Probability

Sampling
with replacement: n^k

without replacement, ordered: $\frac{n!}{(n-k)!} = nPk$

without replacement, unordered: $\frac{n!}{k!(n-k)!} = \binom{n}{k}$

Inclusion-Exclusion Principle
Given two events A and B,

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

Three events:

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

Conditional Probability
$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ where $P(B) > 0$

Statistical Independence
Two events A and B are independent iff

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

Three events A, B, C are independent if

1) $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$
2) Each set of pairwise events are indep.

Mutually Exclusive Events
Two events A and B are mutually exclusive iff

$$P(A \cap B) = 0$$

Law of Total Probability
Let $S = \sum_{k=1}^{n} B_k$ where $B_i \cap B_j = \emptyset$ if $i \neq j$ then

$$P(A) = \sum_{R=1}^{n} P(A \mid B_R) \cdot P(B_R)$$

Bayes' Theorem
Let $B_1, \ldots, B_n$ be a partition of S and for event A,

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{P(A)} = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{R=1}^{n} P(A \mid B_R) \cdot P(B_R)}$$

Discrete RVs

pmf
$$f(x) = P(X = x)$$
$$f(x) \geq 0 \text{ for all } x$$
$$\sum_{x \in X} f(x) = 1$$

cdf
$$F(x) = P(X \leq x)$$
$$F(x) \geq 0, \text{ non-decreasing}$$
$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$$

pmf ↔ cdf
$$\sum_{y \leq x} f(y) = F(x)$$
$$f(x) = F(x) - F(x-1)$$

Expectation
$$\mu = E(X) = \sum_{x \in X} x \cdot f(x)$$
$$E(g(X)) = \sum_{x \in X} g(x) \cdot f(x)$$
$$E(aX + bY) = aE(X) + bE(Y)$$
where X and Y don't have to be independent

$E(k) = k$ for constant k

$E(X - \mu) = 0$ for any r.v. X

$E(XY) = E(X) \cdot E(Y)$

where X and Y are independent

Memorylessness
X is memoryless if $\forall\, m, n \in \mathbb{Z}$,

$$P(X > m + n \mid X > m) = P(X > n)$$

Memoryless distributions:

- Geometric
- Exponential

Standard Deviation
$$\sigma = \sqrt{Var(X)}$$

Continuous RVs

pdf
$$P(a < X < b) = \int_a^b f(x)\,dx$$

cdf
$$F(x) = P(X \leq x)$$
(same properties as discrete cdf)

pdf ↔ cdf
$$F(X) = \int_{-\infty}^{x} f(y)\,dy$$
$$\frac{dF}{dx} = f(x)$$

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

Variance
$$Var(x) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left[\int_{-\infty}^{\infty} x f(x)\,dx\right]^2$$

Variance
$$Var(X) = E\left(\left(X - E(X)\right)^2\right) = E((X - \mu)^2) = E(X^2) - \mu^2$$

$Var(k) = 0$ for constant k

$$Var(aX + b) = a^2 Var(X)$$
$$Var(X + Y) = Var(X) + Var(Y)$$

where X and Y are independent

(for sample variance, see **Data Analysis** section)

Independence
X and Y are independent if $\forall x, y$,

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

or

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

Quadratic Formula
Given $ax^2 + bx + c = 0$,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Useful Antiderivatives
$$\int x^n dx = \frac{1}{n+1} x^{n+1} + C \qquad \int \frac{1}{x} dx = \ln\ |x| + C$$
$$\int e^x dx = e^x + C$$

Log Properties
$$\ln(a^b) = b \ln(a)$$
$$\ln(ab) = \ln(a) + \ln(b)$$
$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$$

Discrete Distributions

Distribution	Probability Function $f(x)$	Expectation $E(x)$	Variance $Var(x)$
Bernoulli $X \sim Ber(p)$ where X takes values {0, 1} with $P(X=1)=p$ $\{0, 1\}$	$\begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$	p	$p(1-p)$
Binomial $X \sim Bin(n,p)$ where X is the number of successes in n (Ber.) trials with a p chance of success each trial $\{0, 1, 2, \dots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
Geometric $X \sim Geo(p)$ where X is the number of (Ber.) trials required to observe the first success $\{1, 2, 3, \dots\}$	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial $X \sim NB(r,p)$ where X is the number of (Ber.) trials required to observe the r^{th} success $\{r, r+1, r+2, \dots\}$	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric $X \sim HGeo(N,r,n)$ where X is the number of successes when drawing n objects from N total objects containing r successes without replacement $\{\max(0, n+r-N), \dots, \min(r,n)\}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\frac{nr}{N}$	$\frac{nr}{N}(1-\frac{r}{N})\frac{N-n}{N-1}$
Poisson $X \sim Poi(\lambda)$ where X is the number of events to occur in a space given an average rate of λ events in the space $\{0, 1, 2, \dots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ

Properties

Poisson Process

An experiment is a Poisson process if it has the following:

- **independence**: # of occurrences in non-overlapping spaces is independent
- **individuality**: probability that 2 or more events occurring in a short interval is close to zero
- **homogeneity/uniformity**: probability of one event in a short interval is approximately λ

Addition of Binomial Distributions

If $X \sim Bin(n_1, p), Y \sim Bin(n_2, p)$ and X and Y are indep., then
$X + Y \sim Bin(n_1 + n_2, p)$

Poisson vs Binomial

If $n \to \infty, p \to 0$ and $np = \lambda$ then $Bin(n,p) \to Poi(\lambda)$

Bernoulli vs Binomial

If $X_1, \dots, X_n$ are independent $Ber(p)$ and
$X = X_1 + \cdots + X_n$, then $X \sim Bin(n,p)$

Geometric vs Negative Binomial

If $X_1, \dots, X_r$ are independent $Geo(p)$ and
$X = X_1 + \cdots + X_r$, then $X \sim NB(r,p)$

Hypergeometric vs Binomial

If N is large and n is small,
$HGeo(N,r,n) \sim Bin(n, r/N)$

Continuous Distributions

Distribution	Probability Density Function $f(x)$	Cumulative Distribution Function F(x)	Mean $E(x)$	Variance $Var(x)$
Uniform Continuous $X \sim Uniform(a,b)$ where X is a real number in the interval $[a,b]$ $[a,b]$	$\begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$	$\begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential $X \sim Exp(\lambda)$ where X is the "waiting time" between events given an average rate of λ events (Poisson) $[0,\infty)$	$\begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal $X \sim N(\mu, \sigma^2)$ $(-\infty,\infty)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$	μ	σ^2
Chi-Squared $W \sim \chi^2(n)$ where $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ such that $Z_i \sim N(0,1)$ and indep. **n: degrees of freedom (d.f.)** $(0,\infty)$	$\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} w^{\frac{n}{2}-1} e^{-\frac{w}{2}}$ $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$	$\int_{-\infty}^{x} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} w^{\frac{n}{2}-1} e^{-\frac{w}{2}} dw$	n	$2n$
Student's t $T \sim T(n)$ where $T = \frac{Z}{\sqrt{\frac{W}{n}}}$ where $Z \sim N(0,1)$ and $W \sim \chi^2(n)$ and independent **n: degrees of freedom (d.f.)** $(-\infty,\infty)$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	$\int_{-\infty}^{x} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt$	0	N/A

Properties

Universality of the Uniform Distribution

Let X be a continuous R.V. with an invertible cdf F, then

$$F(X) \sim U(0,1)$$

To generate a random variable for a given distribution, uniformly generate values and put them into the inverted cdf to get R.V. values.

Exponential vs Geometric

Exponential is to Poisson as Geometric is to Bernoulli
Bernoulli: seq. of 0 or 1 → "waiting time" for 1st success ~ Geo
Poisson: cont. vers. of Ber. → "waiting time" for 1st success ~ Exp

Special Cases of Chi-Squared Distribution

For $n = 1$: $W = Z^2$
e.g. $P(W \ge a) = P(Z^2 \ge a)$
For $n = 2$: $W \sim Exp(\lambda = \frac{1}{2})$
For intermediate n $(2 < n < 30)$: use χ^2 chart
For large n $(n \ge 30)$: $W \sim N(n, 2n)$ by CLT

Student's t vs Normal

As $n \to \infty, T \to Z \sim N(0,1)$

Normal and Standard Normal

Standard Normal

$$Z \sim N(0,1)$$

where the pdf ϕ is $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Normal ↔ Standard Normal

If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0,1)$

68-95-99 Rule

If $X \sim N(\mu, \sigma^2)$,

- ~68% are between $\mu \pm \sigma$
- ~95% are between $\mu \pm 2\sigma$
- ~99% are between $\mu \pm 3\sigma$

Central Limit Theorem

The Sampling Distribution of the Sample Mean

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, then

1. $S_n \sim N(n\mu, n\sigma^2)$
2. $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$

where $S_n = \sum X_i$ and $\bar{X}_n = \frac{1}{n}\sum X_i$

Standard Error of the Mean: $\frac{\sigma}{\sqrt{n}}$ (std. dev. of the mean)

Law of Large Numbers

Consider a sequence of n i.i.d random variables; $X_1, X_2, \ldots, X_n$, where $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for all i,

$$\text{Let } \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \text{, then } \bar{X}_n \to \mu \text{ as } n \to \infty$$

for any distribution with a finite mean and finite variance

Central Limit Theorem

Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ for some sequence of n i.i.d rvs $X_1, X_2, \ldots, X_n$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

As $n \to \infty$,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ or } Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

If X_i are normal distr., then $\bar{X}_n$ is normal

If X_i are not normal distr., then $\bar{X}_n$ approaches normal

CLT and Binomial

If $X \sim Bin(n,p)$, then $X \sim N(np, np(1-p))$

CLT and Poisson

If $X_1, X_2, \ldots, X_n \sim Poi(\lambda)$, then $\bar{X}_n \sim N\left(\lambda, \frac{\lambda}{n}\right)$

Data Analysis

Datatypes

Categorical

- Ordinal (implied order)

Numerical

- Continuous
- Discrete

Quantiles/Quartiles

Q_1(1st quartile) $= P_{25}$

Q_2(2nd quartile) = median $= P_{50}$

Q_3(3rd quartile) $= P_{75}$

5-number summary:

(min, Q_1, median, Q_3, max)

Population Symbols vs Sample Symbols

Symbol	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Std. deviation	σ	s
Std. error	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$

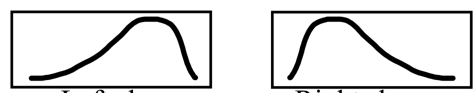
Measures of dispersion and symmetry

Range = max – min IQR $= Q_3 - Q_1$

Sample variance:

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$
$$= \frac{1}{n-1}\left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right]$$

Skewness:



Left skew Right skew

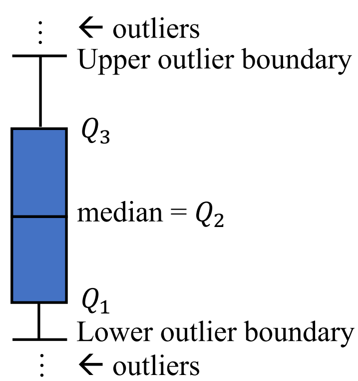
Density Histogram

Graph of relative frequencies of groups where rel. freq. = freq / total

Empirical CDF

$\hat{F}(y) = \sum_{Y \le y}$ rel. freq. of Y

Box Plots



upper outlier boundary: $Q_3 + 1.5\text{IQR}$

lower outlier boundary: $Q_1 - 1.5\text{IQR}$

Maximum Likelihood Estimates

Likelihood Function

If $Y_i \sim f(y_i, \theta), i = 1, 2, \ldots, n$, where Y_i are i.i.d. RVs with observations $\{y_1, y_2, \ldots, y_n\}$, then

$$L(\theta; y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = \prod_{i=1}^{n} f(y_i, \theta)$$

L is a function equal to the likelihood of observing outcomes $\{y_1, y_2, \ldots, y_n\}$ given θ as a parameter

For discrete distributions,

- $f(y_i, \theta)$ is the pmf of the distribution
- The likelihood function L is the probability of observing our data given θ

For continuous distributions,

- $f(y_i, \theta)$ is the pdf of the distribution
- The likelihood function L is NOT a probability

The Maximum Likelihood Estimate (MLE)

$\hat{\theta}$ is the MLE if $\hat{\theta}$ maximizes $L(\theta; y_1, y_2, \ldots, y_n)$

To find the MLE, set the derivative of the likelihood function L (or derivative of the log likelihood function $\ell = \ln L$) w.r.t the parameter equal to 0. Solve for θ.

Properties:

1. Consistency: As $n \to \infty, \hat{\theta} \to \theta$
   "As the size of our data increases, the MLE will be more accurate"
2. Efficiency: Variance is minimized when finding the MLE
3. Invariance: If $\hat{\theta}$ is the MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$

Distributions and their MLEs

Binomial	Poisson	Exponential	Normal
$\hat{\theta} = \frac{y}{n}$	$\hat{\lambda} = \bar{y}$	$\hat{\lambda} = \frac{1}{\bar{y}}$	$\hat{\mu} = \bar{y}$ $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

The Relative Likelihood Function

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$$

where L is the likelihood function and $\hat{\theta}$ is the MLE of θ

$0 \le R(\theta) \le 1$ for all θ, $R(\theta) = 1$ when $\theta = \hat{\theta}$

Interpretation:

$R(x) = k$ means:

"Given the observed data, a parameter value of x is k times as likely compared to the MLE value $\hat{\theta}$"

Likelihood Intervals

A $100p\%$ Likelihood Interval is $\{\theta: R(\theta) \ge p\}$

e.g. a 50% Likelihood Interval is $\{\theta: R(\theta) \ge 0.5\}$

Plausibility

- If $R(\theta) \ge 0.5$, it is **very plausible**
- If $R(\theta) \ge 0.1$, it is **plausible**
- If $R(\theta) < 0.1$, it is **implausible**
- If $R(\theta) < 0.01$, it is **very implausible**

Estimators

The estimator $\tilde{\theta}$ is a RV that has $\hat{\theta}$ as an outcome

Unbiased estimator: The expected value of the estimator is equal to the true population value

e.g. $E(S^2) = \sigma^2$

Biased estimator: The expected value of the estimator is **not** equal to the true population value

e.g. $E(\hat{\sigma}^2) \neq \sigma^2$

Confidence Intervals

Confidence Intervals

A 100p% Confidence Interval for θ is an estimate of the interval $[L, U]$ for RVs, $L(Y_1, \ldots, Y_n)$ and $U(Y_1, \ldots, Y_n)$, such that

$$P\big(L(Y_1, \ldots, Y_n) < \theta < U(Y_1, \ldots, Y_n)\big) = p$$

Interpretation:

WRONG: "There is a 100p% chance the interval contains θ"

RIGHT: "100p% of constructed confidence intervals contain θ"

RIGHT: "We are 100p% **confident** that θ is contained in the interval"

Choosing a sample size for a Binomial CI

To find a sample size n such that

$$z^* \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}} < \text{error}$$

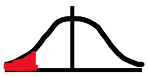
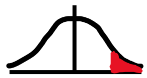
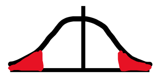
use $\hat{\theta} = 0.5$ and solve for n.

(see next page for CI formulas)

Hypothesis Testing

Hypothesis: some claim about a population (i.e. parameter)
Null Hypothesis (H_0): the "current belief"; status quo
Alternative Hypothesis (H_A): a challenge to H_0
p-value: the probability of observing our evidence given H_0

Types of Tests

One-tailed (left)	One-tailed (right)	Two-tailed
$H_0: \theta \geq k$ $H_A: \theta < k$	$H_0: \theta \leq k$ $H_A: \theta > k$	$H_0: \theta = k$ $H_A: \theta \neq k$
		

Type I and Type II errors

Type I error: Rejecting H_0 when H_0 is actually true

$P(\text{Rejecting } H_0 \text{ when } H_0 \text{ is actually true}) = \alpha$
where α is the significance level

Type II error: Failing to reject H_0 when H_0 is actually false

$P(\text{Failing to reject } H_0 \text{ when } H_0 \text{ is actually false}) = 1 - \beta$
where β is the power

Relationship between $p, \alpha, z_{test}, z_{crit}$

p-val vs α	z_{test} vs z_{crit}	Conclusion
$p < \alpha$	$z_{test} \notin [-z_{crit}, z_{crit}]$	Can reject $<H_0>$
$p > \alpha$	$z_{test} \in [-z_{crit}, z_{crit}]$	Can't reject $<H_0>$

Guidelines for t-tests and t-based CIs

Given that σ^2 is unknown and normality is not specified,
If $n > 40$: they perform well in most practical situations
If $15 < n < 40$: they perform reasonably well, but outliers or strong skewness can cause problems
If $n < 15$: we need to be confident the population is approximately normal before proceeding.

Confidence Intervals and Two-tailed Hypothesis Tests

A $100p\%$ CI consists of all values of θ_0 for which $H_0: \theta = \theta_0$ would not be rejected for a significance level $\alpha = 1 - p$
e.g. If our mean lies within the 95% CI, we would not reject H_0 with a significance level of 0.05

Issues with Hypothesis Testing

1. Testing Unknowns
   (We "know" that $\theta = \cdots$ is false)
2. Existence vs Effect Size
   (CIs est. effect size rather than stating an effect exists)
3. Prior Information
4. Publication Bias

Distribution	Parameter	Pivotal Quantity & Distr.	Confidence Interval	Test Statistic
Normal (σ^2 is known)	μ	$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\bar{y} \pm z^* \frac{\sigma}{\sqrt{n}}$	$z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
Normal (σ^2 is unknown)	μ	$\frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim T(n-1)$	$\bar{y} \pm t^* \frac{s}{\sqrt{n}}$	$t_{test} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$
Normal (σ^2 is unknown)	σ^2	$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\frac{(n-1)s^2}{a}, \frac{(n-1)s^2}{b}\right]$	$w_{test} = \frac{(n-1)S^2}{\sigma_0^2}$
Binomial	θ	$\frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \sim N(0,1)$	$\hat{\theta} \pm z^* \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$	$\frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$
Poisson	θ	$\frac{\bar{Y} - \theta}{\sqrt{\frac{\bar{Y}}{n}}} \sim N(0,1)$	$\bar{y} \pm z^* \sqrt{\frac{\bar{y}}{n}}$	$\frac{\bar{y} - \mu_0}{\sqrt{\frac{\mu_0}{n}}}$

*For CIs using χ^2, a and b are chosen so that the tails are equal size
**For CIs, z^* is a value of z such that $[-z^*, z^*]$ encloses an area p under the curve (likewise for t^*)

Test Statistics for the difference of two means ($\mu_1 - \mu_2 = d_0$)

μ's indep., σ^2 unknown	μ's indep., σ^2 unknown equal variance	μ's indep., σ^2 unknown, unequal variance	μ's not indep., σ^2 unknown
$Z = \frac{(\overline{X_1} - \overline{X_2}) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$t = \frac{(\overline{X_1} - \overline{X_2}) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ d.f. $= n_1 + n_2 - 2$	$t = \frac{(\overline{X_1} - \overline{X_2}) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ d.f. $= \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$ (round d.f. down)	$t = \frac{\overline{X_d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$ d.f. $= n - 1$ $\overline{X_d}$: mean sample diff μ_d: mean pop. diff. s_d: std dev. sample diff n: number of pairs