Summary of Formulas and Concepts

Descriptive Statistics (Ch. 1-4)

Definitions

Population: The complete set of numerical information on a particular quantity in which an investigator is interested. We assume a population consists of $N$ values.

Sample: An observed subset of population values. We assume a sample consists of $n$ values.

Graphic Summaries

Dotplot - A graphic to display the original data. Draw a number line, and put a dot on it for each observation. For identical or really close observations, stack the dots.

Histogram - A graphic that displays the shape of numeric data by grouping it in intervals.
1. Choose evenly spaced categories
2. Count the number of observations in each group
3. Draw a bar graph with the height of each bar equal to the number of observations in the corresponding interval.

Stem and leaf plot Similar to a dotplot. Data are grouped according to their leading digits, and the last digit is used as a plotting symbol (like a dot in the dotplot).

The left digits are a cumulative count on each side of the middle. The bracketed number is how many observations are in the middle. The middle column of digits are the first digits of the number, and the “bars” are the last digit.

2 4 57
5 5 133
13 5 5677899
23 6 022334444
39 6 566777788888899
(18) 7 00000111222222234
38 7 555555666668889999999
16 8 0022333444
Numeric Summaries of Data

Suppose that we have \( n \) observations, labeled \( x_1, x_2, \ldots, x_n \). Then \( \sum_{i=1}^{n} x_i \) means

\[
\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n
\]

Some other relations are:

\[
\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + f(x_3) + \ldots + f(x_n), \text{ for any function } f, \\
\sum_{i=1}^{n} cx_i = c \sum_{i=1}^{n} x_i \text{ for any constant } c, \\
\sum_{i=1}^{n} (ax_i + by_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i, \text{ for any constants } a \text{ and } b.
\]

Measures of Location - numeric summaries of the center of a distribution.

Mean (or average)

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \text{(population)} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{(sample)}
\]

Median - the middle observation

The middle observation of the sorted data if \( n \) is odd, otherwise the average of the two middle values.

Measures of Dispersion - numeric summaries of the spread or variation of a distribution.

Variance and Standard Deviation

\[
\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \quad \text{(population)} \quad s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

Population and sample standard deviations (\( \sigma \) and \( s \)) are just the square roots of these quantities.

Interpreting \( \sigma \)

Chebyshev’s rule: for any population

at least 75% of the observations lie within \( 2\sigma \) of \( \mu \),

at least 89% of the observations lie within \( 3\sigma \) of \( \mu \),

at least \( 100(1 - 1/m^2)\% \) of the observations lie within \( m \times \sigma \) of the mean \( \mu \).
**IQR** (Interquartile Range): The distance between the \((n + 1)/4\)th and \(3 \times (n + 1)/4\)th observations in an ordered dataset. These two values are called the first and third quartiles.

**Measure of symmetry : Skewness**

\[
\text{skewness} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3}{n s^3}
\]

Negative means left skewed, 0 means symmetric, positive means right skewed.

**Measure of heavy tails : Kurtosis**

\[
\text{kurtosis} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{n s^4} - 3
\]

A normal distribution has a kurtosis of 3, so if we subtract 3, interpretations are relative to that. Positive values mean a sharper peak, and negative values mean a flatter top than a normal distribution.

**Box-and-whisker plot:** A graphic that summarizes the data using the median and quartiles, and displays outliers. Good for comparing several groups of data.

```
   o
           |
           |
           |
   first quartile       median        third quartile       largest value below outlier

   Q3 + 1.5 IQR
```
Probability (Ch. 5)

Definitions and Set Theory:

**Random experiment:** A process leading to at least two possible outcomes with uncertainty as to which will occur.

**Basic outcome:** A possible outcome of the random experiment.

**Sample space:** The set of all basic outcomes.

**Event:** A set of basic outcomes from the sample space. An event is said to occur if any one of its constituent basic outcomes occurs.

**Combining events:** let $A$ and $B$ be two events.

<table>
<thead>
<tr>
<th>Technical</th>
<th>Symbol</th>
<th>Pronounced</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union</td>
<td>$A \cup B$</td>
<td>$A$ or $B$</td>
<td>$A$ occurs or $B$ occurs or both occur</td>
</tr>
<tr>
<td>Intersection</td>
<td>$A \cap B$</td>
<td>$A$ and $B$</td>
<td>$A$ occurs and $B$ occurs</td>
</tr>
<tr>
<td>Complement</td>
<td>$\overline{A}$</td>
<td>not $A$</td>
<td>$A$ does not occur</td>
</tr>
</tbody>
</table>

**Venn Diagrams:**

- $\text{shaded} = A \cap B$
- $\text{shaded} = A \cup B$
- $\overline{A}$ shaded
- $A, \text{B mutually exclusive}$
- $\text{A,B,C collectively exhaustive}$

**Probability Postulates:**

1. If $A$ is any event in the sample space $S$, $0 \leq P(A) \leq 1$

2. Let $A$ be an event in $S$ and let $O_i$ denote the basic outcomes. Then $P(A) = \sum P(O_i)$, where the notation implies that the summation extends over all the basic outcomes in $A$.

3. $P(S) = 1$

**Probability rules for combining events:**

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A) + P(B)$ if $A$ and $B$ mutually exclusive.
- $P(\overline{A}) = 1 - P(A)$

**Conditional Probability:**

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ provided $P(B) > 0$.
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
Independence:

Two events are \textit{Statistically Independent} if and only if

\[ P(A \cap B) = P(A)P(B) \]

or equivalently \( P(A|B) = P(A) \) and \( P(B|A) = P(B) \).

General case: events \( E_1, E_2, \ldots, E_k \) are independent if and only if

\[ P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1)P(E_2) \ldots P(E_k) \]

Bivariate Probability

Probabilities of outcomes for bivariate events:

<table>
<thead>
<tr>
<th></th>
<th>( B_1 )</th>
<th>( B_2 )</th>
<th>( B_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( P(A_1 \cap B_1) )</td>
<td>( P(A_1 \cap B_2) )</td>
<td>( P(A_1 \cap B_3) )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( P(A_2 \cap B_1) )</td>
<td>( P(A_2 \cap B_2) )</td>
<td>( P(A_2 \cap B_3) )</td>
</tr>
<tr>
<td></td>
<td>( P(B_1) )</td>
<td>( P(B_2) )</td>
<td>( P(B_3) )</td>
</tr>
</tbody>
</table>

\( P(A_1 \cap B_2) \) is a probability of \( A_1 \) and \( B_2 \) occurring.

\( P(A_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2) + P(A_1 \cap B_3) \) is the \textit{marginal probability} that \( A_1 \) occurs.

If we think of \( A_1, A_2 \) as a group of attributes \( A \), and \( B_1, B_2, B_3 \) as a group of attributes \( B \), then \( A \text{ and } B \text{ are independent} \) only if every one of \( \{A_1, A_2\} \) are independent of every one of \( \{B_1, B_2, B_3\} \).

Discrete Random Variables (Ch. 6-7)

Definitions

\textbf{Random Variable:} (r.v.) A variable that takes on numerical values determined by the outcome of a random experiment.

\textbf{Discrete Random Variable:} A r.v. that can take on no more than a countable number of values.

\textbf{Continuous Random Variable:} A r.v. that can take any value in an interval.

Notation: An upper case letter (e.g. \( X \)) will represent a r.v.; a lower case letter (e.g. \( x \)) will represent one of its possible values.
Discrete Probability Distributions

The **probability function**, \( P_X(x) \), of a discrete r.v. \( X \) gives the probability that \( X \) takes the value \( x \):

\[
P_X(x) = P(X = x)
\]

where the function is evaluated at all possible values of \( x \).

**Properties:**
1. \( P_X(x) \geq 0 \) for any value \( x \)
2. \( \sum_x P_X(x) = 1 \)

**Cumulative probability function**, \( F_X(x_0) \) of a r.v. \( X \):

\[
F_X(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} P_X(x).
\]

**Properties:**
1. \( 0 \leq F_X(x) \leq 1 \) for any \( x \)
2. If \( a < b \), then \( F_X(a) \leq F_X(b) \).

Expectation of Discrete Random Variables

**Expected value** of a discrete r.v.:

\[
E(X) = \mu_X = \sum_x xP_X(x).
\]

For any function \( g(X) \),

\[
E(g(X)) = \sum_x g(x)P_X(x).
\]

**Variance** of a discrete r.v.:

\[
\text{Var}(X) = \sigma_X^2 = E((X - \mu_X)^2) = \sum_x (x - \mu_X)^2 P_X(x) = E(X^2) - \mu_X^2.
\]

The **standard deviation** of \( X \) is \( \sigma_X \).

**Plug-In Rules:** let \( X \) be a r.v., and \( a \) and \( b \) constants. Then

\[
E(a + bX) = a + bE(X)
\]

\[
\text{Var}(a + bX) = b^2\text{Var}(X).
\]

This only works for linear functions.
Jointly Distributed Discrete Random Variables

**Joint Probability Function:** Suppose $X$ and $Y$ are r.v.’s. Their joint probability function gives the probability that simultaneously $X = x$ and $Y = y$:

$$P_{X,Y}(x, y) = P\{X = x \cap Y = y\}$$

**Properties:**
1. $P_{X,Y}(x, y) \geq 0$ for any pair $(x, y)$
2. $\sum_x \sum_y P_{X,Y}(x, y) = 1$.

**Marginal probability function:**

$$P_X(x) = \sum_y P_{X,Y}(x, y).$$

**Conditional probability function:**

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

**Independence:** $X$ and $Y$ are independent if and only if

$$P_{X,Y}(x, y) = P_X(x)P_Y(y) \quad \text{for all possible } (x, y) \text{ pairs}$$

**Expectation:** Let $X$ and $Y$ be r.v.’s, and $g(X, Y)$ any function. Then

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P_{X,Y}(x, y).$$

**Conditional Expectation:** Let $X$ and $Y$ be r.v.’s, and suppose we know the conditional distribution of $X$ for $Y = y$, labeled $P_{X|Y}(x|y)$. Then

$$E(X|Y = y) = \sum_x xP_{X|Y}(x|y).$$

**Covariance:** If $E(X) = \mu_X$ and $E(Y) = \mu_Y$,

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)P_{X,Y}(x, y)$$

$$= E(XY) - \mu_X\mu_Y = \left[ \sum_x \sum_y xyP_{X,Y}(x, y) \right] - \mu_X\mu_Y$$

If two r.v.’s are independent, their covariance is zero. The converse is not necessarily true.
Correlation:

\[ \rho_{XY} = \text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \]

\(-1 \leq \rho_{XY} \leq 1\) always.
\(\rho_{XY} = \pm 1\) if and only if \(Y = a + bX\) \((a,b \text{ constants})\).

Plug-in rules: Let \(X\) and \(Y\) be r.v.’s, and \(a, b\) constants. Then

\[ E(aX + bY) = aE(X) + bE(Y) \]
\[ \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y) \]

Binomial distribution

Bernoulli Trials:
A sequence of repeated experiments are Bernoulli trials if:
1. The result of each trial is either a success or failure.
2. The probability \(p\) of a success is the same for all trials.
3. The trials are independent.

If \(X\) is the number of successes in \(n\) Bernoulli trials, \(X\) is a **Binomial Random Variable**. It has probability function:

\[ P_X(x) = \binom{n}{x}p^x(1 - p)^{n-x} \]

Where \(\binom{n}{x}\) counts the number of ways of getting \(x\) successes in \(n\) trials. The formula for \(\binom{n}{x}\) is

\[ \binom{n}{x} = \frac{n!}{x!(n-x)!} \]

where \(n! = n \times (n - 1) \times (n - 2) \times ... \times 2 \times 1\).

**Mean and Variance:** \(E(X) = np, \text{Var}(X) = np(1 - p)\).

Continuous Random Variables (Ch. 7)

Probability Distributions

**Probability density function:** A function \(f_X(x)\) of the continuous r.v. \(X\) with the following properties:
1. \( f_X(x) \geq 0 \) for all values of \( x \).
2. \( P(a \leq X \leq b) = \) the area under \( f_X(x) \) between \( a \) and \( b \), if \( a < b \).
3. The total area under the curve is 1
4. The area under the curve to the left of any value \( x \) is \( F_X(x) \), the probability that \( X \) does not exceed \( x \).

**Cumulative distribution function:** Same as before.

\[ P(a \leq X \leq b) = F_X(b) - F_X(a) \quad (\text{provided } a < b). \]

**Expectations, Variances, Covariances, etc.**

Same rules as for discrete r.v.’s. The summation \((\sum)\) is replaced by the integral \((\int)\), which is not necessary for this course.

**Normal Distribution**

**Probability Density function:**

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x - \mu)^2}{2\sigma^2}} \]

for constants \( \mu \) and \( \sigma \) such that \(-\infty < \mu < \infty\) and \( 0 < \sigma < \infty \).

**Mean and Variance:** \( E(X) = \mu \quad \text{Var}(X) = \sigma^2 \)

**Notation:** \( X \sim N(\mu, \sigma^2) \) means \( X \) is normal with mean \( \mu \) and variance \( \sigma^2 \).

If \( Z \sim N(0,1) \) we say it has a **standard normal distribution**.
If \( X \sim N(\mu, \sigma^2) \) then \( Z = (X - \mu)/\sigma \sim N(0,1) \). Thus

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

**Central Limit Theorem**

Let \( X_1, X_2, \ldots, X_n \) be \( n \) independent r.v.’s, each with identical distributions, mean \( \mu \) and variance \( \sigma^2 \). As \( n \) becomes large,

\[ \overline{X} \sim N(\mu, \sigma^2/n) \]

\[ \sum_{i=1}^{n} X_i = n \overline{X} \sim N(n\mu, n\sigma^2) \]
Sampling & Sampling distributions

Simple random sample: (or random sample) A method of randomly drawing \( n \) objects which are Independent and Identically Distributed (I.I.D.).

Statistic: A function of the sample information.

Sampling distribution of a statistic: The probability distribution of the values a statistic can take, over all possible samples of a fixed size \( n \).

Sampling distribution of the mean: Suppose \( X_1, \ldots, X_n \) are a random sample from some population with mean \( \mu_X \) and variance \( \sigma^2_X \). The **sample mean** is

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

It has the following properties:
1. \( E(\bar{X}) = \mu_X \)
2. It has standard deviation \( \sigma_{\bar{X}} = \sigma_X / \sqrt{n} \).
3. If the population distribution is normal,
   \[
   \bar{X} \sim N(\mu_X, \sigma^2_X) = N(\mu_X, \sigma^2_X / n).
   \]
4. If the population distribution is not normal, but \( n \) is large, then (3) is roughly true.

Sampling distribution of a proportion: Suppose the r.v. \( X \) is the number of successes in a binomial sample of \( n \) trials, whose probability of success is \( p \). The **sample proportion** is

\[
\hat{p} = X/n
\]

It has the following properties:
1. \( E(\hat{p}) = p \)
2. It has standard deviation \( \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \).
3. If \( n \) is large \((np(1-p) > 9 \text{ or roughly } n \geq 40)\),
   \[
   \hat{p} \sim N(p, \sigma^2_{\hat{p}}) = N(p, p(1-p)/n).
   \]
Point Estimation

**Estimator:** A random variable that depends on the sample information and whose realizations provide approximations to an unknown population parameter.

**Estimate:** A specific realization of an estimator.

**Point estimator:** An estimator that is a single number.

**Point estimate:** A specific realization of a point estimator.

**Bias:** Let \( \hat{\theta} \) be an estimate of the parameter \( \theta \). The bias in \( \hat{\theta} \) is

\[
\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.
\]

If the bias is 0, \( \hat{\theta} \) is an **unbiased estimator**.

**Efficiency:** Let \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) be two estimators of \( \theta \), based on the same sample. Then \( \theta_1 \) is **more efficient** than \( \hat{\theta}_2 \) if

\[
\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2).
\]

Interval Estimation

**Confidence Interval:** Let \( \theta \) be an unknown parameter. Suppose that from sample information, we can find random variables \( A \) and \( B \) such that

\[
P(A < \theta < B) = 1 - \alpha.
\]

If the observed values are \( a \) and \( b \), then \((a, b)\) is a 100\((1 - \alpha)\)% confidence interval for \( \theta \). The quantity \((1 - \alpha)\) is called the **probability content** of the interval.

**Student’s t distribution:** Given a random sample of \( n \) observations with mean \( \bar{X} \) and standard deviation \( s \), from a normal population with mean \( \mu \), the random variable

\[
T = \frac{\bar{X} - \mu}{s/\sqrt{n}}
\]

follows the Student’s t distribution with \((n - 1)\) degrees of freedom. For \( n > 30 \), the t distribution is quite close to a \( N(0, 1) \) distribution.


<table>
<thead>
<tr>
<th>Data</th>
<th>Parameter</th>
<th>100(1 − α)% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N(\mu, \sigma^2), \sigma^2 \text{ known})</td>
<td>(\mu)</td>
<td>(\bar{x} \pm \frac{z_{\alpha/2} \sigma}{\sqrt{n}})</td>
</tr>
<tr>
<td>mean (\mu, \sigma^2 \text{ unknown}, n &gt; 30)</td>
<td>(\mu)</td>
<td>(\bar{x} \pm \frac{z_{\alpha/2} s}{\sqrt{n}})</td>
</tr>
<tr>
<td>(N(\mu, \sigma^2), \sigma^2 \text{ unknown})</td>
<td>(\mu)</td>
<td>(\bar{x} \pm \frac{t_{n-1,\alpha/2} s}{\sqrt{n}})</td>
</tr>
<tr>
<td>Binomial((n, p)), (np(1-p) &gt; 9), or roughly (n \geq 40)</td>
<td>(p)</td>
<td>(\hat{p} \pm \frac{z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})}}{n})</td>
</tr>
<tr>
<td>(n) matched pairs, difference (\sim N(\mu_X - \mu_Y, \sigma^2))</td>
<td>(\mu_X - \mu_Y)</td>
<td>(\bar{d} \pm \frac{t_{n-1,\alpha/2} s_d}{\sqrt{n}})</td>
</tr>
<tr>
<td>2 independent samples, means (\mu_X, \mu_Y) variances unknown, (n &gt; 30)</td>
<td>(\mu_X - \mu_Y)</td>
<td>(\bar{x} - \bar{y} \pm \frac{z_{\alpha/2} \sqrt{s_x^2/n_x + s_y^2/n_y}}{\sqrt{n_x+n_y}})</td>
</tr>
<tr>
<td>2 independent samples, means (\mu_X, \mu_Y) variances unknown</td>
<td>(\mu_X - \mu_Y)</td>
<td>(\bar{x} - \bar{y} \pm \frac{t_{n^*,\alpha/2} \sqrt{s_x^2/n_x + s_y^2/n_y}}{\sqrt{n_x+n_y}})</td>
</tr>
<tr>
<td>2 independent samples, Binomial((n_X, p_X), \text{Binomial}(n_Y, p_Y))</td>
<td>(p_X - p_Y)</td>
<td>(\hat{p}_x - \hat{p}<em>y \pm \frac{z</em>{\alpha/2} \sqrt{\hat{p}_x(1-\hat{p}_x)/n_x + \hat{p}_y(1-\hat{p}_y)/n_y}}{\sqrt{n_x+n_y}})</td>
</tr>
</tbody>
</table>

Notes for the table:

1. All quantities in the C.I. column are either known constants or observed sample quantities.
2. \(P(Z > z_{\alpha/2}) = \alpha/2\) for \(Z \sim N(0, 1)\).
3. \(P(T > t_{n-1,\alpha/2}) = \alpha/2\) for \(T \sim \text{Student's t}\) with \((n - 1)\) d.f.
4. \(s, s_x, s_y, s_d\) are observed sample standard deviations corresponding to \(x_i, x_i, y_i, d_i = x_i - y_i\) respectively.
5. \(\hat{p}, \hat{p}_x, \hat{p}_y\) are the observed sample proportions corresponding to \(x_i, x_i, y_i\) respectively.
6. \(\bar{d}\) is the sample mean corresponding to \(d_i = x_i - y_i\).
7. \(n, n_x, n_y\) are the total, \(x\) and \(y\) sample sizes.
8. \( n^* = \left( \frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^2 \left/ \left[ \frac{(s_x^2/n_x)^2}{n_x - 1} + \frac{(s_y^2/n_y)^2}{n_y - 1} \right] \right. \)

Also note: The text gives a different formula for comparing two means with small sample sizes. It requires that the two variances be the same, which may not be the case. Unless you’re sure the variances are equal, it’s safer to use the approximation given here (the formula with a \( n^* \)). If you are sure that the variances are equal, using the book’s formula is ok.

**Estimating the sample size:** If you want a \( 100(1-\alpha)\% \) interval of \( \pm L \) (i.e. length \( 2L \)), choose \( n \) so

<table>
<thead>
<tr>
<th>situation</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal, ( \sigma ) known</td>
<td>( n = \frac{z_{\alpha/2}^2 \sigma^2}{L^2} )</td>
</tr>
<tr>
<td>Bernoulli, worst case</td>
<td>( n = \frac{0.25 z_{\alpha/2}^2}{L^2} )</td>
</tr>
</tbody>
</table>

**Hypothesis Testing**

**Null Hypothesis** (\( H_0 \)): The hypothesis we assume to be true unless there is sufficient evidence to the contrary.

**Alternative Hypothesis** (\( H_1 \)): The hypothesis we test the null against. If there is evidence that \( H_0 \) is false, we accept \( H_1 \).

**Type I Error:** Rejecting a true \( H_0 \).

**Type II Error:** Not rejecting a false \( H_0 \).

**Significance Level:** \( P(\text{reject} H_0 | H_0 \text{true}) = P(\text{type I error}) \).

**Power:** The probability of rejecting a null hypothesis that is false. Note that this depends on the true value of the parameter.

**P-value:** The smallest significance level at which a null hypothesis can be rejected. This is a measure of how likely the data is, if \( H_0 \) is true.

Notes for the following table (In addition to the comments for CI’s):

1. The first three tests are examples of one (> and < alternatives) and two sided (\( \neq \) alternative) tests. The remaining tests all have a > alternative, but are easily adaptable to either of the other two alternatives.
2. In the last test,

\[ \hat{p}_0 = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y} \]

3. In all the tests, we are comparing the unknown parameter (such as \( \mu, p \) or \( \mu_X - \mu_Y \)) to constants (\( \mu_0, p_0 \), and \( D_0 \)).

**Hypothesis tests with significance level \( \alpha \)**

<table>
<thead>
<tr>
<th>Data</th>
<th>( H_0 )</th>
<th>( H_1 )</th>
<th>reject ( H_0 ) if</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N(\mu, \sigma^2), \sigma^2 ) known</td>
<td>( \mu = \mu_0 ) or ( \mu \leq \mu_0 )</td>
<td>( \mu &gt; \mu_0 )</td>
<td>( \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} &gt; z_\alpha )</td>
</tr>
<tr>
<td>same</td>
<td>( \mu = \mu_0 ) or ( \mu \geq \mu_0 )</td>
<td>( \mu &lt; \mu_0 )</td>
<td>( \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} &lt; -z_\alpha )</td>
</tr>
<tr>
<td>same</td>
<td>( \mu = \mu_0 )</td>
<td>( \mu \neq \mu_0 )</td>
<td>not in ( (-z_{\alpha/2}, z_{\alpha/2}) )</td>
</tr>
<tr>
<td>mean ( \mu, \sigma^2 ) unknown</td>
<td>( \mu = \mu_0 ) or ( \mu \leq \mu_0 )</td>
<td>( \mu &gt; \mu_0 )</td>
<td>( \frac{\bar{x} - \mu_0}{s/\sqrt{n}} &gt; z_\alpha )</td>
</tr>
<tr>
<td>( n &gt; 30 )</td>
<td>( \mu = \mu_0 ) or ( \mu \leq \mu_0 )</td>
<td>( \mu &gt; \mu_0 )</td>
<td>( \frac{\bar{x} - \mu_0}{s/\sqrt{n}} &gt; t_{n-1, \alpha} )</td>
</tr>
<tr>
<td>( N(\mu, \sigma^2), \sigma^2 ) unknown</td>
<td>( \mu = \mu_0 ) or ( \mu \leq \mu_0 )</td>
<td>( \mu &gt; \mu_0 )</td>
<td>( \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} &gt; z_\alpha )</td>
</tr>
<tr>
<td>Binomial((n, p))</td>
<td>( p = p_0 ) or ( p \leq p_0 )</td>
<td>( p &gt; p_0 )</td>
<td>( \frac{\overline{d} - D_0}{s_d/\sqrt{n}} &gt; t_{n-1, \alpha} )</td>
</tr>
<tr>
<td>( n(1-p) &gt; 9 ) or roughly ( n &gt; 40 )</td>
<td>( \mu_X - \mu_Y = D_0 ) or ( \mu_X - \mu_Y \leq D_0 )</td>
<td>( \mu_X - \mu_Y &gt; D_0 )</td>
<td>( \mu_X - \mu_Y \leq D_0 )</td>
</tr>
</tbody>
</table>
Hypothesis tests with significance level $\alpha$

<table>
<thead>
<tr>
<th>Data</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>reject $H_0$ if</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 independent samples, means $\mu_X, \mu_Y$ variances unknown, $n_x &gt; 30, n_y &gt; 30$</td>
<td>$\mu_X - \mu_Y = D_0$ or $\mu_X - \mu_Y \leq D_0$</td>
<td>$\mu_X - \mu_Y &gt; D_0$</td>
<td>$\frac{\bar{x} - \bar{y} - D_0}{\sqrt{s_x^2/n_x + s_y^2/n_y}} &gt; z_\alpha$</td>
</tr>
<tr>
<td>2 independent normal samples, means $\mu_X, \mu_Y$ variances unknown</td>
<td>$\mu_X - \mu_Y = D_0$ or $\mu_X - \mu_Y \leq D_0$</td>
<td>$\mu_X - \mu_Y &gt; D_0$</td>
<td>$\frac{\bar{x} - \bar{y}}{s_x^2/n_x + s_y^2/n_y} &gt; t_{n^*, \alpha}$</td>
</tr>
<tr>
<td>2 independent samples, Binomial($n_x, p_x$) and Binomial($n_y, p_y$)</td>
<td>$p_x - p_y = 0$ or $p_x - p_y \leq 0$</td>
<td>$p_x - p_y &gt; 0$</td>
<td>$\frac{\hat{p}_x - \hat{p}_y}{\hat{p}_0(1 - \hat{p}<em>0)\left(\frac{n_x + n_y}{n_xn_y}\right)} &gt; z</em>\alpha$</td>
</tr>
</tbody>
</table>