Equations List for Statistical Inference

Hadas Lapid, Spring Semester 2020

General Terminology:

Sample Mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Sample Variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Binomial Distribution:

probability of x successes: $P(X = x) = \binom{n}{x} \rho^x q^{n-x}$

Expectation Value: $\mu = np$

Variance: $\sigma^2 = npq$

Poisson Distribution:

population mean: $\mu = \lambda$

variance: $\boldsymbol{\sigma^2} = \lambda$

probability of x occurrences per unit: $\mathbf{P_x} = \frac{\mathbf{e^{-\lambda}\lambda^x}}{\mathbf{x!}}$

Normal Distribution:

$\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$, Distribution function: $\mathbf{f}(x|\mu, \sigma) = \frac{\mathbf{1}}{\boldsymbol{\sigma\sqrt{2\pi}}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Statistical estimate: $\mathbf{z} = \frac{\bar{\mathbf{x}}-\boldsymbol{\mu}}{\boldsymbol{\sigma}/\sqrt{\mathbf{n}}}$

Confidence Interval: $\mathbf{C.I.} = \bar{\mathbf{X}} \pm \mathbf{Z}_{\mathbf{1}-\frac{\alpha}{2}} \cdot \frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}}$

Critical value for 1-sided rejection region $x_c = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ (right-sided)

or $x_c = \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ (left-sided)

Critical value for 2-sided rejection region $\mathbf{x_c} = \boldsymbol{\mu_0} \pm \mathbf{Z_{1-\alpha/2}} \cdot \frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}}$

Single sample t-test:

Statistical estimate $\mathbf{t} = \frac{\bar{\mathbf{x}}-\boldsymbol{\mu}}{\mathbf{S}/\sqrt{\mathbf{n}}}$

Confidence interval: $\mathbf{C.I.} = \bar{\mathbf{X}} \pm \mathbf{t}_{\mathbf{1}-\frac{\alpha}{2}} \cdot \frac{\mathbf{S}}{\sqrt{\mathbf{n}}}$

χ^2 Distribution:

$$\mathbf{X^2} = \sum_{\mathbf{i=1}}^{\mathbf{k}} \frac{(\mathbf{Observed - Expected})^2}{\mathbf{Expected}} = \sum_{\mathbf{i=1}}^{\mathbf{m}} \sum_{\mathbf{j=1}}^{\mathbf{k}} \frac{\left(\mathbf{O_{i,j} - E_{i,j}}\right)^2}{\mathbf{E_{i,j}}}$$

j is all possible groups, i is all possible categories, Probability of i category: $P_i = \frac{t_i}{n} = \frac{\sum_{j=1}^{k} O_{i,j}}{n}$

expected number of cases for i category in j group: $\mathbf{E_{ij}} = \mathbf{n_j} \cdot \frac{\mathbf{t_i}}{\mathbf{n}}$

Two-Sample t-test:

Alternative Hypothesis	Rejection Region
H_a: $\mu_1 \neq \mu_2$	$\lvert T\rvert > t_{1-\alpha/2,df}$
H_a: $\mu_1 > \mu_2$	$T > t_{1-\alpha,df}$
H_a: $\mu_1 < \mu_2$	$T < t_{\alpha,df}$

Mutual variance: $S^2 = \frac{\sum_{i=1}^{2}(n_i-1)\cdot S_i^2}{\sum_{i=1}^{2}(n_i-1)}$

standard error of the mean: $stderr = S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Statistical estimate: $t = \frac{\overline{Y_1}-\overline{Y_2}}{s\cdot\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Wilcoxon Rank-Sum Test:

T = rank sum, i = sample group

$$\mu(T_i) = \frac{n_i}{2}(N+1)$$

$$\sigma(T_i) = \sqrt{\frac{n_1 \cdot n_2(N+1)}{12}}$$

$$Z = \frac{T-\mu(T)}{\sigma(T)}$$

Multiple Hypothesis Testing:

$\alpha <<$ FWER $< 1-(1-\alpha)^m$

Bonferroni Correction: $P_v < \frac{\alpha}{m}$

Holmes Correction: $P_v < \frac{\alpha}{m+1-i}$

Benjamini-Hochberg Correction: $P_v \leq \frac{i}{m}\cdot\alpha$

Analysis of Variance:

$$SS_{\text{between groups}} = \sum_{i=1}^{k} n_i\left(\overline{Y}_i - \overline{\overline{Y}}\right)^2$$

$$MS_{\text{between groups}} = SS_{\text{between groups}}/(k-1)$$

$$SS_{\text{error}} = \sum_{i=1}^{k}(n_i-1)\cdot S_i^2$$

$$MS_{\text{error}} = \frac{\sum_{i=1}^{k}(n_i-1)\cdot S_i^2}{\sum_{i=1}^{k}(n_i-1)} = \frac{SS_{\text{error}}}{n-k}$$

$$F = \frac{MS_{\text{between groups}}}{MS_{\text{error}}}$$

Contrast t-test:

$$t_{i,j} = \frac{\overline{Y}_i - \overline{Y}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Pairwise Comparisons with Studentized Distribution:

$$q_{i,j} = \frac{\overline{Y}_i - \overline{Y}_j}{\sqrt{\frac{1}{2} \cdot MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}, \quad q^* = \frac{\overline{Y}_{max} - \overline{Y}_{min}}{\sigma/\sqrt{n}}$$

Pearson Correlation:

$$z_{xi} = \frac{(x_i - \bar{x})}{s_x}, \quad r_i = z_{xi} \cdot z_{yi}$$

Correlation Coefficient: $r = \frac{1}{n-1}\sum_{i=1}^{n} r_i$

Statistical estimate: $t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$