Useful Definitions and Results

(1) For sample $\{x_1, x_2, \cdots, x_n\}$, the sample mean, sample variance, and sample standard deviation are

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad s = \sqrt{s^2}.$$

(2) When n is large, binomial distribution $b(n, p)$ can be approximated by a Poisson distribution with $\lambda = np$.

(3) Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2, and $\overline{X}$ be the sample mean. Then

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}.$$

(4) (Chebyshev's inequality) Let X be a random variable with mean μ and variance σ^2, then for every $k \geq 1$,

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad \text{and} \quad P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

(5) If X has $N(\mu, \sigma^2)$ distribution, then $Z = \dfrac{X-\mu}{\sigma}$ has $N(0, 1)$ distribution.

(6) If Z has $N(0, 1)$ distribution, then $W = Z^2$ has $\chi^2(1)$ distribution.

(7) If $X_1, X_2, \cdots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$ distribution, then

$$W = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

has $\chi^2(n)$ distribution, and

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}$$

has $\chi^2(n-1)$ distribution, where S^2 is the sample variance.

(8) (A form of the Central Limit Theorem) Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a distribution with finite mean μ and finite positive variance σ^2, and let

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad \text{and} \quad Y = X_1 + X_2 + \cdots + X_n.$$

When n is sufficiently large, $\overline{X}$ will have approximate distribution $N(\mu, \dfrac{\sigma^2}{n})$ and Y will have approximate distribution $N(n\mu, n\sigma^2)$.

(9) Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution depends on a parameter θ with p.m.f. or p.d.f. $f(x;\theta)$. Then the joint p.m.f. or p.d.f. of $X_1, X_2, \ldots, X_n$, namely,

$$L(\theta) = f(x_1, \theta)f(x_2, \theta)\ldots f(x_n, \theta),$$

when regarded as a function of θ, is called the likelihood function of θ.

(10) (Confidence Interval for Mean) If $\overline{x}$ is the value of the mean of a random sample of size n from $N(\mu, \sigma^2)$ with the known variance σ^2, then a $100(1-\alpha)\%$ confidence interval for the population mean μ is given by

$$\Big(\overline{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,, \quad \overline{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\Big).$$

(11) (Small-Sample Confidence interval for Mean) If $\overline{x}$ and s are the values of the mean and standard devistion of a random sample of size n from $N(\mu, \sigma^2)$ with the unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval for the population mean μ is given by

$$\Big(\overline{x} - t_{\alpha/2}(n-1)\,\frac{s}{\sqrt{n}}\,, \quad \overline{x} + t_{\alpha/2}(n-1)\,\frac{s}{\sqrt{n}}\Big).$$

This confidence interval is used mainly when n is small, less than 30. When N is larger than 30, normal distribution can be used.

(12) (Confidence interval for Variance) If s^2 is the value of the variance of a random sample of size n from $N(\mu, \sigma^2)$, then a $100(1-\alpha)\%$ confidence interval for the population variance σ^2 is given by

$$\Big(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\,, \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\Big).$$

(13) (Confidence interval for difference of 2 means)If $\overline{x_1}$ and $\overline{x_2}$ are the values of the means of independent random samples of size n_1 and n_2 from normal populations with the known variances σ_1^2 and σ_2^2, then a $100(1-\alpha)\%$ confidence interval for the difference berween the two population means is given by

$$\Big((\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,, \quad (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\Big).$$

(14) (Small-sample confidence interval for difference of 2 means) If $\overline{x_1}$, $\overline{x_2}$, s_1, and s_2 are the values of the means and standard deviations of independent random samples of size n_1 and n_2 from normal populations with equal variances $\sigma_1^2 = \sigma_2^2$, then a $100(1-\alpha)\%$ confidence interval for the difference berween the two population means is given by

$$\Big((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2}(n_1+n_2-2)\,s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,, \quad (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2}(n_1+n_2-2)\,s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\Big),$$

where s_p is the value of the pooled estimate of sample standard deviation defined by

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}.$$

(15) (Confidence interval for proportion) If Y is a binomial random variable with the parameters n and p, n is large, and $\hat{p} = y/n$, then an approximate $100(1-\alpha)\%$ confidence interval for the population proportion p is given by

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

(16) (Confidence interval for difference of 2 proportions) If Y_1 is a binomial random variable with the parameters n_1 and p_1, Y_2 is a binomial random variable with the parameters n_2 and p_2, n_1 and n_2 are large, and $\hat{p_1} = y_1/n_1$ and $\hat{p_2} = y_2/n_2$, then an approximate $100(1-\alpha)\%$ confidence interval for difference in the population proportions, $p_1 - p_2$, is given by

$$\left((\hat{p_1} - \hat{p_2}) - z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}, \quad (\hat{p_1} - \hat{p_2}) + z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right).$$

(17) (Sample size required for finite population) In estimating population proportion p, suppose n is the sample size required such that the maximum error is ϵ with $100(1-\alpha)\%$ confidence. If the underlying population is finite and the required sample size n is significant compared to the population size N , then the required sample sized should be adjust to m given by

$$m = \frac{n}{1 + \frac{n-1}{N}}.$$

(18) A summary of hypothesis testing.

(a) **Assumption:** $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with σ known.

Null hypothesis: $H_0 : \mu = \mu_0$

Test Statistics: $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ follows $N(0,1)$ distribution.

Alternative Hypothesis H_1	Rejection Region	p-value
$\mu \neq \mu_0$	$\{z : \lvert z\rvert \geq z_{\alpha/2}\}$	$2P(Z > \lvert z_{\text{observed}}\rvert)$
$\mu < \mu_0$	$\{z : z \leq -z_\alpha\}$	$P(Z < z_{\text{observed}})$
$\mu > \mu_0$	$\{z : z \geq z_\alpha\}$	$P(Z > z_{\text{observed}})$

(b) **Assumption:** $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with σ unknown, and the sample size n is small ($n < 30$).

Null hypothesis: $H_0 : \mu = \mu_0$

Test Statistics: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$ follows a t- distribution with degree of freedom $n - 1$.

Alternative Hypothesis H_1	Rejection Region	p-value
$\mu \neq \mu_0$	$\{t : \lvert t\rvert \geq t_{\alpha/2}(n-1)\}$	$2P(T(n-1) > \lvert t_{\text{observed}}\rvert)$
$\mu < \mu_0$	$\{t : t \leq -t_\alpha(n-1)\}$	$P(T(n-1) < t_{\text{observed}})$
$\mu > \mu_0$	$\{t : t \geq t_\alpha(n-1)\}$	$P(T(n-1) > t_{\text{observed}})$

(c) **Null hypothesis:** $H_0 : p = p_0$

Test Statistics: $Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$ approximately follows $N(0,1)$ distribution.

Alternative Hypothesis H_1	Rejection Region	p-value
$p \neq p_0$	$\{z : \|z\| \geq z_{\alpha/2}\}$	$2P(Z > \|z_{\text{observed}}\|)$
$p < p_0$	$\{z : z \leq -z_\alpha\}$	$P(Z < z_{\text{observed}})$
$p > p_0$	$\{z : z \geq z_\alpha\}$	$P(Z > z_{\text{observed}})$

(d) **Null hypothesis:** $H_0 : p_1 = p_2$

Test Statistics: $Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$ approximately follows $N(0,1)$ distribution, where $\hat{p}$ is the pooled sample proportion ($\hat{p} = (Y_1 + Y_2)/(n_1 + n_2)$). Rejection region and p-value same as the above.

(e) **Assumption:** $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution.

Null hypothesis: $H_0 : \sigma^2 = \sigma_0^2$.

Test Statistics: $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ follows Chi-square distribution with degree of freedom $n-1$.

Alternative Hypothesis H_1	Rejection Region
$\sigma \neq \sigma_0$	$\{\chi^2 : \chi^2 \geq \chi^2_{\alpha/2}(n-1)$, or $\chi^2 \leq \chi^2_{1-\alpha/2}(n-1)\}$
$\sigma < \sigma_0$	$\{\chi^2 : \chi^2 \leq \chi^2_{1-\alpha}(n-1)\}$
$\mu > \mu_0$	$\{\chi^2 : \chi^2 \geq \chi^2_{\alpha}(n-1)\}$

(19) A table for some well-known distributions. (Note: In the following table $q = 1 - p$.)

Name	p.d.f	m.g.f	mean	variance
Bernoulli(p)	$f(x) = p^x q^{1-x}, \quad x = 0, 1$	$M(t) = q + pe^t$	p	pq
Binomial(n, p)	$f(x) = \binom{n}{x} p^x q^{n-x}$ $x = 0, 1, 2, \ldots, n$	$M(t) = (q + pe^t)^n$	np	npq
Geometric(p)	$f(x) = q^{x-1} p$ $x = 1, 2, \ldots$	$M(t) = \frac{pe^t}{1 - qe^t}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Negative Binomial (r, p)	$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$ $x = r, r+1, r+2, \ldots$	$M(t) = \frac{(pe^t)^r}{(1 - qe^t)^r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Poisson(λ)	$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$M(t) = e^{\lambda(e^t - 1)}$	λ	λ
Exponential(θ)	$f(x) = \frac{1}{\theta} e^{-x/\theta}$ $0 \leq x < \infty$	$M(t) = \frac{1}{1 - \theta t}$	θ	θ^2
Gamma (α, θ)	$f(x) = \frac{1}{\Gamma(\alpha) \theta^\alpha} x^{\alpha - 1} e^{-x/\theta}$ $0 \leq x < \infty$	$M(t) = \frac{1}{(1 - \theta t)^\alpha}$	$\alpha\theta$	$\alpha\theta^2$
Chi-Square(r)	$f(x) = \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} x^{r/2 - 1} e^{-x/2}$ $0 \leq x < \infty$	$M(t) = \frac{1}{(1 - 2t)^{r/2}}$	r	$2r$
Normal $N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ $-\infty < x < \infty$	$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$	μ	σ^2