Sample Variance: $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$. Equivalent alternative formula: $s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$

Sample z-score: $z = \frac{x - \bar{x}}{s}$

If we transform the data using the linear transformation $x^* = a + bx$:

$\bar{x}^* = a + b\bar{x}$, $s_{x^*} = |b|s_x$, $s^2_{x^*} = b^2 s^2_x$

Probability

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

The conditional probability of A, given B has occurred is $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Two events A and B are independent if and only if:

$P(A|B) = P(A), P(B|A) = P(B), P(A \cap B) = P(A)P(B)$.

Expected Value and Variance of a Discrete Random Variable

$E(X) = \mu = \sum xp(x)$, $\sigma^2 = E[(X-\mu)^2] = \sum (x-\mu)^2 p(x)$.

Alternative method: $E[(X-\mu)^2] = E(X^2) - [E(X)]^2$.

Properties of Expectation and Variance

$E(a + bX) = a + bE(X)$, $\sigma^2_{a+bX} = b^2\sigma^2_X$, $\sigma_{a+bX} = |b|\sigma_X$

If X and Y are two random variables then $E(X+Y) = E(X) + E(Y)$.

If X and Y are independent: $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$ and $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$

Binomial and Poisson Distributions

If X is a binomial random variable then:

$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, ..., n$. $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. $\mu = np, \sigma^2 = np(1-p)$.

Poisson distribution: $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \lambda = \mu = \sigma^2$.

Normal Distribution

If X is normally distributed with a mean of μ and standard deviation σ, then $Z = \frac{X-\mu}{\sigma}$ has the standard normal distribution.

If $\bar{X}$ is the mean of n independent observations from a normal distribution with mean μ and standard deviation σ, then $Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has the standard normal distribution.

Inference Procedures for Means
(When sampling from a normally distributed population)

Inference for μ

If σ is known:
Confidence interval for μ: $\bar{X} \pm z_{\alpha/2}\sigma_{\bar{X}}$, where $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

To test H_0: $\mu = \mu_0$: $Z = \dfrac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$

If σ is unknown:
Confidence interval for μ: $\bar{X} \pm t_{\alpha/2}SE(\bar{X})$, where $SE(\bar{X}) = \dfrac{s}{\sqrt{n}}$

To test H_0: $\mu = \mu_0$: $t = \dfrac{\bar{X} - \mu_0}{SE(\bar{X})}$

Inference for $\mu_1 - \mu_2$

The pooled-variance method:
$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$, $SE(\bar{X}_1 - \bar{X}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Confidence interval for $\mu_1 - \mu_2$: $\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}SE(\bar{X}_1 - \bar{X}_2)$

To test H_0: $\mu_1 = \mu_2$: $t = \dfrac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$. The degrees of freedom are $n_1 + n_2 - 2$.

The Welch Method:
$SE_W(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Confidence interval for $\mu_1 - \mu_2$: $\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}SE_W(\bar{X}_1 - \bar{X}_2)$

To test H_0: $\mu_1 = \mu_2$: $t = \dfrac{\bar{X}_1 - \bar{X}_2}{SE_W(\bar{X}_1 - \bar{X}_2)}$

Approximate $df = \dfrac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2-1}(\frac{s_2^2}{n_2})^2}$ (You won't have to calculate these degrees of freedom by hand)

Minimum Sample Size

Means: $n \geq (\dfrac{z_{\alpha/2}\sigma}{m})^2$, where m is the desired margin of error.