NUS ST2334 Final Cheat Sheet

Counting and Probability

Permutation

order is taken into consideration

${}_nP_r = n(n-1)(n-2)\cdots(n-(r-1)) = n!/(n-r)!$

When not all objects are distinct,

$nP_{n_1,n_2,\cdots,n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$. In circle: $(n-1)!$

Combination

order not considered. $\left(\begin{array}{c} n \\ r \end{array}\right) = \frac{n!}{r!(n-r)!}$

Axioms of Probability

Axiom 1: $0 \leq Pr(A) \leq 1$

Axiom 2: $Pr(S) = 1$

Axiom 3: If $A_1, A_2, \cdots$ are mutually exclusive (disjoint) events, then

$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$

Inclusion-Exclusion Principle

$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n)$
$= \sum_{i=1}^{n} \Pr(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr(A_i \cap A_j)$
$+ \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \Pr(A_i \cap A_j \cap A_k)$
$- \cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)$

The Law of Total Probability

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S. That is $A_1, A_2, \cdots, A_n$ are mutually exclusive and exhaustive events such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{n} A_i = S$.

Then for any event B

$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i) \Pr(B|A_i)$

Bayes' Theorem

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S. Then

$\Pr(A_k|B) = \frac{\Pr(A_k)\Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)}$

for $k = 1, \cdots, n$. Or

$\Pr(A_k|B) = \frac{\Pr(A_k)\Pr(B|A_k)}{\Pr(B)}$

Independence

Independent Events:

Two events A and B are independent iff

$\Pr(A \cap B) = \Pr(A)\Pr(B)$

Pairwise Independence:

A set of events $A_1, A_2, \cdots, A_n$ are pairwise independent iff $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$ for $i \neq j$ and $i, j = 1, \cdots, n$

Mutual Independence:

A set of events $A_1, A_2, \cdots, A_n$ are mutually independent iff for any subset $\{A_{i1}, A_{i2}, \cdots, A_{ik}\}$ of $A_1, A_2, \cdots, A_n$,

$\Pr\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) =$
$\Pr\left(A_{i_1}\right)\Pr\left(A_{i_2}\right)\cdots\Pr\left(A_{i_k}\right)$

Note: their complements are also mutually independent.

Concepts of Random Variables

Probability (Mass) Function

The probability of $X = x_i$ denoted by $f(x_i)$ (i.e. $f(x_i) = \Pr(X = x_i)$, must satisfy the following two conditions.

(1) $f(x_i) \geq 0$ for all x_i.

(2)$\sum_{i=1}^{\infty} f(x_i) = 1$

Probability Density Function

Let X be a **continuous** random variable.

(1) $f(x) \geq 0$ for all $x \in R_X$

(2) $\int_{R_X} f(x)dx = 1$ or $\int_{-\infty}^{\infty} f(x)dx = 1$

since $f(x) = 0$ for x not in R_X

(3) For any c and d such that $c < d$, $(i.e. (c,d) \subset \boldsymbol{R}_X)$, $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$

Cumulative Distribution Function

Defined as: $F(x) = \Pr(X \leq x)$

If X is a **discrete** random variable, then its c.d.f is a step function: $F(x) = \sum_{t \leq x} f(t) = \sum_{t \leq x} \Pr(X = t)$

If X is a **continuous** random variable, then $F(x) = \int_{-\infty}^{x} f(t)dt$

For a **continuous** random variable X, $f(x) = \frac{dF(x)}{dx}$ if the derivative exists.

Mean

Discrete random variable:

$\mu_X = E(X) = \sum_i x_i f(x_i) = \sum_x x f(x)$

Continuous random variable:

$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)dx$

For any function $g(X)$,

(a) $E[g(X)] = \sum_x g(x) f_X(x)$

(b) $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$

Property: $E(aX + b) = aE(X) + b$

In general,

$E[a_1 g_1(X) + a_2 g_2(X) + \cdots + a_k g_k(X)]$
$= a_1 E[g_1(X)] + a_2 E[g_2(X)] + \cdots + a_k E[g_k(X)]$

Variance

$$\sigma_X^2 = V(X) = E\left[(X - \mu_X)^2\right]$$

$$= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Remarks:

(a) $V(X) \geq 0$

(b) $V(X) = E(X^2) - [E(X)]^2$

(c) **Standard deviation** is the **positive square root** of the variance.

Property:$V(aX + b) = a^2 V(X)$

Moment

The **k-th moment** of X is defined by $E(X^k)$.

Chebyshev's Inequality

Let X be a random variable (discrete or continuous, with any distribution with finite mean and var) with $E(X) = \mu$ and $V(X) = \sigma^2$. For any $k > 0$,

$\Pr(|X - \mu| \geq k\sigma) \leq 1/k^2$

That is, the probability that the value of X lies at least k standard deviation from its mean is at most $\frac{1}{k^2}$. Alternatively, $\Pr(|X - \mu| < k\sigma) \geq 1 - 1/k^2$

Discrete vs. Continuous 2D-RVs

Discrete: the possible values of $(X(s), Y(s))$ are **finite or countably infinite**.

Continuous: the possible values of $(X(s), Y(s))$ can assume all values in some region of the Euclidean plane $\mathbb{R}^2$.

Joint Probability Function (discrete)

$f_{X,Y}(x_i, y_i)$ represents $\Pr(X = x_i, Y = y_i)$ and satisfies the following conditions:

1. $f_{X,Y}(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in R_{X,Y}$.
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) =$
$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$

The function $f_{X,Y}(x, y)$ is called the **joint probability function** for (X, Y).

$$\Pr((X,Y) \in A) = \underbrace{\sum\sum}_{(x,y) \in A} f_{X,Y}(x,y)$$

Joint Probability Density Function

$f_{X,Y}(x,y)$ (continuous) satisfies the following conditions:

(1) $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in R_{X,Y}$

(2) $\iint_{(x,y) \in R_{X,Y}} f_{X,Y}(x,y)dxdy = 1$

or

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

Marginal Probability Distribution

Discrete:

$f_X(x) = \sum_y f_{X,Y}(x,y)$ and

$f_Y(y) = \sum_x f_{X,Y}(x,y)$

Continuous:

$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ and

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

Conditional Distribution

The conditional distribution of Y given that $X = x$ is given by $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, if $f_X(x) > 0$ for each x within the range of X. Vice versa for X given Y.

Independent Random Variables

Random variables X and Y are independent iff $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y. This can be extended to n random variables where $n \geq 2$.

Expectation

$E[g(X,Y)] =$

$$\begin{cases} \sum_x \sum_y g(x,y) f_{X,Y}(x,y), & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy, & \text{cont.} \end{cases}$$

Covariance

Let (X,Y) be a bivariate random vector with joint p.f. (or p.d.f.) $f_{X,Y}(x,y)$ then the covariance of (X,Y) is defined as

$Cov(X,Y) = E[(x - \mu_x)(y - \mu_y)] =$

$$\begin{cases} \sum_x \sum_y (x - \mu_x)(y - \mu_y) f_{X,Y}(x,y), & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f_{X,Y}(x,y)dxdy, & \text{cont.} \end{cases}$$

Remarks: 1. $Cov(X,Y) = E(XY) - \mu_x\mu_y$

2. If X and Y are independent, $Cov(X,Y)$ is 0. The converse may not be true.

3. $Cov(aX + b, cY + d) = acCov(X,Y)$

4. $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$

Correlation Coefficient

It measures the degree of linear relationship between X and Y. $-1 \leq \rho_{X,Y} \leq 1$.

The correlation coefficient of X and Y, denoted by $\operatorname{Cor}(X,Y), \rho_{X,Y}$ or ρ, is defined by

$$\rho_{X,Y} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Special Distributions

Discrete Uniform Distribution

$f_X(x) = \frac{1}{k}, forx = x_1, x_2, ..., x_k,$, and 0 otherwise.

Bernoulli Distribution

Bernoulli Experiment: a random experiment with only two possible outcomes, say "success" or "failure".

Bernoulli Distribution: A random variable X is defined to have a Bernoulli distribution if the probability function of X is given by

$f_X(x) = p^x(1-p)^{1-x}, x = 0, 1$, and 0 otherwise.

$Pr(X = 1) = p$ and $Pr(X = 0) = 1 - p = q$.

Mean: $\mu = E(X) = p$

Variance: $\sigma^2 = V(X) = p(1-p) = pq$

Binomial Distribution

$X \sim B(n,p)$, if the probability function of X is:

$\Pr(X = x) = f_X(x) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x(1-p)^{n-x} =$

$\left(\begin{array}{c} n \\ x \end{array}\right) p^x q^{n-x}$

for $x = 0, 1, ..., n$ and $0 < p < 1$ and $q = 1 - p$.

X is the number of successes that occur in n independent Bernoulli trials.

Mean: $\mu = E(X) = np$

Variance: $\sigma^2 = V(X) = np(1-p) = npq$

Negative Binomial Distribution

Let X be a random variable represents the number of trials to produce the k successes in a sequence of independent Bernoulli trials. The random variable X is said to follow a Negative Binomial distribution with parameters k and p (i.e. $NB(k,p)$). The probability function of X is given by

$\Pr(X = x) = f_X(x) = \left(\begin{array}{c} x-1 \\ k-1 \end{array}\right) p^k q^{x-k}$

for $x = k, k+1, k+2, \cdots$

Mean: $\mu = E(X) = \frac{k}{p}$

Variance: $\sigma^2 = V(X) = \frac{(1-p)k}{p^2}$

Geometric Distribution

The number of trials that are required to have the first success is known to follow **geometric distribution**. Let X be the number of attempts necessary for the first success. Therefore X follows a Negative Binomial Distribution with parameters k = 1 and p. (or X follows a Geometric Distribution with p = p). That is $X \sim NB(1,p)$ or $X \sim Geom(p)$.

Poisson Distribution

$f_X(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, 3, \cdots$

Mean: $\mu = E(X) = \lambda$

Variance: $\sigma^2 = V(X) = \lambda$

Poisson approx. to Binomial Dist.

Let $X \sim B(n,p)$. Suppose that $n \to \infty$ and $p \to 0$ such that $\lambda = np$ stays constant. Then $X \sim P(np)$ approximately. That is

$\lim_{\substack{p \to 0 \\ n \to \infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$

Remark: If p is close to 1, we can interchange success and failure to make p close to zero.

Continuous Uniform Distribution

$X \sim U(a,b)$ over an interval $[a,b]$ if

$f_X(x) = \frac{1}{b-a}$, for $a \leq x \leq b$, and 0 otherwise.

(a.k.a. rectangular distribution)

Mean: $\mu = E(X) = \frac{a+b}{2}$

Variance: $\sigma^2 = V(X) = \frac{1}{12}(b-a)^2$

Exponential Distribution

$f_X(x) = \alpha e^{-\alpha x}$, parameter $\alpha > 0$ and $x > 0$; 0 otherwise. $X \sim Exp(\alpha)$

(frequently used as a model for the distribution of times between the occurrence of successive events)

Mean: $\mu = E(X) = \frac{1}{\alpha}$

Variance: $\sigma^2 = V(X) = \frac{1}{\alpha^2}$

OR: $f_X(x) = \frac{1}{\mu} e^{-x/\mu}$, for $x > 0$

Mean: $\mu = E(X) = \mu$

Variance: $\sigma^2 = V(X) = \mu^2$

No Memory Property: for any two positive numbers s and t, $\Pr(X > s + t | X > s) = \Pr(X > t)$
Upper-tailed cdf: $\Pr(X > x) = e^{-ax}$, for $x > 0$

Normal Distribution

$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $-\infty < x < \infty$
denoted by $N(\mu, \sigma^2)$

Standardised Normal Distribution

If $X \sim N(\mu, \sigma^2)$, and if $Z = \frac{(X-\mu)}{\sigma}$, then $Z \sim N(0, 1)$.

* Linear Interpolation (e.g.)

Let $Pr(Z > a) = 0.12$. From the normal table, we have $Pr(Z \geq 1.17) = 0.121$ and $Pr(Z \geq 1.18) = 0.119$. Hence,
$\frac{a-1.17}{1.18-1.17} = \frac{0.12-0.121}{0.119-0.121} \Rightarrow a = 1.175$

Normal approx. to Binomial Dist.

When $n \to \infty$ and $p \to 1/2$. **Rule of thumb**: use this only when $np > 5$ and $n(1-p) > 5$.
If $X \sim B(n, p)$ ($\mu = np, \sigma^2 = np(1-p)$), as $n \to \infty$, $Z = \frac{X-np}{\sqrt{npq}}$ is approximately $\sim N(0, 1)$

Continuity Correction

(for norm. approx. to binom.)
(a) $Pr(X = k) \approx Pr(k - 1/2 < X < k + 1/2)$
(b) $\Pr(a < X \leq b) \approx \Pr(a + 1/2 < X < b + 1/2)$
(c) $\Pr(X \leq c) = \Pr(0 \leq X \leq c) \approx \Pr(-1/2 < X < c + 1/2)$
(d) $\Pr(X > c) = \Pr(c < X \leq n) \approx \Pr(c + 1/2 < X < n + 1/2)$

Sampling

Sample Mean

If $(X_1, X_2, ..., X_n)$ represent a random sample of size n, then the sample mean is defined by the statistic $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sampling Dist. of Sample Mean

For random samples of size n taken from an inf. pop. or from a finite pop. with replacement having pop. mean μ and pop. s.d. σ, the sampling distribution of the sample mean $\bar{X}$: $\mu_{\bar{X}} = \mu_X$ and $\sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n}$

Law of Large Numbers

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2. Then for any $\epsilon \in \mathbb{R}$, $P(|\bar{X} - \mu| > \epsilon) \to 0$ as $n \to \infty$

Central Limit Theorem

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2. The sampling distribution of the sample mean $\bar{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ if n is sufficiently large (**Rule of thumb**: at least 30).
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Sampling distribution of the difference of two sample means

If independent samples of sizes $n_1 (\geq 30)$ and $n_2 (\geq 30)$ are drawn from two populations, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of the sample means, $\bar{X}_1$and $\bar{X}_2$, is approximately normally distributed with mean and standard deviation given by

$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ and $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Chi-square Distribution

$Y \sim \chi^2(n)$ with n degrees of freedom (n is a positive integer).
Mean: $\mu = E(Y) = n$
Variance: $\sigma^2 = V(Y) = 2n$
(1) For large n, $\chi^2(n)$ approx $\sim N(n, 2n)$
(2) If $Y_1, Y_2, ..., Y_k$ are **independent** chi-square random variables with $n_1, n_2, ..., n_k$ degrees of freedom respectively, then $\sum_{i=1}^{k} Y_i \sim \chi^2\left(\sum_{i=1}^{k} n_i\right)$

Theorem regarding Chi-square and random sample

1. If $X \sim N(0, 1)$, then $X^2 \sim \chi^2(1)$.
2. Let $X \sim N(\mu, \sigma^2), then [(x - \mu)/\sigma]^2 \sim \chi^2(1)$.
3. Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a normal population with mean μ and variance σ^2. Define $Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$. Then $Y \sim \chi^2(n)$

Sample Variance

Let $X_1, X_2, ..., X_n$ be a random sample from a population. Sample variance:
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

Sampling distribution of $(n-1)S^2/\sigma^2$

If S^2 is the variance of a random sample of size n taken from a **normal** population having the variance σ^2, then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

t-distribution

Suppose $Z \sim N(0, 1)$, and $U \sim \chi^2(n)$. If Z and U are independent, then let $T = \frac{Z}{\sqrt{U/n}} \sim t(n)$
(t-distribution with n degrees of freedom)
Mean: $\mu = E(T) = 0$
Variance: $\sigma^2 = V(T) = n/(n-2)$ for $n > 2$
Remark: If the random sample was selected from a normal population, then
$Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$ and $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
It can be shown that $\bar{X}$ and S^2 are independent, and so are Z and U. Therefore,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}$$

T (t-value) has a t-distribution with $n - 1$ d.f..

F-distribution

Let U and V be independent random variables having $\chi^2(n_1)$ and $\chi^2(n_2)$, respectively, then the distribution of the random variable, $F = \frac{U/n_1}{V/n_2}$ is called an F-distribution with (n_1, n_2) degrees of freedom. Its p.d.f. $f_F(x) > 0$ for $x > 0$ and 0 otherwise.
Mean: $\mu = E(X) = n_2/(n_2 - 2)$ with $n_2 > 2$
Variance: $\sigma^2 = V(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ with $n_2 > 4$
Remarks:
(1) Suppose that random samples of sizes n_1 and n_2 are selected from two **normal populations** with variances σ_1^2 and σ_2^2 respectively.

$$U = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$$

$$V = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

are independent random variables. Therefore,

$$F = \frac{U/(n_1-1)}{V/(n_2-1)} = \frac{\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_1-1)}}{\frac{(n_2-1)S_2^2/\sigma_2^2}{(n_2-1)}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

(2) If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.
(3) $F(n_1, n_2; 1 - \alpha) = 1/F(n_2, n_1; \alpha)$ (useful for statistical table)

Estimation based on NormDist.

Unbiased Estimator

A statistic $\widehat{\Theta}$ is an unbiased estimator of parameter θ if $E(\widehat{\Theta}) = \theta$. E.g.: $E(\bar{X}) = \mu$, $E(S^2) = \sigma^2$

Margin of Error e

We want $\Pr(|\bar{X} - \mu| \leq e) \geq 1 - \alpha$.
$e \geq z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$. Hence, for a given margin of error e, the sample size is given by $n \geq \left(z_{\alpha/2}\frac{\sigma}{e}\right)^2$

Confidence Interval

Suppose $(\widehat{\Theta}_L, \widehat{\Theta}_U)$ and $Pr(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U) = 1 - \alpha$. Then this interval is called a $(1 - \alpha)100\%$ **confidence interval** for θ. $(1 - \alpha)$ is called **confidence coefficient** or **degree of confidence**.
3 conditions: A. normal distributions; B. other parameters known; C. large sample sizes

One sample; CI for pop. mean μ

(1) B(known σ^2) AND (A OR C):
Test statistic: $Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$
$(1 - \alpha)100\%$(same below) CI: $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
(2) $\neg B$(unknown σ^2) AND A:
Test statistic: $T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t(n-1)$
CI:$\bar{X} \pm t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}$
(3) $\neg B$(unknown σ^2) AND C:
Test statistic: $T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}}$ approx $\sim N(0, 1)$
CI: $\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$

Two indep. samples; CI for diff. b/w 2 means

(1) B(known and unequal σ_1^2, σ_2^2) AND (A OR C):
Test statistic: $T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$
CI: $\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
(2) $\neg B$(unknown σ_1^2, σ_2^2) AND C:
Same as (1), just replace σ_1^2, σ_2^2 with S_1^2, S_2^2
(3) $\neg B$(unknown but EQUAL σ_1^2, σ_2^2) AND A:
Test statistic: $T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} \sim t(n_1 + n_2 - 2)$
where σ_2 can be estimated by the pooled sample variance: $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
CI: $\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2;\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ (note: for large sample sizes, replace $t_{n_1+n_2-2;\alpha/2}$ by $z_{\alpha/2}$.

Two paired /dep. samples; CI for diff. b/w 2 means

[Just treat the differences as a sample itself]
(1) $\neg B$ AND A:
Test statistic: $T = \frac{\bar{X}_D - \mu_D}{S_D/\sqrt{n}} \sim t(n-1)$
CI: $\bar{X}_D \pm t_{n-1;\alpha/2}\frac{S_D}{\sqrt{n}}$
(2) $\neg B$ AND C:
Test statistic: $T = \frac{\bar{X}_D - \mu_D}{S_D/\sqrt{n}}$ approx $\sim N(0, 1)$
CI: $\bar{X}_D \pm z_{\alpha/2}\frac{S_D}{\sqrt{n}}$

One sample; CI for pop. variance σ^2

[note: for σ, just take square roots on both ends of CI]
(1) B(known μ) AND A:
Test statistic: $T = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$
CI: $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$
(2) $\neg B$(unknown μ) AND A:
Test statistic: $T = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
CI: $\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$

Two indep. samples; CI for ratio of 2 variances

[note: for σ_1/σ_2, just take square roots on both ends of CI]
(1) $\neg B$(unknown means) AND A:
Test statistic: $T = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$
CI: $\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\alpha/2}$

Hypothesis Testing

Errors

Type I Error: Pr(reject H_0 given H_0 is true) = α.
i.e. Pr(Type I error) = **level of significance**
Type II Error: Pr(**do not** reject H_0 given H_0 is false) = Pr(Accept H_0 given H_1) = β.
[Note: Pr(**reject** H_0 given H_0 is false) = $1 - \beta$ is called the **power** of the test].

Steps of Hypo-testing

Step 1: Let μ be the mean of ...
Test $H_0 : \mu = x$ against $H_1 : \mu \neq x$.
Step 2: Set $\alpha = 0.05$
Step 3: State the test statistic used e.g.
$Z = \frac{(\bar{X}-\mu_0)}{\sigma/\sqrt{n}}$
$z_{\alpha/2} = z_{0.025} = 1.96$ (Refer to previous sections to choose appropriate test statistic)
State the critical region(s). (Remember to halve α for two-tailed tests)
Step 4: Substitute in values and calculate the value of the test statistic. z = ... = y.
[OR: H_0 is accepted if confidence interval covers μ_0]
[OR: p-value approach: H_0 is rejected if p-value $< \alpha$]
Step 5: Conclusion: $\because$ observed z value = y falls inside critical region, H_0 is rejected at 5% level of sig.

p-value

Probability of obtaining a test statistic more extreme than the observed sample value **given H_0 is true**. Also called observed level of significance. (Remember to multiply probability by 2 for two-tailed tests)

F-test for variance ratio/equality

Under $H_0 : \sigma_1^2 = \sigma_2^2$, $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$,
hence the **test statistic** is $F = \frac{S_1^2}{S_2^2}$.
$H_0 : \sigma_1^2 = \sigma_2^2$ is rejected if:
$H_1 : \sigma_1^2/\sigma_2^2 \neq 1$,
$F < F_{n_1-1,n_2-1;1-\alpha/2}$ or $F > F_{n_1-1,n_2-1;\alpha/2}$
$H_1 : \sigma_1^2/\sigma_2^2 > 1$, $F > F_{n_1-1,n_2-1;\alpha}$
$H_1 : \sigma_1^2/\sigma_2^2 < 1$,
$F < F_{n_1-1,n_2-1;1-\alpha}\ (= 1/F_{n_2-1,n_1-1;\alpha})$