Probability and Applied Statistics Formula Sheet
Alexis Petito

Definition 1.1 – Mean of n Measured Responses.
The mean of a sample *n* measured responses.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Definition 1.2 – Variance of Sample Measurements
The variance of a sample of measurements.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Definition 1.3 – Standard Deviation of Sample Measurements
Standard deviation of a sample measurements.

$$s = \sqrt{s^2}$$

Definition 2.7 - Permutations
Ordered arrangement of *r* distinct objects called a *permutation***.**

$$P_r^n = \frac{n!}{(n-r)!}$$

Definition 2.8 - Combinations
The number of *combinations* of *n* objects taken *r* at a time

$$C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!\,(n-r)!}$$

Theorem 2.3 – n Objects into K Groups
The number of ways partitioning *n* distinct objects into *k* distinct groups containing *n* objects, respectively where each object appears in exactly one group and $\sum_{i=1}^{k} n_i = n$, is

$$N = \binom{n}{n_1\ n_2\ \dots n_k} = \frac{n!}{n_1!\ n_2\,!\dots n_k!}$$

Definition 2.9 – A Given B
The *conditional probability of an event* A, given that an event B has occurred, is equal to

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

Definition 2.10 – Determining Independence
Two events A and B are said to be *independent* if any one of the following holds:

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

Otherwise, the events are said to be *dependent*.

Theorem 2.5 - The Multiplicative Law of Probability
The probability of the intersection of two events A and B is
$P(A \cap B) = P(A)P(B|A)$
$= P(B)P(A|B)$
If A and B are independent, then
$P(A \cap B) = P(A)P(B)$

Theorem 2.6 - The Additive Law of Probability
The probability of the union of two events A and B is
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
If A and B are mutually exclusive events, $P(A \cap B) = 0$ and
$P(A \cup B) = P(A) + P(B)$

Theorem 2.7 – Determine A from $\overline{A}$
If A is an event, then
$P(A) = 1 - P(\overline{A})$

Definition 2.11 – Partition of S
For some positive integer k, let the sets $B_1 B_2 B_3 \dots, B_k$ be such that

1. $S = B_1 \cup B_2 \cup \dots \cup B_k$
2. $B_i \cap B_j = \emptyset$, for $i \neq j$

Then the collection of sets $\{B_1 B_2, \dots, B_k\}$ is said to be a *partition* of S

Theorem 2.8
Assume that $\{B_1 B_2, \dots, B_k\}$ is a partition of S (see definition 2.11) such that $P(B_i) > 0$, for $i = 1,2, \dots, k$. Then for any event A

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

Theorem 2.9 - Bayes' Rule
Assume that $\{B_1 B_2, \dots, B_k\}$ is a partition of S (see definition 2.11) such that $P(B_i) > 0$, for $i = 1,2, \dots, k$. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

Bayes' Theorem
Bayes Theorem: For two events A and B in sample space S, with $P(A) > 0$ and $P(B) > 0$,

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

If $0 < P(B) < 1$, we may write by the *Theorem of Total Probability*.

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

Probability and Applied Statistics Formula Sheet
Alexis Petito

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

Probability Mass Function
$p(y) = P(Y = y)$

Expected Value
Let Y be a discrete random variable with the probability function $p(y)$. Then the *expected value* of Y, $E(Y)$, is defined to be,
$\mathrm{E}(Y) = \sum_{y \epsilon Y} yp(y)$

Theorem 3.2 – Binomial Distribution
Let Y be a discrete random variable with probability function $p(y)$ and $g(y)$ be a real-valued function of Y. Then the expected value of $g(Y)$
Is given by,
$\mathrm{E}[g(Y)] = \sum_{all\ y} g(y)p(y)$

Definition 3.5 – Expected Variance of a Random Variable
If Y is a random variable with mean $E(Y) = \mu$, the variance of a random varible Y is defined to be the expected value of $(Y - \mu)^2$. That is,
$V(Y) = E[(Y - \mu)^2]$

The Standard Deviation of Y
$\sqrt{V[Y]}$

Binomial Distribution
$p(y) = P(Y = y) = \binom{n}{y} p^y q^{n-y} \qquad y \epsilon \{0,1,2, \dots, n\}$

Theorem 3.7 – Binomial Distribution Expected Value and Variance
Let Y be a binomial random variable based on *n* trials and success probability *p*. Then,
$\mu = E(Y) = np$ (Mean/Expected)
$\sigma^2 = V(Y) = npq$ (Standard Deviation)

Geometric Distribution
$p(y) = \mathrm{P}(Y = y) = q^{y-1}p$

Probability and Applied Statistics Formula Sheet
Alexis Petito

Extra Geometric Distribution Formulas

A success occurs on or before the nth trial.

$P(X \leq n) = 1 - (1-p)^n$

A success occurs before the nth trial.

$P(X < n) = 1 - (1-p)^{n-1}$

A success occurs on or after the nth trial.

$P(X \geq n) = (1-p)^{n-1}$

A success occurs after the nth trial.

$P(X > n) = (1-p)^n$

Theorem 3.8 – Geometric Distribution Expected Value and Variance

If Y is a random variable with geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \ and \ \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

Hypergeometric Distribution

$$p(y) = P(Y = y) = P(A) = \frac{n_A}{n_S} = \frac{\binom{r}{y} \ x \binom{N-r}{n-y}}{\binom{N}{n}}$$

Theorem 3.10 – Hypergeometric Distribution Expected Value and Variance

If Y is a random variable with hypergeometric distribution,

$$\mu = E(Y) = \frac{nr}{N} \ and \ \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Negative Binomial Distribution

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}$$

Theorem 3.9 – Negative Binomial Distribution Expected Value and Variance.

If Y is a random variable with negative binomial distribution,

$$\mu = E(Y) = \frac{r}{p} \ and \ \sigma^2 = V(Y) = \left(\frac{r(1-p)}{p^2}\right)$$

Definition 3.11 – Poisson Distribution

A random variable Y is said to have *Poisson probability distribution* if and only if,

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \quad y = 0, 1, 2, \dots, \lambda > 0.$$

Theorem 3.11 – Poisson Distribution Expected Value and Variance

If Y is a random variable possessing a Poisson Distribution with parameter λ, then

$\mu = E(Y) = \lambda \ and \ \sigma^2 = V(Y) = \lambda$

Probability and Applied Statistics Formula Sheet
Alexis Petito

Theorem 3.4 – Chebyshev's Theorem
Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \text{ or } P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Definition 4.1 – Distribution Function of Y
Let Y denote any random variable. The *distribution function* of Y, denoted by $F(y)$, is such that $F(y) = P(Y \leq y) \text{ for} -\infty < y < \infty$

Theorem 4.1. – Properties of a distribution function
If $F(y)$ is a distribution function, then

1. $F(-\infty) = \lim_{y \to \infty} F(y) = 0.$
2. $F(\infty) = \lim_{y \to \infty} F(y) = 1.$
3. $F(y)$ is a nondecreasing function of y. [If y_1 and y_2 are *any* values such that $y_1 < y_2$, then $F(y_1) < F(y_2)$.]

Definition 4.2 – Random Variable with Distribution Function Continuity
A random variable Y with distribution function $F(y)$ is said to be *continuous* if $F(y)$ is continuous, for $-\infty < y < \infty$.

Definition 4.3 – Probability Density Function
Let $F(y)$ be the distribution function for a continuous random Variable Y. Then $f(y)$, given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

Wherever the derivative exists is called the *probability density function* for the random variable Y.

Theorem 4.2 – Properties of a Density Function
If $f(y)$ is a density function for a continuous random variable, then

1. $f(y) \geq 0 \text{ for all } y, -\infty < y < \infty.$
2. $\int_{-\infty}^{\infty} f(y)dy = 1.$

Definition 4.4
Let Y denote a random variable. If $0 < p < 1$, the *pth quantile* of Y, is denoted by ϕ_p, is the smallest value such that P(Y ≤ ϕ_p) = F(ϕ_p) $\geq p$. If Y the continuous, ϕ_p is the smallest value such that F(ϕ_p) = P(Y ≤ ϕ_p) = p. Some prefer to call ϕ_p the 100 *p*th *percentile* of Y.

Theorem 4.3
If the random variable Y had the density function $f(y)$ and a $< b$, then the probability tat Y falls in the interval $[a, b]$ is

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

Definition 4.5 – Expected Value of a Continuous Random Variable Y
The expected value of a continuous random variable Y is

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy,$$

provided that the integral exists.

Theorem 4.4 – Expected Value of $g(Y)$(*Function*)
Let $g(Y)$ be a function of Y; then the expected value of g(Y) is given by

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy,$$

Provided that the integral exists.

Variance of Y
Once $E(Y^2)$ is determined you can find the variance by,
$\sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2$

Definition 4.6 – Probability Density Function For Uniform Distribution
If $\theta_1 < \theta_2$, a random variable Y is said to have a continuous *uniform probability distribution* on the interval (θ_1, θ_2) if and only if the density function of Y is

$$f(y) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2, \\ 0, & elsewhere. \end{cases}$$

Theorem 4.6 – Expected Value and Variance of a Uniform Distribution
If $\theta_1 < \theta_2$ and Y is a random variable uniformly distributed on the interval (θ_1, θ_2), then
$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2}$ and $\sigma^2 = V(Y) = \frac{(\theta_2 + \theta_1)^2}{12}$

Definition 4.8 – Normal Probability Distribution
A random variable Y is said to have a *normal probability distribution* if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is,

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

Theorem 4.7 – Normal Distribution Mean and Variance
If Y is a normally distributed random variables with parameters μ and σ, then,

$$E(Y) = \mu \quad \text{and} \quad V(Y) = \sigma^2$$

Definition 4.9 – Gamma Distribution

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}e^{\frac{-y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, & 0 \le y \le 1, \\ 0, & e.w. \end{cases}$$

Where,

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}dy$$

Definition 4.12 – Beta Probability Distribution

A random variable Y is said to have a *beta probability distribution with parameters* $\alpha > 0$ *and* $\beta > 0$ if and only if the density function of Y is,

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, & 0 \le y \le 1, \\ 0, & elsewhere, \end{cases}$$

Where,

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Definition 5.1 – Bivariate Probability Function For Y_1 and Y_2

Let Y_1 and Y_2 be discrete random variables. The *joint* (or bivariate) *probability function* for Y_1 and Y_2 is given by

$p(y_1, y_2) =$ P($p(Y_1 = y_1,\ Y_2 = y_2)$, $\quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$.

Theorem 5.1

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \ge 0\ for\ all\ y_1,\ y_2$.
2. $\sum_{y_1,y_2} p(y_1,\ y_2) = 1$, where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities.

Definition 5.2

For any random variables Y_1 and Y_2, the joint (bivariate) distribution function $F(y_1,\ y_2)$ is

$F(y_1, y_2) = P(Y_1 \le y_1,\ Y_2 \le y_2)$, $\quad -\infty < y_1 < \infty,\ -\infty < y_2 < \infty$.

Definition 5.3

Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1,\ y_2)$**.** If there exists a nonnegative function $f(y_1,\ y_2)$, such that,

$$F(y_1, y_2) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1, t_2)dt_2dt_1,$$

for all $-\infty < y_1 < \infty,\ -\infty < y_2 < \infty$, then Y_1 and Y_2 are said to be *jointly continuous random variables.* The function $f(y_1, y_2)$ is called the *joint probability density function.*

Probability and Applied Statistics Formula Sheet
Alexis Petito

Theorem 5.2
If Y_1 and Y_2 are random variables with joint distribution function, $F(y_1,\ y_2)$, then,
$F(-\infty,-\infty) = F(-\infty,\ y_2) = F(\ y_1,-\infty) = 0.$
2. $F(\infty,\infty) = 1.$
3. If $y_1^* \geq y_1\ and\ y_2^* \geq y_2\ then$

$$F(y_1^*, y_2^*) - F(\ y_1^*, y_2) - F(y_1,\ y_2^*) + F(y_1, y_2) \geq 0.$$

Theorem 5.2
If Y_1 and Y_2 are jointly continuous random variables with a joint density function given by $f(y_1, y_2)$, then
1. $f(y_1, y_2) \geq 0\ for\ all\ y_1,\ y_2.$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)dy_1dy_2 = 1.$

Definition 5.4
a. Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. The *marginal probability functions* of Y_1 and Y_2, respectively, are given by,

$$p_1(y_1) = \sum_{all\ y_2} p(y_1, y_2)\ \ and\ \ p_2(y_2) = \sum_{all\ y_1} p(y_1, y_2).$$

b. Let Y_1 and Y_2 be jointly continuous random variables with joint density function $f(y_1, y_2)$. Then the *marginal density functions* of Y_1 and Y_2, respectively, are given by,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_2\ \ and\ \ f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_1.$$

Definition 5.5
If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively then the *conditional discrete probability function* of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_2, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)},$$

Provided that $p_2(y_2) > 0$.

Definition 5.6
If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$, then the *conditional distribution function* Y_1 given $Y_2 = y_2$ is,
$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2).$

Definition 5.7
Let Y_1 and Y_2 be jointly continuous random variables with joint density $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density of Y_1 given $Y_2 = y_2$ is given by

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

And, for any y_1 such that $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is given by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_2(y_2)}.$$

Definition 5.8
Let Y_1 have distribution function $F_1(y_1)$, Y_2 have distribution function $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be *independent* if and only if

$F(y_1, y_2) = F_1(y_1)F_2(y_2)$

for every pair of real numbers (y_1, y_2).

If Y_1 and Y_2 are not independent, they are said to be *dependent*.

Theorem 5.4
If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$p(y_1, y_2) = p_1(y_1)p_2(y_2)$

For all pairs of real numbers (y_1, y_2).

If Y_1 and Y_2 are continuous random variables with joint density function $f(y_1, y_2)$ and marginal density functions $f_1(y_1)$ and $F_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$f(y_1, y_2) = f_1(y_1)f_2(y_2)$

For all pairs of real numbers (y_1, y_2).

Theorem 5.5
Let Y_1 and Y_2 have a joint density $f(y_1, y_2)$ that is positive If and only if $a \leq y_1 \leq b$ and $c \leq y_2 \leq d$, for constants $a,\ b,\ c$ and d; and $f(y_1, y_2) = 0$ otherwise. Then Y_1 and Y_2 are independent random variables if and only if

$f(y_1, y_2) = g(y_1)h(y_2)$

Where $g(y_1)$ is a nonnegative function of y_1 alone and $h(y_2)$ is a nonnegative function of y_2 alone.