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 for the ith observation: $z_i = \frac{x_i - \bar{x}}{s}$

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

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

Probability

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

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

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

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

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

The Expected Value and Variance of Discrete Random Variables

$E(X) = \mu = \sum xp(x)$.

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

A handy relationship: $E[(X - \mu)^2] = E(X^2) - [E(X)]^2$.

Properties of Expectation and Variance

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

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

If X and Y are independent: $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ and $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$

Discrete Probablity Distributions

Binomial distribution: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. $\mu = np, \sigma^2 = np(1-p)$.

Hypergeometric distribution: $P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$. $\mu = n\frac{a}{N}$.

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

Geometric distribution: $P(X = x) = (1-p)^{x-1}p$. $\mu = \frac{1}{p}$, $\sigma^2 = \frac{1-p}{p^2}$.

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 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

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

If σ is unknown:

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

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

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: $\mu_1 = \mu_2$: $t = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$. 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: $\mu_1 = \mu_2$: $t = \frac{\bar{X}_1 - \bar{X}_2}{SE_W(\bar{X}_1 - \bar{X}_2)}$

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

Inference Procedures for Proportions

Confidence interval for p: $\hat{p} \pm z_{\alpha/2} SE(\hat{p})$, where $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

To test H_0: $p = p_0$, $Z = \frac{\hat{p} - p_0}{SE_0(\hat{p})}$, where $SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}}$

Confidence interval for $p_1 - p_2$: $\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} SE(\hat{p}_1 - \hat{p}_2)$, where
$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

To test H_0: $p_1 = p_2$, $Z = \frac{\hat{p}_1 - \hat{p}_2}{SE_0(\hat{p}_1 - \hat{p}_2)}$, where $SE_0(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$ and $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$

Minimum Sample Size

Means: $n \geq (\frac{z_{\alpha/2}\sigma}{m})^2$, where m is the desired margin of error. Proportions: $n = (\frac{z_{\alpha/2}}{m})^2 p(1-p)$

χ^2 Tests for Count Data

The test statistic is $\sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

For a basic goodness-of-fit test the degrees of freedom are # of cells -1. We also lose a degree of freedom for every parameter that must be estimated from the data.

For a two-way contingency table the expected counts are: $\frac{\text{row total} \times \text{column total}}{\text{overall total}}$

For a two-way contingency table, the degrees of freedom are: (# of rows -1)(# of columns -1).

One-way ANOVA

Suppose we have k treatment groups, with n_i observations in the i^{th} group, and n observations in total.

Source	df	SS	MS	F
Treatments	$k-1$	SST	$SST/(k-1)$	MST/MSE
Error	$n-k$	SSE	$SSE/(n-k)$	—
Total	$n-1$	SS(Total)	—	—

$$SST = \sum n_i(\bar{X}_i - \bar{X})^2, \qquad SSE = \sum (n_i - 1)s_i^2, \qquad SS(\text{Total}) = \sum\sum (X_{ij} - \bar{X})^2$$

Simple Linear Regression

Model: $Y = \beta_0 + \beta_1 X + \epsilon$

$SS_{XX} = \sum (X_i - \bar{X})^2$, $SS_{YY} = \sum (Y_i - \bar{Y})^2$, $SP_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$

$\hat{\beta}_1 = \dfrac{SP_{XY}}{SS_{XX}}$, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$, $r = \dfrac{SP_{XY}}{\sqrt{SS_{XX}SS_{YY}}} = \dfrac{1}{n-1}\sum(\frac{X_i-\bar{X}}{s_X})(\frac{Y_i-\bar{Y}}{s_Y}) = \hat{\beta}_1\dfrac{s_X}{s_Y}$

$e_i = Y_i - \hat{Y}_i$, $s^2 = \dfrac{\sum e_i^2}{n-2} = \dfrac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$, $SE(\hat{\beta}_1) = \dfrac{s}{\sqrt{SS_{XX}}}$

A $(1-\alpha)100\%$ confidence interval for β_1 is: $\hat{\beta}_1 \pm t_{\alpha/2}SE(\hat{\beta}_1)$

To test H_0: $\beta_1 = 0$, the appropriate test statistic is $t = \dfrac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$