Four important **features of distributions** are

1. central location (mean, median)
2. spread (variance, standard deviation)
3. shape (symmetry: skew, multi-modality)
4. gaps (outliers)

The **numerical summary measures** are

$Q1$		middle value between minimum and median
median M	$\begin{cases} x_{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\ \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{if } n \text{ is even} \end{cases}$	middle value between minimum and maximum
$Q3$		middle value between medium and maximum
Inner Quartile Range IQR	$Q3 - Q1$	
outliers	$[Q1 - 1.5 \cdot IQR,\ Q3 + 1.5 \cdot IQR]$	outliers lie outside this range

The **boxplot** has a box with lines at $Q1$, M, and $Q3$ and *whiskers* at the last data points inside the outliers range.

Probability

A **probability density function f** (y) defines the probability of a continuous random variable Y taking on some value y over an interval. For discrete probability, we switch to a probability mass function and take summations instead of integrals.

$$\mathbb{P}(a \leq Y \leq b) = \int_a^b \mathbf{f}(y) \cdot dy$$

The **expection** $\mathbb{E}[g(Y)]$ of a function g of random variable Y is

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) \cdot \mathbf{f}(y) \cdot dy$$

The **cumulative distribution function** $F(a)$ is the $\mathbb{P}(Y < a)$

$$\mathbb{P}(Y < a) = \int_{-\infty}^{a} \mathbf{f}(y) \cdot dy$$

Get the **marginal probability f** (y) from a **joint probability distribution f** (x, y) of random variables X, Y with

$$\mathbf{f}(y) = \int_{-\infty}^{\infty} \mathbf{f}(x, y) \cdot dx$$

The **conditional probability distribution f** $(x|y)$ is

$$\mathbf{f}(x|y) = \frac{\mathbf{f}(x, y)}{\mathbf{f}(y)}$$

Random variables X and Y are **independent** $X \perp Y$ if

$$\mathbb{P}(X|Y) = \mathbb{P}(X) \quad \text{which means} \quad \mathbf{f}(x|y) = \mathbf{f}(x)$$

The **variance** $\mathbb{V}(X)$ of a single random variable X is

$$\mathbb{V}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2 \right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

and linear combination of random variables $Y = a_1 \cdot X_1, a_2 \cdot X_2, \ldots, a_n \cdot X_n + b$ has variance

$$\mathbb{V}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \mathbb{V}(X_i) + \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} a_i \cdot a_j \cdot \mathbf{Cov}(X_i, X_j)$$

where the **covariance** $\mathbf{Cov}(X, Y)$ of two random variables X and Y is

$$\begin{aligned}\mathbf{Cov}(X,Y) &= \mathbb{E}\left[(X-\mathbb{E}[X])\cdot(Y-\mathbb{E}[Y])\right] \\ &= \mathbb{E}[X\cdot Y]-\mathbb{E}[X]\cdot\mathbb{E}[Y] \\ \mathbb{E}[X\cdot Y] &= \sum_{x\,\epsilon\,X}\sum_{y\,\epsilon Y} x\cdot y\cdot\mathbb{P}(X=x,Y=x) \\ \mathbf{Cov}(X,Y) &= 0 \ \text{ if } X\perp Y\end{aligned}$$

The **correlation coefficient** of two random variables X and Y is always between -1 and 1

$$-1 \leq \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\cdot\mathbb{V}(Y)}} \leq 1$$

Moment Generating Functions

The **moment generating function** $M_Y(t)$ of random variable Y is

$$M_Y(t) = \mathbb{E}\left[e^{t\cdot Y}\right] = \int_{-\infty}^{\infty} e^{t\cdot y}\cdot\mathbf{f}(y)\cdot y$$

and the **linear transformation** $Z = a\cdot Y + b$ has moment generating function

$$M_Z(t) = e^{b\cdot t}\cdot M_Y(a\cdot t)$$

and sum $W = \sum_{i=1}^{n} X_i$ of n independent random variables $X_1 \perp X_2 \perp \ldots \perp X_n$ has moment generating function

$$M_W(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Central Limit Theorem

A **random sample** $X_1, X_2, \ldots, X_n$ of size n has random variables that are independent $X_1 \perp X_2 \perp \ldots \perp X_n$ and identically distributed $X_i \sim \mathbf{f}(x|\theta)\ \forall\ 1 \leq i \leq n$.

The **central limit theorem** says the mean $\bar{X} = \frac{1}{n}\cdot\sum_{i=1}^{n} X_i$ is normally distributed as $n \to \infty$

$$\lim_{n\to\infty} \bar{X} \sim Normal\left(\mathbb{E}\left[\bar{X}\right]=\mu, \mathbb{V}\left(\bar{X}\right)=\frac{\sigma^2}{n}\right)$$

The **margin of error** $1-\alpha$ is the probability $|\bar{X}-\mu|$ is within a units

$$\mathbb{P}\left(-a \leq \bar{X}-\mu \leq a\right) = 1-\alpha$$

and middle term can be **converted to standard normal** Z to find a sufficient sample size n for the margin of error

$$\mathbb{P}\left(\frac{-a}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{a}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha$$

$$\text{then use}\quad \mathbb{P}(Z \geq z) = \frac{\alpha}{2} \quad \text{from a table to solve} \quad \frac{a}{\frac{\sigma}{\sqrt{n}}} = z \quad \text{for } n$$

Estimators

An **estimator** (a statistic) $\hat{\theta}(X_1, X_2, \ldots, X_n)$ for parameter θ is a function of random variables.

An **estimate** $\hat{\theta}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ is an instance of an estimator.

The **bias** $Bias\left(\hat{\theta}\right)$ of estimator $\hat{\theta}$ for parameter θ is below. **Unbiased** if $Bias\left(\hat{\theta}\right) = 0$. **Biased** if $Bias\left(\hat{\theta}\right) \neq 0$

$$Bias\left(\hat{\theta}\right) = \mathbb{E}\left[\hat{\theta}\right] - \theta$$

Unbiased estimator $\hat{\theta}_1$ is more **efficient** than unbiased estimator $\hat{\theta}_2$ if

$$\mathbb{V}\left(\hat{\theta}_1\right) < \mathbb{V}\left(\hat{\theta}_2\right)$$

The **relative efficiency** $\text{Eff}\left(\hat{\theta}_1, \hat{\theta}_2\right)$ of two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ is

$$\text{Eff}\left(\hat{\theta}_1, \hat{\theta}_2\right) = \frac{\mathbb{V}\left(\hat{\theta}_2\right)}{\mathbb{V}\left(\hat{\theta}_1\right)}$$

The **mean squared error** $\text{MSE}\left(\hat{\theta}\right)$ of estimator $\hat{\theta}$ is

$$\text{MSE}\left(\hat{\theta}\right) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \mathbb{V}\left(\hat{\theta}\right) - \left[Bias\left(\hat{\theta}\right)\right]^2$$

Estimator $\hat{\theta}_n\left(X_1, X_2, \ldots, X_n\right)$ of n random variables is **consistent** if

$$\lim_{n\to\infty} \text{MSE}\left(\hat{\theta}_n\right) = 0$$

The **Cramer-Rao lower bound** is a lower limit to the variance of estimator $\hat{\theta}$

$$\mathbb{V}\left(\hat{\theta}\right) \geq \frac{1}{I_Y(\theta)}$$
$$I_Y(\theta) = -n \cdot \mathbb{E}\left[\frac{\delta}{\delta\theta} ln\left(\mathbf{f}\left(y|\theta\right)\right)\right]$$

A **confidence interval** of estimator $\hat{\theta}$ is a random interval that we are $1-\alpha$ (a percent) confident contains the target parameter θ

$$\mathbb{P}\left(\hat{\theta}\left(Y_1, Y_2, \ldots, Y_n\right) - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\mathbb{V}\left(\hat{\theta}\right)} \leq \theta \leq \hat{\theta}\left(Y_1, Y_2, \ldots, Y_n\right) + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\mathbb{V}\left(\hat{\theta}\right)}\right) = 1 - \alpha$$

where $Z \sim Normal(0, 1)$ and $\mathbb{P}\left(Z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(Z < \frac{\alpha}{2}\right) = \mathbb{P}\left(Z > 1 - \frac{\alpha}{2}\right)$.

Also, the **confidence level** is $1 - \alpha$ and **confidence coefficient** is α. In **practice**, we use a sample estimate instead of estimator and do not take the probability (since estimates are constant).

Common populations and estimators are

Population	Estimator	Estimate	Variance
θ	$\hat{\theta}\left(Y_1, Y_2, \ldots, Y_n\right)$	$\hat{\theta}\left(y_1, y_2, \ldots, y_n\right)$	
μ	$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^{n} Y_i$	$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$	$\mathbb{V}\left(\bar{Y}\right) = \frac{\sigma^2}{n}$
π	$\hat{\pi} = \frac{Y}{n}$	$\hat{\pi} = \frac{y}{n}$	$\mathbb{V}\left(\hat{\pi}\right) = \frac{\pi \cdot (1-\pi)}{n}$
$\mu_1 - \mu_2$	$\hat{\theta} = \bar{Y}_1 - \bar{Y}_2$	$\hat{\theta} = \bar{y}_1 - \bar{y}_2$	$\mathbb{V}\left(\bar{Y}_1 - \bar{Y}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

t-Test

For **unknown variance** σ^2 and **small number of samples** n in $Y_1 \perp Y_2 \perp \ldots \perp Y_n \sim Normal\left(\mu, \sigma^2\right)$

$$\frac{(n-1) \cdot s^2}{\sigma^2} = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right) \sim \chi^2\left(df = n - 1\right)$$

and

$$\frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}} \sim \mathbf{t}\left(df = n - 1\right)$$

where $\mathbf{t}$ is the t distribution and $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$

The $1 - \alpha$ **confidence interval** for μ is

$$\left[\bar{y} \pm \mathbf{t}\left(df = n - 1\right) \cdot \frac{s}{\sqrt{n}}\right]$$

To **estimate mean difference** $\mu_1 - \mu_2$ for $X_1 \perp X_2 \perp \ldots \perp X_m \sim Normal\left(\mu_1, \sigma_1^2\right)$ and $Y_1 \perp Y_2 \perp \ldots \perp Y_n \sim Normal\left(\mu_2, \sigma_2^2\right)$ we **estimate sample variances** $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$

$$\hat{\sigma}_1^2 = s_1^2 = \frac{1}{m-1} \cdot \sum_{i=1}^{m} \left(X_i - \bar{X}\right)^2$$
$$\hat{\sigma}_2^2 = s_2^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

The **pooled estimate** s_p^2 of variance is

$$s_p^2 = \frac{(m-1) \cdot s_1^2 + (n-1) \cdot s_2^2}{m + n - 2}$$

Because the result

$$\frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_p^2}{m} + \frac{s_p^2}{n}}} \sim \mathbf{t}\left(df = m + n - 2\right)$$

then the $1 - \alpha$ **confidence interval** is

$$\left[\left(\bar{x} - \bar{y}\right) \pm t_{1-\frac{\alpha}{2}}\left(df = m + n - 2\right) \cdot \sqrt{\frac{s_p^2}{m} + \frac{s_p^2}{n}}\right]$$

Maximum Likelihood Estimation

The **likelihood function** $L\left(\theta\right) = L\left(\theta | y_1, y_2, \ldots, y_n\right)$ for random sample $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y = n$ is

$$\begin{aligned} L\left(\theta\right) &= \mathbf{f}\left(y_1, y_2, \ldots, y_n | \theta\right) \\ &= \prod_{i=1}^{n} \mathbf{f}\left(y_i | \theta\right) \end{aligned}$$

The **log likelihood function** $\ell\left(\theta\right)$ is

$$\begin{aligned} \ell\left(\theta\right) &= log\left(L(\theta)\right) \\ &= \sum_{i=1}^{n} log\left(\mathbf{f}\left(y_i | \theta\right)\right) \end{aligned}$$

The **maximum likelihood estimator** (MLE) $\hat{\theta}$ for parameter θ is

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

To **find maximum likelihood estimator** $\hat{\theta}$ for parameter θ

1. do $L(\theta)$
2. do $\ell(\theta) = log\left(L(\theta)\right)$
3. solve $\frac{d}{d\theta}\ell(\theta) = 0$ for θ to get $\hat{\theta}$
4. check $L(\hat{\theta})$ is the maximum of $L(\theta)$

The **observed information** $i(\hat{\theta})$ gives an estimate of the variance $\mathbb{V}\left(\hat{\theta}\right) = \frac{1}{i(\hat{\theta})}$ of maximum likelihood estimator $\hat{\theta}$

$$i(\hat{\theta}) = -\frac{d^2}{d\theta^2}\ell\left(\theta\right)\Big|_{\theta=\hat{\theta}}$$

The **relative likelihood** $R\left(\theta\right)$ is

$$R\left(\theta\right) = \frac{L\left(\theta\right)}{L\left(\hat{\theta}\right)}$$

The **relative log likelihood** $r\left(\theta\right)$ is

$$r\left(\theta\right) = log\left(L\left(\theta\right)\right) - log\left(L(\hat{\theta})\right)$$

The 95% **likelihood interval** is a vaguely defined region where we are likely to find $\hat{\theta}$ defined by

$$-2 \leq r\left(\theta\right) \leq 0$$

A likelihood interval quickly converges to a confidence interval as $n \to \infty$.

Multinomial Distribution

The **multinomial distribution** $Y_1, Y_2, \ldots, Y_n \sim Multinomial(\pi_1, \pi_2, \ldots, \pi_k)$ for k categories with $\pi_1, \pi_2, \ldots, \pi_k$ probability of each where Y_i is the number of n trials that fall into category $i \; \forall \; 1 \leq i \leq k$. Note $\sum_{i=1}^{k} \pi_i = 1$ and $\sum_{i=1}^{k} Y_i = n$.

The **probability of data** $Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k$ is

$$\mathbb{P}\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k\right) = \frac{n!}{\prod_{i=1}^{k} y_i!} \cdot \prod_{i=1}^{k} \pi_i^{y_i}$$

and **expectation** $\mathbb{E}\left[Y_i\right]$, **variance** $\mathbb{V}\left(Y_i\right)$, and **covariance** $\mathbf{Cov}\left(Y_i, Y_j\right)$ for each $1 \leq i \leq k$, $1 \leq j \leq k$ category is

$$\begin{aligned}
\mathbb{E}\left[Y_i\right] &= n \cdot \pi_i \\
\mathbb{V}\left(Y_i\right) &= n \cdot \pi_i \cdot (1 - \pi_i) \\
\mathbf{Cov}\left(Y_i, Y_j\right) &= -n \cdot \pi_i \cdot \pi_j \qquad i \neq j
\end{aligned}$$

and **maximum likelihood estimator** $\hat{\pi}_i$ for parameter π_i is

$$\hat{\pi}_i = \frac{Y_i}{n}$$

When $\theta_1, \theta_2, \ldots, \theta_k$ **are parameters to** k **distributions** get each $\hat{\theta}_i$ by solving

$$\frac{d}{d\theta_i}\ell\left(\theta_1, \theta_2, \ldots, \theta_k\right) = 0$$

Bayesian Inference

In **bayesian inference**, unknown parameters θ are treated as random variables.

The **posterior distribution** $\mathbb{P}\left(\theta|D\right)$ is given by the likelihood $\mathbb{P}\left(D|\theta\right)$, parameter θ marginal distribution $\mathbb{P}\left(\theta\right)$, and data D marginal distribution $\mathbb{P}\left(D\right)$ as

$$\mathbb{P}\left(\theta|D\right) = \frac{\mathbb{P}\left(D|\theta\right) \cdot \mathbb{P}\left(\theta\right)}{\mathbb{P}\left(D\right)}$$

To **learn posterior** $\mathbb{P}\left(\theta|D\right)$ use knowledge of **prior** $\mathbb{P}\left(\theta\right)$, observe data D, and calculate likelihood $\mathbb{P}\left(D|\theta\right)$ to update $\mathbb{P}\left(\theta|D\right)$.

$Prior_1(\theta)$ is **more informative** than $Prior_2(\theta)$ if

$$\mathbb{V}(Prior_1(\theta)) < \mathbb{V}(Prior_2(\theta))$$

Hypothesis Testing

A **null hypothesis** H_0 is the assumed model generated the data.

A **composite hypothesis** is when we estimate a parameter for H_0.

A measure of **goodness of fit** X^2 with observed frequency O_i and expected frequency E_i for each $1 \leq i \leq k$ category is

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(df = k - 1 - e)$$

where e is the number of parameters estimated in the expected model.

A **p-value** p is the probability the model could generate the data.

$$p = \mathbb{P}\left(\chi^2(df) > X^2\right)$$

With a **small p-value** ($p \leq 0.05$) we reject H_0 and say the model could not generate the data.

A **rejection region** RR is values of data $D = X_1 \perp X_2 \perp \ldots \perp X_n \sim f(x_i|\theta)$ where we reject $H_0 : \theta = \theta_0$.

The **type I error** $W^* \in RR|H_0$ for a given test statistic $W^* = W(X_1, X_2, \ldots, X_n)$ where $W^* \sim g(w|\theta)$ has probability

$$\alpha = \mathbb{P}(W^* \in RR|H_0) = \begin{cases} \int_{RR} g(w|\theta = \theta_0) \cdot dw & \text{if } g \text{ is continuous} \\ \sum_{RR} g(w|\theta = \theta_0) \cdot dw & \text{if } g \text{ is discrete} \end{cases}$$

while **type II error** $W^* \notin RR|H_A$ has probability

$$\beta = \mathbb{P}(W^* \notin RR|H_A)$$

The **large sample rejection region** is

$$RR = \begin{cases} Z^* > Z_\alpha & \text{if } H_A : \theta > \theta_0 \\ Z^* < -Z_\alpha & \text{if } H_A : \theta < \theta_0 \\ Z^* < -Z_{\frac{\alpha}{2}} \text{ or } Z^* > Z_{\frac{\alpha}{2}} & \text{if } H_A : \theta \neq \theta_0 \end{cases}$$

where test statistic $Z^* = \frac{\hat{\theta} - \theta_0}{\mathbb{V}(\hat{\theta})}$ depends on estimator $\hat{\theta}$. Large sample **p-value** $p = \mathbb{P}(W^* \in RR|H_0)$ is

$$p = \begin{cases} \mathbb{P}(Z > Z^*) & \text{if } H_A : \theta > \theta_0 \\ \mathbb{P}(Z < -Z^*) & \text{if } H_A : \theta < \theta_0 \\ 2 \cdot \mathbb{P}(Z > |Z^*|) & \text{if } H_A : \theta \neq \theta_0 \end{cases}$$

where $Z \sim Normal(0, 1)$.

The **single variable small sample** $Y \sim Normal(\mu, \sigma^2)$ for unknown μ, σ^2 for $H_0 : \mu = \mu_0$ has test statistic t^*

$$t^* = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t(df = n - 1)$$

The **two variable small sample** for $H_0 : \mu_1 - \mu_2 = \Delta$ has test statistic t^*

$$t^* = \frac{(\bar{y}_1 - \bar{y}_2) - \Delta}{\sqrt{S_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(df = n_1 + n_2 - 2)$$

A **contrast** is a linear combination of k means $\mu_1, \mu_2, \ldots, \mu_k$ where

$$\theta = \sum_{i=1}^{k} a_i \cdot \mu_i$$
$$0 = \sum_{i=1}^{k} a_i$$

has $1-\alpha$ confidence interval

$$\left[\hat{\theta} \pm t_{1-\frac{\alpha}{2}}\left(df = \sum_{i=1}^{k}(n_i) - k\right) \cdot S_p \cdot \sqrt{\sum_{i=1}^{k} \frac{a_i^2}{n_i}}\right]$$

The **power** $W^* \in RR|H_A$ has probability $1-\beta$. The **power function** $Pow\left(\theta\right) = \mathbb{P}\left(\text{reject } H_0|\theta\right) = 1-\beta\left(\theta\right)$.

A test is **uniformly most powerful test** $Pow_1\left(\theta\right)$ if

$$Pow_1\left(\theta\right) \geq Pow_i\left(\theta\right) \;\; \forall \; i, \, \theta$$

Note: $Pow\left(\theta = \theta_0\right) = \mathbb{P}\left(\text{reject } H_0|H_0 : \theta = \theta_0\right) = \alpha$.

The test that **maximizes power** $Pow\left(\theta\right)$ for data $D = Y_1 \perp Y_2 \perp \ldots \perp Y_n \sim f(y_i|\theta)$ has rejection region RR

$$RR = \left\{\frac{\mathcal{L}\left(\theta_0|D\right)}{\mathcal{L}\left(\theta_A|D\right)} < k_\alpha\right\}$$

Likelihood Ratio

Let Ω be the set of all values parameters can take. Define

$$H_0 : \Theta \subseteq \Omega_0$$
$$H_A : \Theta \subseteq \Omega_A \text{ or } \Theta \not\subseteq \Omega_0$$

Step 1: find likelihood of H_0 parameters given data D by

$$\mathcal{L}\left(\hat{\Omega}_0\right) = \max_{\Theta \subseteq \Omega_0} \mathcal{L}\left(\Theta|D\right)$$

Step 2: find likelihood of H_A parameters given data D by

$$\mathcal{L}\left(\hat{\Omega}_A\right) = \max_{\Theta \subseteq \Omega_A} \mathcal{L}\left(\Theta|D\right)$$

Step 3: the **likelihood ratio test** (LTR) defines rejection region RR as

$$RR = \left\{\lambda = \frac{\mathcal{L}\left(\hat{\Omega}_0\right)}{\mathcal{L}\left(\hat{\Omega}_A\right)} < k_\alpha\right\}$$

We then choose k_α so $\mathbb{P}\left(0 \leq \lambda \leq k_\alpha|H_0\right) = \alpha$.

To **compare variances** use F-distribution so

$$\frac{W_1/v_1}{W_2/v_2} \sim F\left(v_1, v_2\right)$$

where $W_1 \sim \chi^2\left(df = v_1\right)$, $W_2 \sim \chi^2\left(df = v_2\right)$, and $W_1 \perp W_2$.

The **general result** for likelihood ratio test with large sample size is

$$-2 \cdot ln\left(\lambda\right) \sim \chi^2\left(df = k - q\right)$$

where H_A estimates k parameters and H_0 estimates q parameters.

Linear Models

The **measurement model** given $Y_1, Y_2, \ldots, Y_n$ is

$$Y_i = \mu + \epsilon_i \ \forall \ i = 1, 2, \ldots, n$$
$$\epsilon_i^2 = (Y_i - \mu)^2$$

where each Y_i and ϵ_i are functions of θ and $\epsilon_i \sim Normal(0, \sigma^2)$.

The **sum of squared errors** SSE is

$$SSE = \sum_{i=1}^{n} [\epsilon_i(\theta)]^2$$

The **least squares estimation (LSE)** $\hat{\theta}$ (not MLE) finds θ that minimizes squared residual ϵ_i^2

$$\hat{\theta} = \min_{\theta} SSE = \min_{\theta} \sum_{i=1}^{n} [\epsilon_i(\theta)]^2$$

The **linear regression model** is defined by

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i \ \forall \ i = 1, 2, \ldots, n$$

The **predicted value** $\hat{y}_i$ is found from filling in x_i^* in

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i^*$$

The **LSE solution** to the **linear regression model** for $D = Y_1, Y_2, \ldots, Y_n | X_i = x_i \sim Normal(\beta_0 + \beta_1 \cdot x_i, \sigma^2)$ is

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{x}$$
$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i \cdot Y_i \text{ where } w_i = \frac{(x_i - \bar{x})}{\sum_{i=1}(x_i - \bar{x})^2}$$

with probability distributions

$$\hat{\beta}_0 \sim Normal\left(\mathbb{E}\left[\hat{\beta}_0\right] = \beta_0, \mathbb{V}\left(\hat{\beta}_0\right) = \frac{\sigma^2}{n} \cdot \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$
$$\hat{\beta}_1 \sim Normal\left(\mathbb{E}\left[\hat{\beta}_1\right] = \beta_1, \mathbb{V}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

Note: confidence intervals and hypothesis testing use $t_{\frac{\alpha}{2}}$ $(df = n - 2)$ because β_0 and β_1 are being estimated.

The **observed residual** $\hat{e}_i$ is $\hat{e}_i = y_i - \hat{y}_i$

The **observed sum of squared errors** $S\hat{S}E$ is $S\hat{S}E = \sum_{i=1}^{n} \hat{e}_i^2$

The **variance estimator** $\hat{\sigma}^2$ is $\hat{\sigma}^2 = \frac{1}{n-2} \cdot S\hat{S}E$.

Sampling with Subgroups Example

The **data** D contains n subgroups $Y_1 \perp Y_2 \perp \ldots \perp Y_n \sim Binomial(k, \theta)$ each with k people.

The **indicator** I_i that sample i has an infection $\forall \ 1 \leq i \leq n$ is

$$I_i = \begin{cases} 1 & Y_i > 0 \\ 0 & Y_i = 0 \end{cases}$$

with probability

$$\mathbb{P}(I_i = 0) = \binom{k}{0} \cdot \theta^0 \cdot (1 - \theta)^k = (1 - \theta)^k$$
$$\mathbb{P}(I_i = 1) = 1 - \mathbb{P}(I_i = 0) = 1 - (1 - \theta)^k$$

The **number of groups not infected** X is

$$X = \sum_{i=1}^{n} I_i$$
$$X \sim Binom\left(n, (1 - \theta)^k\right)$$