Chapter 1 Summary

Numerical data can be discrete or continuous.

Categorical data can be dichotomous (attribute), nominal or ordinal.

Data can be presented either in tabular form (using a frequency table, a cumulative frequency table or a stem and leaf diagram) or in graphical form (using a lineplot, a dotplot, a boxplot, a bar chart or a histogram).

The *location* of a data set can be summarised using the mean, the median or the mode.

The *spread* of a data set can be summarised using the standard deviation, the range or the interquartile range.

The variance measures the spread squared.

Third moments can be used to summarise the *skewness* (*ie* the degree of asymmetry) of a data set.

Chapter 1 Formulae

Measures of Location

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{or} \quad \overline{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$M = \left(\tfrac{1}{2}n + \tfrac{1}{2}\right)^{th} \text{ value}$$

Measures of Spread

$$R = \max_i(x_i) - \min_i(x_i)$$

$$IQR = Q_3 - Q_1, \quad Q_3 = \left(\tfrac{3}{4}n + \tfrac{1}{2}\right)^{th} \text{ value} \quad Q_1 = \left(\tfrac{1}{4}n + \tfrac{1}{2}\right)^{th} \text{ value}$$

$$\textit{alternatively} \qquad Q_3 = \left(\tfrac{3}{4}n + \tfrac{3}{4}\right)^{th} \text{ value} \quad Q_1 = \left(\tfrac{1}{4}n + \tfrac{1}{4}\right)^{th} \text{ value}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^{\,2} - n\overline{x}^2\right]$$

$$\text{or} \quad \frac{1}{\sum f_i - 1}\sum_{i=1}^{n} f_i(x_i - \overline{x})^2 = \frac{1}{\sum f_i - 1}\left[\sum f_i x_i^{\,2} - n\overline{x}^2\right]$$

Measures of Skewness

$$\text{skewness} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3$$

$$\text{coeff of skew} = \frac{\text{skewness}}{s_n^3} \qquad \text{where} \quad s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Sample Moments

$$k\text{th moment} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$$k\text{th moment about } \alpha = \frac{1}{n}\sum_{i=1}^{n}(x_i - \alpha)^k$$

$$k\text{th central moment} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$$

Chapter 2 Summary

A *set* is a collection of objects, called *elements*. *A* is a *subset* of *B*, written $A \subset B$, if all the elements in *A* are contained in *B*. The *complement* of *A*, written A', is the set of all the elements *not* in *A*. The empty set is denoted $\varnothing$.

The *union* of *A* and *B*, written $A \cup B$, is the set of all elements in *A* or *B* or both. The *intersection* of *A* and *B*, written $A \cap B$, is the set of all elements in *A* and *B*.

Venn diagrams are used to represent sets and the relationships between them.

A *sample space*, *S*, is the set of all the possible outcomes from an experiment. An *event* is anything we might wish to observe from our experiment.

Probabilities are a numerical way of describing how likely an event is to happen. A formula for equally likely elements is given overleaf. Probabilities lie between 0 (impossible) and 1 (certain).

We can use the *addition rule* and the *multiplication rule* (see overleaf) to calculate probabilities. Tree diagrams are a helpful way of working out probabilities.

The conditional probabilities of *A* occurring given that *B* has already occurred is written $P(A \mid B)$. The formula is given overleaf.

Events *A* and *B* are *mutually exclusive* if $A \cap B = \varnothing$. Events *A* and *B* *independent* if $P(A \mid B) = P(A)$.

$E_1, \dots, E_n$ is a *partition* of *S* if the E_i's are mutually exclusive and together make up the whole set *S*.

Bayes' Theorem (see overleaf) allows us to 'turnaround' conditional probabilities, *ie* calculate $P(E_i \mid A)$ given only information about $P(A \mid E_i)$.

Chapter 2 Formulae

Probabilities

For equally likely elements:

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } S}$$

Addition rule

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

For mutually exclusive events $A \cap B = \varnothing$

$$P(A \cup B) = P(A) + P(B)$$

Multiplication rule

$$P(A \cap B) = P(A)P(B \mid A) \quad \text{or} \quad P(B)P(A \mid B)$$

For independent events $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$, so:

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

Conditional probabilities

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

Bayes' Theorem

For a partition E_i, $i = 1, 2, \dots, n$

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{j=1}^{n} P(E_j)P(A \mid E_j)}, \qquad i = 1, 2, 3, \dots, n$$

Chapter 3 Summary

Random variables are used to model features of a population using probabilities.

A discrete random variable has a probability function (PF), $P(X=x)$. This defines how the probability is split between the values the variable can take. The PF satisfies:

$$\sum_x P(X=x)=1 \qquad \text{and} \qquad P(X=x)\geq 0$$

A continuous random variable has a probability density function (PDF), $f_X(x)$. The PDF satisfies:

$$\int_x f_X(x)\,dx=1 \qquad \text{and} \qquad f_X(x)\geq 0$$

We can use the PDF to find probabilities as follows:

$$P(a<X<b)=\int_a^b f_X(x)\,dx\,.$$

The cumulative distribution function (CDF), for both discrete and continuous random variables is given by:

$$F_X(x)=P(X\leq x)$$

For continuous random variables $F_X'(x)=f_X(x)$.

Using formulae given overleaf, we can calculate the:

- mean μ
- variance σ^2
- skewness μ_3

and other central and non-central moments of a random variable.

Chapter 3 Formulae

Population Mean (Expectation)

$$\mu = E(X) = \sum_x xP(X = x) \quad \text{or} \quad \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$E[g(X)] = \sum_x g(x)P(X = x) \quad \text{or} \quad \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

Population Variance

$$\sigma^2 = \text{var}(X) = E[(X-\mu)^2] = E(X^2) - E^2(X)$$

Population Skewness

$$\mu_3 = skew(X) = E[(X-\mu)^3] = E[X^3] - 3\mu E[X^2] + 2\mu^3$$

$$\text{coefficient of skewness } = \frac{\mu_3}{\sigma^3}$$

Population Moments

kth moment $= E[X^k]$

kth moment about $c = E[(X-c)^k]$

kth central moment $= E[(X-\mu)^k]$

Population median and IQR

m such that $P(X < m) = 0.5$

$IQR = q_3 - q_1$ where $P(X < q_1) = 0.25$ and $P(X < q_3) = 0.75$

Linear Functions of X

$E[aX + b] = aE[X] + b$

$\text{var}[aX + b] = a^2 \,\text{var}[X]$

Functions of a Random Variable

$$Y = u(X) \quad \Rightarrow \quad f_Y(y) = f_X\left[u^{-1}(y)\right]\left|\frac{du^{-1}(y)}{dy}\right|$$

Chapter 4 Summary

Standard discrete distributions covered in this course are the discrete uniform, Bernoulli, binomial, geometric, negative binomial, hypergeometric and Poisson.

Waiting times between events in a $Poi(\lambda)$ distribution have a $Exp(\lambda)$ distribution.

Standard continuous distributions covered in this course are the continuous uniform, gamma, exponential, chi-square, normal, lognormal, beta, *t* and *F*.

The geometric and exponential distributions have the "memoryless" property:

$$P(X > x + n \mid X > n) = P(X > x)$$

The properties of the distributions are summarised overleaf.

The *t*-distribution with *k* degrees of freedom is defined as:

$$t_k \equiv \frac{N(0,1)}{\sqrt{\chi_k^2 / k}}$$

The *F*-distribution with *m*, *n* degrees of freedom is defined as:

$$F_{m,n} = \frac{\chi_m^2 / m}{\chi_n^2 / n}$$

The Poisson process counts events occurring up to and including time *t*:

$$N(t) \sim Poi(\lambda t)$$

It can be derived by considering events occurring in a small time interval *h*.

The waiting time between events in a Poisson process has an exponential distribution.

Random variables can be simulated by using the inverse transform method. First we take a random number, *u*, from $U(0,1)$ then we set:

continuous	$x = F^{-1}(u)$		
discrete	$x = x_j$	where	$F(x_{j-1}) < u \le F(x_j)$

	Distribution	PF or PDF	Mean	Variance
Discrete Distributions	Discrete uniform	$\frac{1}{k}$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$
	Bernouilli	$\theta^x(1-\theta)^{1-x}$	θ	$\theta(1-\theta)$
	Binomial	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$	$n\theta$	$n\theta(1-\theta)$
	Geometric	$\theta(1-\theta)^{x-1}$	$\frac{1}{\theta}$	$\frac{1-\theta}{\theta^2}$
	Negative binomial	$\binom{x-1}{k-1}\theta^k(1-\theta)^{x-k}$	$\frac{k}{\theta}$	$\frac{k(1-\theta)}{\theta^2}$
	Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Continuous Distributions	Continuous uniform	$\frac{1}{\beta-\alpha}$	$\frac{1}{2}(\alpha+\beta)$	$\frac{1}{12}(\beta-\alpha)^2$
	Gamma	$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
	Exponential	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
	Chi-square	$\frac{\left(\frac{1}{2}\right)^{\frac{\upsilon}{2}}}{\Gamma(\frac{\upsilon}{2})}x^{\frac{\upsilon}{2}-1}e^{-\frac{1}{2}x}$	υ	2υ
	Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
	Lognormal	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}$	$e^{\mu+\frac{1}{2}\sigma^2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
	Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2

Chapter 5 Summary

Generating functions are used to make finding moments of distributions easier.

The probability generating function (PGF) of a counting random variable is defined to be:

$$G_X(t) = E\left[t^X\right]$$

The series expansion for PGFs is:

$$G_X(t) = P(X = 0) + tP(X = 1) + t^2 P(X = 2) + t^3 P(X = 3) + \cdots$$

The moment generating function (MGF) of a random variable is defined to be:

$$M_X(t) = E\left[e^{tX}\right]$$

The series expansion for MGFs:

$$M_X(t) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

The cumulant generating function (CGF) of a random variable is defined to be:

$$C_X(t) = \ln M_X(t)$$

Moments of a random variable can be found from its PGF, MGF or CGF using the formulae listed overleaf.

The *uniqueness property* means that if two variables have the same PGF, MGF or CGF then they have the same distribution.

If $Y = a + bX$, then:

$$G_Y(t) = t^a G_X(t^b),\ M_Y(t) = e^{at} M_X(bt) \text{ and } C_Y(t) = at + C_X(bt)$$

Chapter 5 Formulae

Probability Generating Functions

$$G_X(t) = E\left(t^X\right) = \sum_x t^x P(X = x)$$

$$E(X) = G'_X(1)$$

$$\text{var}(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2$$

$$G_X(t) = P(X = 0) + tP(X = 1) + t^2 P(X = 2) + t^3 P(X = 3) + \cdots$$

Moment Generating Functions

$$M_X(t) = E\left(e^{tX}\right) = \sum_x e^{tx} P(X = x) \qquad \text{or} \qquad \int_x e^{tx} f(x)\, dx$$

$$E(X) = M'_X(0)$$

$$\text{var}(X) = M''_X(0) - \left(M'_X(0)\right)^2$$

$$M_X(t) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

Cumulant Generating Functions

$$C_X(t) = \ln M_X(t)$$

$$E(X) = C'_X(0)$$

$$\text{var}(X) = C''_X(0)$$

$$skew(X) = C'''_X(0)$$

Linear Transformations

$$Y = aX + b \quad \Rightarrow \quad M_Y(t) = e^{bt} M_X(at)$$

Chapter 6 Summary

Two discrete random variables X and Y have joint probability function (PF), $P(X=x,Y=y)$. This defines how the probability is split between the different combinations of the variables. The joint PF satisfies:

$$\sum_x \sum_y P(X=x,Y=y)=1 \qquad \text{and} \qquad P(X=x,Y=y)\geq 0$$

Two continuous random variables X and Y have joint probability density function (PDF), $f_{X,Y}(x,y)$. The joint PDF satisfies:

$$\int_x \int_y f_{X,Y}(x,y)\,dx\,dy=1 \qquad \text{and} \qquad f_{X,Y}(x,y)\geq 0$$

We can use the joint PDF to find probabilities as follows:

$$P(x_1<X<x_2, y_1<Y<y_2)=\int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x,y)\,dx\,dy$$

The joint distribution function, for both discrete and continuous random variables is given by:

$$F(x,y)=P(X<x,Y<y)$$

For continuous random variables $f(x,y)=\dfrac{\partial^2}{\partial x\partial y}F(x,y)$.

Using the formulae overleaf, we can calculate the:

- Marginal distributions, *eg* $P(X=x)$ or $f_X(x)$
- Conditional distributions, *eg* $P(X=x\mid Y=y)$ or $f_{X\mid Y=y}(x\mid y)$
- Expectation of any function, $E[g(X,Y)]$
- Covariance, $\text{cov}(X,Y)$
- Correlation coefficient, $\rho(X,Y)=\text{corr}(X,Y)$

The random variables X and Y are uncorrelated if and only if:

$$\text{corr}(X,Y)=0 \quad \Leftrightarrow \quad \text{cov}(X,Y)=0 \quad \Leftrightarrow \quad E(XY)=E(X)E(Y)$$

The random variables X and Y are independent if, and only if:

$$P(X=x,Y=y)=P(X=x)P(Y=y)$$

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

Independent random variables are always uncorrelated. Uncorrelated random variables are not necessarily independent.

There are rules connecting sums and products of expectations and sums of variances.

The convolution of the marginal probability (density) functions of X and Y is the probability (density) function of $Z=X+Y$. $P(Z=z)$ or $f_Z(z)$ is given using the formulae on the formulae sheet. The convolution is written $f_Z = f_X * f_Y$.

Sums of independent random variables make other random variables. The full list is given on the formulae sheet.

Chapter 6 Formulae

Marginal probability (density) function

$$P(X=x)=\sum_y P(X=x,Y=y) \qquad f_X(x)=\int_y f_{X,Y}(x,y)\,dy$$

Conditional probability (density) function

$$P(X=x \mid Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)} \qquad f_{X|Y=y}(x,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Expectation

$$E[g(X,Y)]=\sum_x\sum_y g(x,y)P(X=x,Y=y) \quad \text{or} \quad \int_y\int_x g(x,y)f_{X,Y}(x,y)\,dxdy$$

Covariance

$$\text{cov}(X,Y)=E\Big[\big(X-E(X)\big)\big(Y-E(Y)\big)\Big]=E(XY)-E(X)E(Y)$$

Correlation

$$\text{corr}(X,Y)=\rho(X,Y)=\frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}}$$

Sums and products of moments

$$\begin{aligned} E(X+Y)&=E(X)+E(Y) \\ E(XY)&=E(X)E(Y)+Cov(X,Y) \\ &=E(X)E(Y) \qquad \text{if } X,Y \text{ independent} \end{aligned}$$

The above are also true for functions $g(X)$ and $h(Y)$ of the random variables.

$$\begin{aligned} \text{var}(X+Y)&=\text{var}(X)+\text{var}(Y)+2\text{cov}(X,Y) \\ &=\text{var}(X)+\text{var}(Y) \qquad \text{if } X,Y \text{ independent} \end{aligned}$$

Convolutions

$$f_Z = f_X * f_Y = \sum_x P(X=x)P(Y=z-x) \qquad \text{or} \qquad \int_x f_X(x) f_Y(z-x)\,dx$$

Linear Combinations

For independent random variables $X_1, \dots, X_n$:

$$E\left(c_1X_1 + \cdots + c_nX_n\right) = c_1E\left(X_1\right) + \cdots + c_nE\left(X_n\right)$$
$$\text{var}\left(c_1X_1 + \cdots + c_nX_n\right) = c_1^2\,\text{var}\left(X_1\right) + \cdots + c_2^2\,\text{var}\left(X_n\right)$$

Linear Combinations of PGFs and MGFs

For independent random variables $X_1, \dots, X_n$:

$$\begin{aligned} Y = X_1 + \cdots + X_n \quad \Rightarrow \quad G_Y(s) &= G_{X_1}(s) \dots G_{X_n}(s) \\ &= \left[G_X(s)\right]^n \qquad X_i\text{'s identical} \\ \Rightarrow \quad M_Y(t) &= M_{X_1}(t) \dots M_{X_n}(t) \\ &= \left[M_X(t)\right]^n \qquad X_i\text{'s identical} \end{aligned}$$

Linear Combinations of random variables

For independent distributions:

"$Bernoulli(p) + \cdots + Bernoulli(p) \sim Bin(n,p)$"
"$Bin(n,\theta) + Bin(m,\theta) \sim Bin(n+m,\theta)$"

"$Geo(p) + \cdots + Geo(p) \sim NBin(k,p)$"
"$NBin(k,\theta) + NBin(m,\theta) \sim NBin(k+m,\theta)$"

"$Exp(\lambda) + \cdots + Exp(\lambda) \sim Ga(\alpha,\lambda)$"
"$Ga(\alpha,\lambda) + Ga(\delta,\lambda) \sim Ga(\alpha+\delta,\lambda)$"

"$\chi^2_m + \chi^2_n \sim \chi^2_{m+n}$"

"$Poi(\lambda) + Poi(\quad) \sim Poi(\lambda + \quad)$"

"$N(\ _1,\sigma_1^2) \pm N(\ _2,\sigma_2^2) \sim N(\ _1 \pm \ _2, \sigma_1^2 + \sigma_2^2)$"

Please note that some of the notation used for the linear combinations of random variables is non-standard and is used simply to convey the results in a concise format.

Chapter 7 Summary

$E(Y \mid X)$ is the mean of the conditional distribution of Y given X (which was defined in Chapter 6). The formulae for discrete and continuous distributions are given overleaf.

$\text{var}(Y \mid X)$ is the variance of the conditional distribution of Y given X, it is given by:

$$\text{var}(Y \mid X) = E(Y^2 \mid X) - E^2(Y \mid X)$$

The unconditional mean and variance can be found from the conditional mean and variance using the formulae given overleaf and on page 16 of the *Tables*.

A quantity that is the sum of a random number of random quantities has a compound distribution:

$$S = X_1 + \cdots + X_N$$

We can find the mean, variance and MGF of a compound distribution using the formulae given overleaf.

We can find the skewness using the CGF.

Chapter 7 Formulae

Conditional Expectation

$$E[Y \mid X = x] = \sum_i y_i P[Y = y_i \mid X = x] = \sum_i y_i \frac{P[Y = y_i, X = x]}{P[X = x]}$$

$$E[Y \mid X = x] = \int_y y\, f(y \mid x)\, dy = \int_y y \frac{f(x, y)}{f(x)}\, dy$$

Conditional Variance

$$\text{var}[Y \mid X = x] = E[Y^2 \mid X = x] - E^2[Y \mid X = x]$$

Relationships between unconditional and conditional moments

$$E[Y] = E[E(Y \mid X)]$$

$$\text{var}[Y] = E[\text{Var}(Y \mid X)] + \text{var}[E(Y \mid X)]$$

Compound Distributions

$$E(S) = E(N)E(X)$$

$$\text{var}(S) = E(N)\,\text{Var}(X) + \text{var}(N)E^2(X)$$

$$M_S(t) = M_N\{\log M_X(t)\}$$

Chapter 8 Formulae

The Central Limit Theorem

For $X_1, \dots, X_n$ iid RV's with mean μ and variance σ^2:

$$\sum X_i \div N\left(n\mu, n\sigma^2\right) \quad \Rightarrow \quad \frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} \div N(0,1) \quad \text{as } n \to \infty$$

$$\bar{X} \div N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \div N(0,1) \qquad \text{as } n \to \infty$$

Normal Approximations

$$\left.\begin{array}{ll} Bin(n,p) \div N(np, npq) & np > 5,\ nq > 5 \\ Poi(\lambda) \div N(\lambda, \lambda) & \lambda \text{ large} \end{array}\right\} \text{with continuity correction}$$

$$Ga(\alpha, \lambda) \div N\left(\tfrac{\alpha}{\lambda}, \tfrac{\alpha}{\lambda^2}\right) \qquad \alpha \text{ large}$$

$$\chi_k^2 \div N(k, 2k) \qquad k \text{ large}$$

Chapter 9 Formulae

Moments of Statistics

$$E(\bar{X}) = \qquad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \qquad E(S^2) = \sigma^2$$ any distribution

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$ normal distribution only

t-distribution

$$t_k \equiv \frac{N(0,1)}{\sqrt{\chi_k^2 / k}}$$

F-distribution

$$F_{m,n} \equiv \frac{\chi_m^2 / m}{\chi_n^2 / n}$$

Sampling Distributions

$$\bar{X} \doteq N\left(\;\; , \frac{\sigma^2}{n} \right)$$ any distribution large n (or normal any n)

$$\Rightarrow \quad \frac{\bar{X} - }{\sigma / \sqrt{n}} \sim N(0,1)$$ σ^2 known

$$\Rightarrow \quad \frac{\bar{X} - }{S / \sqrt{n}} \sim t_{n-1}$$ σ^2 unknown

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$ normal distribution only

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{n_1 - 1, n_2 - 1}$$ normal distributions only

Chapter 10 Summary

We have covered two methods here for estimating parameters.

The method of moments technique equates the population moments to the sample moments using the formulae detailed overleaf.

The method of maximum likelihood:

- find the likelihood $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$
- $\log L$
- find θ that solves $\dfrac{\partial}{\partial\theta}\ln L(\theta) = 0$
- check for maximum $\dfrac{\partial^2}{\partial\theta^2}\ln L(\theta) < 0$.

If the range of the distribution is a function of the parameter the maximum must be found from first principles.

Three properties of estimators are bias, mean square error (MSE) and consistency:

The bias of an estimator is given by $E[g(\underline{X})]-\theta$ where $g(\underline{X})$ is the estimator.

$g(\underline{X})$ is an unbiased estimator of θ if $E[g(\underline{X})]=\theta$.

The mean square error of an estimator is given by $E[(g(\underline{X})-\theta)^2]$ where $g(\underline{X})$ is the estimator. An easier formula is $\text{var}[g(\underline{X})]+bias^2[g(\underline{X})]$. An estimator is consistent if $MSE \to 0$ as $n \to \infty$, where n is the size of the sample.

A good estimator has a small MSE, is unbiased and consistent.

The Cramér-Rao lower bound gives a lower bound for the variance of an unbiased estimator. It can be found using the formulae overleaf. It can be used to obtain confidence intervals.

The value of the CRLB depends on the parameter you are estimating. To use this formula, the likelihood must be expressed in terms of the correct parameter.

Chapter 10 Formulae

Method of Moments

1 parameter $E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$

2 parameters $E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ or $\text{var}(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

alternatively $E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{var}(X) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Method of Maximum Likelihood

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

θ that solves $\frac{\partial}{\partial\theta}\ln L(\theta) = 0$

Bias

$$bias[g(\underline{X})] = E[g(\underline{X})] - \theta$$

Mean Square Error

$$MSE[g(\underline{X})] = E[(g(\underline{X}) - \theta)^2] = \text{var}[g(\underline{X})] + bias^2[g(\underline{X})]$$

Cramér-Rao Lower Bound

$$CRLB(\theta) = -\frac{1}{E\left[\frac{\partial^2}{\partial\theta^2}\ln L(\theta, \underline{X})\right]}$$

Asymptotic Distribution of MLE

$$\hat{\theta} \doteq N(\theta, CRLB)$$

Chapter 11 Formulae

One sample normal distribution

$$\frac{\bar{X} - \ _0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \sigma^2 \text{ known} \qquad\qquad \frac{\bar{X} - \ _0}{s/\sqrt{n}} \sim t_{n-1} \quad \sigma^2 \text{ unknown}$$

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Two sample normal distribution

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\ _1 - \ _2\right)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1) \qquad\qquad \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\ _1 - \ _2\right)}{s_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$$

σ^2 known $\qquad\qquad$ σ^2 unknown

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

One sample binomial

$$\frac{\hat{p} - p}{\sqrt{\hat{p}\hat{q}/n}} \div N(0,1) \quad \text{or} \quad \frac{X - np}{\sqrt{n\hat{p}\hat{q}}} \div N(0,1)$$

Two sample binomial

$$\frac{\left(\hat{p}_1 - \hat{p}_2\right) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}} \div N(0,1) \qquad \text{where} \quad \hat{p}_1 = \frac{x_1}{n_1},\ \hat{p}_2 = \frac{x_2}{n_2}$$

One sample Poisson

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} \div N(0,1) \quad \text{or} \quad \frac{\sum X - n\lambda}{\sqrt{n\hat{\lambda}}} \div N(0,1)$$

Two sample Poisson

$$\frac{(\hat{\lambda}_1 - \hat{\lambda}_2) - (\lambda_1 - \lambda_2)}{\sqrt{\dfrac{\hat{\lambda}_1}{n_1} + \dfrac{\hat{\lambda}_2}{n_2}}} \div N(0,1) \qquad \text{where} \quad \hat{\lambda}_1 = \bar{X}_1,\ \hat{\lambda}_2 = \bar{X}_2$$

Chapter 12 Summary

Statistical tests can be used to test assertions about populations.

The process of statistical testing involves setting up a null hypothesis and an alternative hypothesis, calculating a test statistic and using this to determine a *p*-value.

The probability of a Type I error is the probability of rejecting H_0 when it is true. This is also called the size (or level) of the test. The probability of a Type II error is the probability of accepting H_0 when it is false. The power of a test is the probability of rejecting H_0 when it is false.

The "best" test can be found using the likelihood ratio criterion. This leads to the tests detailed on the formulae summary sheet.

The test for two normal means (unknown variances) requires that the variances are the same and uses the pooled sample variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

χ^2 tests can be carried out to test for goodness of fit or to test whether two factors are independent (using contingency tables).

The statistic is $\sum \frac{(O_i - E_i)^2}{E_i}$.

To find the number of degrees of freedom for the goodness of fit test, take the number of cells, subtract 1 if the total of the observed figures has been used in the calculation of the expected numbers (which is usually the case), and then subtract the number of parameters estimated.

To find the number of degrees of freedom for a contingency table calculate $(r-1)(c-1)$. If the expected numbers in some cells are small, these should be grouped. One degree of freedom is lost for each cell that is "lost".

Chapter 12 Formulae

One sample normal distribution

$$\frac{\bar{X} - {}_{0}}{\sigma/\sqrt{n}} \sim N(0,1) \quad \sigma^2 \text{ known} \qquad \frac{\bar{X} - {}_{0}}{s/\sqrt{n}} \sim t_{n-1} \quad \sigma^2 \text{ unknown}$$

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$$

Two sample normal distribution

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left({}_{1} - {}_{2}\right)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1) \qquad \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left({}_{1} - {}_{2}\right)}{s_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$$

σ^2 known $\qquad$ σ^2 unknown

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

One sample binomial

$$\frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}} \doteq N(0,1) \quad \text{or} \quad \frac{X - np_0}{\sqrt{np_0 q_0}} \doteq N(0,1) \qquad \text{with continuity correction}$$

Two sample binomial

$$\frac{\left(\hat{p}_1 - \hat{p}_2\right) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}\hat{q}}{n_1} + \dfrac{\hat{p}\hat{q}}{n_2}}} \doteq N(0,1) \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ is the overall sample proportion}$$

One sample Poisson

$$\frac{\bar{X} - \lambda_0}{\sqrt{\lambda_0/n}} \doteq N(0,1) \quad \text{or} \quad \frac{\sum X - n\lambda_0}{\sqrt{n\lambda_0}} \doteq N(0,1) \qquad \text{with continuity correction}$$

Two sample Poisson

$$\frac{(\hat{\lambda}_1 - \hat{\lambda}_2) - (\lambda_1 - \lambda_2)}{\sqrt{\dfrac{\hat{\lambda}}{n_1} + \dfrac{\hat{\lambda}}{n_2}}} \doteq N(0,1) \qquad \hat{\lambda} = \frac{n_1\hat{\lambda}_1 + n_2\hat{\lambda}_2}{n_1 + n_2} \text{ is the overall sample mean}$$

Chapter 13 Summary

A regression model, such as the simple linear regression model, can be used to model the response when an explanatory variable operates at a given level, or to model bivariate data points.

The sample correlation coefficient, *r*, measures the strength of the linear relationship between *x* and *y*. The formula is given overleaf and on page 25 of the *Tables*.

We can carry out hypothesis tests on the population correlation coefficient, ρ, using the *t* result or the Fisher-Z test. Both of these results are given overleaf and on page 25 of the *Tables*.

The linear regression model is given by:

$$Y_i = \alpha + \beta x_i + e_i \qquad \text{where } e_i \sim N(0,\sigma^2)$$

The parameters α, β and σ^2 can be estimated using the formulae overleaf and on page 24 of the *Tables*.

Confidence intervals can be obtained for β and the predicted individual (or mean) response *y* using the formulae given overleaf and on pages 24 and 25 of the *Tables*.

The fit of the linear regression model can be analysed by:

Partitioning the total variance, SS_{TOT}, into that which is explained by the model, SS_{REG}, and that which is not, SS_{RES}. The coefficient of determination, R^2, gives the percentage of this variance which *is* explained by the model. The formula is given overleaf but not in the *Tables*.

Examining the residuals, $\hat{e}_i = y_i - \hat{y}_i$. We would expect them to be normally distributed about zero and to have no relationship with the *x* values. Both of these can be examined using diagrams.

Chapter 13 Formulae

Correlation

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$H_0 : \rho = 0 \qquad \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

$$\text{otherwise} \qquad \frac{1}{2}\ln\frac{1+r}{1-r} \sim N\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho}, \frac{1}{n-3}\right)$$

Regression

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \text{ and } \hat{\alpha} = \overline{y} - \hat{\beta}\,\overline{x}$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum (y_i - \hat{y}_i)^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$$

$$\frac{\hat{\mu}_0 - \mu_0}{\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\hat{\sigma}^2}} \sim t_{n-2} \quad \text{and} \quad \frac{\hat{y}_0 - y_0}{\sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\hat{\sigma}^2}} \sim t_{n-2}$$

Fit of Model

$$SS_{TOT} = SS_{RES} + SS_{REG}$$

$$R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

$$\hat{e}_i = y_i - \hat{y}_i$$

Chapter 14 Summary

Analysis of variance (ANOVA) is testing for a difference between treatment means. The model is:

$$Y_{ij} = \mu + \tau_i + e_{ij}$$

where e_{ij}'s are independent and identically distributed $N(0, \sigma^2)$.

The parameters μ (total mean), τ_i (treatment effect) and σ^2 can be estimated using the formulae overleaf. The variance estimate is given on page 26 of the *Tables*.

The total variance (SS_T) is split into variance within each treatment (SS_R) and variance between treatment means (SS_B):

$$SS_T = SS_R + SS_B$$

These can be calculated using the formulae given overleaf and on page 26 of the *Tables*.

The significance of the variance between treatment means (SS_B) is established using the *F* test given overleaf and on page 26 of the *Tables*.

Residuals can be plotted to check the adequacy of the model – *ie* the normality and equality of variances assumptions.

Confidence intervals for means should use the results of Chapter 11, but with the variance estimate $\hat{\sigma}^2$ instead.

If it is found under ANOVA that the treatment means are *not* the same, we can analyse them further using the least significant difference approach. This is essentially testing each of the pairs of treatments to see if they have the same mean or not. The results of this can be shown on a diagram like the one below:

$$\underline{\bar{y}_{2\bullet} < \bar{y}_{3\bullet}} < \bar{y}_{1\bullet} < \bar{y}_{4\bullet}$$

Chapter 14 Formulae

Parameter Estimates

$$\hat{\mu} = \bar{y}_{\bullet\bullet} = \frac{1}{n}\sum_i \sum_j y_{ij}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet} = \frac{1}{n_i}\sum_j y_{ij} - \frac{1}{n}\sum_i \sum_j y_{ij}$$

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

Sum of Squares

$$SS_T = \sum\sum y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{n}$$

$$SS_B = \sum \frac{y_{i\bullet}^2}{n_i} - \frac{y_{\bullet\bullet}^2}{n}$$

$$SS_R = SS_T - SS_B$$

Statistical Test

$$\left.\frac{SS_B}{k-1}\right/ \frac{SS_R}{n-k} \sim F_{k-1,n-k}$$