

Applied Bioostatistics I - formulas

Axioms of Kolmogorov:

$A(\text{probability measure}) \subset \Omega(\text{sample space}) : P[A] \in [0, 1] :$

- $0 \leq P[A] \leq 1$ for every event $A \subset \Omega$
- $P[\Omega] = 1$
- $P[A \cup B] = P[A] + P[B]$ for *disjoint* event A and B.

De Morgan's laws

Let A and B be events. Then, $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$

Probability of unions

Let A and B be events. Then, $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ More general: let A_1, A_2, \dots, A_n be events. Then, $P[A_1 \cup A_2 \cup \dots \cup A_n] = \sum_{i_1=1}^n P[A_{i_1}] - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n P[A_{i_1} \cap A_{i_2}] + \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n P[A_{i_1} \cap A_{i_2} \cap A_{i_3}] - \dots$

Independence

Two events A and B are called independent if $P[A \cap B] = P[A] \cdot P[B]$

Conditional probability

Let A and B be events (with $P[B] > 0$) .

The conditional probability of A given B is defined as $P[A|B] = \frac{P[A \cap B]}{P[B]}$

Law of total probability

Assume B_1, B_2, \dots, B_k are disjoint events with $B_1, B_2, \dots, B_k = \Omega$. Then we can calculate the probability of any event A as $P[A] = \sum_{i=1}^k P[A \cap B_i] = \sum_{i=1}^k P[A|B_i]P[B_i]$

Bayes' theorem

Let A and B be events with $P[A] > 0$ and $P[B] > 0$. Then we have: $P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{P[A|B]P[B]}{P[A]}$ In the setting of the law of total probability, we have $P[B_i|A] = \frac{P[A \cap B_i]}{P[A]} = \frac{P[A|B_i]P[B_i]}{\sum_{j=1}^k P[A|B_j]P[B_j]}$

Cumulative distribution function

The cumulative distribution function (CDF) of a random variable X is defined as $F_X(x) := P[X \leq x]$ continuous $F(x) = \int_{-\infty}^x f(u)du$

Discrete random variables

$X : \Omega \rightarrow \{x_1, x_2, \dots\}$; probability mass function $p(x_k) := P[X = x_k]$; $A \subset \{x_1, x_2, \dots\} :$

- $P[X \in A] = \sum_{k: x_k \in A} p(x_k)$
- $\sum_k p(x_k) = 1$
- CDF : $F_X(x) = P[X \leq x] = \sum_{k: x_k \leq x} p(x_k)$

Expectation value

$E[X] := \sum_k x_k p(x_k)$ continuous $E[X] = \int_{-\infty}^{\infty} x f(x) dx$

Variance

$\text{Var}(X) := \sum_k (x_k - E[X])^2 p(x_k)$ continuous $\text{Var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$

Bernoulli distribution $X \in \{0, 1\}$ $X \sim \text{Bernoulli}(\pi)$

- $\pi := P[X = 1]$

Binomial distribution $X \in \{0, 1, \dots, n\}$ $X \sim \text{Bin}(n, \pi), n \in \mathbb{N}, \pi \in (0, 1)$

- $p(x) = P[X = x] = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$
- $E[X] = n\pi, \text{Var}(X) = n\pi(1 - \pi)$

Poisson distribution $X \in \mathbb{N}$ $X \sim \text{Pois}(\lambda), \lambda > 0$

- $p(x) = P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}$
- $E[X] = \lambda, \text{Var}(X) = \lambda$
- CDF: $F(x; \lambda) = \sum_{i=0}^x \frac{e^{-\lambda} \lambda^i}{i!}$

Uniform distribution $X \sim \mathcal{U}([a, b])$

- $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$
- $E[X] = \frac{b+a}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}$

Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$

- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, x \in \mathbb{R}$

Standard normal distribution $Z \sim \mathcal{N}(0, 1)$

- $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \int_{-\infty}^z \varphi(t) dt$
- $Z = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
- $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

Exponential distribution $X \sim \text{Exp}(\lambda), \lambda > 0$

- $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$
- $E[X] = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$

discrete	continuous
$E[X] = \sum_{k \geq 1} x_k p(x_k)$	$E[X] = \int_{-\infty}^{\infty} x f(x) dx$
$\text{Var}(X) = \sum_{k \geq 1} (x_k - E[X])^2 p(x_k) = E[X^2] - (E[X])^2$	$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$

R function naming

- p or “probability”, the cumulative distribution function (c. d. f.)
- q for “quantile”, the inverse c. d. f.
- d for “density”, the density function (p. f. or p. d. f.)
- r for “random”, a random variable having the specified distribution

Discrete multivariate distributions

Let $X : \Omega \rightarrow W_x$ and $Y : \Omega \rightarrow W_Y$ be discrete random variables
Joint Cumulative Distribution Function: $F_{X,Y}(x, y) := P[X \leq x, Y \leq y]$
Joint Probability Mass Function: $p_{X,Y}(x, y) := P[X = x, Y = y], x \in W_X, y \in W_Y$
Marginal Probability Mass Function: $p_X(x) = P[X = x] = \sum_{y \in W_Y} p_{X,Y}(x, y)$
Independence IF: $p_{X,Y}(x, y) = p_X(x)p_Y(y)$
Conditional Probability Mass function: $p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$

Continuous multivariate distributions

Let $X \rightarrow \mathbb{R}$ and $Y \rightarrow \mathbb{R}$ be continuous random variables
Joint cumulative distribution function: $F_{X,Y}(x, y) := P[X \leq x, Y \leq y]$
Joint probability density: $f_{X,Y}(x, y) := \frac{\partial}{\partial x} \frac{\partial}{\partial y} F_{X,Y}(x, y)$
 $P[a \leq X \leq b, c \leq Y \leq d] = \int_a^b \int_c^d f_{X,Y}(x, y) dy dx \ (a < b, c < d)$

Marginal probability density: $f_X(x) := \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
Independence IF: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
Conditional probability density: $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Covariance	Correlation
$\text{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])]$	$\rho_{XY} := \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$

- if X, Y independent $\Rightarrow \text{Cov}(X, Y) = 0$, $\rho_{XY} = 0$, $E[XY] = E[X] \cdot E[Y]$ and $\text{Cov}(X, Y) = 0$ (the other direction is not true!)
- $-1 \leq \rho_{XY} \leq 1$
- $\rho_{XY} = 1$ if $Y = a + bX$ for some $b > 0$
- $\rho_{XY} = -1$ if $Y = a + bX$ for some $b < 0$
- $E[X + Y] = E[X] + E[Y]$
- $E[aX] = aE[X]$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- $\text{Var}(aX) = a^2 \text{Var}(X)$

Descriptive Statistics

Sample Mean: $\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow \mu = E[X]$ if $n \rightarrow \infty$ (consistent/unbiased estimator for the true mean)

Sample Variance: $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ (s_x : sample standard deviation) $s_x^2 \rightarrow \sigma^2 = \text{Var}(X)$ if $n \rightarrow \infty$
 $E[s_x^2] = \sigma^2$ (consistent/unbiased estimator for the true variance)

Median ($x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$): $m = \begin{cases} x_{(n+1)/2}, & n \text{ is odd,} \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}), & \text{otherwise} \end{cases}$

Empirical α quantile: $q_\alpha = x_{(\alpha(n-1)+1)}$ if $\alpha \cdot (n-1)$ is an integer; otherwise $(x_{(\lfloor \alpha(n-1) \rfloor + 1)} + x_{(\lceil \alpha(n-1) \rceil + 1)})/2$
random variable X : value m such that $P[X \leq m] \geq \alpha$ and $P[X \geq m] \geq 1 - \alpha$

Kernel density estimation

Given a set of points x_1, x_2, \dots, x_n , the kernel density estimator for the generating distribution is

$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$ (kernel function: arbitrary positive symmetric, h : bandwidth)

* Uniform/rectangular kernel: $K \sim \mathcal{U}\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right)$ (same weight for all points)

* Gaussian kernel: $K \sim \mathcal{N}(0, 1)$ (less weight to far apart points)

Empirical cumulative distribution function (ECDF): $\hat{F}(x) = \frac{\#\{k | x_k \leq x\}}{n}$

Empirical correlation: $r = \frac{s_{xy}}{s_x s_y} \in [-1, 1]$, $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

Linear dependence between 2 samples $\{x_i\}$ and $\{y_i\}$ * $r = +1$ if $y_i = a + bx_i$ for some $b > 0$ * $r = -1$ if $y_i = a + bx_i$ for some $b < 0$

Central Limit Theorem

Let X be random variable with expectation value μ and variance σ^2 , and X_1, X_2, \dots, X_n i.i.d. copies of X .

Then $\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ for large n . : $E[\bar{X}_n] = \mu, \sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}} \rightarrow 0$ as $n \rightarrow \infty$

$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \approx \mathcal{N}(0, 1)$ for large n

Standard error of the mean (SEM)

Natural estimator for $\sigma(\bar{X}_n)$: $\text{se}_{\bar{x}} = \frac{s_x}{\sqrt{n}}$; s_x is the empirical standard deviation

Law of large numbers

Let X be random variable with expectation value μ , and X_1, X_2, \dots, X_n i.i.d. copies of X . Then, $\bar{X}_n \rightarrow \mu$ as $n \rightarrow \infty$

Confidence interval with confidence level $1 - \alpha$, $\frac{1}{2} < \alpha < 1$

$$\left[\bar{X}_n - \Phi^{-1}(1 - \alpha/2) \cdot \frac{s_x}{\sqrt{n}}, \bar{X}_n + \Phi^{-1}(1 - \alpha/2) \cdot \frac{s_x}{\sqrt{n}} \right]$$

Approximation of a Binomial distribution

$X \sim \text{Bin}(n, \pi)$ (if $n\pi > 5$ and $n(1 - \pi) > 5$) $\Rightarrow X \approx \mathcal{N}(n\pi, n\pi(1 - \pi))$

Maximum likelihood estimation (MLE) for discrete distributions with measurements X_1, X_2, \dots, X_n : i.i.ds

probability mass function $p(x; \theta)$: parameterized by θ

Likelihood $L(\theta) := \prod_{i=1}^n p(x_i; \theta)$

Log-likelihood $\ell(\theta) := \log(L(\theta)) = \sum_{i=1}^n \log(p(x_i; \theta))$

Maximum likelihood estimator (MLE) for θ : $\hat{\theta}$ = value of θ for which ℓ attains its maximum

MLE for continuous distributions with probability density $f(x; \theta)$: parameterized by θ

$L(\theta) := \prod_{i=1}^n f(x_i; \theta)$

$\ell(\theta) := \log(L(\theta)) = \sum_{i=1}^n \log(f(x_i; \theta))$

MLE for Poisson distribution

$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$

$\ell(\lambda) = \sum_{i=1}^n [x_i \log(\lambda) - \lambda - \log(x_i!)]$

$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$

confidence intervals: $\left[\hat{\lambda} - \Phi^{-1}(0.975) \frac{s_{\bar{x}}}{\sqrt{n}}, \hat{\lambda} + \Phi^{-1}(0.975) \frac{s_{\bar{x}}}{\sqrt{n}} \right]$

MLE for Normal distribution $\mathcal{N}(\mu, \sigma^2)$

$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

MLE for Exponential distribution $E \times p(\lambda)$

$\hat{\lambda} = \frac{1}{\bar{x}}$

confidence interval: $\left[\hat{\lambda} \left(1 - \frac{\Phi^{-1}(0.975)}{\sqrt{n}} \right), \hat{\lambda} \left(1 + \frac{\Phi^{-1}(0.975)}{\sqrt{n}} \right) \right]$

Bayesian estimation approach: parameter θ as random

Likelihood as conditional probability: $L(\theta) = p_{X|\Theta=\theta}(x) = P[X = x | \Theta = \theta]$

$P[\Theta = \theta | X = x] = \frac{P[X=x|\Theta=\theta] \cdot P[\Theta=\theta]}{P[X=x]} : \text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$

Maximum a posteriori (MAP) estimator: $\hat{\theta}$ that maximizes the *posterior* $P[\Theta = \theta | X = x]$

Bayesian estimation of continuous parameter with density $f_{\Theta}(\theta)$

$f_{\Theta|X=x}(\theta) = \frac{f_{X|\Theta=\theta}(x) \cdot f_{\Theta}(\theta)}{f_X(x)}$

In large sample limit, $n \rightarrow \infty$: MAP estimate converges to ML estimate

Statistical Hypothesis Testing

- 1) Model: choose distribution describing your data. Formulate claim you want to prove.
- 2) Null hypothesis: choose the H_0 (*null hypothesis*) , H_A (*alternative hypothesis*) and their distribution parameters
- 3) Test statistic: based on your sample data
- 4) Choose significance level: e.g. $\alpha = 5\%$
- 5) Range of rejection K such that $P[X \in K] \leq \alpha$ under H_0
reject H_0 if $X \in K$
- 6) Test decision: reject H_0 if $X \in K$ otherwise keep it.

Decision			
Truth	H_0	H_A	
	H_0	H_A	
	true negative	type I error (FP)	
	type II error (FN)	true positive	

- Significance level α : probability of type I error given that H_0 is true
- Power $1 - \beta$: β is probability of type II error given that H_1 is true

P-value

(Def.) The p -value is the smallest significance level α for which we reject a null hypothesis for the given data set.
(Alt.) The p -value is the probability under the null hypothesis to find the actual outcome or a more extreme one.

Test using the normal approximation $X \approx \mathcal{N}(n\pi_0, n\pi_0(1-\pi_0))$

Test statistic: $Z = \frac{X-n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$ Distribution of Z under $H_0 : Z \approx \mathcal{N}(0, 1)$

Paired-samples (or one-sample) t test with model $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$

Test statistic: $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{s_x}$

Student's t distribution $T \sim t_m$ with m “degrees of freedom”, symmetry $t_{m,\alpha} = -t_{m,1-\alpha}$

Range of rejection: $K = (-\infty, -t_{n-1,1-\frac{\alpha}{2}}] \cup [t_{n-1,1-\frac{\alpha}{2}}, \infty)$

Confidence Interval for μ with confidence level $1-\alpha$

$I = \{\mu_0 \mid \text{null hypothesis } H_0 : \mu = \mu_0 \text{ is not rejected} \}$

$H_A : \mu \neq \mu_0 \Rightarrow I = \left[\bar{x} - t_{n-1,1-\alpha/2} \frac{s_x}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2} \frac{s_x}{\sqrt{n}} \right]$

$H_A : \mu < \mu_0 \Rightarrow I = \left(-\infty, \bar{x} + t_{n-1,1-\alpha} \frac{s_x}{\sqrt{n}} \right]$

$H_A : \mu > \mu_0 \Rightarrow I = \left[\bar{x} - t_{n-1,1-\alpha} \frac{s_x}{\sqrt{n}}, \infty \right)$

Sign Test: consider differences $X_i = Z_i - Y_i$ i. i. d. with median m

$H_0 : m = m_0 = 0$, $H_A : m \neq m_0$

Test statistic: $V = \#\{i \mid X_i > m_0\}$, V under $H_0 : V \sim \text{Bin}(n, 0.5)$

Range of rejection: $K = [0, c] \cup [n-c, n]$ such that $P_{H_0}[V \in K] \leq \alpha$ (significance level)

c determined by binomial distribution: $P_{H_0}[V \in K] = 2P_{H_0}[V \leq c]$

Wilcoxon Signed-Rank Test (wilcox.test): consider differences $X_i = Z_i - Y_i$ i. i. d. with median m

$H_0 : m = 0$, $H_A : m \neq 0$

Test statistic: $W = \sum_{i=1}^n \text{sign}(X_i) R_i$, where R_i : rank of X_i order by absolute value $|X_i|$

Range of rejection: $K = (-\infty, 0.5 - c] \cup [0.5 + c, \infty)$ such that $P_{H_0}[W \in K] \leq \alpha$

Permutation Test: nonparametric test

$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F_X(\cdot)$, $Y_1, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} F_Y(\cdot)$

$H_0 : F_X = F_Y$, $H_A : F_X \neq F_Y$

Test statistic: $D = \bar{X} - \bar{Y}$

Resampling: choose number of repetitions $N > 1000$

Randomly assign n values of $\{X_i\} \cup \{Y_i\}$ to “type I” and the rest m values to “type II”

Repeat N times

Range of rejection: $K = (-\infty, c_l] \cup [c_u, \infty)$, c_l : empirical $\alpha/2$ -quantile of resampling distribution, c_u : empirical $1-\alpha/2$ -quantile of resampling distribution

Effect size

Two samples: experimental group $\{X_i\}_i$, control group $\{Y_i\}_i$, effect size = $\frac{\bar{X}-\bar{Y}}{s_{\text{pool}}}$

False Positive Rate: $FPR = E \left[\frac{FP}{FP+TN} \right] = E \left[\frac{V}{m_0} \right]$ controlled by significance level $\alpha = FPR$

		Decision		Total
Truth	H_0	true negative U	type I error (FP): V	m_0
	H_A	type II error (FN)	true positive: S	$m - m_0$
Total		$m - R$	R	m

Family-Wise Error Rate: FWER = $P[1 \text{ or more type I errors}] = P[V \geq 1]$: **n:20-50, errors are critical**

FWER controlled by experiment-wise type I error rate $\bar{\alpha}$

Test procedure that guarantees a FWER of (at most) $\bar{\alpha}$:

1. for each test case (e.g. gene), calculate p-value
2. adjust p-value
3. reject null hypotheses whose adjusted p-value is smaller than $\bar{\alpha}$; accept others

Controlling FWER: Bonferroni method

$$P_{\text{adj},i} = \min \{m \cdot P_i, 1\}$$

guarantees FWER $\leq \bar{\alpha}$

Controlling FWER: Holm method

order p-values: $P_{(1)} \leq P_{(2)} \leq P_{(3)} \leq \dots \leq P_{(m)}$

adjust p-values: $P_{\text{adj},(i)} = \max \left\{ \min \left\{ (m - i + 1) \cdot P_{(i)}, 1 \right\}, P_{\text{adj},(i-1)} \right\}$

reject null hypotheses whose adjusted p-value is smaller than $\bar{\alpha}$; accept others

Procedure guarantees FWER $\leq \bar{\alpha}$

Adjusted p-value

The adjusted p-value of a certain null hypothesis is the smallest experiment-wise type I error rate $\bar{\alpha}$ for which we reject this hypothesis for the given data set.

False discovery rate: $\text{FDR} = E \left[\frac{\text{FP}}{\text{FP} + \text{TP}} \right] = E \left[\frac{V}{R} \right]$: **n > 500 ; looking for discovery**

1. for each test case, calculate p-value
2. adjust p-values to get corresponding q-values
3. reject null hypotheses whose q-value is smaller than \bar{q} ; accept others

Controlling FDR: Benjamini-Hochberg method

order p-values: $P_{(1)} \leq P_{(2)} \leq P_{(3)} \leq \dots \leq P_{(m)}$

adjust p-values to get q-values: $Q_{(i)} = \max \left\{ \min \left\{ \frac{m}{i} \cdot P_{(i)}, 1 \right\}, Q_{(i-1)} \right\}$

reject null hypotheses whose q-value is smaller than \bar{q} ; accept others

Procedure guarantees $\text{FDR} \leq \bar{q}$

Simple linear regression: $Y_i = \beta_0 + \beta_1 x_i + E_i$, $E_1, \dots, E_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, $i = 1, \dots, n$

Y_i : response variable ; x_i : explanatory variable ; E_i : error or noise variables

Residuals: $R_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

Residual Sum of Squares: $\text{RSS} = \sum_{i=1}^n R_i^2$

Minimizers $\hat{\beta}_0$ and $\hat{\beta}_1$ of RSS: unbiased estimators for the true β_0 and β_1

Estimate $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n R_i^2$

Significance of explanatory variable: t-test

$H_0 : \beta_1 = 0$; $H_A : \beta_1 \neq 0$

$T = \frac{\hat{\beta}_1 - 0}{\text{se}(\hat{\beta}_1)}$; $H_0 : T \sim t_{n-2}$

$K = (-\infty, -t_{n-2, 1-\frac{\alpha}{2}}] \cup [t_{n-2, 1-\frac{\alpha}{2}}, \infty)$, $|T| > t_{n-2, 1-\frac{\alpha}{2}} : \text{reject}$

Coefficient of determination $R^2 = \left(\frac{s_{\hat{y}y}}{s_{\hat{y}} s_y} \right)^2$

$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} = \text{Cor}(Y, X) \frac{\text{Sd}(Y)}{\text{Sd}(X)}$, $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Fitting linear model to transformed data

$\log(Y_i) = \hat{\beta}_0 + \hat{\beta}_1 \log(X_i) + E_i$