SAMPLE STATISTICS

Statistic	Formula
Sample mean	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Sample variance	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^{\,2} - n\overline{X}^2\right]$
Sample proportion	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , X_i = 0, 1$
Coefficient of variation	$CV = \frac{S}{\overline{X}} \times 100$
Covariance	$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$
Pearson's correlation coefficient	$r = \frac{S_{XY}}{S_X S_Y}$

PROPERTIES OF THE PROBABILITY DISTRIBUTIONS' PARAMETERS

Expected value (Mean)	Variance	Covariance
$E(a) = a$ $E(aX) = aE(X)$ $E(X + Y) = E(X) + E(Y)$	Def.: $V(X) = E[(X - \mu)^2]$ $V(X) = E(X^2) - E(X)^2$ $V(a) = 0$ $V(aX) = a^2V(X)$ $V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X, Y)$	Def.: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ $Cov(X, Y) = E(XY) - E(X)E(Y)$ $Cov(aX, bY) = abCov(X, Y)$ $Cov(X+a, Y+b) = Cov(X, Y)$ If X and Y are independent then $Cov(X, Y) = 0$

THEOREMS

- If $X_1, X_2,\ldots,X_n$ are iid $N(\mu,\sigma)$ then $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n} Z_i^{\,2} \sim \chi^2_{(n)}$ where $Z_i \sim N(0,1)$
- If $Z \sim N(0, 1)$ and $Y \sim \chi^2_{(n)}$ are independent random variables then $T = \frac{Z}{\sqrt{Y/n}} \sim t_{(n)}$
- If $X \sim \chi^2_{(n)}$ and $Y \sim \chi^2_{(m)}$ are independent random variables then $T = \frac{X/n}{Y/m} \sim F_{(n,\,m)}$
- If $X_1, X_2,\ldots,X_n$ are iid $N(\mu,\sigma)$ then $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2_{(n-1)}$

SAMPLING DISTRIBUTIONS

Statistic	σ^2	Population	Sample size	Sampling distribution
$\bar{X}$	σ^2 known	Normal	n	$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0;1)$
		Any	Large*	
	σ^2 unknown	Normal	n	$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$
		Any	Large	$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \overset{a}{\sim} N(0;1)$
$\hat{p}$	-	Bernoulli	np ≥ 5 and nq ≥ 5	$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} N(0;1)$

* When the sample is large and the statistic's distribution is indicated as Normal, this distribution is asymptotic.

POINT ESTIMATION

Property	Definition
Unbiasedness	$bias(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$
Mean squared error (efficiency)	$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = V(\hat{\theta}) + \left(bias(\hat{\theta})\right)^2$
Consistency	The *unbiased* estimator $\hat{\theta}_n$ is consistent if $\lim\limits_{n \to +\infty} V(\hat{\theta}_n) = 0$
Consistency in Mean Square Error	The estimator $\hat{\theta}_n$ is consistent if $\lim\limits_{n \to +\infty} MSE(\hat{\theta}_n) = 0$

INTERVAL ESTIMATION*

Parameter	σ^2	Population	Sample size	C.I. $(1-\alpha)100\%$
μ	σ^2 known	Normal	n	$\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
		Any	Large	
	σ^2 unknown	Normal	n	$\bar{X} \pm t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$
		Any	Large	$\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

* When the sample is large and the statistic's distribution is indicated as Normal, the confidence level is approximate.

INTERVAL ESTIMATION* (continued)

Parameter	σ^2	Population	Sample size	C.I. $(1-\alpha)100\%$
$\mu_1-\mu_2$	σ_1^2, σ_2^2 known	Normal	n_1, n_2	$(\bar{X}_1-\bar{X}_2) \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
		Any	Large	
	$\sigma_1^2=\sigma_2^2$ unknown	Normal	n_1, n_2	$(\bar{X}_1-\bar{X}_2) \pm t_{n_1+n_2-2;1-\frac{\alpha}{2}}\, S^* \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ $S^*=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$
		Any	Large	$(\bar{X}_1-\bar{X}_2) \pm z_{1-\frac{\alpha}{2}}\, S^* \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ $S^*=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$
	$\sigma_1^2 \neq \sigma_2^2$ unknown	Normal	n_1, n_2	$(\bar{X}_1-\bar{X}_2) \pm t_{r;1-\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$ where *r* is the integer part of $r^*=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$
		Any	Large	$(\bar{X}_1-\bar{X}_2) \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$
$\mu_1-\mu_2$	σ_D^2 known	Normal	Paired	$\bar{D} \pm z_{1-\frac{\alpha}{2}}\frac{\sigma_D}{\sqrt{n}}$
	σ_D^2 unknown	Normal	Paired	$\bar{D} \pm t_{n-1;1-\frac{\alpha}{2}}\frac{S_D}{\sqrt{n}}$
	σ_D^2 unknown	Any	Paired & Large	$\bar{D} \pm z_{1-\frac{\alpha}{2}}\frac{S_D}{\sqrt{n}}$
p	-	Bernoulli	$np \geq 5$ and $nq \geq 5$	$\hat{P} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$
p_1-p_2	-	Bernoulli	$n_1p_1 \geq 5$; $n_1q_1 \geq 5$ $n_2p_2 \geq 5$; $n_2q_2 \geq 5$	$(\hat{p}_1-\hat{p}_2) \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
σ^2	-	Normal or approximately symmetrical	-	$\left]\frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}};\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right[$

* When the sample is large and the statistic's distribution is indicated as Normal, the confidence level is approximate.

PARAMETRIC TESTS*

H_0	H_1	σ^2	Population	Sample size	Test statistic
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 known	Normal	n	$Z_{obs} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0;1)$
$\mu \geq \mu_0$	$\mu < \mu_0$		Any	Large	
$\mu \leq \mu_0$	$\mu > \mu_0$				
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 unknown	Normal	n	$T_{obs} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
$\mu \geq \mu_0$	$\mu < \mu_0$				
$\mu \leq \mu_0$	$\mu > \mu_0$				
$\mu = \mu_0$	$\mu \neq \mu_0$		Any	Large	$Z_{obs} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \overset{a}{\sim} N(0;1)$
$\mu \geq \mu_0$	$\mu < \mu_0$				
$\mu \leq \mu_0$	$\mu > \mu_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	σ_1^2, σ_2^2 known	Normal	n_1, n_2	$Z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$		Any	Large	
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	$\sigma_1^2 = \sigma_2^2$ unknown	Normal	n_1, n_2	$T_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	$\sigma_1^2 = \sigma_2^2$ unknown	Any	Large	$Z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \overset{a}{\sim} N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	$\sigma_1^2 \neq \sigma_2^2$ unknown	Normal	n_1, n_2	$T_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(r)}$ where *r* is the integer part of $r^* = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	$\sigma_1^2 \neq \sigma_2^2$ unknown	Any	Large	$Z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim} N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	σ_D^2 known	Normal	Paired	$Z_{obs} = \frac{\bar{D} - D_0}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	σ_D^2 unknown			$T_{obs} = \frac{\bar{D} - D_0}{\frac{S_D}{\sqrt{n}}} \sim t_{(n-1)}$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	σ_D^2 known	Any	Paired & large	$Z_{obs} = \frac{\bar{D} - D_0}{\frac{\sigma_D}{\sqrt{n}}} \overset{a}{\sim} N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$\mu_1 - \mu_2 = D_0$	$\mu_1 - \mu_2 \neq D_0$	σ_D^2 unknown			$Z_{obs} = \frac{\bar{D} - D_0}{\frac{S_D}{\sqrt{n}}} \overset{a}{\sim} N(0;1)$
$\mu_1 - \mu_2 \leq D_0$	$\mu_1 - \mu_2 > D_0$				
$\mu_1 - \mu_2 \geq D_0$	$\mu_1 - \mu_2 < D_0$				
$p = p_0$	$p \neq p_0$	-	Bernoulli	Large	$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim} N(0;1)$
$p \geq p_0$	$p < p_0$				
$p \leq p_0$	$p > p_0$				

* When the sample is large and the statistic's distribution is indicated as Normal, the significance level is approximate.

PARAMETRIC TESTS* (continued)

H_0	H_1	σ^2	Population	Sample size	Test statistic
$p_1 - p_2 = 0$	$p_1 - p_2 \neq 0$	-	Bernoulli	Large	$Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \overset{a}{\sim} N(0;1), \quad \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$
$p_1 - p_2 \geq 0$	$p_1 - p_2 < 0$				
$p_1 - p_2 \leq 0$	$p_1 - p_2 > 0$				
$p_1 - p_2 = D_0$	$p_1 - p_2 \neq D_0$	-	Bernoulli	Large	$Z_{obs} = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \overset{a}{\sim} N(0;1)$
$p_1 - p_2 \geq D_0$	$p_1 - p_2 < D_0$				
$p_1 - p_2 \leq D_0$	$p_1 - p_2 > D_0$				
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	-	Normal	n	$\chi^2_{obs} = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$				
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$				
$\frac{\sigma_1^2}{\sigma_2^2} = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$	$\frac{\sigma_1^2}{\sigma_2^2} \neq \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$		Normal	n_1, n_2	$F_{obs} = \frac{S_1^2}{S_2^2}\left(\frac{\sigma_2^2}{\sigma_1^2}\right)_0 \sim F_{(n_1-1;n_2-1)}$
$\frac{\sigma_1^2}{\sigma_2^2} \geq \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$	$\frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$				
$\frac{\sigma_1^2}{\sigma_2^2} \leq \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$	$\frac{\sigma_1^2}{\sigma_2^2} > \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0$				
$\rho = 0$	$\rho \neq 0$	-	Bivariate Normal	n	$T_{obs} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)}$
$\rho \geq 0$	$\rho < 0$				
$\rho \leq 0$	$\rho > 0$				
$\rho = \rho_0 (\neq 0)$	$\rho \neq \rho_0$	-	Bivariate Normal	n	$Z_{obs} = \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\frac{1}{\sqrt{n-3}}} \sim N(0;1)$
$\rho \geq \rho_0 (\neq 0)$	$\rho < \rho_0$				

* When the sample is large and the statistic's distribution is indicated as Normal, the significance level is approximate.

ANALYSIS OF VARIANCE (ANOVA)

k → nr. levels of the factor (nr. of groups)

n$_i$ → nr. observations of the level *i* (i=1,...,k)

$$\overline{X}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} = \frac{X_{i\bullet}}{n_i}; \quad X_{i\bullet} = \sum_{j=1}^{n_i} X_{ij}$$

$$n = \sum_{i=1}^{k} n_i$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} = \frac{1}{n}\sum_{i=1}^{k} X_{i\bullet} = \frac{1}{n}\sum_{i=1}^{k} n_i \overline{X}_{i\bullet}$$

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_{i\bullet}\right)^2 = \frac{1}{n_i - 1}\left(\sum_{j=1}^{n_i} X_{ij}^2 - n_i \overline{X}_{i\bullet}^2\right)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 - n\overline{X}^2\right)$$

ANALYSIS OF VARIANCE (ANOVA)

Source of variace	Degrees of freedom (df)	Sum of squares	Mean squares	*F*
Treatments (*between*)	k – 1	$SSTr = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot} - \bar{X})^2 = \sum_{i=1}^{k} \frac{\bar{X}_{i\cdot}{}^2}{n_i} - n\bar{X}^2$	$MSTr = \frac{SSTr}{k-1}$	$F = \frac{MSTr}{MSE}$
Error (*within*)	n – k	$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2 = \sum_{i=1}^{k}(n_i - 1)S_i{}^2$	$MSE = \frac{SSE}{n-k}$	
Total	n – 1	$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X})^2 = (n-1)S^2$		

Tukey's HSD test	Tukey-Kramer test

H_0: $\mu_i = \mu_j$ $(i \neq j)$

$$W = \frac{\left|\bar{X}_{i\bullet} - \bar{X}_{j\bullet}\right|}{\sqrt{\frac{S^2}{b}}} \sim q_{(k;n-k)}$$

$b = n_1 = n_2 = \ldots = n_k$

$$W = \frac{\left|\bar{X}_{i\bullet} - \bar{X}_{j\bullet}\right|}{\sqrt{\frac{S^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \sim q_{(k;n-k)}$$

S^2 = MSE; $q_{(k;\,n-k)}$ → quantile of the *Studentized Range* distribution

Reject H_0 if: $W_{obs} \geq q_{(k;\,n-k);1-\alpha}$

Bartlett's test

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$

$$Q = \frac{(n-k)\ln S^2 - \sum_{i=1}^{k}(n_i - 1)\ln S_i^2}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right)} \sim \chi^2_{(k-1)}$$

S^2 = MSE

Reject H_0 if: $Q_{obs} \geq \chi^2_{(k-1;\,1-\alpha)}$

NOVA IMS Information Management School

NONPARAMETRIC TESTS

Chi-Square test

H_0: $X \sim f(x)$

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim} \chi^2_{(k-p-1)}$$

Reject H_0 if $Q_{obs} > Q_{crit}$

k → nr. categories or classes A_i

p → nr. estimated parameters

$O_i = n_i$ → observed absolute frequency of A_i

E_i → estimated absolute frequency of A_i, under H_0

$E_i = n\pi_i = n \times P(X \in A_i)$

Not more than 20% of classes with $E_i < 5$

Shapiro-Wilk test

H_0: $X \sim N(\mu, \sigma)$, μ and σ unknown

$$W = \frac{b^2}{(n-1)S^2}$$

Ordered sample: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

Reject H_0 if $W_{obs} < W_{crit}$

$$b = \begin{cases} \sum_{i=1}^{n/2} a_i \left(x_{(n-i+1)} - x_{(i)}\right), & \text{if } n \text{ is even} \\ \sum_{i=1}^{(n+1)/2} a_i \left(x_{(n-i+1)} - x_{(i)}\right), & \text{where } a_{[n/2]+1} = 0 \text{ if } n \text{ is odd} \end{cases}$$

Wilcoxon-Mann-Whitney test

R_i → *ranks* attached to the observations in the combined sorted sample (from 1 to N=n+m)

Small samples with a few ties ($n \leq 20$, $m \leq 20$)

W^* = Sum of R_i from sample X

Wilcoxon signed-ranks test

$D_i = X_i - Y_i$

n' → number of differences D_i that are different from zero

R_i → *rank* value of the absolute difference $|D_i|$ that is affected by the sign of D_i

W_{crit} of the inferior quantile → value of w_α in the table for n'

W_{crit} of the superior quantile → $w_{(1-\alpha)} = n'(n'+1)/2 - w_\alpha$

Small samples with a few ties

$$W^+ = \sum_{\substack{i=1 \\ (D_i>0)}}^{n'} R_i$$

NOVA IMS Information Management School

NONPARAMETRIC TESTS (continued)

Spearman's correlation test

H_0: X_i and Y_i are mutually independent, $\forall i$

$R(X_i) \rightarrow$ *rank* of X_i
$R(Y_i) \rightarrow$ *rank* of Y_i

$n \leq 30$ and few ties $\rightarrow$ superior quantiles $[w_{(1-\alpha)}]$ of the exact distribution of ρ are tabled; the inferior quantiles are symmetrical to the superior ones $[w_\alpha = -w_{(1-\alpha)}]$

$n > 30$ or many ties $\rightarrow$ use approximation to N(0,1): $w_\alpha \approx z_\alpha \big/ \sqrt{n-1}$

p-value of the two-sided test: $p - \text{value} \approx 2P\left[Z > |\rho_{obs}|\sqrt{n-1}\right]$

$$\rho = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{n}[R(X_i)]^2 - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^{n}[R(Y_i)]^2 - n\left(\frac{n+1}{2}\right)^2\right)}}$$

For few ties

$$\rho = 1 - \frac{6\sum_{i=1}^{n}[R(X_i) - R(Y_i)]^2}{n(n^2-1)}$$

DISCRETE DISTRIBUTIONS

Distribution	Probability function	Parameters' domain	Mean (μ)	Variance (σ^2)	Moment generating function
Uniform	$\frac{1}{n}$ $x = 1, 2, \ldots, n$	$n \in \mathbb{N}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^t(1-e^{nt})}{n(1-e^t)}$
Bernoulli	$p^x (1-p)^{1-x}$ $x = 0, 1$	$0 \le p \le 1$ $(q = 1 - p)$	p	pq	$q + pe^t$
Binomial	$C_x^n p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$0 \le p \le 1$ $n = 1, 2, \ldots$ $(q = 1 - p)$	np	npq	$(q + pe^t)^n$
Hypergeometric	$\frac{C_x^m \boldsymbol{C_{n-x}^{N-m}}}{\boldsymbol{C_n^N}}$ $x = 0, \ldots, \min\{m; n\}$	$N = 1, 2, \ldots$ $0 \le m \le N$ $n = 1, 2, \ldots, N$	$n\frac{k}{N}$	$\frac{nk(N-k)(N-n)}{N^2(N-1)}$	*Not used*
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$\lambda > 0$	λ	λ	$\exp[\lambda(e^t - 1)]$
Geometric	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$0 \le p \le 1$ $(q = 1 - p)$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$

NOVA IMS Information Management School

CONTINUOUS DISTRIBUTIONS

Distribution	Probability density function	Parameters' domain	Mean (μ)	Variance (σ^2)	Moment generating function
Uniform	$\begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{o.v.} \end{cases}$	$-\infty < a < b < +\infty$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ $-\infty < x < +\infty$	$-\infty < \mu < +\infty$ $\sigma > 0$	μ	σ^2	$\exp\left[\mu t + \frac{1}{2}\sigma^2 t^2\right]$
Gamma	$\begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & x \le 0 \end{cases}$	$\beta > 0$ $\alpha > 0$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta t}\right)^\alpha, t < \frac{1}{\beta}$
Exponential (particular case of Gamma)	$\begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & x \le 0 \end{cases}$	$\beta > 0$	β	β^2	$\frac{1}{1-\beta t}, t < \frac{1}{\beta}$
χ^2 (particular case of Gamma)	$\begin{cases} \frac{x^{(n/2)-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}, & x > 0 \\ 0, & x \le 0 \end{cases}$	$n = 1, 2, \ldots$	n	$2n$	$\frac{1}{(1-2t)^{n/2}}, t < \frac{1}{2}$
t	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$ $-\infty < x < +\infty$	$n > 0$	0 $(n > 1)$	$\frac{n}{n-2}$ $(n > 2)$	*Does not exist*
F	$\frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{\left(\frac{m-2}{2}\right)}}{\left(1+\frac{mx}{n}\right)^{\frac{m+n}{2}}}$	$m = 1, 2, \ldots$ $n = 1, 2, \ldots$	$\frac{n}{n-2}$ $(n > 2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n > 4)$	*Does not exist*