A. Statistical inference.

1. Relative frequency density (RFD).

$$RFD = \frac{\text{frequency (number of data points in given bin)}}{\text{(bin width } B) \times \text{(data set size } n)}$$

2. Sample mean and standard deviation.

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \quad s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1}}.$$

3. Sample mean and standard deviation for grouped data.

$$\overline{x} = \frac{f_1 y_1 + f_2 y_2 + \cdots + f_m y_m}{n}, \quad s = \sqrt{\frac{f_1(y_1 - \overline{x})^2 + f_2(y_2 - \overline{x})^2 + \cdots + f_m(y_m - \overline{x})^2}{n-1}},$$

where $y_1, y_2, \ldots, y_m$ are the distinct values that the data assumes, and each value y_k occurs f_k times.

4. Standard normal probabilities. If X is $N(0,1)$, then

$$P(-L < x < L) = p,$$

where the "critical value" L corresponds to the probability p as follows:

- $L = 1$: $p = 0.683 = 68.3\%$
- $L = 2$: $p = 0.955 = 95.5\%$
- $L = 3$: $p = 0.997 = 99.7\%$
- $L = 1.96$: $p = 0.950 = 95\%$
- $L = 2.33$: $p = 0.980 = 98\%$
- $L = 2.58$: $p = 0.990 = 99\%$

5. NISNID (normal is standard normal in disguise) fact.

If X is $N(\mu, \sigma)$, then $Z = \dfrac{X - \mu}{\sigma}$ is $N(0,1)$.

6. Confidence intervals.

(a) For a population mean:

$$\left(\overline{x} - L\frac{s}{\sqrt{n}}, \overline{x} + L\frac{s}{\sqrt{n}}\right).$$

Here $L = 1.96$ for a 95% confidence interval; $L = 2.33$ for a 98% confidence interval; $L = 2.58$ for a 99% confidence interval.

(b) For a population proportion:

$$\left(\widehat{p} - L\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + L\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).$$

Again, $L = 1.96/2.33/2.58$ for a 95%/98%/99% confidence interval.

In a confidence interval for a population proportion, the quantity

$$L\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

is called the *margin of error*. Since $\widehat{p}(1-\widehat{p}) \leq 1/4$, the margin of error is never larger than

$$\frac{L}{2\sqrt{n}}.$$

7. The "test statistic" or "z-score" for hypothesis testing of a population mean.

$$z = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})}.$$

If $|z|$ is larger than or equal to 1.96/2.33/2.576, then reject H_0 and accept H_A at the 95%/98%/99% level respectively.

B. Random variables, expected value, variance.

1. Definition of random variable.

A random variable X is a function defined on the sample space S of an experiment. That is, a random variable X is a way of assigning a number to each possible outcome of an experiment.

2. Random variable probabilities.

If X is a random variable, then $P(X = x)$ is the probability of X taking the value x.

3. Probability mass function (pmf).

The probability mass function (pmf) of a random variable X just means the function $P(X = x)$.

4. Expected value.

If X is a random variable, then the expected value $E[X]$ is defined by

$$E[X] = \sum_{\substack{\text{values } x \\ \text{of } X}} x \cdot P(X = x),$$

where the sum is over all possible values x of X.

5. Sum rule for expected values.

If $X_1, X_2, X_3, \ldots X_n$ are random variables, then

$$E[X_1 + X_2 + X_3 \cdots + X_n] = E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n].$$

6. Variance.

If X is a random variable, then the variance $\mathrm{Var}[X]$ is defined by

$$\mathrm{Var}[X] = E[(X-\mu)^2] = \sum_x (x-\mu)^2 \cdot P(X = x),$$

where $\mu = E[X]$.

7. Sum rule for variance.

If the random variables $X_1, X_2, X_3, \dots X_n$ are independent, then

$$\mathrm{Var}[X_1 + X_2 + X_3 \cdots + X_n] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \mathrm{Var}[X_3] + \cdots + \mathrm{Var}[X_n].$$

C. Bernoulli and binomial random variables.

1. Mean and variance of a Bernoulli random variable.

Suppose X is a Bernoulli random variable, meaning $X = 1$ if a certain event happens and $X = 0$ if not. Suppose the probability of that event happening (that is, the probability of a "success") is p. Then

$$E[X] = p, \qquad \text{Var}\,[X] = p(1-p).$$

2. Probability mass function, mean, and variance of a binomial random variable. Suppose a binomial experiment – that is, an experiment made up of repeated, independent trials of a Bernoulli experiment – has $P(\text{success}) = p$ (for a single trial). Suppose n is the number of trials. Let X denote the number of successes in the n trials. Then we say "X is $B(n,p)$," and we have

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (0 \le k \le n),$$
$$E[X] = np, \qquad \text{Var}[X] = np(1-p).$$

D. Poisson random variables.

1. Probability mass function. Suppose a certain event happens, on average, λ times in each interval of a given extent. Let X denote the actual number of times it happens in such an interval. Then we say "X is $P(\lambda)$," and for $k = 0, 1, 2, \ldots,$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

2. Mean and variance. If X is $P(\lambda)$, then

$$E[X] = \text{Var}[X] = \lambda.$$

E. Basic probability.

1. Permutations.

(a) The number of ways of arranging n objects in order is

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1.$$

(b) The number of ways of arranging r objects (in order) out of n objects is

$$n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!}.$$

(c) The number of ways of arranging n objects, where n_1 of them are the same, n_2 of them are the same, $\ldots n_r$ of them are the same, is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

2. Combinations.

(a) (Combinations.) The number of ways of choosing r objects out of n objects, without keeping track of order, is

$$\frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.$$

This number is sometimes called "n choose r," written $\binom{n}{r}$. Note that this is also the number of r-element subsets of a set with n elements.

(b) The number of ways of placing n objects into r distinct groups, of size $n_1, n_2, \ldots, n_r$, is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

(same number as in 1(c) above).

3. Probability axioms.

(a) $P(A) \geq 0$ for any event A.

(b) $P(S) = 1$, where S is the sample space.

(c) If the events $A_1, A_2, A_3, A_4, \ldots$ are mutually exclusive (no two of them can happen together), then

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + P(A_4) + \cdots.$$

(The list $A_1, A_2, A_3, A_4, \ldots$ could be finite or infinite.)

4. Basic probability rules and formulas.

(a) If all outcomes in a sample space S are equally likely, and $|E|$ denotes the number of outcomes in the event E, then

$$P(E) = \frac{|E|}{|S|}$$

(assuming the sample space is a finite set).

(b) For any event E,

$$P(E) = 1 - P(E^c),$$

where E^c denotes the complement of E (meaning all outcomes in the sample space except those in E).

(c) For any events E and F (not necessarily mutually exclusive), we have

$$P(E \cup F) = P(E) + P(F) - P(EF).$$

(d) For any events E, F, and G (not necessarily mutually exclusive), we have

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG).$$

F. Conditional probability.

1. Formulas for $P(E|F)$.

(a) Given any events E and F, we have

$$P(E|F) = \frac{P(EF)}{P(F)}.$$

(b) Suppose all events in the sample space are equally likely. Then for any events E and F, we have

$$P(E|F) = \frac{|EF|}{|F|}.$$

2. Formulas for $P(EF)$.

(a) Given any events E and F, we have

$$P(EF) = P(F) \cdot P(E|F).$$

(b) (Generalization.) Given any events A_1, A_2, and A_3, we have

$$P(A_1A_2A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1A_2).$$

(c) (Further generalization.) Given any finite or infinite list of events, we have

$$P(\text{all events happen}) = P(\text{first one happens}) \cdot P(\text{second happens given that first does}) \\ \cdot P(\text{third does given that first two do}) \cdot P(\text{fourth does given that first three do}) \cdots .$$

3. Independent events.

(a) If events E and F are independent ($P(E) = P(E|F)$), we have

$$P(EF) = P(E) \cdot P(F).$$

(b) (Generalization.) If events $A_1, A_2, A_3, A_4, \ldots$ are independent (they don't affect each other), then

$$P(A_1A_2A_3A_4\cdots) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) \cdots .$$

(The list $A_1, A_2, A_3, A_4, \ldots$ could be finite or infinite.)

4. Bayes's Formula (also known as the law of total probability).

(a) For any events E and F,

$$P(E) = P(F)P(E|F) + P(F^c)P(E|F^c)$$

(again, F^c denotes the complement of F).

(b) (Generalization.) Suppose $F_1, F_2, F_3, \ldots, F_n$ are mutually exclusive and exhaustive events. Then

$$P(E) = P(F_1)P(E|F_1) + P(F_2)P(E|F_2) + P(F_3)P(E|F_3) + \cdots + P(F_n)P(E|F_n).$$
