Statistics Cheat Sheet

Ch 1: Overview & Descriptive Stats

Populations, Samples and Processes

`Population:` well-defined collection of objects
`Sample:` a subset of the population
`Descriptive Stats:` summarize & describe features of data
`Inferential Stats:` generalizing from sample to population
`Probability:` bridge btwn descriptive & inferential techniques.
In probability, properties of the population are assumed known & questions regarding a sample taken from the population are posed and answered.
`Discrete and Continuous Variables:` A numerical variable is *discrete* if its set of possible values is at most countable.
A numerical value is *continuous* if its set of possible values is an uncountable set.
Probability: pop $\rightarrow$ sample
Stats: sample $\rightarrow$ pop

Measures of Location

For observations $x_1, x_2, \cdots, x_n$

`Sample Mean`	$\bar{x} = \frac{\sum_{1=1}^{n} x_i}{n}$
`Sample Median`	$\tilde{x} = (\frac{n+1}{2})^{\text{nth}}$ observation
`Trimmed Mean`	btwn $\tilde{x}$ and $\bar{x}$, compute by removing smallest and largest observations

Measures of Variability

`Range`	=lgst-smllst observation
`Sample Variance,` σ^2	$= \frac{\sum(x_i - \bar{x}))^2}{n-1} = \frac{S_{xx}}{n-1}$
S_{xx}	$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$
`Sample Standard Deviation,` σ	$= \sqrt{\sigma^2}$

Box Plots

Order the n observations from small to large. Separate the smallest half from the largest (If n is odd then $\tilde{x}$ is in both halves). The lower fourth is the median of the smallest half (upper fourth..largest..). A measure of the spread that is resistant to outliers is the *fourth spread* f_s given by $f_s =$ upper fourth- lower fourth. Box from lower to upper fourth with line at median. Whiskers from smallest to largest x_i.

Ch 2: Probability

Sample Space and Events

`Experiment`	activity with uncertain outcome
`Sample Space(`$\mathcal{S}$`)`	the set of all possible outcomes
`Event`	any collection of outcomes in $\mathcal{S}$

Axioms, Interpretations and Properties of Probability

Given an experiment and a sample space $\mathcal{S}$, the objective probability is to assign to each event A a number $P(A)$, called the probability of event A, which will give a precise measure of the chance that A will occur. Behaves very much like norm.

Axioms & Properties of Probability:

1. $\forall A \in \mathcal{S}, 0 \leq P(A) \leq 1$
2. $P(\mathcal{S}) = 1$
3. If $A_1, A_2, \ldots$ is an infinite collection of disjoint events, $P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$
4. $P(\emptyset) = 0$
5. $\forall A, P(A) + P(A') = 1$ from which $P(A) = 1 - P(A')$
6. For any two events $A, B \in \mathcal{S}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. For any three events $A, B, C \in \mathcal{S}, P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Equally Likely Outcomes : $P(A) = \frac{N(A)}{N}$

Counting Techniques

`Product Rule for Ordered k-Tuples:` If the first element can be selected in n_1 ways, the second in n_2 ways and so on, then there are $n_1 n_2 \cdots n_k$ possible k-tuples.
`Permutations:` An ordered subset. The number of permutations of size k that can be formed from a set of n elements is $P_{k,n}$

$$P_{k,n} = (n)(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

`Combinations:` An unordered subset.

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

Conditional Probability

$P(A|B)$ is the conditional probability of A given that the event B has occurred. B is the conditioning event.

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

Multiplication Rule: $P(A \cap B) = P(A|B) \cdot P(B)$

Baye's Theorem

Let $A_1, A_2, \ldots, A_k$ be disjoint and exhaustive events (that partition the sample space). Then for any other event B
$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$
$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$

Independence

Two events A and B are **independent** if $P(A|B) = P(A)$ and are **dependent** otherwise.
A and B are **independent** iff $P(A \cap B) = P(A) \cdot P(B)$ and this can be generalized to the case of n mutually independent events.

Random Variables

`Random Variable:`	any function $X : \Omega \rightarrow \mathbb{R}$
`Prob Dist.:`	describes how the probability of Ω is distributed along the range of X
`Discrete rv:`	rv whose domain is at most countable
`Continuous rv:`	rv whose domain is uncountable and where $\forall c \in \mathbb{R}, P(X = c) = 0$
`Bernoulli rv:`	discrete rv whose range is $\{0, 1\}$

The *probability distribution of* X says how the total probability of 1 is distributed among the various possible X values.

1. Distributions

Discrete RVs

Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular rv X.

`Probability Mass Fxn/Probability Distribution,(pmf):`

$$p(x) = P(X = x) = P(\forall w \in \mathcal{W} : X(w) = x)$$

Gives the probability of observing $w \in \mathcal{W} : X(w) = x$
The conditions $p(x) \geq 0$ and $\sum_{\text{all possible x}} p(x) = 1$ are required for any pmf.
`parameter:` Suppose p(x) depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a parameter of distribution. The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Cumulative Distribution Function

(To compute the probability that the observed value of X will be at most some given x)
`Cumulative Distribution Function(cdf):` $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

For any number $x, F(x)$ is the probability that the observed value of X will be at most x.
For discrete rv, the graph of $F(x)$ will be a step function- jump at every possible value of X and flat btwn possible values.
For any two number a and b with $a \leq b$:
$P(a \leq X \leq b) = F(b) - F(a^-)$
$P(a < X \leq b) = F(b) - F(a)$
$P(a \leq X \leq a) = F(a) - F(a^-) = p(a)$
$P(a < X < b) = F(b^-) - F(a)$
(where a^- is the largest possible X value strictly less than a)
Taking $a = b$ yields $P(X = a) = F(a) - F(a-1)$ as desired.
`Expected value or Mean Value`

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Describes where the probability distribution is centered and is just a weighted average of the possible values of X given their distribution. However, the sample average of a sequence of X values may not settle down to some finite number (harmonic series) but will tend to grow without bound. Then the distribution is said to have a *heavy tail*. Can make if difficult to make inferences about μ.
`The Expected Value of a Function`: Sometimes interest will focus on the expected value of some function $h(x)$ rather than on just $E(x)$.
If the RV X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(x)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$ is computed by

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

`Properties of Expected Value`:

$$E(aX+b) = a \cdot E(X) + b$$

`Variance of X`: Let X have pmf $p(x)$ and expected value μ. Then the $V(X)$ or σ_X^2 is

$$V(X) = \sum_D (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2]$$

The standard deviation (SD) of X is $\sigma = \sqrt{\sigma}$
Alternatively,

$$V(X) = \sigma^2 = [\sum_D x^2 \cdot p(x)] - \mu^2 = E(X^2) - [E(X)]^2$$

`Properties of Variance`

1. $V(aX+b) = a^2 \cdot \sigma^2$
2. In particular, $\sigma_{aX} = |a| \cdot \sigma_x$
3. $\sigma_{X+b} = \sigma_X$

Continuous RVs

Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular rv X. Recall: an rv X is continuous if its set of possible values is uncountable and if $P(X=c)=0 \quad \forall c.//$

`Probability Density Fxn/Probability Distribution,(pdf)`: $\forall a, b \in \mathbb{R}, a \leq b$

$$P(\forall w \in \mathcal{W} : a \leq X(w) \leq b) = \int_a^b f(x)dx$$

Gives the probability that X takes values between a and b. The conditions $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x) = 1$ are required for any pdf.

`Cumulative Distribution Function(cdf)`:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

For any number $x, F(x)$ is the probability that the observed value of X will be at most x.
By the continuity arguments for continuous RVs we have that

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b)$$

Other probabilities can be computed from the cdf $F(x)$:

$$P(X > a) = 1 - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a)$$

Furthermore, if X is a cont rv with pdf $f(x)$ and cdf $F(x)$, then at every x at which $F'(x)$ exists, $F'(x) = f(x)$.
`Median`($\tilde{\mu}$): is the 50th percentile st $F(\tilde{\mu}) = .5$. That is half the area under the density curve. For a symmetric curve, this is the point of symmetry.
`Expected/Mean Value`(μ or $E(X)$): of cont rv with pdf $f(x)$

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

If X is a cont rv with pdf $f(x)$ and $h(X)$ is any function of X then

$$E[h(X)] = \mu = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

`Variance`: of a cont rv X with pdf $f(x)$ and mean value μ is

$$\sigma_x^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx = E[(X-\mu)^2]$$

Alternatively,

$$V(X) = E(X^2) - [E(X)]^2$$

Discrete Distributions

The Binomial Probability Distribution

1) The experiment consists of n *trials* where n is fixed
2) Each trial can result in either success (S) or failure (F)
3) The trials are independent
4) The probability of success $P(S)$ is constant for all trials

Note that in general if the sampling is without replacement, the experiment will not yield independent trials. However, if the sample size (number of trials) n is at most 5% of the population, then the experiment can be analyzed as though it were exactly a binomial experiment.
`Binomial rv X:` = no of S's among the n trials
`pmf of a Binomial RV:`,

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad : x = 0, 1, 2, \dots$$

`cdf for Binomal RV:` Values in Tble A.1

$$B(x; n, p) = P(X \leq x) = \sum_{y=0}^{x} b(y; n, p)$$

`Mean & Variance of X` If $X \sim Bin(n, p)$ then

$$E(X) = np \quad V(X) = npq$$

Negative Binomial Distribution

1) The experiment consists of independent trials
2) Each trial can result in either Success(S) or Failure(F)
3) The probability of success is constant from trial to trial
4) The experiment continues until a total of r successes have been observed, where r is a specified integer.

`RV Y:` = the no of trials before the rth success.
`Negative Binomial rv:` $X = Y - r$ the number of failures that precede the *rth* success. In contrast to the binomial rv, the number of successes is fixed while the number of trials is random.
`pmf of the negative binomial rv` : with parameters r = number of S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0, 1, 2, \dots$$

`Mean & Variance of negative binomial rv X:` with pmf $nb(x; r, p)$

$$E(X) = \frac{r(1-p)}{p} \qquad V(X) = \frac{r(1-p)}{p^2}$$

Geometric Distribution

`RV X:` = the no of trials before the 1st success.
`pmf of the geometric rv` :

$$p(x) = q^{x-1}p$$

$$E(X) = \sum xq^{x-1}p = 1/p$$

The Poisson Probability Distribution

Useful for modeling rare events
1) independent: no of events in an interval is independent of no of events in another interval
2) Rare: no 2 events at once
3) Constant Rate: average events/unit time is constant ($\mu > 0$)
`RV X=` no of occurrence in unit time interval
`Possion distribution/ Poisson pmf`: of a random variable X with parameter $\mu > 0$ where

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \qquad x = 0, 1, 2, \dots$$

`Binomial Approximation`: Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n, p) \to p(x; \mu)$.
That is to say that in any binomial experiment in which n(the number of trials) is large and p(the probability of success) is small, then $b(x; n, p) \approx p(x; \mu)$, where $\mu = np$.
`Mean and Variance of X`: If X has probability distribution with parameter μ, then $E(X) = V(X) = \mu$

Continuous Distributions

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

`PDF`: with parameters μ and σ where $-\infty < \mu < \infty$ and $0 < \sigma$

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

We can then easily show that $E(X) = \mu$ and $V(X) = \sigma^2$.
`Standard Normal Distribution`: The specific case where $\mu = 0$ and $\sigma = 1$. Then

$$\text{pdf}: \quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{cdf}: \quad \Phi(z) = \int_{-\infty}^{z} \phi(u)du$$

`Standardization`: Suppose that $X \sim N(\mu, \sigma^2)$. Then

$$Z = (X - \mu)/\sigma$$

transforms X into standard units. Indeed $Z \sim N(0, 1)$.

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

`Independence`: If $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ and X and Y are independent, then $X \pm Y \sim N(\mu_x \pm \mu_y, \sigma_x^2 + \sigma_y^2)$
NOTE: By symmetry of the standard normal distribution, it follows that $\Phi(-z) = 1 - \Phi(z) \quad \forall z \in \mathbb{R}$

Normal Approx to Binomial Dist: Let $X \sim Bin(n,p)$. As long as a binomial histogram is not too skewed, Binomial probabilities can be well approximated by normal curve areas.

$$P(X \leq x) = B(x;n,p) \approx \Phi\left(\frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$$

As a rule, the approx is adequate provided that both $np \geq 10$ and $n(1-p) \geq 10$.

The Exponential Distribution, $X \sim Exp(\lambda)$

Model for lifetime of firms/products/humans

Exponential Distribution: A cont rv X has exp distribution if its pdf is given by

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0 \quad \lambda > 0$$

$F(x,\lambda)=$	$P(X \leq x) = 1 - e^{\lambda x} \quad x \geq 0$
$E(X)=$	$1/\lambda$
$V(X)=$	$1/\lambda^2$
Memoryless Prop:	$P(X > a + x \mid X > a) = P(X > x)$ for $x \in D, a > 0$

Note: If Y is an rv distributed as a Poisson $p(y;\lambda)$, then the time between consecutive Poisson events is distributed as an exponential rv with parameter λ

Joint Probability Dist

Joint Range: Let $X : \mathcal{S} \to \mathbb{D}_1$ and $Y : \mathcal{S} \to \mathbb{D}_2$ be 2 rvs with a common sample space. We define the joint range of the vector (X,Y) of the form

$$\mathbb{D} = \mathbb{D}_1 \times \mathbb{D}_2 = \{(x,y) : x \in \mathbb{D}_1, y \in \mathbb{D}_2\}$$

Random Vector: A 2-D random vector(X,Y) is a function from $\mathcal{S} \to \mathbb{R}^2$. It is defined $\forall \omega \in \mathcal{S}$ such that

$$(X,Y)(\omega) = (X(\omega), Y(\omega)) = (x,y) \in \mathbb{D}$$

Joint Probability Mass Fxn: For two discrete rv's X and Y. The joint pmf of (X,Y) is defined $\forall (x,y) \in \mathbb{D}$

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

It must be that $p(x,y) \geq 0$ and $\sum_i \sum_j p(x_i, y_j) = 1$.

Marginal Prob Mass Fxn: of X and of Y, denoted $p_X(x)$ and $p_Y(y)$ respectively,

$$p_X(x) = \sum_{y : p(x,y) > 0} p(x,y) \quad \forall x \in \mathbb{D}_1$$

Joint Probability Density Fxn: For two continuous rv's X and Y. The joint pdf of (X,Y) is defined $\forall A \subseteq \mathbb{R}^2$

$$P((X,Y) \in A) = \iint_A f(x,y)dxdy$$

It must be that $f(x,y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$. Note also that this integration is commutative.

Marginal Prob Density Fxn: of X and of Y, denoted $f_X(x)$ and $f_Y(y)$ respectively,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \forall x \in \mathbb{D}_1$$

Note that if $f(x,y)$ is the joint density of the random vector (X,Y) and $A \in \mathbb{R}^2$ is of the form $A = [a,b] \times [c,d]$ we have that

$$P((X,Y) \in A) = \int_c^d \int_a^b f(x,y)dxdy = \int_a^b \int_c^d f(x,y)dxdy$$

Independence: Two rvs are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad f(x,y) = f_X(x)f_Y(y)$$

Conditional Distribution(discrete): For two discrete rv's X and Y with joint pmf $p(x_i, y_j)$ and marginal X pmf $p_X(x)$, then for any realized value x in the range of X, the conditional mass function of Y, given that $X = x$ is

$$p_{Y|X}(y|x) = \frac{p(x_i, y_j)}{p_X(x)}$$

Conditional Distribution(cont): For two continuous rv's X and Y with joint pdf $f(x,y)$ and marginal X pdf $f_X(x)$, then for any realized value x in the range of X, the conditional density function of Y, given that $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

Expected Values, Covariance & Correlation

Expected value: The expected value of a function $h(X,Y)$ of two jointly distributed random variables is

$$E(g(X,Y)) = \sum_{x \in \mathbb{D}_1} \sum_{y \in \mathbb{D}_2} g(x,y)p(x,y)$$

and can be generalized to the continuous case with integrations.//

Covariance: Measures the strength of the relation btwn 2 RVs, however very

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Shortcut Formula:

$$Cov(X,Y) = E(XY) - \mu_x \mu_y$$

The defect of the covariance however is that its value depends critically on the units of measurement.

Correlation: Cov after standardization. Helps interpret Cov.

$$\rho = \rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

Has the property that $Corr(aX + b, cY + d) = Corr(X,Y)$ and that for any rvs X, Y $-1 \leq \rho \leq 1$.

Note also that ρ is independent of units, the larger $|\rho|$ the stronger the linear association, considered strong linear relationship if $|\rho| \geq 0.8$.

Caution though: if X and Y are independent then $\rho = 0$ but $\rho = 0$ does not imply that X,Y are independent.

Also that $\rho = 1$ or -1 iff $Y = aX + b$ for some a,b with $a \neq 0$.

Statistic: Any quantity whose value can be calculated with sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. Therefore, a statistic is a random variable and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

Sampling Distribution: probability distribution of a statistic, it describes how the statis- tic varies in value across all samples that might be selected

Stats & Their Distributions

Fxns of Observed Sample Observ

Obs Sample Mean $\bar{x} = \frac{1}{n}\sum x_i$

Obs Sample Var $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

Obs Sample Max $x_{(n)} = max(x_i)$

A statistic is a random variable and the most common are listed above.

Simple Random Samples: The random variables $X_1, \ldots, X_n$ are said to form a simple random sample of size n if each X_i is an independent random variable, every X_i has the same probability distribution.

Sampling Distribs: Every statistic has a probability distribution (a pmt or pdf) which we call its sampling distribution. To determine its distrib can be hard but we use simulations and the CLT to do so.

Simulation Experiments: we must specify the statistic of interest, the population distribution, the sample size(n) and the number of samples (k). Use a computer to simulate each different simple random sample, construct a histogram which will give approx sampling distribution of the statistic.

The Dist % Sample Mean

Prop: Let $X_1, \ldots, X_n$ be a simple random sample from a distribution with mean μ and variance σ^2. Then $E(\bar{X}) = \mu_{\bar{X}} = \mu$ and $V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n$. Also if $S_n = X_1 + \ldots + X_n$ then $E(S_n) = n\mu$ and $V(S_n) = n\sigma^2$.

Prop: Let $X_1, \ldots, X_n$ be a simple random sample from a normal distribution with mean μ and variance σ^2. Then for any n, $\bar{X}$ is normal distributed with mean μ and variance σ^2/n. Also S_n is normal distributed with mean $n\mu$ and variance $n\sigma^2$.

Prop: Let $X_1, \ldots, X_n$ be a simple random sample from Bernoulli(p), then $S_n \sim$ Binomial(n,p).

Distribution of The Sample Mean $\bar{X}$

Let $X_1, \ldots, X_n$ be a simple random sample from a distribution with mean μ and variance σ^2. Then $E(\bar{X}) = \mu_{\bar{X}} = \mu$ and $V(\bar{X}) = \sigma^2_{\bar{X}} = \sigma^2/n$

The standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ is often called the *standard error of the mean.*

For a NORMAL random sample with the same mean and std as above, then for any n, $\bar{X}$ is normally distributed with the same mean and std.

Central Limit Theorem: Let $X_1, \ldots, X_n$ be a simple random sample from a distribution with mean μ and variance σ^2. Then if n is sufficiently large, $\bar{X}$ has approximately a normal dis with mean μ and variance σ^2/n. Also S_n is normal distributed with mean $n\mu$ and variance $n\sigma^2$. No matter which population we sample from, the probability histogram of the sample mean follow closely a normal curve when n is sufficiently large. *Rule of thumb: if $n \geq 30$ CLT can be used.*

It follows from CLT that is $X \sim Bin(n,p)$ and n is large, then n can be distributed by a $N(np, npq)$.

Dist of a Linear Combination

`Linear Comb:` Let $X_1, \ldots, X_n$ be a collxn of n random variables and let $a_1 \ldots a_n$ be n numerical constants. Then the random variable $Y = a_1X_1 + \ldots a_nX_n$ is a linear comb of the $X_i's$.

1. Regardless of whether the $X_i's$ are independent or not
$$E(Y) = a_1E(X_1) + \ldots + a_nE(X_n) = a_1\mu_1 + \ldots + a_n\mu_n$$
2. If $X_1, \ldots, X_n$ are independent
$$V(Y) = V(a_1X_1 + \ldots a_nX_n) + a_1^2\sigma_1^2 + \ldots$$
3. For any $X_1, \ldots, X_n$,
$$V(Y) = \sum_{i=1}\sum_{j=1} a_ia_jCov(X_i, X_j)$$
4. If $X_1, \ldots, X_n$ are independent, normally distributed rvs, then any linear combination of the rvs also has a normal distribution- as does their difference.

$E(X_1 - X_2) = E(X_1) - E(X_2), \forall X, Y$ while $V(X_1 - X_2) = V(X_1) + V(X_2)$ if X1, X2 independent,

2. Estimators

`Parameter of Interest` (θ) true yet unknown pop parameter

`Point Estimate:` ($\hat{\theta}$) Our guess for θ based on sample data

`Point Estimator:` ($\hat{\theta}$) statistic selected to get a sensible pt est

A sensible way to quantity the idea of $\hat{\theta}$ being close to θ is to consider the least squared error $(\hat{\theta} - \theta)^2$. A good measure of the accuracy is the expected or mean square error MSE $= E[(\hat{\theta} - \theta)^2]$. It is often not possible to find the estimator with the smallest MSE so we often restrict our attention to *unbiased* estimators and find the best estimator of this group.

`Unbiased:` Pt Est $\hat{\theta}$ if $E(\hat{\theta}) = \theta$ for all θ.

Then $\hat{\theta}$ has a prob distribution that is always "centered" at the true θ value.

When choosing estimators, select the unbiased and the one that has the minimum variance.

Estimators

-When $X \sim Bin(n, p)$, the sample proportion $\hat{p} = X/n$ is an unbiased est of p.

- Let $X_1, \ldots, X_n$ be a SRS from a distribution with mean μ and variance σ^2.Then $\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ is unbiased for σ^2.

-Let $X_1, \ldots, X_n$ be a SRS from a distribution with mean μ, then $\bar{X}$ is MVUE for μ.

`Standard Error:` of an estimator is its standard deviation $\sigma_{\hat{\theta}} - \sqrt{V(\hat{\theta})}$

`Estimated Standard Error:` If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields $\hat{\sigma\theta} = s_{\hat{\theta}}$.

Method of Moments

Let $X_1, \ldots, X_n$ be a SRS from a pdf $f(x)$. For $k = 1, 2, \ldots$ the kth population moment, or kth moment of the distribution $f(x)$, is $E(X^k)$. The kth sample moment is $(1/n)\sum_{i=1}^n X_i^k$.
Let $X_1, \ldots, X_n$ be a SRS from a distribution with pdf $f(x; \theta_1 \ldots \theta_m)$ where θ_i's are unknown. Then the moment estimators $\hat{\theta}_i$'s are obtained from the first m sample moments to the corresponding first m population moments and solving for the θ_i's.

Maximum Likelihood Estimator

Works best when the sample size is large!
Let $X_1, \ldots, X_n$ have joint pmf or pdf
$$f(x_1, \ldots, x_n; \theta_1 \ldots \theta_m)$$
where the θ_i's have unknown values.
When $x_1, \ldots, x_n$ are observed sample values, the above is considered a fxn of the θ_i's and is called the **likelihood function**.
The maximum likelihood estimates (mles) $\hat{\theta}_i$'s are those θ_i's that maximize the likelihood function such that
$$f(x_1, \ldots, x_n; \hat{\theta_1} \ldots \hat{\theta_m}) \geq f(x_1, \ldots, x_n; \theta_1 \ldots \theta_m) \quad \forall \theta_1 \ldots \theta_m$$
When $X_1, \ldots, X_n$ substituted in, the **maximum likelihood estimators** result.

3. Confidence Intervals

Tests in a single sample

When measuring n random variables $Y_i \sim i.i.d.$

Hypotheses about the population mean $E[Y_i]$

`Z-test` (when $n > 40$ or if normality with known variances could be assumed)
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
`CI for Normal Population:` A $100(1-\alpha)\%$ CI for the mean μ of a population when σ is known is
$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$
`T -test` (normality must be assured; for large n this is the same as the z-test). When $\bar{X}$ is the sample mean of a SRS of size n from a Normal(μ, σ^2) population then the RV
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$
has a probability distribution-t with n-1 degrees of freedom.
Note: the density of t_ν is symmetric around 0. t_ν is more spread out than a normal, indeed the few dof the more spread. When dof is large (< 40), the t and normal curve are close. In addition we have that
$$P(|\frac{\bar{X} - \mu}{S/\sqrt{n}}| \leq t_{\alpha/2, n-1}) = 1 - \alpha$$
As a result, the $(1-\alpha)100\%$ CI for the population mean μ under the normal model is
$$\bar{X} \pm t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}$$
Note that here we make the assumption that the observations are realizations of a SRS from a Normal distribution with unknown mean and variance.// `Large Sample Test` for the population proportion (proportions are just means; only valid for $np_0 \geq 10$ and $n(1-p_0) \geq 10$. The $(1-\alpha)$ confidence interval for a population mean μ is
$$\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$$
For a population proportion
$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \qquad \hat{p} = \bar{X}$$
Hypotheses about the population variance $V[X_i]$
The $(1-\alpha)100\%$ CI for the variance σ^2 of a normal population has a lower limit:
$$(n-1)s^2/\chi^2_{\alpha/2, n-1}$$
and Upper limit:
$$(n-1)s^2/\chi^2_{1-\alpha/2, n-1}$$
A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in the interval for σ^2. An upper or a lower confidence bound results from replacing $\alpha/2$ with α in the corresponding limit of the CI.

When measuring two variables for each unit $(X_i, Y_i) \sim i.i.d.$

`Paired t-test` about the difference of population means:

`Test about parameters` β_1 and β_0

Tests in two non-paired, independent samples

4. Hypothesis Testing

In it hard to example the evidence of such a strong count as a lucky draw. The p-value or observed significance level determines whether or not a hypothesis will be rejected- the smaller it is, the stronger evidence against he null hypothesis. The plausibility of statistical models determined by the null hypothesis is based on the sample data and their distributions. The idea is that the null is not rejected unless it is testified implausible overwhelmingly by data.

`Possible Errors:` Type I: reject the null hypothesis when it is true; Type II: fail to reject the null even though it is false.