U \sim \chi^2(m), V \sim \chi^2(n)$ has (m,n) degrees of freedom

i. $\mu_F = \frac{n}{n-2}$ for $n > 2$
   $\sigma_F^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ for $n > 4$
ii. $\frac{1}{F} \sim F(n,m)$
iii. Define: $F(m,n;\alpha)$: $P(F > F(m,n;\alpha)) = \alpha$
iv. $F(m,n;1-\alpha) = \frac{1}{F(n,m;\alpha)}$

9. Estimation

Estimators are rules, usually formulas, used to compute an estimate from the sample.

i. Point Estimator: A single number is calculated
   - Unbiased Estimator: An estimator $\hat{\theta}$ of a parameter θ is unbiased if $E(\hat{\theta}) = \theta$.
ii. Interval Estimation: An interval is calculated for some confidence level

Maximum error E for estimating μ using $\bar{X}$ when σ is known for confidence level $(1-\alpha)$ is: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Sample size to achieve maximum error E_0 with confidence level $(1-\alpha)$ is: $n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{E_0}\right)^2$

10. Hypothesis Testing

Hypothesis test can be used given a null hypothesis H_0, a alternative hypothesis H_1, and level of significance α.

	Do not reject H_0	**Reject H_0**
H_0 true	Correct	**Type I Error**
H_0 false	**Type II Error**	Correct

i. LOS $\alpha = P(\text{Type I}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$
ii. $\beta = P(\text{Type II}) = P(\text{Do not reject } H_0 \mid H_0 \text{ is false})$
iii. Power $1-\beta$ is given by $P(\text{Reject } H_0 \mid H_0 \text{ is false})$

Test statistic (e.g. z, t) is a function of sample data.

p-value can be defined as:

i. Probability of obtaining a sample statistic as extreme or more extreme than the observed statistic, assuming H_0 is true.
ii. Smallest level of significance at which H_0 is rejected, assuming H_0 is true

where we reject H_0 in favour of H_1 when p-value $< \alpha$
or not reject H_0 (doesn't imply H_0 true) when p-value $\geq \alpha$

Test Statistics for Population Mean

Case	Population	σ	n	CI	Statistic	n for desired E_0, α
I	Normal	known	any	$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$
II	any	known	≥ 30	$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$
III	Normal	unknown	< 30	$\overline{x} \pm t_{n-1;\alpha/2} \cdot \frac{s}{\sqrt{n}}$	$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$	$\left(\frac{t_{n-1;\alpha/2} \cdot s}{E_0}\right)^2$
IV	any	unknown	≥ 30	$\overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot s}{E_0}\right)^2$

For example, if $T \sim t_{n-1}$ under H_0:

- Right-tailed test $H_1 : \mu > \mu_0$: reject H_0 if $T > t_{n-1;\alpha}$.
- Left-tailed test $H_1 : \mu < \mu_0$: reject H_0 if $T < -t_{n-1;\alpha}$.
- Two-tailed test $H_1 : \mu \neq \mu_0$: reject H_0 if $|T| > t_{n-1;\alpha/2}$.

For a given observed test statistic t_{obs}:

- Right-tailed: p-value $= P(T \geq t_{\text{obs}} \mid H_0)$.
- Left-tailed: p-value $= P(T \leq t_{\text{obs}} \mid H_0)$.
- Two-tailed: p-value $= P(|T| \geq |t_{\text{obs}}| \mid H_0)$.

Given CI for $H_0 : \mu = \mu_0$:

- do not reject H_0 if μ_0 in CI
- reject H_0 if μ_0 not in CI

Test Statistics for Independent Samples

Population	Variance	σ_1, σ_2	n	CI	Statistic
any	known	unequal	≥ 30	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
Normal	known	unequal	any	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
any	unknown	unequal	≥ 30	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$
Normal	unknown	equal	< 30	$(\bar{x} - \bar{y}) \pm t_{n_1+n_2-2;\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$T = \frac{\bar{X} - \bar{Y}}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
any	unknown	equal	≥ 30	$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$Z = \frac{\bar{X} - \bar{Y}}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$

*Variance assumed equal if $\frac{1}{2} < \frac{s_1}{s_2} < 2$

** For dependent samples, consider sample $D_i - X_i - Y_i$ and use results for Population Mean.

Pooled Estimator: $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

Maclaurin series for $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

Geometric series: $a + ar + ar^2 + \cdots + ar^{n-1} = \frac{a(1-r^n)}{1-r}$