![mathcentre]

For the help you need to support your course

Guide to Statistics:

Probability & Statistics Facts, Formulae and Information

mathcentre is a project offering students and staff free resources to support the transition from school mathematics to university mathematics in a range of disciplines.

www.mathcentre.ac.uk

This leaflet has been produced in conjunction with and is distributed by the Higher Education Academy Maths, Stats & OR Network.

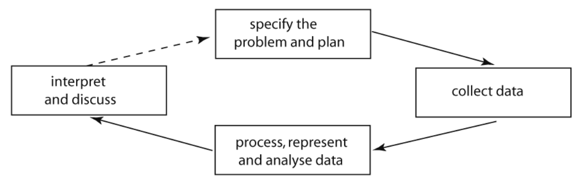
For more copies contact the Network at info@mathstore.ac.uk

sigma centre for excellence in mathematics & statistics support

The Higher Education Academy Maths, Stats & OR Network

The statistical problem solving cycle

Data are numbers in context and the goal of statistics is to get information from those data, usually through *problem solving*. A procedure or paradigm for statistical problem solving and scientific enquiry is illustrated in the diagram. The dotted line means that, following discussion, the problem may need to be re-formulated and at least one more iteration completed.



Descriptive statistics

Given a sample of n observations, $x_1, x_2, \ldots, x_n$, we define the **sample mean** to be

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

and the *corrected* sum of squares by

$$S_{xx} = \sum (x_i - \bar{x})^2 \equiv \sum x_i^2 - n\bar{x}^2 \equiv \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$\frac{S_{xx}}{n}$ is sometimes called the *mean squared deviation*. An **unbiased estimator** of the population variance, σ^2, is $s^2 = \frac{S_{xx}}{(n-1)}$. The **sample standard deviation** is s. In calculating s^2, the divisor $(n-1)$ is called the **degrees of freedom (df)**. Note that s is also sometimes written $\hat{\sigma}$.

If the sample data are ordered from smallest to largest then the:

minimum (Min) is the smallest value;
lower quartile (LQ) is the $\frac{1}{4}(n+1)$-th value;
median (Med) is the middle [or the $\frac{1}{2}(n+1)$ -th] value;
upper quartile (UQ) is the $\frac{3}{4}(n+1)$-th value;
maximum (Max) is the largest value.

These five values constitute a **five-number summary** of the data. They can be represented diagrammatically by a *box-and-whisker plot*, commonly called a *boxplot*.

Grouped Frequency Data

If the data are given in the form of a grouped frequency distribution where we have f_i observations in an interval whose mid-point is x_i then, if $\sum f_i = n$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{n} \quad \text{and}$$

$$S_{xx} = \sum f_i (x_i - \bar{x})^2 = \sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{n}.$$

Events & probabilities

The *intersection* of two events A and B is $A \cap B$. The union of A and B is $A \cup B$. A and B are **mutually exclusive** if they cannot both occur, denoted $A \cap B = \emptyset$ where $\emptyset$ is called the **null event**. For an event A, $0 \le P(A) \le 1$. For two events A and B

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

If A and B are mutually exclusive then

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

Equally likely outcomes

If a complete set of n elementary outcomes are all equally likely to occur, then the probability of each elementary outcome is $\frac{1}{n}$. If an event A consists of m of these n elements, then $P(A) = \frac{m}{n}$.

Independent events

A, B are *independent* if and only if $P(A \cap B) = P(A)P(B)$.

Conditional Probability of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if} \quad P(B) \neq 0.$$

Bayes' Theorem: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$.

Theorem of Total Probability

The k events $B_1, B_2, \ldots B_k$ form a *partition* of the sample space S if $B_1 \cup B_2 \cup B_3 \ldots \cup B_k = S$ and no two of the B_i's can occur together. Then $P(A) = \sum_i P(A|B_i)P(B_i)$. In this case Bayes' Theorem generalizes to

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)} \qquad (i = 1, 2, \ldots k)$$

If B' is the *complement* of the event B , $P(B') = 1 - P(B)$ and $P(A) = P(A|B)P(B) + P(A|B')P(B')$ is a special case of the theorem of total probability. The complement of the event B is commonly denoted $\overline{B}$.

43799 D&PS Jan11

A **hypothesis test** involves testing a claim, or **null hypothesis** H_0, about a parameter against an alternative, H_1. A decision to **reject** H_0 or **not reject** H_0 uses sample evidence to *calculate* a **test statistic** which is judged against a **critical value**. H_0 is maintained unless it is made untenable by sample evidence. Rejecting H_0 when we should not is a **Type I error**. The probability (we are prepared to accept) of making a Type I error is called the **significance level** α and yields the critical value. The *smallest* α at which we can just reject H_0 is the **p-value** of the test. Not rejecting H_0 when we should is a **Type II error**, with probability β. The **power** of a hypothesis test is $1-\beta$. An **interval estimate** for a parameter is a *calculated* range within which it is deemed likely to fall. Given α, the set of intervals from infinitely repeated random samples of size n will contain the parameter $(100-\alpha)\%$ of the time: each interval is a $(100-\alpha)\%$ **confidence interval**.

One sample hypothesis tests

1. For $X \sim N(\mu, \sigma^2)$, σ^2 *known; random sample evidence* $\bar{x}$ *and* n. Null hypothesis, $H_0 : \mu = \mu_0$; 2-sided alternative $H_1 : \mu \neq \mu_0$. Test statistic $z_{\text{calc}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$. Reject H_0 (at the α level) if $|z_{\text{calc}}| \ge z_{\alpha/2}$, the critical value of z.

2. For $X \sim N(\mu, \sigma^2)$, σ^2 *unknown; random sample evidence* $\bar{x}$, s *and* n. Null hypothesis, $H_0 : \mu = \mu_0$; 2-sided alternative $H_1 : \mu \neq \mu_0$. Test statistic $t_{\text{calc}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{(n-1)}$, the t distribution with $(n-1)$ df. For $n > 30$ and if X has *any* distribution, $t \sim N(0,1)$. Reject H_0 if $|t_{\text{calc}}| \ge t_{\alpha/2}$, the critical value of t with $(n-1)$ df.

3. For $X \sim N(\mu, \sigma^2)$, σ^2 *unknown; random sample evidence* s *and* n. Null hypothesis, $H_0 : \sigma^2 = \sigma_0^2$; alternative $H_1 : \sigma^2 > \sigma_0^2$. Test statistic $\chi^2_{\text{calc}} = (n-1)s^2/\sigma_0^2 \sim \chi^2_{n-1}$. Reject H_0 if $\chi^2_{\text{calc}} > \chi^2_\alpha$, the critical value of χ^2 with $(n-1)$ df.

In each case the p-value is the tail area outside the calculated statistic.

Two sample hypothesis tests

For $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, σ_1^2, σ_2^2 *unknown; random sample evidence* $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2, n_1 *and* n_2.

1. Null hypothesis, $H_0 = \mu_1 - \mu_2 = c$; 2-sided alternative $H_1 : \mu_1 - \mu_2 \neq c$. Test statistic $t_{\text{calc}} = \frac{(\bar{x}_1 - \bar{x}_2 - c)}{s\sqrt{1/n_1 + 1/n_2}} \sim t_{(n_1+n_2-2)}$, and $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}$, assuming $\sigma_1^2 = \sigma_2^2$. Reject H_0 if $|t_{\text{calc}}| \ge t_{\alpha/2}$ the critical value of t with (n_1+n_2-2) df.

2. Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$; alternative $H_1 : \sigma_1^2 > \sigma_2^2$. Test statistic $F_{\text{calc}} = \frac{(n_1-1)s_1^2}{(n_2-1)s_2^2} \sim F_{n_1-1, n_2-1}$. Reject H_0 if $F_{\text{calc}} > F_\alpha$ the critical value of F with $n_1 - 1$ and $n_2 - 1$ df.

Confidence interval for a population mean - σ^2 unknown

If X has mean μ and variance σ^2, with $n > 30$ an approximate $100(1-\alpha)\%$ confidence interval for μ is $\bar{x} - \frac{t_{\alpha/2}s}{\sqrt{n}}$ to $\bar{x} + \frac{t_{\alpha/2}s}{\sqrt{n}}$.

If $X \sim N(\mu, \sigma^2)$ the interval is exact for all n.

Standard statistical distributions

Name/parameters	Conditions/application	pdf/pmf	Mean	Variance	mgf	Notes
Binomial $\text{Bin}(n,p)$ Positive integer n Probability p, $0 \le p \le 1$	n independent success/fail trials each with probability p of success. X = number of successes.	$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	np	$np(1-p)$	$(1-p+pe^t)^n$	$X \sim \text{Bin}\ (n,p)$ $\Rightarrow n - X \sim \text{Bin}(n, 1-p)$
Geometric $\text{Geom}(p)$ Probability p, $0 \le p \le 1$	Repeated independent success/fail trials each with probability p of success. X = number of trials up to and including the first success.	$P(X=x) = (1-p)^{x-1}p$ $x = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$	Has the "lack of memory" property $P(X > a+b \mid X > b) = P(X > a)$
Poisson $\text{Po}(\lambda)$ λ a positive number	Events occur randomly at a constant rate. X = number of occurrences in some interval. λ is the expected number of occurrences	$P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	λ	λ	$\exp(\lambda(e^t - 1))$	Useful as approximation to $\text{Bin}(n,p)$ if n is large and p is small
Normal $N(\mu, \sigma^2)$ μ, σ both real; $\sigma > 0$	A widely used distribution for symmetrically distributed random variables with mean μ and standard deviation σ.	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ all real x	μ	σ^2	$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$	Can approximate Binomial, Poisson Pascal and Gamma distributions (see Central Limit Theorem)
Exponential $\text{Expon}(\theta)$	Events are occurring at rate θ per unit time. X = time to first occurrence.	$f(x) = \theta\exp(-\theta x)$ $x > 0$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$	$\frac{\theta}{\theta - t}$, $t < \theta$	Has the "lack of memory" property $P(X > a+b \mid X > b) = P(X > a)$
Negative-binomial or Pascal $\text{Pasc}(r,p)$ Positive integer n Probability p, $0 \le p \le 1$	Repeated independent success/fail trials each with probability p of success. X = number of trials up to and including the r-th success.	$P(X=x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, r+2, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$	$\text{Pasc}(1,p) \equiv \text{Geom}(p)$
Gamma $\text{Ga}(\alpha, \beta)$ $\alpha, \beta > 0$	Generalization of the exponential distribution; if α is an integer it represents the waiting time to the α-th occurrence of a random event where β is the expected number of events.	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ $x > 0$	$\frac{\alpha}{\beta}$ $\alpha > 1$	$\frac{\alpha}{\beta^2}$	$\left(\frac{\beta}{\beta - t}\right)^\alpha$, $t < \beta$	$\text{Ga}(1,\lambda) \equiv \text{Expon}(\lambda)$ If ν is an integer, $\text{Ga}(\nu/2, 2)$ is χ^2_ν, the Chi-squared distribution with ν df.

Permutations and combinations

The number of ways of selecting r objects out of a total of n, where the order of selection is important, is the number of **permutations**: ${}^nP_r = \dfrac{n!}{(n-r)!}$. The number of ways in which r objects can be selected from n when the order of selection is not important is the number of **combinations**: ${}^nC_r = \binom{n}{r} = \dfrac{n!}{r!(n-r)!}$. ${}^nC_n$ must equal 1, so $0! = 1$ and ${}^nC_0 = 1$; ${}^nC_r = {}^nC_{n-r}$. Also

$${}^nC_0 + {}^nC_1 + \ldots {}^nC_{n-1} + {}^nC_n = 2^n$$

$${}^{n+1}C_r = {}^nC_r + {}^nC_{r-1}$$

Random variables

Data arise from observations on variables that are **measured** on different **scales**. A *nominal* scale is used for named categories (e.g. race, gender) and an *ordinal* scale for data that can be ranked (e.g. attitudes, position) - no arithmetic operations are valid with either. *Interval* scale measurements can be added and subtracted only (e.g. temperature), but with *ratio* scale measurements (e.g. age, weight) multiplication and division can be used meaningfully as well. Generally, random variables are either *discrete* or *continuous*. Note: in reality, all data are discrete because the accuracy of measuring is always limited.

A **discrete** random variable X can take one of the values $x_1, x_2, \ldots,$, the probabilities $p_i = P(X = x_i)$ must satisfy $p_i \geq 0$ and $p_1 + p_2 + \ldots = 1$. The pairs (x_i, p_i) form the **probability mass function** (pmf) of X.

A **continuous** random variable X takes values x from a continuous set of possible values. It has a **probability density function** (**pdf**) $f(x)$ that satisfies $f(x) \geq 0$ and $\int f(x)\mathrm{d}x = 1$, with $P(a < x \leq b) = \int_a^b f(x)\mathrm{d}x$.

Expected values

The expected value of a function $H(X)$ of a random variable X is defined as

$$E[H(X)] = \begin{cases} \sum H(x_i)P(X = x_i), & X \text{ discrete.} \\ \int H(x)f(x)\mathrm{d}x, & X \text{ continuous.} \end{cases}$$

Expectation is linear in that the expectation of a linear combination of functions is the same linear combination of expectations. For example,

$$E[X^2 + \log X + 1] = E[X^2] + E[\log X] + 1$$

but

$$E[\log X] \neq \log E[X] \text{ and } E[1/X] \neq 1/E[X]$$

Variance

The variance of a random variable is defined as

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] \equiv \mathrm{E}[X^2] - \mu^2$$

Properties:

$\mathrm{Var}(X) \geq 0$ and is equal to 0 only if X is a constant.
$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, where a and b are constants.

Moment generating functions

The moment generating function (mgf) of a random variable is defined as

$$M_X(t) = \mathrm{E}[\exp(tX)] \qquad \text{if this exists.}$$

$\mathrm{E}[X^k]$ can be evaluated as the:

(i) coefficient of $\dfrac{t^r}{r!}$ in the power expansion of $M_X(t)$.

(ii) r-th derivative of $M_X(t)$ evaluated at $t = 0$.

Measures of location

The **mean** or **expectation** of the random variable X is $\mathrm{E}[X]$, the long-run average of realisations of X. The **mode** is where the **pmf** or **pdf** achieves a maximum (if it does so). For a random variable, X, the **median** is such that $P(X \leq \text{median}) = \frac{1}{2}$, so that 50% of values of X occur above and 50% below the median.

Percentiles

x_p is the 100-p-th percentile of a random variable X if $P(X \leq x_p) = p$. For example, the 5th percentile, $x_{0.05}$, has 5% of the values smaller than or equal to it. The **median** is the 50-th percentile, the **lower quartile** is the 25th percentile, the **upper quartile** is the 75th percentile.

Measures of dispersion

The **inter-quartile range** is defined to be the difference between the upper and lower quartiles, UQ - LQ. The **standard deviation** is defined as the square root of the variance, $\sigma = \sqrt{\mathrm{Var}(X)}$, and is in the same units as the random variable X.

Cumulative Distribution Function

This is defined as a function of any real value t by

$$F(t) = P(X \leq t)$$

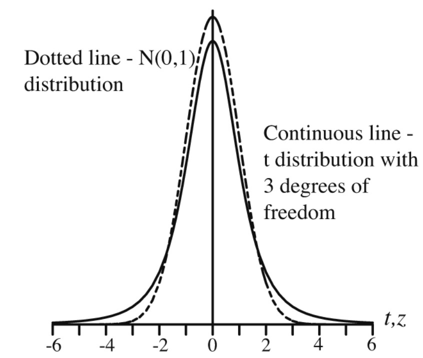
If X is a continuous random variable, F is a continuous function of t; if X is discrete, then F is a step function.

v1. Mar.07

The Central Limit Theorem

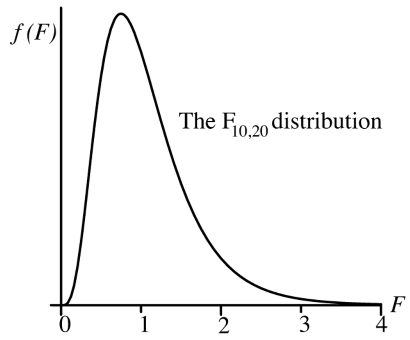
If a random sample of size n is taken from *any* distribution with mean μ and variance σ^2, the sampling distribution of the mean will be *approximately* $\sim \mathrm{N}(\mu, \sigma^2/n)$, where $\sim$ means 'is distributed as'. The larger n is, the better the approximation.

The standard normal and Student's t distributions



If a random variable $X \sim \mathrm{N}(\mu, \sigma^2)$, $z = (X-\mu)/\sigma \sim \mathrm{N}(0,1)$, the *standard normal distribution*. The t distribution with $(n-1)$ degrees of freedom is used in place of z for small samples size n from normal populations when σ^2 is unknown. As n increases the distribution of t converges to $\mathrm{N}(0,1)$. These distributions are used, e.g., for inference about means, differences between means and in regression.

Fisher's F distribution



If $X_1 \sim \chi^2_{\nu_1}$ and $X_2 \sim \chi^2_{\nu_2}$ are independent random variables then

$$\frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}$$

the F distribution with (ν_1, ν_2) degrees of freedom. This distribution is used, for example, for inference about the ratio of two variances, in Analysis of Variance (ANOVA) and in simple and multiple linear regression.

Statistics & Sampling Distributions

Population and samples

A (statistical) **population** is the complete set of all possible measurements or values, corresponding to the entire collection of units, for which inferences are to be made from taking a **sample** - the set of measurements or values that are actually collected from a population.

Simple random sample: every item in the population is equally likely to be in the sample, independently of which other members of the population are chosen.

Parameter: a quantity that describes an aspect of a population, eg. the population mean, μ, or variance, σ^2.

Statistic: a quantity calculated from the sample, e.g. the sample mean, $\bar{x}$, or variance, s^2.

Sampling distributions: The value of a statistic will in general vary from sample to sample, in which case it will have its own probability distribution, called its **sampling distribution**. A statistic used to estimate the value of a *parameter* θ in a distribution is called an **estimator** (the random variable) or an **estimate** (the value).

If $\hat{\theta}$ is an estimator of θ , the mean of its sampling distribution, $\mathbf{E}[\hat{\theta}]$, is called the *sampling mean*. The variance, $\mathrm{Var}(\hat{\theta})$, is called the *sampling variance*.

$\sqrt{\mathrm{Var}(\hat{\theta}\ )}$ is called the *standard error* of $\hat{\theta}$. If $\mathbf{E}[\hat{\theta}] = \theta$, then $\hat{\theta}$ is an unbiased estimator of θ e.g. $\bar{X}$ is an unbiased estimator for μ and has sampling variance $\frac{\sigma^2}{n}$ where $\mathrm{Var}(X_i\ ) = \sigma^2$, $(i = 1, 2, \ldots, n)$.

Corrected sum of squares

$$S_{xx} = \sum(x_i - \bar{x})^2 \equiv \sum x_i^2 - n\bar{x}^2 \equiv \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

has expectation $(n-1)\sigma^2$ so that dividing S_{xx} by $(n-1)$ will give an unbiased estimator of σ^2, denoted s^2.

Normal and Chi-squared distributions

If $X_1, X_2, \ldots X_n$ are independently and identically $\sim \mathrm{N}(\mu, \sigma^2)$, then $\sum \left(\dfrac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$, a Chi-squared distribution with n **degrees of freedom**.

Also $\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ independently of $\dfrac{S_{xx}}{\sigma^2} \sim \chi^2_{(n-1)}$.

Simple Linear Regression

To fit the straight line $y = \alpha + \beta x$ to data (x_i, y_i), $i = 1, 2, \ldots n$ by the method of **least squares** the estimates of slope, β, and intercept, α , are given by:

$$b = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i \sum y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} = \frac{S_{xy}}{S_{xx}}, \qquad a = \bar{y} - b\bar{x}$$

If we assume that the x_i are known and that the y_i have normal distributions with means $\alpha + \beta x_i$, and constant variance σ^2, written as $y_i \sim N(\alpha + \beta x_i, \sigma^2)$, then if x_0 is a fixed value

$$b \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$$

$$a \sim N\left(\alpha, \sigma^2\left\{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right\}\right)$$

$$a + bx_0 \sim N\left(\alpha + \beta x_0, \sigma^2\left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\}\right)$$

A common alternative is to use $\hat{\alpha}$ for a and $\hat{\beta}$ for b.

Correlation

Given observations (x_i, y_i), $i = 1, 2, \ldots, n$ on two random variables X and Y the **Pearson (product moment**) correlation between them is given by:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i \sum y_i\right)}{\sqrt{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}\sqrt{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}}$$

We use r to estimate the correlation, ρ , between X and Y. For large n, r is approximately $\sim N\left(\rho, \frac{1}{n-2}\right)$. The (**Spearman**) Rank Correlation Coefficient is given by

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where d_i is the difference between the *ranks* of (x_i, y_i), $i = 1, 2, \ldots, n$. If ranks are tied, see further reading.

Further reading: Kotz, S., and Johnson,L. (1988) Encyclopedia of Statistical Sciences, Vols.1-9. New York: John Wiley and Sons.

Time Series

A time series Y_t $(t = 1, 2, \ldots, n)$ is a set of n observations recorded through time t, (e.g. days, weeks, months). The arithmetic mean of blocks of k successive values

$$\frac{Y_1 + Y_2 + \ldots + Y_k}{k}, \frac{Y_2 + Y_3 + \ldots + Y_{k+1}}{k}, \ldots$$

is a **simple moving average** of order k, itself a time series which is *smoother* than Y_t and can be used to track, or estimate, the underlying level, μ_t, of Y_t. If $0 < \alpha < 1$ an **exponentially weighted moving average** (EWMA) at time t uses a discounted weighted average of current and past data to estimate μ_t with

$$\hat{\mu}_t = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots$$

This is equivalent to the recurrence relation

$$\hat{\mu}_t = \alpha Y_t + (1-\alpha)\hat{\mu}_{t-1}$$

Moving averages are often plotted on the same graph as Y_t. If Y_t additionally contains trend, R_t, the rate of change of data per unit time, and $\mu_t = \mu_{t-1} + R_{t-1}$, then the recurrence relation is

$$\hat{\mu}_t = \alpha Y_t + (1-\alpha)(\hat{\mu}_{t-1} + \hat{R}_{t-1})$$

If $0 < \beta < 1$ the *trend smoothing equation* is

$$\hat{R}_t = \beta(\hat{\mu}_t - \hat{\mu}_{t-1}) + (1-\beta)\hat{R}_{t-1}$$

known as *Holt's Linear Exponential Smoothing*. If Y_t also contain *seasonality*, S_t, a smoothing constant γ, $(0 < \gamma < 1)$ is used in a *seasonal smoothing equation*, $\hat{S}_t = \gamma Y_t/\hat{\mu}_t + (1-\gamma)\hat{S}_{t-k}$, assuming the periodicity is k, with *multiplicative* seasonality. For monthly data $k = 12$.

Forecasting from time n (now) to time $n+h$ $(h = 1, 2, \ldots)$

Level only, $\hat{Y}_{n+h} = \hat{\mu}_n$, the latest EWMA.

Level and constant trend, $\hat{Y}_{n+h} = a + b(n+h)$, the simple linear regression trend line of Y_t against t.

Level and changing trend, $\hat{Y}_{n+h} = \hat{\mu}_n + h\hat{R}_n$.

Level, changing trend and seasonality $\hat{Y}_{n+h} = \hat{\mu}_n + h\hat{R}_n$, where $\hat{\mu}_n = \alpha Y_n/\hat{S}_{n-12} + (1-\alpha)(\hat{\mu}_{n-1} + \hat{R}_{n-1})$.