# ST2334 Probability and Statistics

AY 25/26 Sem 1 — github/omgeta

## 1. Counting

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

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. $P(E) = \dfrac{|E|}{|S|}$, where $0 \le P(E) \le 1$

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

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. $P(A) = \displaystyle\sum_{i=1}^{n} P(A \cap B_i)$
$= \displaystyle\sum_{i=1}^{n} P(A \mid B_i)P(B_i)$

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

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

i. $P(B \mid A) = \dfrac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B')P(B')}$

ii. $P(B_k \mid A) = \dfrac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)}$

iii. $P(B_k \mid A \cap C) = \dfrac{P(A \mid B_k \cap C)P(B_k \cap C)}{P(A \cap C)}$

iv. $\dfrac{P(B \mid A)}{P(B' \mid A)} = \dfrac{P(A \mid B)}{P(A \mid B')}\dfrac{P(B)}{P(B')}$ (Odds)

## 3. Random Variables

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

i. $f(x) = P(X = x)$

ii. $0 \le f(x_i) \le 1, \forall x_i \in R_x$ 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. $\int_a^b f(x)\, dx = P(a \le X \le b)$

ii. $f(x) \ge 0, \forall x \in R_x$ and $f(x) = 0, \forall x \notin R_x$

iii. $\int_a^b f(x)\, dx \ge 0$ but not necessarily $\le 1$

iv. $\int_{R_x} f(x)\, dx = 1$

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

i. $F(x) = P(X \le x)$

ii. $F(x) = \int_{-\infty}^{x} f(t)dt$ and $f(x) = F'(x)$

iii. Non-decreasing and right continuous

iv. $0 \le F(x) \le 1$

### Expectation and Variance

Expectation of random variable $X$, $E(X)$ or $\mu_X$, 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 $\sigma_X^2$, 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. $\forall X, V(X) \ge 0$

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 is:

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

Conditional distribution of $Y$ given $X$ is:

i. $f_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

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

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. $f_X(x) = p^x(1-p)^{1-x}, \quad x = 0$ (fail), $1$ (success)

ii. $\mu_X = p, \quad \sigma_X^2 = p(1-p)$

**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. $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots n$

ii. $\mu_X = np, \quad \sigma_X^2 = np(1-p)$

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

i. $f_X(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, \ldots$

ii. $\mu_X = \frac{k}{p}, \quad \sigma_X^2 = \frac{(1-p)k}{p^2}$

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

i. $f_X(x) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots$

ii. $\mu_X = \frac{1}{p}, \quad \sigma_X^2 = \frac{1-p}{p^2}$

**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 $\alpha$, such that:
- expected occurences in interval $T$ is $\alpha T$
- 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 $\mu$ and flattens out as $\sigma$ 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. Upper $\alpha$ quartile $x_\alpha$ is s.t. $P(X \ge x_\alpha) = \alpha$

iv. Standard normal: $Z \sim N(0,1) = \frac{X-\mu}{\sigma}$
- $f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$
- Upper $\alpha$ quartile $z_\alpha$ is s.t. $P(Z \ge z_\alpha) = P(Z \le -z_\alpha) = \alpha$

v. As $n \to \infty$ and $p \to 0$, $X \sim \text{Bin}(n,p)$ converges to $X \sim \text{N}(np, np(1-p))$ or $Z = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0,1)$.
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 entire group of interest.
Population parameter is a population's numerical fact.
Sample of a population is used to make inferences.

Probability sampling:

i. Simple Random Sampling: sample is chosen s.t. every subset of $n$ observations of the population has the same probability of being selected.

Statistic is a function of sample data:

i. Sampling Mean, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

ii. Sampling Variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Standard Deviation, $\lambda_{\overline{X}}$ describes how much $\overline{x}$ tends to vary from sample to sample of size $n$

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:

i. Sampling distribution of sample mean $\overline{X}$ is approximately normal if $n$ is sufficiently large

ii. $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ follows approximately $N(0,1)$

## 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**: $Y \sim \chi^2(n) = \sum^n Z_i^2$ 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. $\chi^2(n;\alpha) = k \implies P(Y > k) = \alpha$

iii. $Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2) \implies Y_1 + Y_2 \sim \chi^2(n_1+n_2)$

iv. As $n$ increases, $\chi^2(n)$ is approximately $N(n, 2n)$

v. If $S^2$ is sample variance of size $n$ from normal population of variance $\sigma^2$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

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

i. $\mu_T = 0, \quad \sigma_T^2 = \frac{n}{n-2}$ for $n > 2$

ii. $t(n;\alpha) = k \implies P(T > k) = \alpha$

iii. When $n \geq 30$, can be replaced by $N(0,1)$

iv. If random sample selected from normal population, $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

**F-Distribution**: $F \sim Fn, m = \frac{U}{n} / \frac{V}{m}$ for independent random variables $U \sim \chi^2(n), V \sim \chi^2(m)$

i. $\mu_F = \frac{m}{m-2}$ for $m > 2$

ii. $\sigma_T^2 = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ for $n > 4$

iii. $F(n,m;\alpha) = k \implies P(F > k) = \alpha$

iv. $\frac{1}{F} \sim F(m,n)$

v. $F(n,m;\alpha) = \frac{1}{F(m,n;1-\alpha)}$

## 9. Estimation

Estimators are rules 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 $\theta$ is unbiased if $E(\hat{\theta}) = \theta$.

ii. Interval Estimation: An interval is calculated for some confidence level

Maximum error $E$ for estimating $\mu$ using $\bar{X}$ when $\sigma$ 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 a significance value $\alpha$.

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

Level of significance $\alpha$ is the probability of Type I error:

$$\alpha = P(\text{Type I Error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

Power is the probability of correctly rejecting a false $H_0$. Let $\beta$ denote the probability of a Type II error:

$$\beta = P(\text{Type II Error}) = P(\text{Do not reject } H_0 \mid H_0 \text{ is false})$$

$$\text{Power} = 1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ is false})$$

$p$-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 | $\sigma$ | $n$ | CI | Statistic |
|---|---|---|---|---|---|
| I | Normal | known | any | $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ | $Z = \frac{X - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ |
| II | any | known | $\geq 30$ | $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ | $Z = \frac{X - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ |
| III | Normal | unknown | $< 30$ | $\bar{x} \pm t_{n-1;\alpha/2} \cdot \frac{s}{\sqrt{n}}$ | $T = \frac{X - \mu}{s/\sqrt{n}} \sim t_{n-1}$ |
| IV | any | unknown | $\geq 30$ | $\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ | $Z = \frac{X - \mu}{s/\sqrt{n}} \sim N(0,1)$ |

## Test Statistics for Independent Samples

| Population | Variance | $\sigma_1, \sigma_2$ | $n$ | CI | Statistic |
|---|---|---|---|---|---|
| any | known | unequal | $\geq 30$ | $(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ | $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ |
| 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 | $\geq 30$ | $(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ | $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$ |
| 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 | $\geq 30$ | $(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ | $Z = \frac{\bar{X} - \bar{Y}}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$ |

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

## Pooled Estimator

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