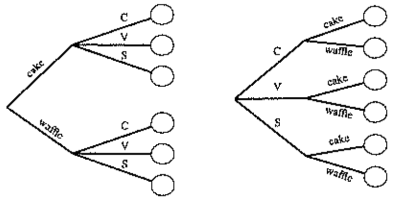
Probability Cheatsheet v2.0

Compiled by William Chen (`http://wzchen.com`) and Joe Blitzstein, with contributions from Sebastian Chiu, Yuan Jiang, Yuqi Hou, and Jessy Hwang. Material based on Joe Blitzstein's (`@stat110`) lectures (`http://stat110.net`) and Blitzstein/Hwang's Introduction to Probability textbook (`http://bit.ly/introprobability`). Licensed under `CC BY-NC-SA 4.0`. Please share comments, suggestions, and errors at `http://github.com/wzchen/probability_cheatsheet`.

Last Updated September 4, 2015

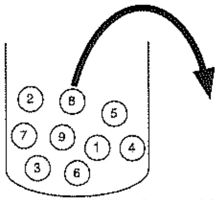
Counting

Multiplication Rule



Let's say we have a compound experiment (an experiment with multiple components). If the 1st component has n_1 possible outcomes, the 2nd component has n_2 possible outcomes, $\dots$, and the rth component has n_r possible outcomes, then overall there are $n_1 n_2 \dots n_r$ possibilities for the whole experiment.

Sampling Table



The sampling table gives the number of possible samples of size k out of a population of size n, under various assumptions about how the sample is collected.

	Order Matters	Not Matter
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Naive Definition of Probability

If all outcomes are equally likely, the probability of an event A happening is:

$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes}}$$

Thinking Conditionally

Independence

Independent Events A and B are independent if knowing whether A occurred gives no information about whether B occurred. More formally, A and B (which have nonzero probability) are independent if and only if one of the following equivalent statements holds:

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

Conditional Independence A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C)P(B|C)$. Conditional independence does not imply independence, and independence does not imply conditional independence.

Unions, Intersections, and Complements

De Morgan's Laws A useful identity that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. Analogous results hold with more than two sets.

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Joint, Marginal, and Conditional

Joint Probability $P(A \cap B)$ or $P(A, B)$ – Probability of A and B.

Marginal (Unconditional) Probability $P(A)$ – Probability of A.

Conditional Probability $P(A|B) = P(A, B)/P(B)$ – Probability of A, given that B occurred.

Conditional Probability *is* Probability $P(A|B)$ is a probability function for any fixed B. Any theorem that holds for probability also holds for conditional probability.

Probability of an Intersection or Union

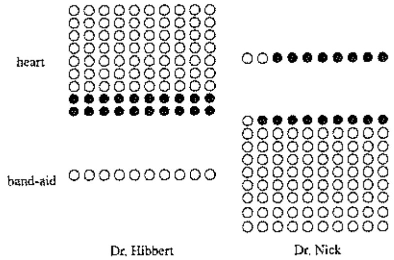
Intersections via Conditioning

$$P(A, B) = P(A)P(B|A)$$
$$P(A, B, C) = P(A)P(B|A)P(C|A, B)$$

Unions via Inclusion-Exclusion

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

Simpson's Paradox



It is possible to have

$$P(A \mid B, C) < P(A \mid B^c, C) \text{ and } P(A \mid B, C^c) < P(A \mid B^c, C^c)$$
$$\text{yet also } P(A \mid B) > P(A \mid B^c).$$

Law of Total Probability (LOTP)

Let $B_1, B_2, B_3, \dots B_n$ be a *partition* of the sample space (i.e., they are disjoint and their union is the entire sample space).

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$

For **LOTP with extra conditioning**, just add in another event C!

$$P(A|C) = P(A|B_1, C)P(B_1|C) + \dots + P(A|B_n, C)P(B_n|C)$$
$$P(A|C) = P(A \cap B_1|C) + P(A \cap B_2|C) + \dots + P(A \cap B_n|C)$$

Special case of LOTP with B and B^c as partition:

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

Bayes' Rule

Bayes' Rule, and with extra conditioning (just add in C!)

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

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

We can also write

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

Odds Form of Bayes' Rule

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

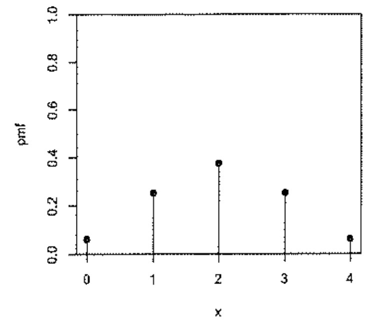
The *posterior odds* of A are the *likelihood ratio* times the *prior odds.*

Random Variables and their Distributions

PMF, CDF, and Independence

Probability Mass Function (PMF) Gives the probability that a *discrete* random variable takes on the value x.

$$p_X(x) = P(X = x)$$

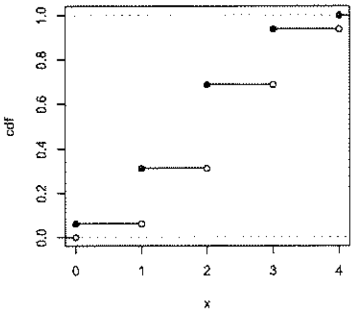


The PMF satisfies

$$p_X(x) \geq 0 \text{ and } \sum_x p_X(x) = 1$$

Cumulative Distribution Function (CDF) Gives the probability that a random variable is less than or equal to x.

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



The CDF is an increasing, right-continuous function with

$$F_X(x) \to 0 \text{ as } x \to -\infty \text{ and } F_X(x) \to 1 \text{ as } x \to \infty$$

Independence Intuitively, two random variables are independent if knowing the value of one gives no information about the other. Discrete r.v.s X and Y are independent if for *all* values of x and y

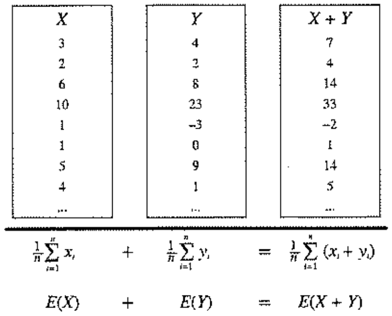
$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Expected Value and Indicators

Expected Value and Linearity

Expected Value (a.k.a. *mean*, *expectation*, or *average*) is a weighted average of the possible outcomes of our random variable. Mathematically, if $x_1, x_2, x_3, \ldots$ are all of the distinct possible values that X can take, the expected value of X is

$$E(X) = \sum_i x_i P(X = x_i)$$



X	Y	$X+Y$
3	4	7
2	2	4
6	8	14
10	23	33
1	–3	–2
1	0	1
5	9	14
4	1	5
...	...	...

$$\frac{1}{n}\sum_{i=1}^{n} x_i \quad + \quad \frac{1}{n}\sum_{i=1}^{n} y_i \quad = \quad \frac{1}{n}\sum_{i=1}^{n} (x_i + y_i)$$

$$E(X) \quad + \quad E(Y) \quad = \quad E(X+Y)$$

Linearity For any r.v.s X and Y, and constants a, b, c,

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

Same distribution implies same mean If X and Y have the same distribution, then $E(X) = E(Y)$ and, more generally,

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

Conditional Expected Value is defined like expectation, only conditioned on any event A.

$$E(X|A) = \sum_x xP(X = x|A)$$

Indicator Random Variables

Indicator Random Variable is a random variable that takes on the value 1 or 0. It is always an indicator of some event: if the event occurs, the indicator is 1; otherwise it is 0. They are useful for many problems about counting how many events of some kind occur. Write

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{if } A \text{ does not occur.} \end{cases}$$

Note that $I_A^2 = I_A, I_A I_B = I_{A \cap B}$, and $I_{A \cup B} = I_A + I_B - I_A I_B$.

Distribution $I_A \sim \text{Bern}(p)$ where $p = P(A)$.

Fundamental Bridge The expectation of the indicator for event A is the probability of event A: $E(I_A) = P(A)$.

Variance and Standard Deviation

$$\text{Var}(X) = E\left(X - E(X)\right)^2 = E(X^2) - (E(X))^2$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Continuous RVs, LOTUS, UoU

Continuous Random Variables (CRVs)

What's the probability that a CRV is in an interval? Take the difference in CDF values (or use the PDF as described later).

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

For $X \sim \mathcal{N}(\mu, \sigma^2)$, this becomes

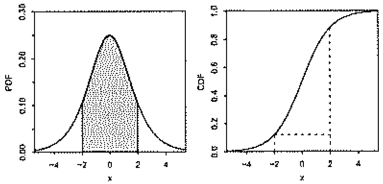
$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

What is the Probability Density Function (PDF)? The PDF f is the derivative of the CDF F.

$$F'(x) = f(x)$$

A PDF is nonnegative and integrates to 1. By the fundamental theorem of calculus, to get from PDF back to CDF we can integrate:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



To find the probability that a CRV takes on a value in an interval, integrate the PDF over that interval.

$$F(b) - F(a) = \int_a^b f(x)dx$$

How do I find the expected value of a CRV? Analogous to the discrete case, where you sum x times the PMF, for CRVs you integrate x times the PDF.

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

LOTUS

Expected value of a function of an r.v. The expected value of X is defined this way:

$$E(X) = \sum_x xP(X = x) \text{ (for discrete } X\text{)}$$

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

The **Law of the Unconscious Statistician (LOTUS)** states that you can find the expected value of a *function of a random variable*, $g(X)$, in a similar way, by replacing the x in front of the PMF/PDF by $g(x)$ but still working with the PMF/PDF of X:

$$E(g(X)) = \sum_x g(x)P(X = x) \text{ (for discrete } X\text{)}$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx \text{ (for continuous } X\text{)}$$

What's a function of a random variable? A function of a random variable is also a random variable. For example, if X is the number of bikes you see in an hour, then $g(X) = 2X$ is the number of bike wheels you see in that hour and $h(X) = \binom{X}{2} = \frac{X(X-1)}{2}$ is the number of *pairs* of bikes such that you see both of those bikes in that hour.

What's the point? You don't need to know the PMF/PDF of $g(X)$ to find its expected value. All you need is the PMF/PDF of X.

Universality of Uniform (UoU)

When you plug any CRV into its own CDF, you get a Uniform(0,1) random variable. When you plug a Uniform(0,1) r.v. into an inverse CDF, you get an r.v. with that CDF. For example, let's say that a random variable X has CDF

$$F(x) = 1 - e^{-x}, \text{ for } x > 0$$

By UoU, if we plug X into this function then we get a uniformly distributed random variable.

$$F(X) = 1 - e^{-X} \sim \text{Unif}(0,1)$$

Similarly, if $U \sim \text{Unif}(0,1)$ then $F^{-1}(U)$ has CDF F. The key point is that for any continuous random variable X, we can transform it into a Uniform random variable and back by using its CDF.

Moments and MGFs

Moments

Moments describe the shape of a distribution. Let X have mean μ and standard deviation σ, and $Z = (X - \mu)/\sigma$ be the *standardized* version of X. The kth moment of X is $\mu_k = E(X^k)$ and the kth standardized moment of X is $m_k = E(Z^k)$. The mean, variance, skewness, and kurtosis are important summaries of the shape of a distribution.

Mean $E(X) = \mu_1$

Variance $\text{Var}(X) = \mu_2 - \mu_1^2$

Skewness $\text{Skew}(X) = m_3$

Kurtosis $\text{Kurt}(X) = m_4 - 3$

Moment G[illegible]ating Functions

MGF For any random variable X, the function

$$M_X(t) = E(e^{tX})$$

is the **moment generating function (MGF)** of X, if it exists for all t in some open interval containing 0. The variable t could just as well have been called u or v. It's a bookkeeping device that lets us work with the *function* M_X rather than the *sequence* of moments.

Why is it called the Moment Generating Function? Because the kth derivative of the moment generating function, evaluated at 0, is the kth moment of X.

$$\mu_k = E(X^k) = M_X^{(k)}(0)$$

This is true by Taylor expansion of e^{tX} since

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} \frac{E(X^k)t^k}{k!} = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}$$

MGF of linear functions If we have $Y = aX + b$, then

$$M_Y(t) = E(e^{t(aX+b)}) = e^{bt}E(e^{(at)X}) = e^{bt}M_X(at)$$

Uniqueness *If it exists, the MGF uniquely determines the distribution.* This means that for any two random variables X and Y, they are distributed the same (their PMFs/PDFs are equal) if and only if their MGFs are equal.

Summing Independent RVs by Multiplying MGFs. If X and Y are independent, then

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX})E(e^{tY}) = M_X(t) \cdot M_Y(t)$$

The MGF of the sum of two random variables is the product of the MGFs of those two random variables.

Joint PDFs and CDFs

Joint Distributions

The **joint CDF** of X and Y is

$$F(x, y) = P(X \leq x, Y \leq y)$$

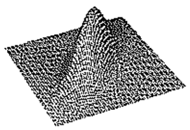
In the discrete case, X and Y have a **joint PMF**

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

In the continuous case, they have a **joint PDF**

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

The joint PMF/PDF must be nonnegative and sum/integrate to 1.



Conditional Distributions

Conditioning and Bayes' rule for discrete r.v.s

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

Conditioning and Bayes' rule for continuous r.v.s

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

Hybrid Bayes' rule

$$f_X(x|A) = \frac{P(A|X = x)f_X(x)}{P(A)}$$

Marginal Distribution

To find the distribution of one (or more) random variables from a joint PMF/PDF, sum/integrate over the unwanted random variables.

Marginal PMF from joint PMF

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

Marginal PDF from joint PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

Independence of Random Variables

Random variables X and Y are independent if and only if any of the following conditions holds:

- Joint CDF is the product of the marginal CDFs
- Joint PMF/PDF is the product of the marginal PMFs/PDFs
- Conditional distribution of Y given X is the marginal distribution of Y

Write $X \perp\!\!\!\perp Y$ to denote that X and Y are independent.

Multivariate LOTUS

LOTUS in more than one dimension is analogous to the 1D LOTUS. For discrete random variables:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

For continuous random variables:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$$

Covariance and Transformations

Covariance and Correlation

Covariance is the analog of variance for two random variables.

$$\text{Cov}(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$$

Note that

$$\text{Cov}(X, X) = E(X^2) - (E(X))^2 = \text{Var}(X)$$

Correlation is a standardized version of covariance that is always between -1 and 1.

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

Covariance and Independence If two random variables are independent, then they are uncorrelated. The converse is not necessarily true (e.g., consider $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$).

$$X \perp\!\!\!\perp Y \longrightarrow \text{Cov}(X, Y) = 0 \longrightarrow E(XY) = E(X)E(Y)$$

Covariance and Variance The variance of a sum can be found by

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

If X and Y are independent then they have covariance 0, so

$$X \perp\!\!\!\perp Y \Longrightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

If $X_1, X_2, \ldots, X_n$ are identically distributed and have the same covariance relationships (often by **symmetry**), then

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = n\text{Var}(X_1) + 2\binom{n}{2}\text{Cov}(X_1, X_2)$$

Covariance Properties For random v[illegible]les W, X, Y, Z and constants a, b:

$$\begin{aligned}
\text{Cov}(X, Y) &= \text{Cov}(Y, X) \\
\text{Cov}(X + a, Y + b) &= \text{Cov}(X, Y) \\
\text{Cov}(aX, bY) &= ab\text{Cov}(X, Y) \\
\text{Cov}(W + X, Y + Z) &= \text{Cov}(W, Y) + \text{Cov}(W, Z) + \text{Cov}(X, Y) \\
&\quad + \text{Cov}(X, Z)
\end{aligned}$$

Correlation is location-invariant and scale-invariant For any constants a, b, c, d with a and c nonzero,

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

Transformations

One Variable Transformations Let's say that we have a random variable X with PDF $f_X(x)$, but we are also interested in some function of X. We call this function $Y = g(X)$. Also let $y = g(x)$. If g is differentiable and strictly increasing (or strictly decreasing), then the PDF of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

The derivative of the inverse transformation is called the **Jacobian**.

Two Variable Transformations Similarly, let's say we know the joint PDF of U and V but are also interested in the random vector (X, Y) defined by $(X, Y) = g(U, V)$. Let

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

be the **Jacobian matrix.** If the entries in this matrix exist and are continuous, and the determinant of the matrix is never 0, then

$$f_{X,Y}(x, y) = f_{U,V}(u, v)\left|\left|\frac{\partial(u, v)}{\partial(x, y)}\right|\right|$$

The inner bars tells us to take the matrix's determinant, and the outer bars tell us to take the absolute value. In a 2×2 matrix,

$$\left|\left|\begin{matrix} a & b \\ c & d \end{matrix}\right|\right| = |ad - bc|$$

Convolutions

Convolution Integral If you want to find the PDF of the sum of two independent CRVs X and Y, you can do the following integral:

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t - x)dx$$

Example Let $X, Y \sim \mathcal{N}(0, 1)$ be i.i.d. Then for each fixed t,

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-(t-x)^2/2}dx$$

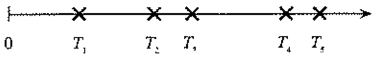
By completing the square and using the fact that a Normal PDF integrates to 1, this works out to $f_{X+Y}(t)$ being the $\mathcal{N}(0, 2)$ PDF.

Poisson Process

Definition We have a **Poisson process** of rate λ arrivals per unit time if the following conditions hold:

1. The number of arrivals in a time interval of length t is $\text{Pois}(\lambda t)$.
2. Numbers of arrivals in disjoint time intervals are independent.

For example, the numbers of arrivals in the time intervals $[0, 5]$, $(5, 12)$, and $[13, 23)$ are independent with $\text{Pois}(5\lambda), \text{Pois}(7\lambda), \text{Pois}(10\lambda)$ distributions, respectively.



Count-Time Duality Consider a Poisson process of emails arriving in an inbox at rate λ emails per hour. Let T_n be the time of arrival of the nth email (relative to some starting time 0) and N_t be the number of emails that arrive in $[0, t]$. Let's find the distribution of T_1. The event $T_1 > t$, the event that you have to wait more than t hours to get the first email, is the same as the event $N_t = 0$, which is the event that there are no emails in the first t hours. So

$$P(T_1 > t) = P(N_t = 0) = e^{-\lambda t} \longrightarrow P(T_1 \le t) = 1 - e^{-\lambda t}$$

Thus we have $T_1 \sim \text{Expo}(\lambda)$. By the memoryless property and similar reasoning, the interarrival times between emails are i.i.d. $\text{Expo}(\lambda)$, i.e., the differences $T_n - T_{n-1}$ are i.i.d. $\text{Expo}(\lambda)$.

Order Statistics

Definition Let's say you have n i.i.d. r.v.s $X_1, X_2, \dots, X_n$. If you arrange them from smallest to largest, the ith element in that list is the ith order statistic, denoted $X_{(i)}$. So $X_{(1)}$ is the smallest in the list and $X_{(n)}$ is the largest in the list.

Note that the order statistics are *dependent*, e.g., learning $X_{(4)} = 42$ gives us the information that $X_{(1)}, X_{(2)}, X_{(3)}$ are ≤ 42 and $X_{(5)}, X_{(6)}, \dots, X_{(n)}$ are ≥ 42.

Distribution Taking n i.i.d. random variables $X_1, X_2, \dots, X_n$ with CDF $F(x)$ and PDF $f(x)$, the CDF and PDF of $X_{(i)}$ are:

$$F_{X_{(i)}}(x) = P(X_{(i)} \le x) = \sum_{k=i}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$$

$$f_{X_{(i)}}(x) = n\binom{n-1}{i-1} F(x)^{i-1}(1 - F(x))^{n-i} f(x)$$

Uniform Order Statistics The jth order statistic of i.i.d. $U_1, \dots, U_n \sim \text{Unif}(0, 1)$ is $U_{(j)} \sim \text{Beta}(j, n - j + 1)$.

Conditional Expectation

Conditioning on an Event We can find $E(Y|A)$, the expected value of Y given that event A occurred. A very important case is when A is the event $X = x$. Note that $E(Y|A)$ is a *number*. For example:

- The expected value of a fair die roll, given that it is prime, is $\frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 5 = \frac{10}{3}$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success. Let A be the event that the first 3 trials are all successes. Then
$$E(Y|A) = 3 + 7p$$
since the oer of successes among the last 7 trials is $\text{Bin}(7, p)$.
- Let $T \sim \text{Expo}(1/10)$ be how long you have to wait until the shuttle comes. Given that you have already waited t minutes, the expected additional waiting time is 10 more minutes, by the memoryless property. That is, $E(T|T > t) = t + 10$.

Discrete Y	Continuous Y
$E(Y) = \sum_y yP(Y = y)$	$E(Y) = \int_{-\infty}^{\infty} yf_Y(y)dy$
$E(Y\|A) = \sum_y yP(Y = y\|A)$	$E(Y\|A) = \int_{-\infty}^{\infty} yf(y\|A)dy$

Conditioning on a Random Variable We can also find $E(Y|X)$, the expected value of Y given the random variable X. This is a *function of the random variable X*. It is *not* a number except in certain special cases such as if $X \perp\!\!\!\perp Y$. To find $E(Y|X)$, find $E(Y|X = x)$ and then plug in X for x. For example:

- If $E(Y|X = x) = x^3 + 5x$, then $E(Y|X) = X^3 + 5X$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success and X be the number of successes among the first 3 trials. Then $E(Y|X) = X + 7p$.
- Let $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$. Then $E(Y|X = x) = x^2$ since if we know $X = x$ then we know $Y = x^2$. And $E(X|Y = y) = 0$ since if we know $Y = y$ then we know $X = \pm\sqrt{y}$, with equal probabilities (by symmetry). So $E(Y|X) = X^2, E(X|Y) = 0$.

Properties of Conditional Expectation

1. $E(Y|X) = E(Y)$ if $X \perp\!\!\!\perp Y$
2. $E(h(X)W|X) = h(X)E(W|X)$ **(taking out what's known)** In particular, $E(h(X)|X) = h(X)$.
3. $E(E(Y|X)) = E(Y)$ (**Adam's Law**, a.k.a. Law of Total Expectation)

Adam's Law (a.k.a. Law of Total Expectation) can also be written in a way that looks analogous to LOTP. For any events $A_1, A_2, \dots, A_n$ that partition the sample space,

$$E(Y) = E(Y|A_1)P(A_1) + \dots + E(Y|A_n)P(A_n)$$

For the special case where the partition is A, A^c, this says

$$E(Y) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$$

Eve's Law (a.k.a. Law of Total Variance)

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

MVN, LLN, CLT

Law of Large Numbers (LLN)

Let $X_1, X_2, X_3 \dots$ be i.i.d. with mean μ. The **sample mean** is

$$\bar{X}_n = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

The **Law of Large Numbers** states that as $n \to \infty$, $\bar{X}_n \to \mu$ with probability 1. For example, in flips of a coin with probability p of Heads, let X_j be the indicator he jth flip being Heads. Then LLN says the proportion of Heads c ges to p (with probability 1).

Central Limit Theorem (CLT)

Approximation using CLT

We use $\dot{\sim}$ to denote *is approximately distributed*. We can use the **Central Limit Theorem** to approximate the distribution of a random variable $Y = X_1 + X_2 + \dots + X_n$ that is a sum of n i.i.d. random variables X_i. Let $E(Y) = \mu_Y$ and $\text{Var}(Y) = \sigma_Y^2$. The CLT says

$$Y \dot{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$$

If the X_i are i.i.d. with mean μ_X and variance σ_X^2, then $\mu_Y = n\mu_X$ and $\sigma_Y^2 = n\sigma_X^2$. For the sample mean $\bar{X}_n$, the CLT says

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \dot{\sim} \mathcal{N}(\mu_X, \sigma_X^2/n)$$

Asymptotic Distributions using CLT

We use $\xrightarrow{D}$ to denote *converges in distribution to* as $n \to \infty$. The CLT says that if we standardize the sum $X_1 + \dots + X_n$ then the distribution of the sum converges to $\mathcal{N}(0, 1)$ as $n \to \infty$:

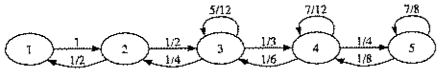
$$\frac{1}{\sigma\sqrt{n}}(X_1 + \dots + X_n - n\mu_X) \xrightarrow{D} \mathcal{N}(0, 1)$$

In other words, the CDF of the left-hand side goes to the standard Normal CDF, Φ. In terms of the sample mean, the CLT says

$$\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma_X} \xrightarrow{D} \mathcal{N}(0, 1)$$

Markov Chains

Definition



A Markov chain is a random walk in a **state space**, which we will assume is finite, say $\{1, 2, \dots, M\}$. We let X_t denote which element of the state space the walk is visiting at time t. The Markov chain is the sequence of random variables tracking where the walk is at all points in time, $X_0, X_1, X_2, \dots$. By definition, a Markov chain must satisfy the **Markov property**, which says that if you want to predict where the chain will be at a future time, if we know the present state then the entire past history is irrelevant. *Given the present, the past and future are conditionally independent.* In symbols,

$$P(X_{n+1} = j|X_0 = i_0, X_1 = i_1, \dots, X_n = i) = P(X_{n+1} = j|X_n = i)$$

State Properties

A state is either recurrent or transient.

- If you start at a **recurrent state**, then you will always return back to that state at some point in the future. ♪*You can check-out any time you like, but you can never leave.* ♪
- Otherwise you are at a **transient state**. There is some positive probability that once you leave you will never return. ♪*You don't have to go home, but you can't stay here.* ♪

A state is either periodic or aperiodic.

- If you start at a **periodic state** of period k, then the GCD of the possible numbers of steps it would take to return back is $k > 1$.
- Otherwise you are at an **aperiodic state**. The GCD of the possible numbers of steps it wo ke to return back is 1.

Transition Matrix

Let the state space be $\{1, 2, \dots, M\}$. The transition matrix Q is the $M \times M$ matrix where element q_{ij} is the probability that the chain goes from state i to state j in one step:

$$q_{ij} = P(X_{n+1} = j | X_n = i)$$

To find the probability that the chain goes from state i to state j in exactly m steps, take the (i, j) element of Q^m.

$$q_{ij}^{(m)} = P(X_{n+m} = j | X_n = i)$$

If X_0 is distributed according to the row vector PMF $\vec{p}$, i.e., $p_j = P(X_0 = j)$, then the PMF of X_n is $\vec{p}Q^n$.

Chain Properties

A chain is **irreducible** if you can get from anywhere to anywhere. If a chain (on a finite state space) is irreducible, then all of its states are recurrent. A chain is **periodic** if any of its states are periodic, and is **aperiodic** if none of its states are periodic. In an irreducible chain, all states have the same period.

A chain is **reversible** with respect to $\vec{s}$ if $s_i q_{ij} = s_j q_{ji}$ for all i, j. Examples of reversible chains include any chain with $q_{ij} = q_{ji}$, with $\vec{s} = (\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M})$, and random walk on an undirected network.

Stationary Distribution

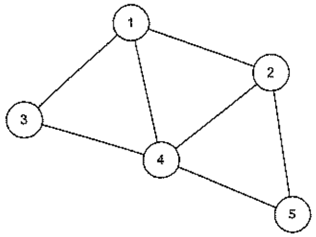
Let us say that the vector $\vec{s} = (s_1, s_2, \dots, s_M)$ be a PMF (written as a row vector). We will call $\vec{s}$ the **stationary distribution** for the chain if $\vec{s}Q = \vec{s}$. As a consequence, if X_t has the stationary distribution, then all future $X_{t+1}, X_{t+2}, \dots$ also have the stationary distribution.

For irreducible, aperiodic chains, the stationary distribution exists, is unique, and s_i is the long-run probability of a chain being at state i. The expected number of steps to return to i starting from i is $1/s_i$.

To find the stationary distribution, you can solve the matrix equation $(Q' - I)\vec{s}\,' = 0$. The stationary distribution is uniform if the columns of Q sum to 1.

Reversibility Condition Implies Stationarity If you have a PMF $\vec{s}$ and a Markov chain with transition matrix Q, then $s_i q_{ij} = s_j q_{ji}$ for all states i, j implies that $\vec{s}$ is stationary.

Random Walk on an Undirected Network



If you have a collection of **nodes**, pairs of which can be connected by undirected **edges**, and a Markov chain is run by going from the current node to a uniformly random node that is connected to it by an edge, then this is a random walk on an undirected network. The stationary distribution of this chain is proportional to the **degree sequence** (this is the sequence of degrees, where the degree of a node is how many edges are attached to it). For example, the stationary distribution of random walk on the network shown above is proportional to $(3, 3, 2, 4, 2)$, so it's $(\frac{3}{14}, \frac{3}{14}, \frac{3}{14}, \frac{4}{14}, \frac{2}{14})$.

Continuous Distributions

Uniform Distribution

Let us say that U is distributed $\text{Unif}(a, b)$. We know the following:

Properties of the Uniform For a Uniform distribution, the probability of a draw from any interval within the support is proportional to the length of the interval. See *Universality of Uniform* and *Order Statistics* for other properties.

Example William throws darts really badly, so his darts are uniform over the whole room because they're equally likely to appear anywhere. William's darts have a Uniform distribution on the surface of the room. The Uniform is the only distribution where the probability of hitting in any specific region is proportional to the length/area/volume of that region, and where the density of occurrence in any one specific spot is constant throughout the whole support.

Normal Distribution

Let us say that X is distributed $\mathcal{N}(\mu, \sigma^2)$. We know the following:

Central Limit Theorem The Normal distribution is ubiquitous because of the Central Limit Theorem, which states that the sample mean of i.i.d. r.v.s will approach a Normal distribution as the sample size grows, regardless of the initial distribution.

Location-Scale Transformation Every time we shift a Normal r.v. (by adding a constant) or rescale a Normal (by multiplying by a constant), we change it to another Normal r.v. For any Normal $X \sim \mathcal{N}(\mu, \sigma^2)$, we can transform it to the standard $\mathcal{N}(0, 1)$ by the following transformation:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Standard Normal The Standard Normal, $Z \sim \mathcal{N}(0, 1)$, has mean 0 and variance 1. Its CDF is denoted by Φ.

Exponential Distribution

Let us say that X is distributed $\text{Expo}(\lambda)$. We know the following:

Story You're sitting on an open meadow right before the break of dawn, wishing that airplanes in the night sky were shooting stars, because you could really use a wish right now. You know that shooting stars come on average every 15 minutes, but a shooting star is not "due" to come just because you've waited so long. Your waiting time is memoryless; the additional time until the next shooting star comes does not depend on how long you've waited already.

Example The waiting time until the next shooting star is distributed Expo(4) hours. Here $\lambda = 4$ is the **rate parameter**, since shooting stars arrive at a rate of 1 per 1/4 hour on average. The expected time until the next shooting star is $1/\lambda = 1/4$ hour.

Expos as a rescaled Expo(1)

$$Y \sim \text{Expo}(\lambda) \rightarrow X = \lambda Y \sim \text{Expo}(1)$$

Memorylessness The Exponential Distribution is the only continuous memoryless distribution. The memoryless property says that for $X \sim \text{Expo}(\lambda)$ and any positive numbers s and t,

$$P(X > s + t | X > s) = P(X > t)$$

Equivalently,

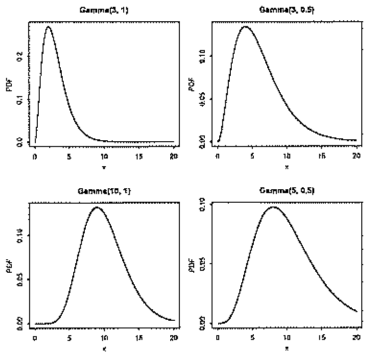
$$X - a | (X > a) \sim \text{Expo}(\lambda)$$

For example, a product with an $\text{Expo}(\lambda)$ lifetime is always "as good as new" (it doesn't experience wear and tear). Given that the product has survived a years, the additional time that it will last is still $\text{Expo}(\lambda)$.

Min of Expos If we have independent $X_i \sim \text{Expo}(\lambda_i)$, then $\min(X_1, \dots, X_k) \sim \text{Expo}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$.

Max of Expos If we have i.i.d. $X_i \sim \text{Expo}(\lambda)$, then $\max(X_1, \dots, X_k)$ has the same distribution as $Y_1 + Y_2 + \dots + Y_k$, where $Y_j \sim \text{Expo}(j\lambda)$ and the Y_j are independent.

Gamma Distribution

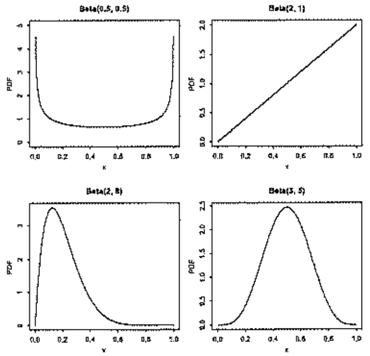


Let us say that X is distributed $\text{Gamma}(a, \lambda)$. We know the following:

Story You sit waiting for shooting stars, where the waiting time for a star is distributed $\text{Expo}(\lambda)$. You want to see n shooting stars before you go home. The total waiting time for the nth shooting star is $\text{Gamma}(n, \lambda)$.

Example You are at a bank, and there are 3 people ahead of you. The serving time for each person is Exponential with mean 2 minutes. Only one person at a time can be served. The distribution of your waiting time until it's your turn to be served is $\text{Gamma}(3, \frac{1}{2})$.

Beta Distribution



Conjugate Prior of the Binomial In the Bayesian approach to statistics, parameters are viewed as random variables, to reflect our uncertainty. The *prior* for a parameter is its distribution before observing data. The *posterior* is the distribution for the parameter after observing data. Beta is the *conjugate* prior of the Binomial because if you have a Beta-distributed prior on p in a Binomial, then the posterior distribution on p given the Binomial data is also Beta-distributed. Consider the following two-level model:

$$X | p \sim \text{Bin}(n, p)$$
$$p \sim \text{Beta}(a, b)$$

Then after observing $X = x$, we get the posterior distribution

$$p | (X = x) \sim \text{Beta}(a + x, b + n - x)$$

Order statistics of the Uniform See *Order Statistics*.

Beta-Gamma relationship If $X \sim \text{Gamma}(a, \lambda)$, $Y \sim \text{Gamma}(b, \lambda)$, with $X \perp\!\!\!\perp Y$ then

- $\frac{X}{X+Y} \sim \mathrm{Beta}(a, b)$
- $X + Y \perp\!\!\!\perp \frac{X}{X+Y}$

This is known as the **bank–post office result**.

χ^2 (Chi-Square) Distribution

Let us say that X is distributed χ^2_n. We know the following:

Story A Chi-Square(n) is the sum of the squares of n independent standard Normal r.v.s.

Properties and Representations

$$X \text{ is distributed as } Z_1^2 + Z_2^2 + \cdots + Z_n^2 \text{ for i.i.d. } Z_i \sim \mathcal{N}(0, 1)$$

$$X \sim \mathrm{Gamma}(n/2, 1/2)$$

Discrete Distributions

Distributions for four sampling schemes

	Replace	No Replace
Fixed # trials (n)	Binomial (Bern if $n = 1$)	HGeom
Draw until r success	NBin (Geom if $r = 1$)	NHGeom

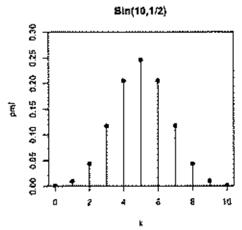
Bernoulli Distribution

The Bernoulli distribution is the simplest case of the Binomial distribution, where we only have one trial ($n = 1$). Let us say that X is distributed Bern(p). We know the following:

Story A trial is performed with probability p of "success", and X is the indicator of success: 1 means success, 0 means failure.

Example Let X be the indicator of Heads for a fair coin toss. Then $X \sim \mathrm{Bern}(\frac{1}{2})$. Also, $1 - X \sim \mathrm{Bern}(\frac{1}{2})$ is the indicator of Tails.

Binomial Distribution



Let us say that X is distributed $\mathrm{Bin}(n, p)$. We know the following:

Story X is the number of "successes" that we will achieve in n independent trials, where each trial is either a success or a failure, each with the same probability p of success. We can also write X as a sum of multiple independent $\mathrm{Bern}(p)$ random variables. Let $X \sim \mathrm{Bin}(n, p)$ and $X_j \sim \mathrm{Bern}(p)$, where all of the Bernoullis are independent. Then

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Example If Jeremy Lin makes 10 free throws and each one independently has a $\frac{3}{4}$ chance of getting in, then the number of free throws he makes is distributed $\mathrm{Bin}(10, \frac{3}{4})$.

Properties Let $X \sim \mathrm{Bin}(n, p), Y \sim \mathrm{Bin}(m, p)$ with $X \perp\!\!\!\perp Y$.

- **Redefine success** $n - X \sim \mathrm{Bin}(n, 1 - p)$
- **Sum** X [illegible] $\mathrm{Bin}(n + m, p)$
- **Conditional** $X|(X + Y = r) \sim \mathrm{HGeom}(n, m, r)$
- **Binomial-Poisson Relationship** $\mathrm{Bin}(n, p)$ is approximately $\mathrm{Pois}(\lambda)$ if p is small.
- **Binomial-Normal Relationship** $\mathrm{Bin}(n, p)$ is approximately $\mathcal{N}(np, np(1 - p))$ if n is large and p is not near 0 or 1.

Geometric Distribution

Let us say that X is distributed $\mathrm{Geom}(p)$. We know the following:

Story X is the number of "failures" that we will achieve before we achieve our first success. Our successes have probability p.

Example If each pokeball we throw has probability $\frac{1}{10}$ to catch Mew, the number of failed pokeballs will be distributed $\mathrm{Geom}(\frac{1}{10})$.

First Success Distribution

Equivalent to the Geometric distribution, except that it includes the first success in the count. This is 1 more than the number of failures. If $X \sim \mathrm{FS}(p)$ then $E(X) = 1/p$.

Negative Binomial Distribution

Let us say that X is distributed $\mathrm{NBin}(r, p)$. We know the following:

Story X is the number of "failures" that we will have before we achieve our rth success. Our successes have probability p.

Example Thundershock has 60% accuracy and can faint a wild Raticate in 3 hits. The number of misses before Pikachu faints Raticate with Thundershock is distributed $\mathrm{NBin}(3, 0.6)$.

Hypergeometric Distribution

Let us say that X is distributed $\mathrm{HGeom}(w, b, n)$. We know the following:

Story In a population of w desired objects and b undesired objects, X is the number of "successes" we will have in a draw of n objects, without replacement. The draw of n objects is assumed to be a **simple random sample** (all sets of n objects are equally likely).

Examples Here are some HGeom examples.

- Let's say that we have only b Weedles (failure) and w Pikachus (success) in Viridian Forest. We encounter n Pokemon in the forest, and X is the number of Pikachus in our encounters.
- The number of Aces in a 5 card hand.
- You have w white balls and b black balls, and you draw n balls. You will draw X white balls.
- You have w white balls and b black balls, and you draw n balls without replacement. The number of white balls in your sample is $\mathrm{HGeom}(w, b, n)$; the number of black balls is $\mathrm{HGeom}(b, w, n)$.
- **Capture-recapture** A forest has N elk, you capture n of them, tag them, and release them. Then you recapture a new sample of size m. How many tagged elk are now in the new sample? $\mathrm{HGeom}(n, N - n, m)$

Poisson Distribution

Let us say that X is distributed $\mathrm{Pois}(\lambda)$. We know the following:

Story There are rare events (low probability events) that occur many different ways (high possibilities of occurences) at an average rate of λ occurrences per unit space or time. The number of events that occur in that unit of space or time is X.

Example A certain busy intersection has an average of 2 accidents per month. Since an accident is a low probability event that can happen many different ways, it is reasonable to model the number of accidents in a month at that intersection as $\mathrm{Pois}(2)$. Then the number of accidents that happen in two months at that intersection is distributed $\mathrm{Pois}(4)$.

Properties Let $X \sim \mathrm{Pois}(\lambda_1)$ and $Y \sim \mathrm{Pois}(\lambda_2)$, with $X \perp\!\!\!\perp Y$.

1. **Sum** $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$
2. **Conditional** $X|(X + Y = n) \sim \mathrm{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$
3. **Chicken-egg** If there are $Z \sim \mathrm{Pois}(\lambda)$ items and we randomly and independently "accept" each item with probability p, then the number of accepted items $Z_1 \sim \mathrm{Pois}(\lambda p)$, and the number of rejected items $Z_2 \sim \mathrm{Pois}(\lambda(1 - p))$, and $Z_1 \perp\!\!\!\perp Z_2$.

Multivariate Distributions

Multinomial Distribution

Let us say that the vector $\vec{X} = (X_1, X_2, X_3, \ldots, X_k) \sim \mathrm{Mult}_k(n, \vec{p})$ where $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Story We have n items, which can fall into any one of the k buckets independently with the probabilities $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Example Let us assume that every year, 100 students in the Harry Potter Universe are randomly and independently sorted into one of four houses with equal probability. The number of people in each of the houses is distributed $\mathrm{Mult}_4(100, \vec{p})$, where $\vec{p} = (0.25, 0.25, 0.25, 0.25)$. Note that $X_1 + X_2 + \cdots + X_4 = 100$, and they are dependent.

Joint PMF For $n = n_1 + n_2 + \cdots + n_k$,

$$P(\vec{X} = \vec{n}) = \frac{n!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

Marginal PMF, Lumping, and Conditionals Marginally, $X_i \sim \mathrm{Bin}(n, p_i)$ since we can define "success" to mean category i. If you lump together multiple categories in a Multinomial, then it is still Multinomial. For example, $X_i + X_j \sim \mathrm{Bin}(n, p_i + p_j)$ for $i \neq j$ since we can define "success" to mean being in category i or j. Similarly, if $k = 6$ and we lump categories 1-2 and lump categories 3-5, then

$$(X_1 + X_2, X_3 + X_4 + X_5, X_6) \sim \mathrm{Mult}_3(n, (p_1 + p_2, p_3 + p_4 + p_5, p_6))$$

Conditioning on some X_j also still gives a Multinomial:

$$X_1, \ldots, X_{k-1}|X_k = n_k \sim \mathrm{Mult}_{k-1}\left(n - n_k, \left(\frac{p_1}{1 - p_k}, \ldots, \frac{p_{k-1}}{1 - p_k}\right)\right)$$

Variances and Covariances We have $X_i \sim \mathrm{Bin}(n, p_i)$ marginally, so $\mathrm{Var}(X_i) = np_i(1 - p_i)$. Also, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

Multivariate Uniform Distribution

See the univariate Uniform for stories and examples. For the 2D Uniform on some region, probability is proportional to area. Every point in the support has equal density, of value $\frac{1}{\text{area of region}}$. For the 3D Uniform, probability is proportional to volume.

Multivariate Normal (MVN) Distribution

A vector $\vec{X} = (X_1, X_2, \ldots, X_k)$ is Multivariate Normal if every linear combination is Normally distributed, i.e., $t_1X_1 + t_2X_2 + \cdots + t_kX_k$ is Normal for any constants $t_1, t_2, \ldots, t_k$. The parameters of the Multivariate Normal are the **mean vector** $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and the **covariance matrix** where the (i, j) entry is $\mathrm{Cov}(X_i, X_j)$.

Properties The Multivariate Normal has the following properties.

- Any subvector is also MVN.
- If any two elements within an MVN are uncorrelated, then they are independent.
- The joint PDF of a Bivariate Normal (X, Y) with $\mathcal{N}(0, 1)$ marginal distributions and correlation $\rho \in (-1, 1)$ is

  $$f_{X,Y}(x, y) = \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2\tau^2}(x^2 + y^2 - 2\rho xy)\right),$$

  with $\tau = \sqrt{1 - \rho^2}$.

Distribution Properties

Important CDFs

Standard Normal Φ

Exponential(λ) $F(x) = 1 - e^{-\lambda x}$, for $x \in (0, \infty)$

Uniform(0,1) $F(x) = x$, for $x \in (0, 1)$

Convolutions of Random Variables

A convolution of n random variables is simply their sum. For the following results, let X and Y be *independent*.

1. $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2) \longrightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p) \longrightarrow X + Y \sim \text{Bin}(n_1 + n_2, p)$. $\text{Bin}(n, p)$ can be thought of as a sum of i.i.d. $\text{Bern}(p)$ r.v.s.
3. $X \sim \text{Gamma}(a_1, \lambda)$, $Y \sim \text{Gamma}(a_2, \lambda)$ $\longrightarrow X + Y \sim \text{Gamma}(a_1 + a_2, \lambda)$. $\text{Gamma}(n, \lambda)$ with n an integer can be thought of as a sum of i.i.d. $\text{Expo}(\lambda)$ r.v.s.
4. $X \sim \text{NBin}(r_1, p)$, $Y \sim \text{NBin}(r_2, p)$ $\longrightarrow X + Y \sim \text{NBin}(r_1 + r_2, p)$. $\text{NBin}(r, p)$ can be thought of as a sum of i.i.d. $\text{Geom}(p)$ r.v.s.
5. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ $\longrightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Special Cases of Distributions

1. $\text{Bin}(1, p) \sim \text{Bern}(p)$
2. $\text{Beta}(1, 1) \sim \text{Unif}(0, 1)$
3. $\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$
4. $\chi_n^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$
5. $\text{NBin}(1, p) \sim \text{Geom}(p)$

Inequalities

1. **Cauchy-Schwarz** $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$
2. **Markov** $P(X \geq a) \leq \frac{E|X|}{a}$ for $a > 0$
3. **Chebyshev** $P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$ for $E(X) = \mu, \text{Var}(X) = \sigma^2$
4. **Jensen** $E(g(X)) \geq g(E(X))$ for g convex; reverse if g is concave

Formulas

Geometric Series

$$1 + r + r^2 + \cdots + r^{n-1} = \sum_{k=0}^{n-1} r^k = \frac{1 - r^n}{1 - r}$$

$$1 + r + r^2 + \cdots = \frac{1}{1 - r} \text{ if } |r| < 1$$

Exponential Function (e^x)

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

Gamma and Beta Integrals

You can sometimes solve complicated-looking integrals by pattern-matching to a gamma or beta integral:

$$\int_0^{\infty} x^{t-1} e^{-x}\, dx = \Gamma(t) \qquad \int_0^1 x^{a-1}(1 - x)^{b-1}\, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

Also, $\Gamma(a + 1) = a\Gamma(a)$, and $\Gamma(n) = (n - 1)!$ if n is a positive integer.

Euler's Approximation for Harmonic Sums

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \log n + 0.577\ldots$$

Stirling's Approximation for Factorials

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Miscellaneous Definitions

Medians and Quantiles Let X have CDF F. Then X has median m if $F(m) \geq 0.5$ and $P(X \geq m) \geq 0.5$. For X continuous, m satisfies $F(m) = 1/2$. In general, the ath quantile of X is $\min\{x : F(x) \geq a\}$; the median is the case $a = 1/2$.

log Statisticians generally use log to refer to natural log (i.e., base e).

i.i.d r.v.s Independent, identically-distributed random variables.

Example Problems

Contributions from Sebastian Chiu

Calculating Probability

A textbook has n typos, which are randomly scattered amongst its n pages, independently. You pick a random page. What is the probability that it has no typos? **Answer:** There is a $\left(1 - \frac{1}{n}\right)$ probability that any specific typo isn't on your page, and thus a $\boxed{\left(1 - \frac{1}{n}\right)^n}$ probability that there are no typos on your page. For n large, this is approximately $e^{-1} = 1/e$.

Linearity and Indicators (1)

In a group of n people, what is the expected number of distinct birthdays (month and day)? What is the expected number of birthday matches? **Answer:** Let X be the number of distinct birthdays and I_j be the indicator for the jth day being represented.

$$E(I_j) = 1 - P(\text{no one born on day } j) = 1 - (364/365)^n$$

By linearity, $\boxed{E(X) = 365\left(1 - (364/365)^n\right)}$. Now let Y be the number of birthday matches and J_i be the indicator that the ith pair of people have the same birthday. The probability that any two specific people share a birthday is 1/365, so $\boxed{E(Y) = \binom{n}{2}/365}$.

Linearity and Indicators (2)

This problem is commonly known as the ***hat-matching problem.*** There are n people at a party, each with hat. At the end of the party, they each leave with a random hat. What is the expected number of people who leave with the right hat? **Answer:** Each hat has a $1/n$ chance of going to the right person. By linearity, the average number of hats that go to their owners is $\boxed{n(1/n) = 1}$.

Linearity and First Success

This problem is commonly known as the ***coupon collector problem.*** There are n coupon types. At each draw, you get a uniformly random coupon type. What is the expected number of coupons needed until you have a complete set? **Answer:** Let N be the number of coupons needed; we want $E(N)$. Let $N = N_1 + \cdots + N_n$, where N_1 is the draws to get our first new coupon, N_2 is the *additional* draws needed to draw our second new coupon and so on. By the story of the First Success, $N_2 \sim \text{FS}((n-1)/n)$ (after collecting first coupon type, there's $(n-1)/n$ chance you'll get something new). Similarly, $N_3 \sim \text{FS}((n-2)/n)$, and $N_j \sim \text{FS}((n - j + 1)/n)$. By linearity,

$$E(N) = E(N_1) + \cdots + E(N_n) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = \boxed{n \sum_{j=1}^{n} \frac{1}{j}}$$

This is approximately $n(\log(n) + 0.577)$ by Euler's approximation.

Orderings of i.i.d. random variables

I call 2 UberX's and 3 Lyfts at the same time. If the time it takes for the rides to reach me are i.i.d., what is the probability that all the Lyfts will arrive first? **Answer:** Since the arrival times of the five cars are i.i.d., all 5! orderings of the arrivals are equally likely. There are 3!2! orderings that involve the Lyfts arriving first, so the probability that the Lyfts arrive first is $\boxed{\frac{3!2!}{5!} = 1/10}$. Alternatively, there are $\binom{5}{3}$ ways to choose 3 of the 5 slots for the Lyfts to occupy, where each of the choices are equally likely. One of these choices has all 3 of the Lyfts arriving first, so the probability is $\boxed{1/\binom{5}{3} = 1/10}$.

Expectation of Negative Hypergeometric

What is the expected number of cards that you draw before you pick your first Ace in a shuffled deck (not counting the Ace)? **Answer:** Consider a non-Ace. Denote this to be card j. Let I_j be the indicator that card j will be drawn before the first Ace. Note that $I_j = 1$ says that j is before all 4 of the Aces in the deck. The probability that this occurs is 1/5 by symmetry. Let X be the number of cards drawn before the first Ace. Then $X = I_1 + I_2 + \ldots + I_{48}$, where each indicator corresponds to one of the 48 non-Aces. Thus,

$$E(X) = E(I_1) + E(I_2) + \ldots + E(I_{48}) = 48/5 = \boxed{9.6}.$$

Minimum and Maximum of RVs

What is the CDF of the maximum of n independent Unif(0,1) random variables? **Answer:** Note that for r.v.s $X_1, X_2, \ldots, X_n$,

$$P(\min(X_1, X_2, \ldots, X_n) \geq a) = P(X_1 \geq a, X_2 \geq a, \ldots, X_n \geq a)$$

Similarly,

$$P(\max(X_1, X_2, \ldots, X_n) \leq a) = P(X_1 \leq a, X_2 \leq a, \ldots, X_n \leq a)$$

We will use this principle to find the CDF of $U_{(n)}$, where $U_{(n)} = \max(U_1, U_2, \ldots, U_n)$ and $U_i \sim \text{Unif}(0, 1)$ are i.i.d.

$$\begin{aligned} P(\max(U_1, U_2, \ldots, U_n) \leq a) &= P(U_1 \leq a, U_2 \leq a, \ldots, U_n \leq a) \\ &= P(U_1 \leq a)P(U_2 \leq a) \ldots P(U_n \leq a) \\ &= \boxed{a^n} \end{aligned}$$

for $0 < a < 1$ (and the CDF is 0 for $a \leq 0$ and 1 for $a \geq 1$).

Pattern-matching with e^x Taylor series

For $X \sim \text{Pois}(\lambda)$, find $E\left(\frac{1}{X+1}\right)$. **Answer:** By LOTUS,

$$E\left(\frac{1}{X+1}\right) = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} = \boxed{\frac{e^{-\lambda}}{\lambda}(e^{\lambda} - 1)}$$

Adam's Law and Eve's Law

William really likes speedsolving Rubik's Cubes. But he's pretty bad at it, so sometimes he fails. On any given day, William will attempt $N \sim \text{Geom}(s)$ Rubik's Cubes. Suppose each time, he has probability p of solving the cube, independently. Let T be the number of Rubik's Cubes he solves during a day. Find the mean and variance of T.
Answer: Note that $T|N \sim \text{Bin}(N, p)$. So by Adam's Law,

$$E(T) = E(E(T|N)) = E(Np) = \boxed{\frac{p(1-s)}{s}}$$

Similarly, by Eve's Law, we have that

$$\text{Var}(T) = E(\text{Var}(T|N)) + \text{Var}(E(T|N)) = E(Np(1-p)) + \text{Var}(Np)$$
$$= \frac{p(1-p)(1-s)}{s} + \frac{p^2(1-s)}{s^2} = \boxed{\frac{p(1-s)(p+s(1-p))}{s^2}}$$

MGF – Finding Moments

Find $E(X^3)$ for $X \sim \text{Expo}(\lambda)$ using the MGF of X. **Answer:** The MGF of an Expo(λ) is $M(t) = \frac{\lambda}{\lambda - t}$. To get the third moment, we can take the third derivative of the MGF and evaluate at $t = 0$:

$$\boxed{E(X^3) = \frac{6}{\lambda^3}}$$

But a much nicer way to use the MGF here is via pattern recognition: note that $M(t)$ looks like it came from a geometric series:

$$\frac{1}{1 - \frac{t}{\lambda}} = \sum_{n=0}^{\infty} \left(\frac{t}{\lambda}\right)^n = \sum_{n=0}^{\infty} \frac{n!}{\lambda^n} \frac{t^n}{n!}$$

The coefficient of $\frac{t^n}{n!}$ here is the nth moment of X, so we have $E(X^n) = \frac{n!}{\lambda^n}$ for all nonnegative integers n.

Markov chains (1)

Suppose X_n is a two-state Markov chain with transition matrix

$$Q = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-\alpha & \alpha \\ 1 & \beta & 1-\beta \end{array}$$

Find the stationary distribution $\vec{s} = (s_0, s_1)$ of X_n by solving $\vec{s}Q = \vec{s}$, and show that the chain is reversible with respect to $\vec{s}$. **Answer:** The equation $\vec{s}Q = \vec{s}$ says that

$$s_0 = s_0(1-\alpha) + s_1\beta \text{ and } s_1 = s_0(\alpha) + s_0(1-\beta)$$

By solving this system of linear equations, we have

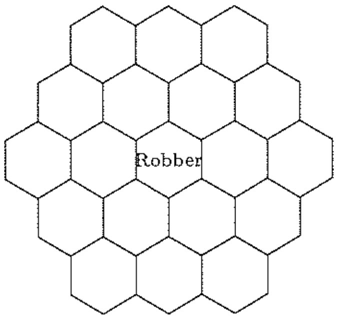
$$\boxed{\vec{s} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)}$$

To show that the chain is reversible with respect to $\vec{s}$, we must show $s_i q_{ij} = s_j q_{ji}$ for all i, j. This is done if we can show $s_0 q_{01} = s_1 q_{10}$. And indeed,

$$s_0 q_{01} = \frac{\alpha\beta}{\alpha+\beta} = s_1 q_{10}$$

Markov chains (2)

William and Sebastian play a modified game of Settlers of Catan, where every tur y randomly move the robber (which starts on the center tile) to c the adjacent hexagons.



(a) Is this Markov chain irreducible? Is it aperiodic? **Answer:** $\boxed{\text{Yes to both.}}$ The Markov chain is irreducible because it can get from anywhere to anywhere else. The Markov chain is aperiodic because the robber can return back to a square in $2, 3, 4, 5, \ldots$ moves, and the GCD of those numbers is 1.

(b) What is the stationary distribution of this Markov chain? **Answer:** Since this is a random walk on an undirected graph, the stationary distribution is proportional to the degree sequence. The degree for the corner pieces is 3, the degree for the edge pieces is 4, and the degree for the center pieces is 6. To normalize this degree sequence, we divide by its sum. The sum of the degrees is $6(3) + 6(4) + 7(6) = 84$. Thus the stationary probability of being on a corner is $3/84 = 1/28$, on an edge is $4/84 = 1/21$, and in the center is $6/84 = 1/14$.

(c) What fraction of the time will the robber be in the center tile in this game, in the long run? **Answer:** By the above, $\boxed{1/14}$.

(d) What is the expected amount of moves it will take for the robber to return to the center tile? **Answer:** Since this chain is irreducible and aperiodic, to get the expected time to return we can just invert the stationary probability. Thus on average it will take $\boxed{14}$ turns for the robber to return to the center tile.

Problem-Solving Strategies

Contributions from Jessy Hwang, Yuan Jiang, Yuqi Hou

1. **Getting started.** Start by *defining relevant events and random variables.* ("Let A be the event that I pick the fair coin"; "Let X be the number of successes.") Clear notion is important for clear thinking! Then decide what it is that you're supposed to be finding, in terms of your notation ("I want to find $P(X = 3|A)$"). Think about what type of object your answer should be (a number? A random variable? A PMF? A PDF?) and what it should be in terms of.

   Try simple and extreme cases. To make an abstract experiment more concrete, try *drawing a picture* or making up numbers that could have happened. Pattern recognition: does the structure of the problem resemble something we've seen before?

2. **Calculating probability of an event.** Use counting principles if the naive definition of probability applies. Is the probability of the complement easier to find? Look for symmetries. Look for something to condition on, then apply Bayes' Rule or the Law of Total Probability.

3. **Finding the distribution of a random variable.** First make sure you need the full distribution not just the mean (see next item). Check the *support* of the random variable: what values can it take on? Use this to rule out distributions that don't fit. Is there a *story* for one of the named distributions that fits the problem at hand? Can y vrite the random variable as a function of an r.v. with own distribution, say $Y = g(X)$?

4. **Calculating expectation.** If it has a named distribution, check out the table of distributions. If it's a function of an r.v. with a named distribution, try LOTUS. If it's a count of something, try breaking it up into indicator r.v.s. If you can condition on something natural, consider using Adam's law.

5. **Calculating variance.** Consider independence, named distributions, and LOTUS. If it's a count of something, break it up into a sum of indicator r.v.s. If it's a sum, use properties of covariance. If you can condition on something natural, consider using Eve's Law.

6. **Calculating $E(X^2)$.** Do you already know $E(X)$ or $\text{Var}(X)$? Recall that $\text{Var}(X) = E(X^2) - (E(X))^2$. Otherwise try LOTUS.

7. **Calculating covariance.** Use the properties of covariance. If you're trying to find the covariance between two components of a Multinomial distribution, X_i, X_j, then the covariance is $-np_ip_j$ for $i \neq j$.

8. **Symmetry.** If $X_1, \ldots, X_n$ are i.i.d., consider using symmetry.

9. **Calculating probabilities of orderings.** Remember that all $n!$ ordering of i.i.d. continuous random variables $X_1, \ldots, X_n$ are equally likely.

10. **Determining independence.** There are several equivalent definitions. Think about simple and extreme cases to see if you can find a counterexample.

11. **Do a painful integral.** If your integral looks painful, see if you can write your integral in terms of a known PDF (like Gamma or Beta), and use the fact that PDFs integrate to 1?

12. **Before moving on.** Check some simple and extreme cases, check whether the answer seems plausible, check for biohazards.

Biohazards

Contributions from Jessy Hwang

1. **Don't misuse the naive definition of probability.** When answering "What is the probability that in a group of 3 people, no two have the same birth month?", it is *not* correct to treat the people as indistinguishable balls being placed into 12 boxes, since that assumes the list of birth months {January, January, January} is just as likely as the list {January, April, June}, even though the latter is six times more likely.

2. **Don't confuse unconditional, conditional, and joint probabilities.** In applying $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, it is *not* correct to say "$P(B) = 1$ because we know B happened"; $P(B)$ is the *prior* probability of B. Don't confuse $P(A|B)$ with $P(A, B)$.

3. **Don't assume independence without justification.** In the matching problem, the probability that card 1 is a match and card 2 is a match is not $1/n^2$. Binomial and Hypergeometric are often confused; the trials are independent in the Binomial story and dependent in the Hypergeometric story.

4. **Don't forget to do sanity checks.** Probabilities must be between 0 and 1. Variances must be ≥ 0. Supports must make sense. PMFs must sum to 1. PDFs must integrate to 1.

5. **Don't confuse random variables, numbers, and events.** Let X be an r.v. Then $g(X)$ is an r.v. for any function g. In particular, X^2, $|X|$, $F(X)$, and $I_{X>3}$ are r.v.s. $P(X^2 < X|X \geq 0)$, $E(X)$, $\text{Var}(X)$, and $g(E(X))$ are numbers. $X = 2$ and $F(X) \geq -1$ are events. It does not make sense to write $\int_{-\infty}^{\infty} F(X)dx$, because $F(X)$ is a random variable. It does not make sense to write $P(X)$, ' se X is not an event.

6. **Don't co e a random variable with its distribution.** To get the PDF of X^2, you can't just square the PDF of X. The right way is to use transformations. To get the PDF of $X+Y$, you can't just add the PDF of X and the PDF of Y. The right way is to compute the convolution.

7. **Don't pull non-linear functions out of expectations.** $E(g(X))$ does not equal $g(E(X))$ in general. The St. Petersburg paradox is an extreme example. See also Jensen's inequality. The right way to find $E(g(X))$ is with LOTUS.

Distributions in R

Command	What it does
`help(distributions)`	shows documentation on distributions
`dbinom(k,n,p)`	PMF $P(X=k)$ for $X \sim \text{Bin}(n,p)$
`pbinom(x,n,p)`	CDF $P(X \leq x)$ for $X \sim \text{Bin}(n,p)$
`qbinom(a,n,p)`	ath quantile for $X \sim \text{Bin}(n,p)$
`rbinom(r,n,p)`	vector of r i.i.d. $\text{Bin}(n,p)$ r.v.s
`dgeom(k,p)`	PMF $P(X=k)$ for $X \sim \text{Geom}(p)$
`dhyper(k,w,b,n)`	PMF $P(X=k)$ for $X \sim \text{HGeom}(w,b,n)$
`dnbinom(k,r,p)`	PMF $P(X=k)$ for $X \sim \text{NBin}(r,p)$
`dpois(k,r)`	PMF $P(X=k)$ for $X \sim \text{Pois}(r)$
`dbeta(x,a,b)`	PDF $f(x)$ for $X \sim \text{Beta}(a,b)$
`dchisq(x,n)`	PDF $f(x)$ for $X \sim \chi_n^2$
`dexp(x,b)`	PDF $f(x)$ for $X \sim \text{Expo}(b)$
`dgamma(x,a,r)`	PDF $f(x)$ for $X \sim \text{Gamma}(a,r)$
`dlnorm(x,m,s)`	PDF $f(x)$ for $X \sim \mathcal{LN}(m,s^2)$
`dnorm(x,m,s)`	PDF $f(x)$ for $X \sim \mathcal{N}(m,s^2)$
`dt(x,n)`	PDF $f(x)$ for $X \sim t_n$
`dunif(x,a,b)`	PDF $f(x)$ for $X \sim \text{Unif}(a,b)$

The table above gives R commands for working with various named distributions. Commands analogous to `pbinom`, `qbinom`, and `rbinom` work for the other distributions in the table. For example, `pnorm`, `qnorm`, and `rnorm` can be used to get the CDF, quantiles, and random generation for the Normal. For the Multinomial, `dmultinom` can be used for calculating the joint PMF and `rmultinom` can be used for generating random vectors. For the Multivariate Normal, after installing and loading the `mvtnorm` package `dmvnorm` can be used for calculating the joint PDF and `rmvnorm` can be used for generating random vectors.

Recommended Res rces

- Introduction to Probability Book (`http://bit.ly/introprobability`)
- Stat 110 Online (`http://stat110.net`)
- Stat 110 Quora Blog (`https://stat110.quora.com/`)
- Quora Probability FAQ (`http://bit.ly/probabilityfaq`)
- R Studio (`https://www.rstudio.com`)
- LaTeX File (`github.com/wzchen/probability_cheatsheet`)

Please share this cheatsheet with friends!
`http://wzchen.com/probability-cheatsheet`

Table of Distributions

Distribution	PMF/PDF and Support	Expected Value	Variance	MGF
Bernoulli Bern(p)	$P(X=1)=p$ $P(X=0)=q=1-p$	p	pq	$q+pe^t$
Binomial Bin(n,p)	$P(X=k)=\binom{n}{k}p^kq^{n-k}$ $k \in \{0,1,2,\dots n\}$	np	npq	$(q+pe^t)^n$
Geometric Geom(p)	$P(X=k)=q^kp$ $k \in \{0,\,1,\,2,\,\dots\}$	q/p	q/p^2	$\frac{p}{1-qe^t}$, $qe^t<1$
Negative Binomial NBin(r,p)	$P(X=n)=\binom{r+n-1}{r-1}p^rq^n$ $n \in \{0,\,1,\,2,\,\dots\}$	rq/p	rq/p^2	$(\frac{p}{1-qe^t})^r$, $qe^t<1$
Hypergeometric HGeom(w,b,n)	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k \in \{0,1,2,\dots,n\}$	$\mu=\frac{nw}{b+w}$	$\left(\frac{w+b-n}{w+b-1}\right)n\frac{\mu}{n}(1-\frac{\mu}{n})$	messy
Poisson Pois(λ)	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k \in \{0,\,1,\,2,\,\dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Uniform Unif(a,b)	$f(x)=\frac{1}{b-a}$ $x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$ $x \in (-\infty,\infty)$	μ	σ^2	$e^{t\mu+\frac{\sigma^2t^2}{2}}$
Exponential Expo(λ)	$f(x)=\lambda e^{-\lambda x}$ $x \in (0,\infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$, $t<\lambda$
Gamma Gamma(a,λ)	$f(x)=\frac{1}{\Gamma(a)}(\lambda x)^ae^{-\lambda x}\frac{1}{x}$ $x \in (0,\infty)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^a$, $t<\lambda$
Beta Beta(a,b)	$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $x \in (0,1)$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{(a+b+1)}$	messy
Log-Normal $\mathcal{LN}(\mu,\sigma^2)$	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}$ $x \in (0,\infty)$	$\theta=e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$	doesn't exist
Chi-Square χ^2_n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x \in (0,\infty)$	n	$2n$	$(1-2t)^{-n/2}$, $t<1/2$
Student-t t_n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$ $x \in (-\infty,\infty)$	0 if $n>1$	$\frac{n}{n-2}$ if $n>2$	doesn't exist

Quick Study ACADEMIC

STATISTICS

parameters.variables.intervals.proportions

THE BASIC PRINCIPLES OF STATISTICS FOR INTRODUCTORY COURSES

DEFINITIONS

☐ **STATISTICS** - A set of tools for collecting, organizing, presenting, and analyzing numerical facts or observations.

1. **Descriptive Statistics** - procedures used to organize and present data in a convenient, useable, and communicable form.
2. **Inferential Statistics** - procedures employed to arrive at broader generalizations or inferences from sample data to populations.

☐ **STATISTIC** - A number describing a sample characteristic. Results from the manipulation of sample data according to certain specified procedures.

☐ **DATA** - Characteristics or numbers that are collected by observation.

☐ **POPULATION** - A complete set of actual or potential observations.

☐ **PARAMETER** - A number describing a population characteristic; typically, inferred from sample statistic.

☐ **SAMPLE** - A subset of the population selected according to some scheme.

☐ **RANDOM SAMPLE** - A subset selected in such a way that each member of the population has an equal opportunity to be selected. **Ex.** ***lottery numbers in a fair lottery***

☐ **VARIABLE** - A phenomenon that may take different values.

☐ **MEAN** - The point in a distribution of measurements about which the summed deviations are equal to zero. ***Average value of a sample or population.***

POPULATION MEAN **SAMPLE MEAN**

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Note: The mean is very sensitive to extreme measurements that are not balanced on both sides.

☐ **WEIGHTED MEAN** - Sum of a set of observations multiplied by their respective weights, divided by the sum of the weights:

$$\textbf{WEIGHTED MEAN} \quad \frac{\sum_{i=1}^{G} w_i x_i}{\sum_{i=1}^{G} w_i}$$

where w_i = weight; x_i = observation; G = number of observation groups. Calculated from a population, sample, or groupings in a frequency distribution.

Ex. ***In the FrequencyDistribution below, the mean is 80.3; calculated by using frequencies for the w_i's. When grouped, use class midpoints for x_i's.***

☐ **MEDIAN** - Observation or potential observation in a set that divides the set so that the same number of observations lie on each side of it. For an odd number of values, it is the middle value; for an even number it is the average of the middle two.

Ex. ***In the Frequency Distribution table below, the median is 79.5.***

☐ **MODE** - Observation that occurs with the greatest frequency. **Ex.** ***In the Frequency Distribution table below, the mode is 88.***

MEASURES OF DISPERSION

☐ **SUM OF SQUARES *(SS)*** - Deviations from the mean, squared and summed:

$$\text{Population SS} = \Sigma(x_i - \mu_x)^2 \text{ or } \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N}$$

$$\text{Sample SS} = \Sigma(x_i - \bar{x})^2 \text{ or } \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

☐ **VARIANCE** - The average of square differences between observations and their mean.

POPULATION VARIANCE **SAMPLE VARIANCE**

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

VARIANCES FOR GROUPED DATA

POPULATION **SAMPLE**

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{G} f_i(m_i-\mu)^2 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{G} f_i(m_i-\bar{x})^2$$

☐ **STANDARD DEVIATION** - Square root of the variance:

Ex. ***Pop. S.D.*** $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$

FREQUENCY DISTRIBUTION

Shows the number of times each observation occurs when the values of a variable are arranged in order according to their magnitudes.

FREQUENCY DISTRIBUTION

Frequency Distribution of student scores on an exam

x	*f*	*x*	*f*	*x*	*f*	*x*	*f*
100	1	83	11	74	111	65	0
99	1	84	11111	75	1111	66	1
98	0	85	1	76	11	67	11
97	0	86	0	77	111	68	1
96	11	87	1	78	1	69	111
95	0	88	1111111	79	11	70	1111
94	0	89	111	80	1	71	0
93	1	90	11	81	11	72	11
92	0	91	1	82	1	73	111

x = observation *f* = frequency

☐ **GROUPED FREQUENCY DISTRIBUTION** - A frequency distribution in which the values of the variable have been grouped into classes.

GROUPED FREQUENCY DISTRIBUTION

CLASS	*f*	*CLASS*	*f*
98-100	2	80-82	4
95-97	2	77-79	6
92-94	1	74-76	9
89-91	6	71-73	5
86-88	8	68-70	8
83-85	8	65-67	3

GROUPING OF DATA

CUMULATIVE FREQUENCY/PERCENTAGE DISTRIBUTIONS

☐ **CUMULATIVE FREQUENCY DISTRIBUTION** - A distribution which shows the total frequency through the upper real limit of each class.

☐ **CUMULATIVE PERCENTAGE DISTRIBUTION**- A distribution which shows the total percentage through the upper real limit of each class.

CUMULATIVE FREQUENCY / PERCENTAGE DISTRIBUTION

CLASS	f	Cum f	%
65-67	3	3	4.84
68-70	8	11	17.74
71-73	5	16	25.81
74-76	9	25	40.32
77-79	6	31	50.00
80-82	4	35	56.45
83-85	8	43	69.35
86-88	8	51	82.26
89-91	6	57	91.94
92-94	1	58	93.55
95-97	2	60	96.77
98-100	2	62	100.00

GRAPHING TECHNIQUES

☐ **BAR GRAPH** - A form of graph that uses bars to indicate the frequency of occurrence of observations.

- **Histogram** - a form of bar graph used with interval or ratio-scaled variables.

  - **Interval Scale-** a quantitative scale that permits the use of arithmetic operations. The zero point in the scale is arbitrary.

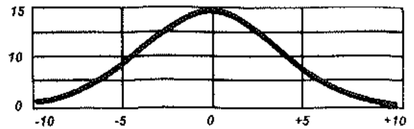
  - **Ratio Scale-** same as interval scale except that there is a true zero point.

☐ **FREQUENCY CURVE** - A form of graph representing a frequency distribution in the form of a continuous line that traces a histogram.

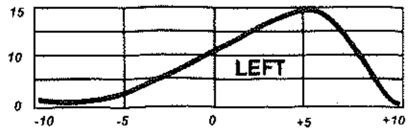
- **Cumulative Frequency Curve** - a continuous line that traces a histogram where bars in all the lower classes are stacked up in the adjacent higher class. It cannot have a negative slope.
- **Normal curve** - bell-shaped curve.
- **Skewed curve** - departs from symmetry and tails-off at one end.

FREQUENCY CURVES

NORMAL CURVE



SKEWED CURVE



PROBABILITY

The long term relative frequency with which an outcome or event occurs.

Probability of occurrence of Event A $\rho(A) = \frac{\textit{Number of outcomes favoring Event A}}{\textit{Total number of outcomes}}$

☐ **SAMPLE SPACE** - All possible outcomes of an experiment.

☐ **TYPE OF EVENTS**

- **Exhaustive** - two or more events are said to be exhaustive if all possible outcomes are considered.
  Symbolically, ρ (A or B or...) = 1.
- **Non-Exhaustive** -two or more events are said to be non-exhaustive if they do not exhaust all possible outcomes.
- **Mutually Exclusive** - Events that cannot occur simultaneously: ρ(A and B) = 0; and ρ(A or B) = ρ(A) + ρ(B).
  Ex. ***males, females***
- **Non-Mutually Exclusive** - Events that can occur simultaneously: ρ(A or B) = ρ(A) + ρ(B) - ρ(A and B).
  Ex. ***males, brown eyes.***
- **Independent** - Events whose probability is unaffected by occurrence or nonoccurrence of each other: $\rho(A|B) = \rho(A)$; $\rho(B|A) = \rho(B)$; and ρ(A and B) = $\rho(A)\,\rho(B)$.
  Ex. ***gender and eye color***
- **Dependent** - Events whose probability changes depending upon the occurrence or non-occurrence of each other: $\rho(A|B)$ differs from $\rho(A)$; $\rho(B|A)$ differs from $\rho(B)$; and ρ(A and B) = $\rho(A)\,\rho(B|A) = \rho(B)\,\rho(A|B)$
  Ex. ***race and eye color.***

☐ **JOINT PROBABILITIES** - Probability that 2 or more events occur simultaneously.

☐ **MARGINAL PROBABILITIES** or Unconditional Probabilities = summation of probabilities.

☐ **CONDITIONAL PROBABILITIES** - Probability of *A* given the existence of ***S***, written, ρ (***A\S***).

☐ **EXAMPLE**- Given the numbers 1 to 9 as observations in a sample space:

- **Events mutually exclusive and exhaustive-**
  Example: ρ ***(all odd numbers)***; ρ ***(all even numbers)***
- **Events mutually exclusive but not exhaustive-**
  Example: ρ ***(an even number)***; ρ ***(the numbers 7 and 5)***
- **Events neither mutually exclusive or exhaustive-**
  Example: ρ ***(an even number or a 2)***

FREQUENCY TABLE

	EVENT C	EVENT D	TOTALS
EVENT E	52	36	87
EVENT F	62	71	133
TOTALS	114	106	220

Ex. Joint Probability Between C and E
ρ (C & E) = 52/220 = 0.24

JOINT, MARGINAL & CONDITIONAL PROBABILITY TABLE

	EVENT C	EVENT D	MARGINAL PROBABILITY	CONDITIONAL PROBABILITY
EVENT E	0.24	0.16	0.40	(C/E)=0.60 (D/E)=0.40
EVENT F	0.28	0.32	0.60	(C/F)=0.47 (D/F)=0.53
MARGINAL PROBABILITY	0.52	0.48	1.00	
CONDITIONAL PROBABILITY	(E/C)=0.46 (F/C)=0.54	(E/D)=0.33 (F/D)=0.67		

☐ **SAMPLING DISTRIBUTION** - A theoretical ***probability distribution*** of a statistic that would result from drawing all possible samples of a given size from some population.

THE STANDARD ERROR OF THE MEAN

A theoretical standard deviation of sample mean of a given sample size, drawn from some specified population.

☐When based on a very large, known population, the standard error is: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

☐When estimated from a sample drawn from very large population, the standard error is:

$$\sigma_{\overline{x}} \approx \frac{s}{\sqrt{\mathbf{n}}}$$

☐The dispersion of sample means decreases as sample

RANDOM VARIABLES

A mapping or function that assigns one and only one numerical value to each outcome in an experiment.

☐ **DISCRETE RANDOM VARIABLES** - Involves rules or probability models for assigning or generating only distinct values (not fractional measurements).

☐ **BINOMIAL DISTRIBUTION** - A model for the sum of a series of **n** independent trials where trial results in a 0 (failure) or 1 (success). **Ex.** ***Coin toss*** $p(s) = \binom{n}{s}\pi^{s}(1-\pi)^{n-s}$

where **p(s)** is the probability of **s** success in **n** trials with a constant π probability per trials, and where $\binom{n}{s} = \frac{n!}{s!(n-s)!}$

Binomial mean: $\mu = n\pi$

Binomial variance: $\sigma^2 = n\pi(1-\pi)$

As n increases, the Binomial approaches the Normal distribution.

☐ **HYPERGEOMETRIC DISTRIBUTION** - A model for the sum of a series of **n** trials where each trial results in a 0 or 1 and is drawn from a small population with **N** elements split between $\mathbf{N_1}$ successes and $\mathbf{N_2}$ failures. Then the probability of splitting the **n** trials between $\mathbf{x_1}$ successes and $\mathbf{x_2}$ failures is:

$$p(x_1 \text{ and } x_2) = \frac{\dfrac{N_1!}{x_1!(N_1-x_1)!}\,\dfrac{N_2!}{x_2!(N_2-x_2)!}}{\dfrac{N!}{n!(N-n)!}}$$

Hypergeometric mean: $\mu_1 = E(x_1) = \frac{nN_1}{N}$

and variance: $\sigma^2 = \frac{N-n}{N-1}\left[\frac{nN_1}{N}\right]\left[\frac{N_2}{N}\right]$

☐ **POISSON DISTRIBUTION** - A model for the number of occurrences of an event x = 0,1,2,..., when the probability of occurrence is small, but the number of opportunities for the occurrence is large, for **x = 0,1,2,3,...** and $\lambda > 0$, otherwise $P(x) = 0$.

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Poisson mean and variance: λ.

For continuous variables, frequencies are expressed in terms of areas under a curve.

☐ **CONTINUOUS RANDOM VARIABLES** - Variable that may take on any value along an uninterrupted interval of a numberline.

☐ **NORMAL DISTRIBUTION** - bell curve; a distribution whose values cluster symmetrically around the mean (also median and mode).

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

where $f(x)$ = frequency at a given value
σ = standard deviation of the distribution
π = approximately 3.1416
e = approximately 2.7183
μ = the mean of the distribution
x = any score in the distribution

☐ **STANDARD NORMAL DISTRIBUTION** - A normal random variable Z, that has a mean of 0, and standard deviation of 1.

☐ **Z-VALUES** - The number of standard deviations a specific observation lies from the mean:

$$z = \frac{x-\mu}{\sigma}$$

TESTING STATISTICAL HYPOTHESES

☐ **LEVEL OF SIGNIFICANCE** - A probabil value considered rare in the sampling distributic specified under the null hypothesis where one willing to acknowledge the operation c factors. Common significance levels ar 5%, 10%. Alpha (α) level = the lowest level for which the null hypothesis can be rejecte The significance level determines the critical regic

☐ **NULL HYPOTHESIS** (H_0) - A stateme that specifies hypothesized value(s) for one more of the population parameter. [**Ex.** H_0 = coin is unbiased. That is $p = 0.5$.]

☐ **ALTERNATIVE HYPOTHESIS** (H_1) - statement that specifies that the populati parameter is some value other than the o specified under the null hypothesis. [**Ex.** H_1= a cc is biased. That is $p \neq 0.5$.]

1. NONDIRECTIONAL HYPOTHESI an alternative hypothesis (H_1) that states or that the population parameter is different frc the one specified under H_0. **Ex.** $H_1: \mu \neq \mu$ Two-Tailed Probability Value is employed wh the alternative hypothesis is non-directional.

2. DIRECTIONAL HYPOTHESIS - alternative hypothesis that states the direction which the population parameter differs from t one specified under H_0. **Ex.** $H_1: \mu > \mu_0$ *or* $H_1: \mu <$ One-Tailed Probability Value is employed wh the alternative hypothesis is directional.

☐ **NOTION OF INDIRECT PROOF** - St interpretation of hypothesis testing reveals that t null hypothesis can never be proved. [**Ex.** If we t a coin 200 times and tails comes up 100 times. i no guarantee that heads will come up exactly h the time in the long run; small discrepancies mi exist. A bias can exist even at a small magnitu We can make the assertion however that N BASIS EXISTS FOR REJECTING TI HYPOTHESIS THAT THE COIN UNBIASED. (*The null hypothesis is not* When employing the 0.05 level of sign reject the null hypothesis when a given res occurs by chance 5% of the time or less.]

☐ **TWO TYPES OF ERRORS**

- <u>Type 1 Error</u> (Type α Error) = the rejection H_0 when it is actually true. The probability a type 1 error is given by α.
- <u>Type II Error</u> (Type β Error) = The acceptai of H_0 when it is actually false. The probabil of a type II error is given by β.

Statistical Hypotheses		True Status of H_0	
		H_0 True	H_0 False
Decision:	Accept H_0	Correct (1-α)	Type II error (β)
	Reject H_0	Type I error (α)	Correct (1-β)

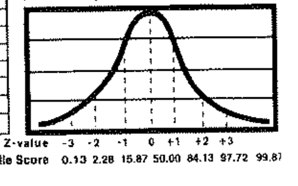
CENTRAL LIMIT THEORE

(for sample mean $\overline{x}$)

- If $x_1, x_2, x_3, \ldots x_n$, is a simple random sample c elements from a large (infinite) population, with me mu(μ) and standard deviation σ, then the distributio $\overline{x}$ takes on the bell shaped distribution of a norr random variable as n increases and the distribution of ratio: $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ approaches the standard normal distribution as n g to infinity. In practice, a normal approximatior acceptable for samples of 30 or larger.

z	% Cum (total area to the left of z)
-3	0.13
-2	2.28
-1	15.87
0	50.00
+1	84.13
+2	97.72
+3	99.87

Percentage Cumulative Distribution
for selected Z values under a normal curv



INFERENCE FOR PARAMETERS

☐ **UNBIASEDNESS** - Property of a reliable estimator being estimated.

- **Unbiased Estimate of a Parameter** - an estimate that equals on the average the value of the parameter.

Ex. ***the sample mean is an unbiased estimator of the population mean.***

- **Biased Estimate of a Parameter** - an estimate that does not equal on the average the value of the parameter.

Ex. ***the sample variance calculated with n is a biased estimator of the population variance, however, when calculated with n-1 it is unbiased.***

☐ **STANDARD ERROR** - The standard deviation of the estimator is called the standard error.

Ex. The standard error for $\bar{x}$'s is. $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

This has to be distinguished from the STANDARD DEVIATION OF THE SAMPLE:

$$s = \sqrt{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The standard error measures the variability in the $\bar{x}$'s around their expected value E($\bar{x}$) while the standard deviation of the sample reflects the variability in the sample around the sample's mean ($\bar{x}$).

☐ **USED WHEN THE STANDARD DEVIATION IS UNKNOWN** -Use of **Student's *t*.** When σ is not known, its value is estimated from [illegible] data.

• [illegible]io- the ratio employed in the testing of hypotheses or determining the significance of a difference between means (two-sample case) involving a sample with a t-distribution. The formula is:

$\frac{\bar{x} - \mu}{s_{\bar{x}}}$ where μ = population mean under H_0

and $s_{\bar{x}} = s/\sqrt{n}$

- **Distribution**-symmetrical distribution with a mean of zero and standard deviation that approaches one as degrees of freedom increases (i.e., approaches the Z distribution).

Assumption and condition required in assuming *t*-distribution: Samples are drawn from a normally distributed population and σ (population standard deviation) is unknown.

- **Homogeneity of Variance-** If 2 samples are being compared, the assumption in using t-ratio is that the variances of the populations from where the samples are drawn are equal.
- Estimated $\sigma_{\bar{X}_1 - \bar{X}_2}$ (that is $s_{\bar{X}_1 - \bar{X}_2}$) is based on the unbiased estimate of the population variance.
- **Degrees of Freedom (*df*)-** the number of values that are free to vary after placing certain restrictions on the data.

Example. ***The sample (43,74,42,65) has n = 4. The sum is 224 and mean = 56. Using these 4 numbers and determining deviations from the mean, we'll have 4 deviations namely (–13,18,–14,9) which sum up to zero. Deviations from the mean is one restriction we have imposed and the natural consequence is that the sum of these deviations should equal zero. For this to happen, we can choose any number but our freedom to choose is limited to only 3 numbers because one is [illegible] by the requirement that the sum of the de[illegible]s should equal zero. We use the equality:***

$$(x_1-\bar{x})+(x_2-\bar{x})+(x_3-\bar{x})+(x_4-\bar{x})=0$$

So given a mean of 56, if the first 3 observations are 43, 74, and 42, the last observation has to be 65. This single restriction in this case helps us determine df. The formula is n less number of restrictions. In this case, it is n–1= 4-1=3df.

- ***t*-Ratio is a robust test-** This means that statistical inferences are likely valid despite fairly large departures from normality in the population distribution. If normality of population distribution is in doubt, it is wise

USING THE Z - STATISTIC

☐ **USED WHEN THE STANDARD DEVIATION IS KNOWN:** When σ is known it is possible to describe the form of the distribution of the sample mean as a Z statistic. The sample must be drawn from a normal distribution or have a sample size (n) of at least 30.

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ where u = population mean (either known or hypothesized under H_o) and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

- **Critical Region** - the portion of the area under the curve which includes those values of a statistic that lead to the rejection of the null hypothesis.

- The most often used significance levels are 0.01, 0.05, and 0.1. For a one-tailed test using *z*-statistic, these correspond to z-values of 2.33, 1.65, and 1.28 respectively. For a two-tailed test, the critical region of 0.01 is split into two equal outer areas marked by z-values of |2.58|.

Example 1. ***Given a population with μ=250 and σ = 50, what is the probability of drawing a sample of n=100 values whose mean (x̄) is at least 255? In this case, Z=1.00. Looking at Table A, the given area for Z=1.00 is 0.3413. To its right is 0.1587(=0.5-0.3413) or 15.85%.***

Conclusion: there are approximately 16 chances in 100 of obtaining a sample mean = 255 from this population when n = 100.

Example 2. ***Assume we do not know the population mean. However, we suspect that it may have been selected from a population with μ = 250 and σ = 50, but we are not sure. The hypothesis to be tested is whether the sample mean was selected from this population. Assume we obtained from a sample (n) of 100, a sample mean of 263. Is it reasonable to assume that this sample was drawn from the suspected population?***

1. H_0:μ = 250 (that the actual mean of the population from which the sample is drawn is equal to 250) H_1: μ not equal to 250 (the alternative hypothesis is that it is greater than or less than 250, thus a two-tailed test).
2. *z*-statistic will be used because the population σ is known.
3. Assume the significance level (α) to be 0.01. Looking at Table A, we find that the area beyond a *z* of 2.58 is approximately 0.005.

To reject H_0 at the 0.01 level of significance, the absolute value of the obtained *z* must be equal to or greater than $|z_{0.01}|$ or 2.58. Here the value of z corresponding to sample mean = 263 is 2.60.

☐ **CONCLUSION-** Since this obtained *z* falls within the critical region, we may reject H_0 at the 0.01 level of significance.

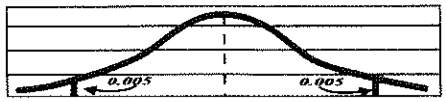
☐ **CONFIDENCE INTERVAL**- Interval within which we may consider a hypothesis tenable. Common confidence intervals are 90%, 95%, and 99%. Confidence Limits: limits defining the confidence interval.

(1- α)100% confidence interval for μ :

$$\bar{x}_i - z_{\alpha/2}(\sigma/\sqrt{n}) \le \mu \le \bar{x} + z_{\alpha/2}(\sigma/\sqrt{n})$$

where $Z_{\alpha/2}$ is the value of the standard normal variable z that puts α/2 percent in each tail of the distribution. The confidence interval is the complement of the critical regions.

A t-statistic may be used in place of the z-statistic when σ is unknown and s must be used as an



Critical region for rejection of H_0 when α = 0.01, two-tailed test

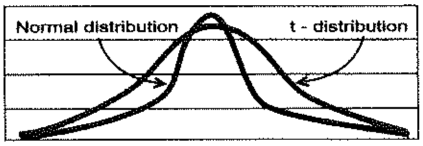
Table A Normal Curve Areas

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	.0040	.0080	.0120	0160	0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	0793	.0832	.0871	0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	1179	1217	1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	1554	1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	3643	3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	3849	3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4516	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Example. Given $\bar{x}$=108, s=15, and n=26 estimate a 95% confidence interval for the population mean. Since the population variance is unknown, the t-distribution is used. The resulting interval, using a t-valve of 2.060 from Table B (row 25 of the middle column), is approximately 102 to 114. Consequently, any hypothesized μ between 102 to 114 is tenable on the basis of this sample. Any hypothesized μ below 102 or above 114 would be rejected at 0.05 significance.

☐ **COMPARISON BETWEEN *t* AND *z* DISTRIBUTIONS**

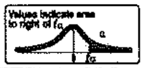
Although both distributions are symmetrical about a mean of zero, the *t*-distribution is more spread out than the normal distribution (*z*-distribution).



Thus a much larger value of *t* is required to mark off the bounds of the critical region of rejection.

As *df* increases, differences between *z*- and *t*- distributions are reduced. Table A (*z*) may be used instead of Table B (*t*) when n>30. To use either table when n<30, the sample must be drawn from a normal population.

Table B Critical Values of *t*



A* =Level of significance for one-tailed test
B* =level of significance for two-tailed test

df A*	0.1	0.05	0.025	0.01	0.005
B*	0.2	0.1	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750

TESTING INDEPENDENT SAMPLES

☐ **SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN MEANS-** If a number of pairs of samples were taken from the same population or from two different populations, then:

- The distribution of differences between pairs of sample means tends to be normal (z-distribution).
- The mean of these differences between means $\mu_{\bar{x}_1-\bar{x}_2}$ is equal to the difference between the population means, that is $\mu_1-\mu_2$.

☐ **Z-DISTRIBUTION: σ_1 and σ_2 are known**

- The standard error of the difference between means

$$\sigma_{\bar{x}_1-\bar{x}_2}=\sqrt{(\sigma_1^2)/n_1+(\sigma_2^2)/n_2}$$

- Where $(\mu_1 - \mu_2)$ represents the hypothesized difference in means, the following statistic can be used for hypothesis tests:

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}}$$

- When n_1 and n_2 are >30, substitue s_1 and s_2 for σ_1 and σ_2, respectively.

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{\left(\frac{SS_1+SS_2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

(To obtain sum of squares (SS) see Measures of Central Tendency on page 1)

☐ **POOLED *t*-TEST**

- **Distribution is normal**
- ***n* < 30**
- **σ_1 and σ_2 are *not* known but assumed equal**

- The hypothesis test may be 2 tailed (= vs. ≠) or 1 tailed: $\mu_1 \leq \mu_2$ and the alternative is $\mu_1 > \mu_2$ (or $\mu_1 \geq \mu_2$ and the alternative is $\mu_1 < \mu_2$.)

- degrees of freedom(***df***): $(n_1-1)+(n_2-1)=n_1+n_2-2$.

- Use the given formula below for estimating $\sigma_{\bar{x}_1-\bar{x}_2}$ to determine $s_{\bar{x}_1-\bar{x}_2}$.

- Determine the critical region for rejection by assigning an acceptable level of significance and looking at the *t*-table with ***df*** $= n_1 + n_2-2$.

- **Use the following formula for the estimated standard error:**

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{\left[\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1 n_2}\right]}$$

F - TEST

☐ **HETEROGENEITY OF VARIANCES** may be determined by using the F-test:

$$F=\frac{s_1^2(\textit{larger variance})}{S_2^2(\textit{smaller variance})}$$

☐ **NULL HYPOTHESIS-** Variances are equal and their ratio is one.

☐ **ALTERNATIVE HYPOTHESIS-** Variances differ and their ratio is not one.

☐ Look at "**Table C**" below to determine if the variances are significantly different from each other. Use degrees of freedom from the 2 samples: (n_1-1, n_2-1).

Table C Critical Values of *F*

Top row=.05, Bottom row=.01 points for distribution of *F*

Degrees of freedom for numerator

Degrees of freedom for denominator

	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	1937	19.38	19.39
	98.49	99.01	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78
	34.12	30.81	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23
4	7.71	6.94	6.59	6.39	6.26	6.16	6.08	6.04	6.00	5.96
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97

CORRELATED SAMPLES

☐ **STANDARD ERROR OF THE DIFFERENCE** between Means for Correlated Groups. The general formula is:

$$S_{\bar{x}_1-\bar{x}_2}=\sqrt{s^2_{\bar{x}_1}+s^2_{\bar{x}_2}-2rs_{\bar{x}_1}s_{\bar{x}_2}}$$

where r is Pearson correlation

- By matching samples on a variable correlated with the criterion variable, the magnitude of the standard error of the difference can be reduced.
- The higher the correlation, the greater the reduction in the standard error of the difference.

ANALYSIS OF VARIANCE (*ANOVA*)

☐ **PURPOSE-** Indicates possibility of overall mean effect of the experimental treatments before investigating a specific hypothesis.

☐ **ANOVA-** Consists of obtaining independent estimates from population subgroups. It allows for the partition of the sum of squares into known components of variation.

☐ **TYPES OF VARIANCES**

- **Between-Group Variance (BGV)-** reflects the magnitude of the difference(s) among the group means.
- **Within-Group Variance (WGV)-** reflects the dispersion within each treatment group. It is also referred to as the error term.

☐ **CALCULATING VARIANCES**

- Following the F-ratio, when the BGV is large relative to the WGV, the F-ratio will also be large.

$$BGV=\frac{n\Sigma(\bar{x}_i-\bar{x}_{tot})^2}{k-1}$$

where x_i = mean of ith treatment group and x_{tot} = mean of all n values across all k treatment groups.

$$WGV=\frac{SS_1+SS_2+...+SS_k}{n-k}$$

where the SS's are the sums of squares (see *Measures of Central Tendency* on page 1) of each subgroup's values around the subgroup mean.

☐ **USING F-RATIO-** $F = BGV/WGV$

- Degrees of freedom are k-1 for the numerator and n-k for the denominator.
- If $BGV > WGV$, the experimental treatments are responsible for the large differences among group means. Null hypothesis: the group means are estimates of a common population mean.

PROPORTIONS

In random samples of size *n*, the sample proportion ***p*** fluctuates around the proportion mean = π with a proportion variance of $\frac{\pi(1-\pi)}{n}$ proportion standard error of $\sqrt{\pi(1-\pi)/n}$

As the sampling distribution of *p* increases, it concentrates more around its target mean. It also gets closer to the normal distribution. In which case: $z=\frac{p-\pi}{\sqrt{\pi(1-\pi)/n}}$

ISBN 157222649-8

9 781572 226494

50495

quickstudy.com

CORRELATION

Definition - Correlation refers to the relationship betw two variables. The Correlation Coefficient is a measure expresses the extent to which two variables are r

☐ **"PEARSON r" METHOD** (Product-Mome Correlation Coefficient) - Corelation coefficie employed with interval- or ratio-scaled variabl

Ex.: Given observations to two variables *X* and *Y*, we can compute their corresponding z values: $Z_x=(x-\bar{x})/s_x$ and $Z_y=(y-\bar{y})/s_y$.

- The formulas for the Pearson correlation (r)

$$r=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{SS_x \cdot SS_y}}$$

- Use the above formula for large samples.

- Use this formula (also known as the ***Mean-Deviati Method*** of computing the Pearson r) for small sampl

$$r=\frac{\Sigma(z_x z_y)}{n}$$

☐ **RAW SCORE METHOD** is quicker and c be used in place of the first formula above wh the sample values are available.

$$r=\frac{\Sigma xy-\frac{(\Sigma x)(\Sigma y)}{N}}{\sqrt{\left[\Sigma x^2-\frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2-\frac{(\Sigma y^2)}{n}\right]}}$$

- Most widely-used non-parametric test.
- The χ^2 mean = its degrees of freedom.
- The χ^2 variance = twice its degrees of freedo
- Can be used to test one or two independent sampl
- The square of a standard normal variable is a chi-square variable.
- Like the t-distibution, it has different distrib tions depending on the degrees of freed

☐ **DEGREES OF FREEDOM (*d.* COMPUTATION**

- If ***chi-square*** tests for the goodness-of-fit to a pothesized distribution,

d.f. = g - 1 - m, where

g = number of groups, or classes, in the frequer distribution.

m = number of population parameters that m be estimated from sample statistics to test t hypothesis.

- If ***chi-square*** tests for homogeneity or contingen

d.f. = (rows-1) (columns-1)

☐ **GOODNESS-OF-FIT TEST-** To apply t chi-square distribution in this manner, t critical chi-square value is expressed as:

$$\Sigma\frac{(f_0-f_e)^2}{f_e}\ \textbf{where}$$

f_0 = observed frequency of the variable

f_e = expected frequency (based on hypothesiz population distribution).

☐ **TESTS OF CONTINGENCY-** Application Chi-square tests to two separate populations to t statistical independence of attributes.

☐ **TESTS OF HOMOGENEITY-** Application Chi-square tests to two samples to test if they ca from populations with like distributions.

☐ **RUNS TEST-** Tests whether a sequence comprise a sample) is random. The followi equations are applied:

$$(\bar{R})=\frac{2n_1n_2}{n_1+n_2}+1 \text{ and } S_R\sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

Where

$\bar{R}$ = mean number of runs

n_1 = number of outcomes of one type

n_2 = number of outcomes of the other type

S_R = standard deviation of the distributio number of runs.

U.S.$4.95 / CAN.$7.50 **February 200**

Customer Hotline # 1.800.230.9522

Manufactured by Laminating Services, Inc., Largo, FL, and sold under license from Las [illegible]

Descriptive Statistics:

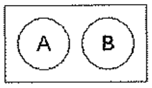
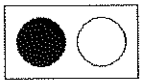
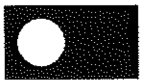
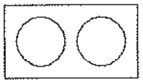
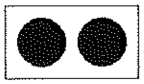
Term	Meaning	Population Formula	Sample Formula	Example {1,16,1,3,9}
Sort	Sort values in increasing order			{1,1,3,9,16}
Mean	Average	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$	6
Median	The middle value – half are below and half are above			3
Mode	The value with the most appearances			1
Variance	The average of the squared deviations between the values and the mean	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$	$(1-6)^2 + (1-6)^2 + (3-6)^2 + (9-6)^2 + (16-6)^2$ divided by 5 values = 168/5 = 33.6
Standard Deviation	The square root of Variance, thought of as the "average" deviation from the mean.	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$	Square root of 33.6 = 5.7966
Coefficient of Variation	The variation relative to the value of the mean		$CV = \frac{s}{\overline{X}}$	5.7966 divided by 6 = 0.9661
Minimum	The minimum value			1
Maximum	The maximum value			16
Range	Maximum minus Minimum			16 – 1 = 15

Probability Terms:

Term	Meaning	Notation	Example* (see footnote)
Probability	For any event A, probability is represented within $0 \le P \le 1$.	P()	0.5
Random Experiment	A process leading to at least 2 possible outcomes with uncertainty as to which will occur.		Rolling a dice
Event	A subset of all possible outcomes of an experiment.		Events A and B
Intersection of Events	Let A and B be two events. Then the intersection of the two events is the event that both A and B occur (logical AND).	$A \cap B$	The event that a 2 appears
Union of Events	The union of the two events is the event that A or B (or both) occurs (logical OR).	$A \cup B$	The event that a 1, 2, 4, 5 or 6 appears
Complement	Let A be an event. The complement of A is the event that A does not occur (logical NOT).	$\overline{A}$	The event that an odd number appears
Mutually Exclusive Events	A and B are said to be mutually exclusive if at most one of the events A and B can occur.		A and B are not mutually exclusive because if a 2 appears, both A and B occur
Collectively Exhaustive Events	A and B are said to be collectively exhaustive if at least one of the events A or B must occur.		A and B are not collectively exhaustive because if a 3 appears, neither A nor B occur
Basic Outcomes	The simple indecomposable possible results of an experiment. One and exactly one of these outcomes must occur. The set of basic outcomes is mutually exclusive and collectively exhaustive.		Basic outcomes 1, 2, 3, 4, 5, and 6
Sample Space	The totality of basic outcomes of an experiment.		{1,2,3,4,5,6}

* Roll a fair die once. Let A be the event an even number appears, let B be the event a 1, 2 or 5 appears

***Probability Rules*:**

If events A and B are mutually **exclusive**		
Term	**Equals**	**Area:** 
$P(A)=$	$P(A)$	
$P(\bar{A})=$	$1 - P(A)$	
$P(A\cap B)=$	0	
$P(A\cup B)=$	$P(A) + P(B)$	

***General probability rules*:**

1) If **P(A|B) = P(A)**, then A and B are **independent events**! (for example, rolling dice one after the other).

2) If there are *n* possible outcomes which are equally likely to occur:

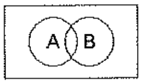
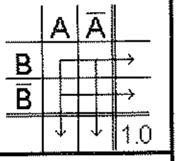
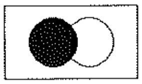
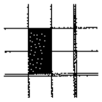
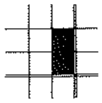
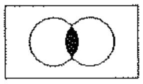
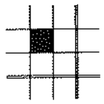
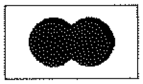
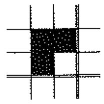
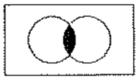
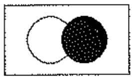
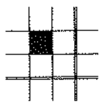
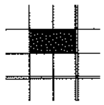
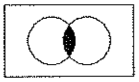
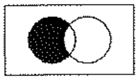
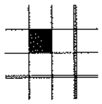
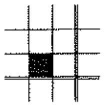
$$P(\text{outcome i occurs}) = \frac{1}{n} \text{ for each } i \in [1, 2, ..., n]$$

**Example: Shuffle a deck of cards, and pick one at random. P(chosen card is a 10 ♦) = 1/52.*

3) If event A is composed of ***n*** **equally likely basic outcomes**:

$$P(A) = \frac{\textit{Number of Basic Outcomes in A}}{n}$$

**Example: Suppose we toss two dice. Let A denote the event that the sum of the two dice is 9. P(A) = 4/36 = 1/9, because there are 4 out of 36 basic outcomes that will sum 9.*

If events A and B are NOT mutually **exclusive**			
Term	**Equals**	**Venn:** 	
$P(A)=$	$P(A)$		
$P(\bar{A})=$	$1 - P(A)$		
$P(A\cap B)=$	$P(A) * P(B)$ *only if A and B are independent*		
$P(A\cup B)=$	$P(A) + P(B) - P(A\cap B)$		
$P(A\|B)=$ ***[Bayes' Law**: P(A holds given that B holds)]*	$\frac{P(A\cap B)}{P(B)}$	 ÷ 	 ÷ 
$P(A\cap B) = P(A\|B) * P(B)$			
$P(A\cap B) = P(B\|A) * P(A)$			
$P(A)=$	$P(A\cap B) + P(A\cap \bar{B})$ = $\mathbf{P(A\|B)P(B) + P(A\|\bar{B})P(\bar{B})}$	 + 	 + 

*Example: Take a deck of 52 cards. Take out 2 cards sequentially, but don't look at the first. The probability that the second card you chose was a ♣ is the probability of choosing a ♣ (event A) after choosing a ♣ (event B), plus the probability of choosing a ♣ (event A) after not choosing a ♣ (event B), which equals (12/51)(13/52) + (13/51)(39/52) = 1/4 = 0.25.

Random Variables and Distributions:

o calculate the **Expected Value** $E(X) = \sum x \cdot P(X = x)$, use the following table:

	∇ *	∇ =	∇
Event	**Payoff**	**Probability**	**Weighted Payoff**
[name of first event]	[payoff of first event in $]	[probability of first event $0 \leq P \leq 1$]	[product of Payoff * Probability]
[name of second event]	[payoff of second event in $]	[probability of second event $0 \leq P \leq 1$]	[product of Payoff * Probability]
[name of third event]	[payoff of third event in $]	[probability of third event $0 \leq P \leq 1$]	[product of Payoff * Probability]
		Total (Expected Payoff):	[total of all Weighted Payoffs above]

* See example in BOOK 1 page 54

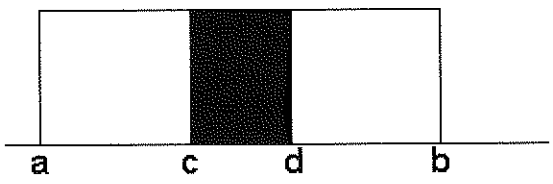
To calculate the **Variance** $\text{Var}(X) = \sum (x - E(X))^2 P(X = x)$ and **Standard Deviation** $\sigma(X) = \sqrt{Var(X)}$, use:

	∇	- ∇	= ∇	^2= ∇	* ∇	= ∇
Event	**Payoff**	**Expected Payoff**	**Error**	**(Error)^2**	**Probability**	**Weighted (Error)^2**
[1st event]	[1st payoff]	[Total from above]	[1st payoff minus Expected Payoff]	1st Error squared	1st event's probability	1st (Error)^2 * 1st event's probability
[2nd event]	[2nd payoff]	[Total from above]	[2nd payoff minus Expected Payoff]	2nd Error squared	2nd event's probability	2nd (Error)^2 * 2nd event's probability
[3rd event]	[3rd payoff]	[Total from above]	[3rd payoff minus Expected Payoff]	3rd Error squared	3rd event's probability	3rd (Error)^2 * 3rd event's probability
					Variance:	[total of above]
					Std. Deviation:	[square root of Variance]

Counting Rules:

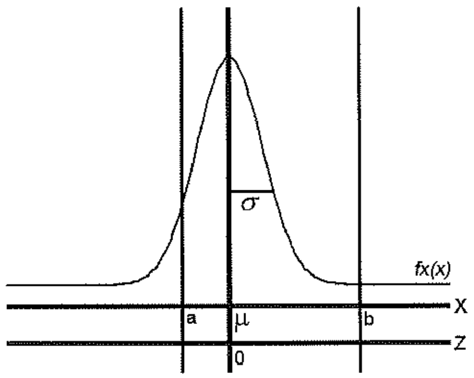
Term	**Meaning**	**Formula**	**Example**
Basic Counting Rule	The *number* of ways to pick x things out of a set of n (with no regard to order). The *probability* is calculated as $1/x$ of the result.	$\binom{n}{x} = \frac{n!}{x!(n-x)!}$	The *number* of ways to pick 4 specific cards out of a deck of 52 is: 52!/((4!)(48!)) = 270,725, and the *probability* is 1/270,725 = 0.000003694
Bernoulli Process	For a sequence of n trials, each with an outcome of either success or failure, each with a probability of p to succeed – the probability to get x successes is equal to the Basic Counting Rule formula (above) times $p^x(1-p)^{n-x}$.	$P(X = x \mid n, p) = \left(\frac{n!}{x!(n-x)!}\right) p^x (1-p)^{n-x}$	If an airline takes 20 reservations, and there is a 0.9 probability that each passenger will show up, then the probability that **exactly** 16 passengers will show is: $\frac{20!}{16!\,4!}(0.9)^{16}(0.1)^4$ = 0.08978
Bernoulli Expected Value	The expected value of a Bernoulli Process, given n trials and p probability.	$E(X) = np$	In the example above, the number of people expected to show is: (20)(0.9) = 18
Bernoulli Variance	The variance of a Bernoulli Process, given n trials and p probability.	$\text{Var}(X) = np(1 - p)$	In the example above, the Bernoulli Variance is (20)(0.9)(0.1) = 1.8
Bernoulli Standard Deviation	The standard deviation of a Bernoulli Process:	$\sigma(X) = \sqrt{np(1-p)}$	In the example above, the Bernoulli Standard Deviation is $\sqrt{1.8}$ = 1.34
Linear Transformation Rule	If X is random and Y=aX+b, then the following formulas apply:	$E(Y) = a^*E(X) + b$ $\text{Var}(Y) = a^2{}^*\text{Var}(X)$ $\sigma(Y) = \lvert a \rvert^*\sigma(X)$	

Uniform Distribution:



Term/Meaning	**Formula**
Expected Value	$\overline{X} = \dfrac{(a+b)}{2}$
Variance	$\sigma_X^2 = \dfrac{(b-a)^2}{12}$
Standard Deviation	$\sigma_X = \dfrac{(b-a)}{\sqrt{12}}$
Probability that X *falls* between *c* and *d*	$P(c \le X \le d) = \dfrac{d-c}{b-a}$

Normal Distribution:



Probability Density Function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}$$

where $\pi \cong 3.1416$ and $e \cong 2.7183$

Standard Deviations away from the mean:

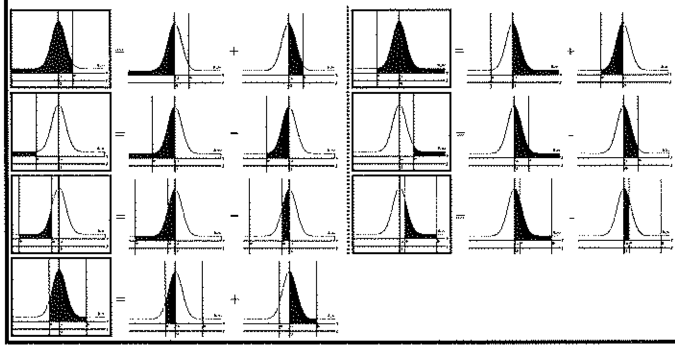
$$Z = \frac{X-\mu}{\sigma}$$

(Z and σ are swappable!)

$P(a \le X \le b)$ = area under $f_X(x)$ between a and b:

$$P(a \le X \le b) = P\left[\left(\frac{a-\mu}{\sigma}\right) \le Z \le \left(\frac{b-\mu}{\sigma}\right)\right]$$

Standard Normal Table - seven usage scenarios:



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.33
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.49
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Correlation:

- If X and Y are two different sets of data, their correlation is represented by $Corr_{(XY)}$, r_{XY}, or ρ_{XY} (rho).
- If Y increases as X increases, $0 < \rho_{XY} < 1$. If Y decreases as X increases, $-1 < \rho_{XY} < 0$.
- The extremes $\rho_{XY} = 1$ and $\rho_{XY} = -1$ indicated perfect correlation – info about one results in an exact prediction about the other.
- If X and Y are completely uncorrelated, $\rho_{XY} = 0$.
- The **Covariance** of X and Y, $Cov_{(XY)}$, has the same sign as ρ_{XY}, has unusual units and is usually a means to find ρ_{XY}.

Term	Formula	Notes
Correlation	$Corr_{(XY)} = \dfrac{Cov_{(XY)}}{\sigma_X \sigma_Y}$	Used with Covariance formulas below
Covariance (2 formulas)	$Cov_{(XY)} = E\left[(X-\overline{X})(Y-\overline{Y})\right]$ *(difficult to calculate)*	Sum of the products of all sample pairs' distance from their respective means multiplied by their respective probabilities
	$Cov_{(XY)} = E(XY) - (\overline{X})(\overline{Y})$	Sum of the products of all sample pairs multiplied by their respective probabilities, minus the product of both means
Finding Covariance given Correlation	$Cov_{(XY)} = \sigma_X \sigma_Y Corr_{(XY)}$	

Portfolio Analysis:

	Term	Formula	Example*
	Mean of any Portfolio "S"	$\overline{S} = a\overline{X} + b\overline{Y}$	$\overline{S}$ = ¾(8.0%)+ ¼(11.0%) = 8.75%
Uncorrelated	Portfolio Variance	$\sigma^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$	$\sigma^2 = (¾)^2(0.5)^2 + (¼)^2(6.0)^2 = 2.3906$
	Portfolio Standard Deviation	$\sigma = \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}$	$\sigma = 1.5462$
Correlated	Portfolio Variance	$\sigma^2_{(aX+bY)} = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov_{(XY)}$	
	Portfolio Standard Deviation	$\sigma_{(aX+bY)} = \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov_{(XY)}}$	

* Portfolio "S" composed of ¾ Stock A (mean return: 8.0%, standard deviation: 0.5%) and ¼ Stock B (11.0%, 6.0% respectively)

The Central Limit Theorem

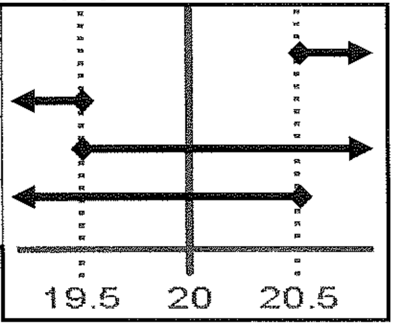
Normal distribution can be used to approximate binominals of more than 30 trials (n≥30):

Term	Formula
Mean	$E(X) = np$
Variance	$Var(X) = np(1 - p)$
Standard Deviation	$\sigma(X) = \sqrt{np(1-p)}$

Continuity Correction

Unlike continuous (normal) distributions (i.e. $, time), discrete binomial distribution of integers (i.e. # people) must be corrected:

Old cutoff	New cutoff
P(X>20)	P(X>20.5)
P(X<20)	P(X<19.5)
P(X≥20)	P(X≥19.5)
P(X≤20)	P(X≤20.5)



Sampling Distribution of the Mean

If the X_i's are normally distributed (*or* $n \geq 30$), then $\overline{X}$ is normally distributed with:

Term	Formula
Mean	μ
Standard Error of the Mean	$\dfrac{\sigma}{\sqrt{n}}$

Sampling Distribution of a Proportion

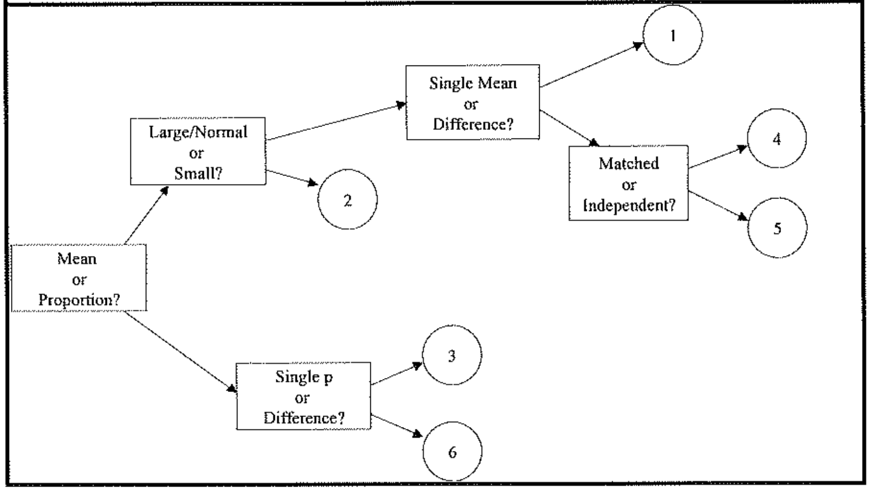
If, for a proportion, $n \geq 30$ then $\hat{p}$ is normally distributed with:

Term	Formula
Mean	p
Standard Deviation	$\sqrt{\dfrac{p(1-p)}{n}}$

Confidence Intervals:

	Parameter	**Confidence Interval**	**Usage**	**Sample**	**σ**
1	μ	$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	Normal		Known σ
	μ	$\overline{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$		Large	Unknown σ
2	μ	$\overline{X} \pm t_{(n-1,\alpha/2)} \frac{s}{\sqrt{n}}$	Normal	Small	Unknown σ
3	p	$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Binomial	Large	
4	$\mu_X - \mu_Y$	$\overline{D} \pm z_{\alpha/2} \frac{s_D}{\sqrt{n}}$	Normal		Matched pairs
5	$\mu_X - \mu_Y$	$\overline{X} - \overline{Y} \pm z_{(\alpha/2)} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$	Normal		Known σ, Independent Samples
	$\mu_X - \mu_Y$	$\overline{X} - \overline{Y} \pm z_{(\alpha/2)} \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$		Large	
6	$p_X - p_Y$	$\hat{p}_X - \hat{p}_Y \pm z_{(\alpha/2)} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}$	Binomial	Large	

Formulae Guide



Confidence Level to Z-Value Guide

Confidence Level		**Z$_{\alpha/2}$ (2-Tail)**	**Z$_{\alpha}$ (1-Tail)**
80%	**α = 20%**	1.28	0.84
90%	**α = 10%**	1.645	1.28
95%	**α = 5%**	1.96	1.645
99%	**α = 1%**	2.575	2.325
c	**α = 1.0-c**	Z$_{(c/2)}$	Z$_{(c-0.5)}$

Determining the Appropriate Sample Size

Term	**Normal Distribution Formula**	**Proportion Formula**
Sample Size (for +/- e)	$n = \frac{(1.96)^2 \sigma^2}{e^2}$	$n \geq \frac{1.96^2}{4e^2}$

t-table

d. f.	0.100	0.050	0.025	0.010	0.005
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763

Hypothesis Testing:

		Two-tailed		**Lower-tail**		**Upper-tail**	
Test Type	**Test Statistic**	H_a	**Critical Value**	H_a	**Critical Value**	H_a	**Critical Value**
Single μ ($n \geq$ 30)	$z_0 = \frac{\overline{X}-\mu_0}{s/\sqrt{n}}$	$\mu \neq \mu_0$	$\pm z_{\alpha/2}$	$\mu < \mu_0$	$-z_{\alpha}$	$\mu > \mu_0$	$+z_{\alpha}$
Single μ ($n <$ 30)	$t_0 = \frac{\overline{X}-\mu_0}{s/\sqrt{n}}$	$\mu \neq \mu_0$	$\pm t_{(n-1,\alpha/2)}$	$\mu < \mu_0$	$-t_{(n-1,\alpha)}$	$\mu > \mu_0$	$+t_{(n-1,\alpha)}$
Single p ($n \geq$ 30)	$z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$p \neq p_0$	$\pm z_{\alpha/2}$	$p < p_0$	$-z_{\alpha}$	$p > p_0$	$+z_{\alpha}$
Diff. between two μs	$z_0 = \frac{(\overline{X}-\overline{Y})-0}{\sqrt{\frac{s_X^2}{n_x}+\frac{s_Y^2}{n_Y}}}$	$\mu_X - \mu_Y \neq 0$	$\pm z_{\alpha/2}$	$\mu_X - \mu_Y < 0$	$-z_{\alpha}$	$\mu_X - \mu_Y > 0$	$+z_{\alpha}$
Diff. between two *p*s	$z_0 = \frac{(\hat{p}_X-\hat{p}_Y)-0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{n_X+n_Y}{n_X n_Y}\right)}}$	$p_X - p_Y \neq 0$	$\pm z_{\alpha/2}$	$p_X - p_Y < 0$	$-z_{\alpha}$	$p_X - p_Y > 0$	$+z_{\alpha}$

Classic Hypothesis Testing Procedure			
Step		**Description**	**Example**
1	**Formulate Two Hypotheses**	The hypotheses ought to be mutually exclusive and collectively exhaustive. The hypothesis to be tested (the null hypothesis) always contains an equals sign, referring to some proposed value of a population parameter. The alternative hypothesis never contains an equals sign, but can be either a one-sided or two-sided inequality.	H_0: $\mu = 0$ H_A: $\mu < 0$
2	**Select a Test Statistic**	The test statistic is a standardized estimate of the difference between our sample and some hypothesized population parameter. It answers the question: "*If the null hypothesis were true, how many standard deviations is our sample away from where we expected it to be*?"	$\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$
3	**Derive a Decision Rule**	The decision rule consists of regions of rejection and non-rejection, defined by critical values of the test statistic. It is used to establish the probable truth or falsity of the null hypothesis.	We reject H_0 if $\overline{X} < \mu_0 - z_{\alpha}\frac{\sigma}{\sqrt{n}}$.
4	**Calculate the Value of the Test Statistic; Invoke the Decision Rule in light of the Test Statistic**	Either reject the null hypothesis (if the test statistic falls into the rejection region) or do not reject the null hypothesis (if the test statistic does not fall into the rejection region.	$\frac{\overline{X}-\mu_0}{s/\sqrt{n}} = \frac{-0.21-0}{0.80/\sqrt{50}}$

Regression:

Statistic	Symbol
Independent Variables	$X_1, \ldots X_k$
Dependent Variable (a random variable)	Y
Dependent Variable (an individual observation among sample)	Y_i
Intercept (or constant); an unknown population parameter	β_0
Estimated intercept; an estimate of β_0	$\hat{\beta}_0$
Slope (or coefficient) for Independent Variable 1 (unknown)	β_1
Estimated slope for Independent Variable 1; an estimate of β_1	$\hat{\beta}_1$

Regression Statistics	
Multiple R	**0.9568**
R Square	**0.9155**
Adjusted R Square	**0.9015**
Standard Error	**6.6220**
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	**5704.0273**	**2852.0137**	**65.0391**	0.0000
Residual	12	**526.2087**	**43.8507**		
Total	14	**6230.2360**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-20.3722	9.8139	**-2.0758**	**0.0601**
Size (100 sq ft)	4.3117	0.4104	10.5059	0.0000
Lot Size (1000 sq ft)	4.7177	0.7646	6.1705	0.0000

Statistic (Mapped to Output Above)	Symbol	Formula
Dependent Variable (sample mean of *n* observations)	$\overline{Y}$	$= \frac{\sum_{i=1}^{n} Y_i}{n}$
Dependent Variable (estimated value for a given vector of independent variables)	$\hat{Y}_i$	$= \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \ldots + \hat{\beta}_k x_{ki}$
Error for observation *i*. The unexplained difference between the actual value of Y_i and the prediction for Y_i based on our regression model.	ε_i	$= Y_i - \hat{Y}_i$
6230.2360 **Total Sum of Squares**	TSS **(or SST)**	$= \sum_{i-1}^{n} \left(Y_i - \overline{Y}\right)^2 = SSR + SSE$
526.2087 **Sum of Squares due to Error**	SSE	$= \sum_{i-1}^{n} \left(Y_i - \hat{Y}_i\right)^2$
43.8507 **Mean Squares due to Error**	MSE	$= \frac{SSE}{n-k-1}$
5704.0273 **Sum of Squares due to Regression**	SSR	$= \sum_{i-1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2$
2852.0137 **Mean Squares due to Regression**	MSR	$= \frac{SSR}{k}$

Statistic (Mapped to Output Above)	Symbol	Formula
0.9155 ***R*-square (Coefficient of Determination)**	R^2	$= 1 - \frac{SSE}{TSS}$
0.9568 **Multiple *R* (Coefficient of Multiple Correlation)**	R	$= \sqrt{R^2}$
0.9015 **Adjusted *R*-square**	$\overline{R^2}$	$= 1 - \frac{\left(\frac{SSE}{n-k-1}\right)}{\left(\frac{SST}{n-1}\right)}$
6.6220 **Standard Error (a.k.a. Standard Error of the Estimate)**	s_ε	$= \sqrt{\frac{SSE}{n-k-1}}$
-2.0758 ***t*-statistic for testing** $H_0 : \beta_1 = 0$ **vs.** $H_A : \beta_1 \neq 0$	t_0	$= \frac{\hat{\beta}_1 - 0}{s_{\beta 1}}$
0.0601 ***p*-value for testing** $H_0 : \beta_1 = 0$ **vs.** $H_A : \beta_1 \neq 0$	***p*-value**	$= P\left(\lvert T \rvert > \lvert t_0 \rvert\right)$
65.0391 ***F***	F	$= \frac{MSR}{MSE}$ $= \frac{n-k-1}{k} \times \frac{R^2}{1-R^2}$