ST102 HIGHLIGHTS

(2) Descriptive statistics

Sample mean

$$\overline{X} = \frac{\sum X_i}{n} = \frac{\sum f_i X_i}{\sum f_i} = \sum_{i=1}^{K} x_i \hat{p}(x_i)$$

Sample median

$$q_{50} = X_{\frac{(n+1)}{2}} \text{ or } \frac{\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right)}{2}$$

Sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} f_i X_i^2 - n\overline{X}^2}{n-1}}$$

Sample variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1} = \frac{\sum_{i=1}^{n} f_i X_i^2 - n\overline{X}^2}{n-1}$$

Linear transformation of random variable

$$\overline{Y} = \frac{\sum(aX_i + b)}{n} = \frac{a\sum X_i + bn}{n} = a\overline{X} + b$$

$$s_Y^2 = \frac{\sum_{i=1}^{n}(aX_i + b - a\overline{X} - b)^2}{n-1} = \frac{a^2\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = a^2 s_X^2$$

(3) Introduction to probability theory

Key properties of set operators

Distributive laws I: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Distributive laws II: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

De Morgan's laws I: $(A \cap B)^c = A^c \cup B^c$

De Morgan's laws II: $(A \cup B)^c = A^c \cap B^c$

Partitions and pairwise disjointedness

$$A_i \cap A_j = \emptyset \text{ and } P\left(\bigcup A_i\right) = \sum P(A_i)$$

Probabilities of equally likely outcomes

$$P(\text{Outcome}) = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes in sample}}$$

Combinatorial counting rules

	With replacement	W/O replacement
Ordered	n^k	$\frac{n!}{(n-k)!}$
Unordered	(‡) $\binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

Birthday problem

$$P(A) = \frac{\frac{n!}{(n-k)!}}{n^k} = \frac{\text{Ordered outcomes w/o replacement}}{\text{Ordered outcomes with replacement}}$$

$$n = 365 \text{ days}, k = \text{Number of people}$$

Independence

$$A \perp\!\!\!\perp B, P(A \cap B) = P(A)P(B)$$

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) = P(A_1)\ldots P(A_n)$$

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

Conditional probability

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

Chain rule of conditional probabilities

$$P(A_1 \cap A_2 \cap A_3 \ldots \cap A_n) = P(A_1)\prod_{i=2}^{n} P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right)$$

$$= P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)\ldots P(A_n|A_{n-1} \cap \ldots A_1)$$

Total probability formula

$$P(A) = \sum_{i=1}^{K} P(A \cap B_i) = \sum_{i=1}^{K} P(A|B_i)P(B_i)$$

E.g. $P(A) = P(A \cap B) + (P(A \cap B^c) = P(A|B)P(B) + (P(A|B^c)P(B^c)$

Bayes' theorem

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} \Longrightarrow \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{K} P(A|B_i)P(B_i)}, j \in \{1 \ldots K\}$$

(4) Discrete random variables (DRV) (Stepwise)

DRV probability density function (PDF)

$$\text{Conditions: } P(X = x) = p(x) \geq 0 \text{ and } \sum_{X_i \in S} p(x_i) = 1$$

$$p(x) = \begin{cases} p(x) \text{ for } x = 0,1,2 \ldots n \\ 0 \text{ otherwise} \end{cases}$$

DRV cumulative distribution function (CDF)

$$F(x) = P(X \leq x), x \in \mathbb{R} \text{ and } F(x) = \sum_{x_i \in S, x_i \leq x} p(x_i)$$

$$F(x) = \begin{cases} 0 \text{ when } x < 0 \\ F(x) \text{ for } x = 0,1,2 \ldots n \\ 1 \text{ when } x > n \end{cases}$$

Key properties of DRV CDF

$F(x)$ is strictly increasing: $x_1 < x_2$, then $F(x_1) < F(x_2)$

$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to +\infty} F(x) = 1$$

For any $x_1 < x_2, P(x_1 < X < x_2) = F(x_2) - F(x_1)$

DRV expected value, $E(X)$ or μ

$$E(X) = x_1 p(x_1) + \cdots + x_K p(x_K) = \sum_{i=1}^{K} x_i p(x_i)$$

$$E(X^2) = x_1^2 p(x_1) + \cdots + x_K^2 p(x_K) = \sum_{i=1}^{K} x_i^2 p(x_i)$$

DRV variance and standard deviation, $Var(X)$ or σ^2

$$Var(X) = E(X^2) - E(X)^2 = \sum_{i=1}^{K}[x_i - E(X)]^2 p(x_i)$$

$$= E\{[X - E(X)]^2\}$$

$$\sigma = \sqrt{E(X^2) - E(X)^2} = \sqrt{\sum_{i=1}^{K}[x_i - E(X)]^2 p(x_i)}$$

Linear transformation of random variable

$$E(aX + b) = a\sum xp(x) + b\sum p(x) = aE(X) + b$$

$$Var(aX + b) = E\{[aX + b - aE(X) - b]^2\}$$

$$= E\{a^2[X - E(X)]^2\} = a^2 E\{[X - E(X)]^2\} = a^2 Var(X)$$

Constants: $E(b) = b, Var(b) = sd(b) = 0$

DRV moment generating function (MGF)

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} p(x)$$

$$M_x'(t) = \frac{\partial}{\partial t} M_x(t) \Longrightarrow M_x'(0) = E(X^1)$$

$$M_x''(t) = \frac{\partial}{\partial t} M_x'(t) \Longrightarrow M_x''(0) = E(X^2)$$

$$Var(X) = E(X^2) - E(X)^2 = M_x''(0) - [M_x'(0)]^2$$

(4) Continuous random variables (CRV) (Continuous)

CRV PDF

$$\text{Conditions: } f(x) \geq 0 \text{ and } \int_{-\infty}^{+\infty} f(x)\,dx = 1$$

$$f(x) \neq P(X = x) \Longrightarrow P(X = x) = 0$$

$$f(x) = \frac{\partial}{\partial x} F(x) = F'(x)$$

CRV CDF

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, \forall x$$

Median: $F(m) = 0.5$

Interquartile range: $q_{75} - q_{25}$

(+) skew: Mean > Median, (-) skew: Mean > Median

Pareto distribution

$$f(x) = \begin{cases} 0 \text{ when } x < k \\ \frac{\alpha k^\alpha}{x^{\alpha+1}} \text{ when } x \geq k, \{k, \alpha\} > 0 \end{cases}$$

(1) $\alpha k^\alpha > 0, x^{\alpha+1} > 0 \Longrightarrow f(x) > 0$

(2) $\int_{-\infty}^{+\infty} \frac{\alpha k^\alpha}{x^{\alpha+1}} dx = \alpha k^\alpha \left(\frac{1}{-\alpha}\right) \int_{k}^{+\infty} -\alpha x^{-\alpha-1} dx$

$$= \left[\frac{\alpha k^\alpha}{-\alpha}\right][x^{-\alpha}]_k^{+\infty} = (-k^\alpha)(-k^{-\alpha}) = 1$$

$$F(x) = \begin{cases} 0 \text{ when } x < k \\ 1 - \left(\frac{k}{x}\right)^\alpha \text{ when } \geq k \end{cases}$$

$$E(X) = \frac{\alpha k}{\alpha - 1}$$

$$Var(X) = \left(\frac{k}{\alpha - 1}\right)^2 \left(\frac{\alpha}{\alpha - 2}\right)$$

CRV expected value, $E(X)$ or μ

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

CRV variance and standard deviation, $Var(X)$ or σ^2

$$Var(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{+\infty}[x - E(X)]^2 f(x)dx$$

$$\sigma = \sqrt{E(X^2) - E(X)^2} = \sqrt{\int_{-\infty}^{+\infty}[x - E(X)]^2 f(x)dx}$$

CRV MGF

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx$$

$$M_x'(t) = \frac{\partial}{\partial t} M_x(t) \Longrightarrow M_x'(0) = E(X^1)$$

$$Var(X) = E(X^2) - E(X)^2 = M_x''(0) - [M_x'(0)]^2$$

Summary

	DRV	CRV
$E(X)$	$\sum_{Lower}^{Upper} x_i p(x_i)$	$\int_{Lower}^{Upper} xf(x)dx$
$Var(X)$	$E(X^2) - E(X)^2$	$E(X^2) - E(X)^2$
$M_x(t)$	$\sum_{Lower}^{Upper} e^{tx} p(x)$	$\int_{Lower}^{Upper} e^{tx} f(x)dx$

(5) Common distributions of RV

Binomial distribution, $X \sim B(n, \pi)$, from (3)

$$p(x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x} \text{ for } x = 0,1,2 \ldots n \\ 0 \text{ otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ when } t < 0 \\ \sum_{t=0}^{x} \binom{n}{t} \pi^t (1-\pi)^{n-t} \text{ for } t = 0,1,2 \ldots x \end{cases}$$

$$E(X) = n\pi$$

$$Var(X) = n\pi(1-\pi)$$

$$M_x(t) = [e^t \pi + (1-\pi)]^n$$

$(1)\ \pi \geq 0, (1-\pi) \geq 0 \Longrightarrow p(x) \geq 0$

$(2)\ \sum_{x=0}^{n} \binom{n}{x} \pi^x (1-\pi)^{n-x} = [\pi + (1-\pi)]^n = 1^n = 1$

Discrete uniform distribution

$$p(x) = P(X = x) = \begin{cases} \frac{1}{K}, \forall x = 1,2 \ldots K \\ 0 \text{ otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ for } x < 0 \\ \sum_{i=1}^{K} \frac{1}{K}, \forall i = 1,2 \ldots K \\ 1 \text{ for } i > K \end{cases}$$

$$E(X) = \frac{K+1}{2}$$

$$Var(X) = \frac{K^2 - 1}{12}$$

$$(‡)\ M_x(t) = \frac{e^t(1 - e^{Kt})}{K(1 - e^t)}$$

$(1)\ K > 0, p(x) > 0$

$(2)\ \sum_{i=1}^{K} \frac{1}{K} = 1$

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

$$p(x) = \begin{cases} \pi \text{ for } x = 1 \\ 1 - \pi \text{ for } x = 0 \\ 0 \text{ otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ for } x < 0 \\ 1 - \pi \text{ for } 0 \leq x < 1 \\ 1 \text{ for } x \geq 1 \end{cases}$$

$$E(X) = \pi$$

$$Var(X) = \pi(1-\pi)$$

$$M_x(t) = (1-\pi) + e^t \pi$$

$(1)\ \pi > 0, 1 - \pi > 0 \Longrightarrow p(x) > 0$

$(2)\ \sum_{i=0}^{1} p(x_i) = \pi + (1-\pi) = 1$

Poisson distribution, $X \sim Po(\lambda)$

$$p(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0,1,2 \ldots \\ 0 \text{ otherwise} \end{cases}$$

$$F(x) = \begin{cases} e^{-\lambda} \sum_{t=0}^{x} \frac{\lambda^t}{t!}, \text{ for } t = 0,1,2 \ldots \\ 0 \text{ otherwise} \end{cases}$$

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

$$M_x(t) = e^{\lambda(e^t - 1)}$$

$(1)\ e^{-\lambda} > 0, \lambda^x \geq 0, x! \geq 0 \Longrightarrow p(x) \geq 0$

$(2)\ \sum_{i=0}^{n} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{i=0}^{n} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$

Note: $\sum_{i=0}^{\infty} \frac{\lambda^x}{x!} \approx e^{\lambda}$ (Taylor series approximation)

Linear combinations of the Poisson distribution

$$0.5X \sim Po(0.5\lambda)$$

$$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2) \Longrightarrow \sum_{i=1}^{n} X_i \sim Po\left(\sum_{i=1}^{n} \lambda_i\right)$$

Continuous uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ for } a \leq x \leq b \\ 0 \text{ otherwise} \end{cases}$$

$$F(x) = P(X \leq x) = \int_a^x f(t)\,dt = \begin{cases} 0 \text{ for } x < a \\ \frac{x-a}{b-a} \text{ for } a \leq x \leq b \\ 1 \text{ for } x > b \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

$$(‡)\ M_x(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$(1)\ b - a > 0 \Longrightarrow f(x) > 0$

$(2)\ \int_a^b \frac{1}{b-a} dx = \left[\frac{x}{b-a}\right]_a^b = \frac{b-a}{b-a} = 1$

Exponential distribution, from (4)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ for } x > 0, \lambda > 0 \\ 0 \text{ for } x \leq 0 \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ for } x \leq 0 \\ 1 - e^{-\lambda x} \text{ for } x > 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

$$M_x(t) = \frac{\lambda}{\lambda - t}$$

$(1)\ \lambda > 0, e^{-\lambda x} > 0 \Longrightarrow f(x) > 0$

$(2)\ \int_{-\infty}^{+\infty} \lambda e^{-\lambda x} dx = (-1) \int_0^{+\infty} -\lambda e^{-\lambda x} dx = -[e^{-\lambda x}]_0^{+\infty} = 1$

(‡) Memoryless: $P(X > s + t | X > t) = P(X > s), s, t \geq 0$

Normal distribution, $X \sim N(\mu, \sigma^2)$

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

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

$$(‡)\ M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$(1)\ \frac{1}{\sigma\sqrt{2\pi}} > 0, e^{-\frac{(x-\mu)^2}{2\sigma^2}} > 0 \Longrightarrow f(x) > 0$

$(‡)(2)\ \int_{-\infty}^{+\infty} f(x)\,dx = 1$

Linear combinations of the normal distribution

$(1)\ X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$

$(2)\ aX \pm b \sim N(a\mu \pm b, a^2\sigma^2)$

$(3)\ aX \pm bY \sim N(a\mu_1 \pm b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Standard normal, $\frac{X-\mu}{\sigma} = Z \sim N(0, 1^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \forall x \in \mathbb{R}$$

$$F(x) = \phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$E(X) = 0$$

$$Var(X) = 1$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

Poisson approximation of the binomial distribution

$$X \sim B(n, \pi) \Longrightarrow n\pi \leq 5 \Longrightarrow X \sim Po(n\pi = \lambda)$$

Normal approximation of the binomial distribution

$$X \sim B(n, \pi) \Longrightarrow n\pi \geq 5 \text{ and } n(1-\pi) \geq 5 \ldots$$

$$\ldots \Longrightarrow X \sim N\big(n\pi, n\pi(1-\pi)\big)$$

$$\begin{aligned} &\text{CC1}: P(X \leq x) \approx P(Y \leq x + 0.5) \\ &\text{CC2}: P(X < x) \approx P(Y < x - 0.5) \\ &\text{CC3}: P(X \geq x) \approx P(Y \geq x - 0.5) \\ &\text{CC4}: P(X > x) \approx P(Y > x + 0.5) \end{aligned}$$

(‡) Normal approximation of the Poisson distribution

$$X \sim Po(\lambda) \Longrightarrow \lambda > 10 \Longrightarrow X \sim N(\lambda, \lambda)$$

$$P(X = x) = P(x - 0.5 < X < x + 0.5) \text{ by CC}$$

(6) Multivariate random variables

Discrete joint probability function

$$p(x, y) = P(X = x \cap Y = y) \equiv p_{XY}(x, y)$$

$$\text{Conditions: } p(x,y) \geq 0 \ \forall (x,y) \text{ and } \sum_{x=0}^{n}\sum_{y=0}^{n} p(x,y) = 1$$

Discrete univariate marginal distribution

$$p_X(x=k) = \sum_{y=0}^{n} p(x=k,y)\,, k \in \{x_1, x_2, \dots, x_n\}$$

$$E(X) = \sum_{x=x_1}^{x_n} x p_X(x)$$

$$E(XY) = \sum_{xy=0}^{k} xy\, p_{XY}(x,y), xy \in \{0, \dots, k\}$$

Discrete conditional distribution

$$P_{Y|X}(y|x) = P(Y=y|X=x) = \frac{P(X=x \cap Y=y)}{P(X=x)} \dots$$

$$\dots \equiv \frac{p_{XY}(x,y)}{p_X(x)}$$

$$\text{Conditions: } P_{Y|X}(y|x) \geq 0 \text{ and } \frac{\sum_{y=y_1}^{y_n} p_{XY}(x,y)}{p_X(x)} = \frac{p_X(x)}{p_X(x)}$$

Covariance: Measure of association/dependence

$$Cov(X,Y) = Cov(Y,X) = E\{[X - E(X)][Y - E(Y)]\} \dots$$

$$\dots = E(XY) - E(X)E(Y)$$

Properties of covariance

$$Cov(X,X) = E(XX) - E(X)E(X) = Var(X)$$

$$Cov(k,X) = E(kX) - E(k)E(X) = kE(X) - kE(X) = 0$$

$$Cov(aX+b, cY+d) = \cdots = acCov(X,Y)$$

Correlation: Measure of linear association

$$Corr(X,Y) = Corr(Y,X) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$\text{Positive coorelation: } Corr(X,Y) \approx +1$$

$$\text{Negative coorelation: } Corr(X,Y) \approx -1$$

$$\text{Not coorelated: } Corr(X,Y) \approx 0$$

Independence

$$\text{Dependent: } P_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)} \neq p_Y(y)$$

$$\text{Independent: } p_{XY}(x,y) = p_X(x)p_Y(y)$$

$$\text{Discrete \& independent: } p(x_1 \cap x_2 \dots x_n) = \prod_{i=1}^{n} p(x_i)$$

$$\text{Continuous \& independent: } f(x_1 \cap x_2 \dots x_n) = \prod_{i=1}^{n} f(x_i)$$

Independent random variables are also uncorrelated

$$\text{ie. } Cov(X,Y) = 0 \text{ and } Corr(X,Y) = 0$$

Joint distributions of random variables

$$E\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i\, E(X_i) + b$$

E.g. $E(X \pm Y) = E(X) \pm E(Y)$

E.g. $E\big(E(X)\big) = E(X)$ and $E(E(X)^2) = E(X)^2$

$$Var\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2\, Var(X_i) + 2\sum\sum_{i<j} a_i a_j Cov(X_i, X_j)$$

E.g. $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$

$$\text{Independent: } Var\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2\, Var(X_i)$$

E.g. $Var(X \pm Y) = Var(X) + Var(Y)$, as $Cov(X,Y) = 0$

$$\text{Independent: } E\left(\prod_{i=1}^{n} a_i X_i\right) = \prod_{i=1}^{n} a_i E(X_i)$$

E.g. $E(XY) = E(X)E(Y)$

(7) Sampling distribution

Sampling distribution

$$\text{Population distribution: } X \sim N(\mu, \sigma^2)$$

$$\text{Sampling distribution: } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Chi square distribution, t-distribution – refer to (26)

Central limit theorem

If $n \geq 30$, then any random sample from any distribution with sample mean μ and sample variance σ^2 is asymptotically normally distributed:

$$\text{ie. } \approx \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

F-distribution – refer to (30)

(22) Point estimation I

Method of moments estimator (MME)

Sample moment	Population moment
$M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \Longrightarrow \hat{\mu}$	$E(X) = \mu$
$M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Longrightarrow \hat{\sigma}^2 + \hat{\mu}^2$	$E(X^2) = \sigma^2 + \mu^2$
$\vdots$	$\vdots$
$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$	$E(X^k)$

Let k denote the number of parameters in the random distribution (ie. Normal = 2, Poisson = 1). Use integrals for continuous variables.

$$\hat{\mu}_{MM} = E(X) = \bar{X}$$

$$\hat{\sigma}^2{}_{MM} = E(X^2) - E(X)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Unbiased sample variance

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}\right]$$

$$E(\sigma^2) = E(M_2) - E(M_1^2) = E(X^2) - E(\bar{X}^2)$$

$$\Longrightarrow E(X^2) - [Var(\bar{X}) + E(\bar{X})^2] = \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} + \cancel{\mu^2}$$

$$E(\sigma^2) - E(S^2) = \frac{\sigma^2}{n} > 0 \Longrightarrow \text{Biased}$$

Mean square error (MSE)

$$\text{Bias}\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - E(\theta) = E\left(\hat{\theta}\right) - \theta$$

$$\Longrightarrow E(\hat{\mu}) = \mu \Leftrightarrow E\left(\sum a_i X_i\right) = E(X_i)\sum a_i = \mu$$

$$\text{Where } \sum a_i = \sum \frac{1}{n} = 1$$

$$\text{MSE} = E\left\{\left(\hat{\theta} - \theta\right)^2\right\} = \left\{\text{Bias}\left(\hat{\theta}\right)\right\}^2 + Var\left(\hat{\theta}\right)$$

$$\Longrightarrow Var(\hat{\mu}) = \frac{\sigma^2}{n} \Leftrightarrow Var\left(\sum a_i X_i\right) = Var(X_i)\sum a_i^2 = \frac{\sigma^2}{n}$$

$$\text{Where } \sum a_i^2 = \sum \frac{1}{n^2} = \frac{1}{n}$$

$$\text{Consistent estimator: } \lim_{n\to\infty} \text{MSE} = 0$$

$$\text{Bias}(\hat{\mu}) = E(\hat{\mu}) - \mu = 0$$

$$Var(\hat{\mu}) = \frac{\sigma^2}{n} = \text{MSE}(\hat{\mu})$$

(23) Point estimation II

Standard error

$$SE\left(\hat{X}\right) = \frac{S}{\sqrt{n}}$$

Maximum likelihood estimator (MLE)

$$L(\theta) = f(X_1, \dots, X_n; \theta) = \prod_{i=1}^{n} f(X_i)$$

$$l(\theta) = lnL(\theta)$$

$$\text{Maximize log-likelihood function: } \frac{dl(\theta)}{d\theta} = 0$$

$$\theta = \hat{\theta} = g(\bar{X})$$

$$\hat{\mu}_{ML} = E(\bar{X}) = h(\bar{X})$$

MLE for non-differentiable (ie. Discrete) distributions

$$\hat{\theta} = X_{i\,max}$$

Invariance property of MLE

$$\text{If } \hat{\theta} \text{ is the MLE for } \theta \text{, then } g\left(\hat{\theta}\right) \text{ is the MLE for } g(\theta)$$

$$\text{ie. Solve for } \hat{\theta} \text{ first and sub } g\left(\hat{\theta}\right) \text{ in}$$

MLE utilizes information about the entire population distribution. It is often more efficient/accurate than MME/LSE which uses E(X) and Var(X).

(24) Point estimation III

MLE with unknown parameters

$$l(\theta_1, \theta_2) = lnL(\theta_1, \theta_2)$$

$$\frac{dl(\theta_1)}{d\theta_1} = 0, \frac{dl(\theta_2)}{d\theta_2} = 0$$

Fisher information

$$\text{Central limit theorem: When } n \text{ is large, } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\text{When } n \text{ is large, } \hat{\theta} \sim N\left(\theta, \frac{1}{nI(\theta)}\right)$$

$$I(\theta) = -\int_{-\infty}^{+\infty} f(X;\theta)\frac{\partial^2 lnf(X;\theta)}{\partial\theta^2}\, dX$$

(25) Interval estimation

Confidence interval

Min. length of interval, max. coverage probability

$$P\left(\frac{|\mu - \bar{X}|}{\frac{\sigma}{\sqrt{n}}} \le C_{\alpha/2}\right) = \alpha$$

$$\Rightarrow P\left(\bar{X} - \frac{C_{\alpha/2}\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{C_{\alpha/2}\sigma}{\sqrt{n}}\right) = \alpha$$

σ^2 known: $[\bar{X} - C_{\alpha/2}SE(\bar{X}), \bar{X} + C_{\alpha/2}SE(\bar{X})] \sim N(0,1)$

$$\Rightarrow P\left(\bar{X} - \frac{C_{\alpha/2}S}{\sqrt{n}} \le \mu \le \bar{X} + \frac{C_{\alpha/2}S}{\sqrt{n}}\right) = \alpha$$

σ^2 unknown: $[\bar{X} - C_{\alpha/2}SE(\bar{X}), \bar{X} + C_{\alpha/2}SE(\bar{X})] \sim t_{n-1}$

MLE-based interval

$$\left[\hat{\theta} - \frac{C_{\alpha/2}}{\sqrt{nI(\hat{\theta})}}, \hat{\theta} + \frac{C_{\alpha/2}}{\sqrt{nI(\hat{\theta})}}\right]$$

(26) χ^2 distribution and t distribution

Chi-squared χ^2 distribution

Let $Z_i, i \in 1,2., \dots k$ be $Z \sim N(0,1)$

$$W = Z_1^2 + Z_2^2 + \dots + Z_k^2 = \sum_{i=1}^{k} Z_i^2$$

Then $W \sim \chi_k^2$ ("χ^2 distribution with k degrees of freedom")

$$E(W) = \sum_{i=1}^{k} E(Z_i^2) = kE(Z^2) = k[Var(Z) + E(Z)^2] = k$$

$$Var(W) = kVar(Z^2) = k[E(Z^4) - E(Z^2)^2] = 2k$$

If W_i are independent random variables and $W_i \sim \chi_i^2$

$$W_1 + W_2 + \dots + W_k = \sum_{i=1}^{k} Z_i^2 = U \sim \chi^2_{1+2+\dots+k}$$

χ^2 values can only be positive

Confidence interval for σ^2

$$\left[\frac{(n-1)S^2}{C_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{C_{\alpha/2}}\right] \sim \chi^2_{n-1}$$

T-distribution

Let $Z \sim N(0,1)$ and $W \sim \chi_k^2$

$$\frac{Z}{\sqrt{\frac{W}{k}}} \sim t_k$$

$$\frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$$

"T-distribution with k degrees of freedom"

$E(T) = 0$, for $k > 1$

$Var(T) = \frac{k}{k-2}$, for $k > 2$

(27/28) Hypothesis testing

(One-tailed normal test)

$H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

Perform a one-tailed Z test at α level of significance

Apply a Z test, $P(\mu \ge \mu_0) = P\left(Z \ge \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right) \sim N(0,1)$

$$\Rightarrow p = C_Z$$

If $p \le \alpha$, reject H_0. Otherwise do not reject H_0

Vice versa for "smaller than" equality

(Two-tailed normal test)

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

Perform a two-tailed Z test at α level of significance

Apply a Z test, $P(|\mu| \ge \mu_0) = P\left(|Z| \ge \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right) \sim N(0,1)$

$$\Rightarrow 2P\left(Z \ge \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right) = 2C_Z = p$$

If $p \le \alpha$, reject H_0. Otherwise do not reject H_0

(One-tailed T test)

$H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

Perform a one-tailed T test at α level of significance

Apply a T test, $P(\mu \ge \mu_0) = P\left(T \ge \frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right) \sim t_{n-1}$

$$\Rightarrow C_{t_1} \le p \le C_{t_2}$$

If $p \le \alpha$, reject H_0. Otherwise do not reject H_0

(Two-tailed T test)

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

Perform a two-tailed T test at α level of significance

Apply a T test, $P(|\mu| \ge \mu_0) = P\left(|T| \ge \frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right) \sim t_{n-1}$

$$\Rightarrow 2P\left(T \ge \frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right) = 2(C_{t_1} \le p \le C_{t_2}) = p$$

If $p \le \alpha$, reject H_0. Otherwise do not reject H_0

(Normal variance test)

$H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$

Perform a one-tailed χ^2test at α level of significance

Apply a chi-square test, $P\left(\chi^2_{n-1} > \frac{S^2}{\sigma_0^2}(n-1)\right) = p$

If $p \le \alpha$, reject H_0. Otherwise do not reject H_0

Type I error: Rejecting the null hypothesis when it is actually not false

$P(\text{Type I error}) = \alpha$

Type II error: Not rejecting the null hypothesis when it is false

$P(\text{Type II error}) = \beta$

$$\beta = P\left(Z < \frac{\bar{X}_{\text{Upper bound initial reject}} - \mu_{\text{True}}}{\sigma/\sqrt{n}}\right)$$

		Decision made	
		H_0 not rejected	H_0 rejected
True state	H_0	$1-\alpha$	α
	H_1	β	$1-\beta$

Power of a test: Probability of rejecting H0 when H1 is true

Determinants for power include: True difference (larger differences in μx-μy, higher power), population variance (smaller population variance, higher power), knowledge of population variance (known, higher power), sample size (larger n, higher power), design of experiment (paired design, higher power)

(29) Inference for two normal samples

Two normal means with paired observations, σ_X^2, σ_Y^2 unknown, $n_X = n_Y$ (Unrelated, independent samples)

$$X - Y = Z \sim (\mu_Z, S_Z^2)$$

$$\mu_Z = \mu_X - \mu_Y$$

$$S_Z^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - Y_i) - n(\bar{X} - \bar{Y})\right]$$

$H_0: \mu_Z = \mu_X - \mu_Y = 0$ vs $H_1: \mu_Z = \mu_X - \mu_Y < 0$

Perform a one-tailed T test at α level of significance

$$P\left(T < \sqrt{\frac{n_X + n_Y - 2}{\frac{1}{n_X} + \frac{1}{n_Y}}} \times \frac{X - Y - (\mu_X - \mu_Y)}{\sqrt{S_Z^2(n_X - 1)}}\right) \sim t_{n_X - 1}$$

$$P\left(T < \frac{\sqrt{n}\bar{Z}}{S_Z}\right) \sim t_{n-1}$$

(Confidence interval)

$$\left(\bar{X} - \bar{Y} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} - \bar{Y} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

Two normal means with σ_X^2, σ_Y^2 known but unequal, $n_X \neq n_Y$

$H_0: \mu_Z = \mu_X - \mu_Y = 0$ vs $H_1: \mu_Z = \mu_X - \mu_Y < 0$

$$P\left(Z < \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}\right) \sim N(0,1)$$

(Confidence interval)

$$\left(\bar{X} - \bar{Y} - C_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}, \bar{X} - \bar{Y} + C_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right)$$

Two normal means with σ_X^2, σ_Y^2 unknown but equal, $n_X \neq n_Y$

$H_0: \mu_Z = \mu_X - \mu_Y = 0$ vs $H_1: \mu_Z = \mu_X - \mu_Y < 0$

Perform a one-tailed T test at α level of significance

$$T = \sqrt{\frac{n_X + n_Y - 2}{\frac{1}{n_X} + \frac{1}{n_Y}}} \times \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2(n_X - 1) + S_Y^2(n_Y - 1)}}$$

$$T \sim t_{n_X + n_Y - 2}$$

If $n_X = n_Y, \hat{\sigma}^2 = \frac{S_X^2(n-1) + S_Y^2(n-1)}{2(n-1)} = \frac{S_X^2 + S_Y^2}{2}$

$$\dots = \frac{\sigma^2}{2(n-1)}\left[\frac{(n-1)S_X^2}{\sigma^2} + \frac{(n-1)S_Y^2}{\sigma^2}\right] \sim \chi^2_{2(n-1)}$$

(Confidence interval)

$$\bar{X} - \bar{Y} \pm C_{\alpha/2,n_X+n_Y-2}\sqrt{\frac{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)[S_X^2(n_X - 1) + S_Y^2(n_Y - 1)]}{n_X + n_Y - 2}}$$

(30) Inference for correlation coefficients and variances

Correlation coefficient

$$\rho = Corr(X,Y) = Corr(Y,X) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$\Rightarrow \frac{E\{[X-E(X)][Y-E(Y)]\}}{\sqrt{[E(X^2)-E(X)^2][E(Y^2)-E(Y)^2]}}$$

Sample correlation coefficient

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

$$\Rightarrow \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{(n-1)^2 S_X^2 S_Y^2}}$$

$$|\hat{\rho}| \geq 1$$

Test if 2 samples are correlated

$$H_0: \rho = 0 \text{ vs } H_1: \rho < 0$$

Perform a one-tailed T test at α level of significance

$$P\left(T < \hat{\rho}\sqrt{\frac{n-2}{1-\hat{\rho}^2}}\right) \sim t_{n-2}$$

F-distribution

Let $U \sim \chi_p^2$ and $V \sim \chi_k^2$

$$\frac{U/p}{V/k} = W \sim F_{p,k}$$

"F-distribution with degrees of freedom (p,k)"

$$E(W) = \frac{k}{k-2}, \text{for } k > 2$$

$$Var(W) = \frac{2k^2(p+k-2)}{p(k-2)^2(k-4)}$$

$$\left(\frac{U/p}{V/k}\right)^{-1} = W^{-1} \sim F_{k,p}$$

Test if 2 variances are equal

$$H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1 \text{ vs } H_1: \frac{\sigma_X^2}{\sigma_Y^2} > 1$$

Perform a one-tailed T test at α level of significance

$$P\left(F > \frac{S_X^2}{S_Y^2}\right) \sim F_{\alpha, n_X-1, n_Y-1}$$

Perform a two-tailed T test at α level of significance

$$P\left(|F| > \frac{S_X^2}{S_Y^2}\right) \sim F_{\alpha, n_X-1, n_Y-1}$$

Do not write |F| or F

(31) Goodness of fit tests

Test if the distribution fits the data

Category (Discrete)	1	2	Total
Category (Continuous)	$(-k_1, k_2]$	$(-k_2, k_3]$	Total
Z_i	Z_1	Z_2	$\sum Z_i = n$
p_i	p_1	p_2	1
E_i	E_1	E_2	$\sum E_i = \sum Z_i = n$
$Z_i - E_i$	$Z_1 - E_1$	$Z_2 - E_2$	$\sum (Z_i - E_i)$
$\frac{(Z_i-E_i)^2}{E_i}$	$\frac{(Z_1-E_1)^2}{E_1}$	$\frac{(Z_2-E_2)^2}{E_2}$	$\sum \frac{(Z_i-E_i)^2}{E_i}$

Always aggregate $Z_i \geq 5$

$$E_i = p_i \sum Z_i$$

(All parameters in PDF known)

$$H_0: Z_i = E_i = np_i \text{ vs } H_1: Z_i \neq E_i = np_i$$

Perform a one-tailed χ^2 test at α level of significance

$$P\left(\chi_{\alpha,n-1}^2 = \sum \frac{(Z_i-E_i)^2}{E_i}\right) \sim \chi_{\alpha,n-1}^2$$

(Not all parameters in PDF known ie. Parameters are estimated from the sample data)

$$H_0: Z_i = E_i = np_i \text{ vs } H_1: Z_i \neq E_i = np_i$$

Perform a one-tailed χ^2 test at α level of significance

$$P\left(\chi_{\alpha,n-1-d}^2 = \sum \frac{(Z_i-E_i)^2}{E_i}\right) \sim \chi_{\alpha,n-1-d}^2$$

Where d = Number of unknown parameters

$$\text{E.g. } \lambda = E(X) = \frac{\sum X_i p(X_i)}{n} = \bar{X}$$

E.g. Poisson PDF with unknown λ in 5 categories $\sim \chi_{5-1-1}^2$

E.g. Normal CDF with unknown μ, σ^2 in 7 categories $\sim \chi_{7-1-2}^2$

Identifying intervals $\bar{X}$ for continuous distributions

$$\text{Number of intervals} = \frac{n}{5} \text{ ie. } (Z_i \geq 5)$$

$$\text{Interval } p = \left(\frac{100}{n/5}\right)\%$$

$$\frac{\bar{X}-\mu_0}{S} = C_{p=0.1\ldots}$$

(32) Contingency tables

Test if two distributions are independent

Z_{ij}

		Category 1		
		1	2	Total
Category 2	1	Z_{11}	Z_{12}	$\sum Z_{1-}$
	2	Z_{21}	Z_{22}	$\sum Z_{2-}$
	Total	$\sum Z_{-1}$	$\sum Z_{-2}$	n

$$E_{11} = \frac{\sum Z_{1-}}{n} \times \sum Z_{-1}$$

E_{ij}

		Category 1		
		1	2	Total
Category 2	1	E_{11}	E_{12}	$\sum E_{1-}$
	2	E_{21}	E_{22}	$\sum E_{2-}$
	Total	$\sum E_{-1}$	$\sum E_{-2}$	n

$Z_{ij} - E_{ij}$

		Category 1		
		1	2	Total
Category 2	1	$Z_{11} - E_{11}$	$Z_{12} - E_{12}$	0
	2	$Z_{21} - E_{21}$	$Z_{22} - E_{22}$	0
	Total	0	0	0

$\frac{(Z_{ij}-E_{ij})^2}{E_{ij}}$

		Category 1		
		1	2	Total
Category 2	1	$\frac{(Z_{11}-E_{11})^2}{E_{11}}$	$\frac{(Z_{12}-E_{12})^2}{E_{12}}$	
	2	$\frac{(Z_{21}-E_{21})^2}{E_{21}}$	$\frac{(Z_{22}-E_{22})^2}{E_{22}}$	
	Total			$\sum \frac{(Z_{ij}-E_{ij})^2}{E_{ij}}$

$$H_0: p_{ij} = p_{i-}p_{-j} \text{ vs } H_1: p_{ij} \neq p_{i-}p_{-j}$$

Perform a one-tailed χ^2 test at α level of significance

$$P\left(\chi_{(r-1)(c-1)}^2 = \sum \frac{(Z_{ij}-E_{ij})^2}{E_{ij}}\right) \sim \chi_{(r-1)(c-1)}^2$$

Test if probability of A from X is the same as B from Y

$$\hat{p} = \frac{Z_{\text{A from X}} + Z_{\text{B from Y}}}{n}$$

$$E_{11} = \hat{p} Z_{-1}$$

$$H_0: p_{AX} = p_{BY} \text{ vs } H_1: p_{AX} \neq p_{BY}$$

Perform a one-tailed χ^2 test at α level of significance

$$P\left(\chi_{rc-r-c+1}^2 = \sum \frac{(Z_{ij}-E_{ij})^2}{E_{ij}}\right) \sim \chi_{(r-1)(c-1)}^2$$

(33) Introduction to linear regression analysis

Linear regression model

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$$

$$\hat{\beta}_1 = \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$\Rightarrow \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Normal linear regression model

$$E(y_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i \Rightarrow E(\bar{y}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$Var(y_i) = \sigma^2$$

$$E(\varepsilon_i) = 0$$

$$Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$$

$$E(\hat{\beta}_1) = \beta_1$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

$$SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum(x_i-\bar{x})^2}}$$

$$E(\hat{\beta}_0) = \beta_0$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} \times \frac{\sum x_i^2}{\sum(x_i-\bar{x})^2}$$

$$SE(\hat{\beta}_0) = \frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{\sum x_i^2}{\sum(x_i-\bar{x})^2}} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

(34) Linear regression: Statistical inference

(Confidence interval for β_0)

$$[\hat{\beta}_0 - t_{\alpha/2,n-2}SE(\hat{\beta}_0), \hat{\beta}_0 + t_{\alpha/2,n-2}SE(\hat{\beta}_0)]$$

(Confidence interval for β_1)

$$[\hat{\beta}_1 - t_{\alpha/2,n-2}SE(\hat{\beta}_1), \hat{\beta}_1 + t_{\alpha/2,n-2}SE(\hat{\beta}_1)]$$

Test if the slope of a regression model is non-zero

$$H_0: \beta_1 = k \text{ vs } H_1: \beta_1 \neq k$$

Perform a one-tailed T test at α level of significance

$$P\left(T = \frac{\hat{\beta}_1 - k}{SE(\hat{\beta}_1)}\right) \sim t_{n-2}$$

Residual analysis

$$\hat{\varepsilon}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

Plot $\hat{\varepsilon}_i$ against x_i should yield no discernible pattern

(35) Linear regression analysis with minitab

ANOVA

$$\text{Total SS} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$\text{Regression SS} = \hat{\beta}_1^2\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

$$\text{Residual SS} = \text{Total SS} - \text{Regression SS} = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{\text{Residual SS}}{n-2}$$

$$F = \frac{\text{Regression SS}}{\frac{\text{Residual SS}}{n-2}} = \frac{(n-2)\hat{\beta}_1^2[\sum_{i=1}^{n}(x_i - \bar{x})^2]}{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2} = \left(\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}\right)^2$$

F-test for slope of a regression model is non-zero

$$H_0: \beta_1 = 0 \text{ vs } H_1: \beta_1 \neq 0$$

Perform a one-tailed F test at α level of significance

$$P\left(F = \left(\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}\right)^2\right) \sim F_{1,n-2}$$

Regression correlation coefficient

$$R = \sqrt{\frac{\text{Regression SS}}{\text{Total SS}}} = \sqrt{1 - \frac{\text{Residual SS}}{\text{Total SS}}}$$

R^2represents % of data closest to the best fit line

$$R_{adj} = \sqrt{1 - \frac{\frac{\text{Residual SS}}{n-2}}{\frac{\text{Total SS}}{n-1}}}$$

(36) Prediction and diagnostics

Confidence interval for $\mu(x) = E(y)$

$$E(\hat{\mu}(x)) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x = \beta_0 + \beta_1 x$$

$$Var(\hat{\mu}(x)) = \frac{\sigma^2}{n} \times \frac{\sum(x_i - x)^2}{\sum(x_i - \bar{x})^2}$$

$$1 - \alpha \text{ interval: } \left(y_i \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}\right)$$

$$\text{Where } \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

The confidence interval only covers the true expectation $\hat{\mu}(x) = E(y)$. It does not span the entire y

Predictive interval for $y - \hat{\mu}(x)$

$$E(y - \hat{\mu}(x)) = y - (\beta_0 + \beta_1 x) = 0$$

$$Var(y - \hat{\mu}(x)) = \sigma^2 + \frac{\sigma^2}{n} \times \frac{\sum(x_i - x)^2}{\sum(x_i - \bar{x})^2}$$

$$1 - \alpha \text{ interval: } \left(y_i \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}\right)$$

Predictive interval is longer than the confidence interval, as it spans the entire y