Statistical Formulas-MATH 332

Regression Analysis

$S_x^2 = \sum X^2 - n\,(\overline{X})^2,\ S_y^2 = \sum Y^2 - n\,(\overline{Y})^2,\ S_{xy} = \sum X\,Y - n\,\overline{X}\,\overline{Y},$

Sum of Squares: $SSR = (\hat{\beta}_1)^2\, S_x^2,\ SST = S_y^2,\ SSE = SST - SSR$

Fitted line: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\, X,\ \hat{\beta}_1 = \dfrac{S_{xy}}{S_x^2},\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\,\overline{X}$

SEs: $\hat{\sigma}^2 = s^2 = MSE = \dfrac{SSE}{n-2},\ S.E.(\hat{\beta}_1) = \dfrac{s}{S_x},\ S.E.(\hat{\beta}_0) = s\sqrt{\dfrac{1}{n} + \dfrac{\overline{X}^2}{S_x^2}},$

$S.E.(\hat{Y})\ \ at\ \ x^* = s\sqrt{\dfrac{1}{n} + \dfrac{(x^* - \overline{x})^2}{S_x^2}}.$

Coefficients of Determination and Correlation: $R^2 = \dfrac{SSR}{SST},\quad Corr. = r = \dfrac{S_{xy}}{S_x\, S_y}.$

Test for Correlation $(H_0 : \rho = 0,\ vs.\ H_1 : \rho \neq 0)$: $t = \dfrac{\sqrt{n-2}\ r}{\sqrt{1-r^2}}.$

Lack-of-Fit Test: SSE = SS_{PE} + SS_{LF}, $F = \dfrac{(SSE - SS_{PE})/(k-2)}{SS_{PE}/(n-k)}.$

Multiple Regression: $\hat{\beta} = (X'\,X)^{-1}\ X'Y.$

Experimental Design

One Way Anova:

$SS_{tot} = \sum_{j=1}^{k}\sum_{i=1}^{n_j} (Y_{ij} - \overline{Y})^2,\ SS_{trt} = \sum_{j=1}^{k} n_j\,(\overline{Y_j} - \overline{Y})^2$, SSE $= SS_{tot} - SS_{trt}$

Multiple -t CIs: of $\mu_j - \mu_i$ is $(\overline{Y_j} - \overline{Y}_i) \pm t_{\alpha/2m}\ s\sqrt{\dfrac{1}{n_i} + \dfrac{1}{n_j}}$, where m is the number of pairs and $s = \sqrt{MSE}$ with the corresponding degrees of freedom being df = n-k.

Two-Way Anova (without interaction):

Factor A has p levels, Factor B has q levels with r = number of replications and N= pqr

$CT = \dfrac{Y_{..}^2}{N},\ SS_{Tot} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} Y_{ijk}^2 - CT,\ SS_A = \sum_{j=1}^{p} \dfrac{Y_{i..}^2}{qr} - CT\quad SS_B = \sum_{j=1}^{q} \dfrac{Y_{.j.}^2}{pr} - CT,$

$SSE = SS_{Tot} - SS_A - SS_B.$

Two-Way Anova (with interaction):

$CT = \dfrac{Y_{...}^2}{N}, SS_{Tot} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} Y_{ijk}^2 - CT,\ SS_A = \sum_{j=1}^{p} \dfrac{Y_{i..}^2}{qr} - CT,\quad SS_B = \sum_{j=1}^{q} \dfrac{Y_{.j.}^2}{pr} - CT,$

$SS_{Subtotals} = \sum_{i=1}^{p}\sum_{j=1}^{q} \dfrac{Y_{ij.}^2}{r} - CT,\ SS_{AB} = SS_{Subtotal} - SS_A - SS_B,$

$SSE = SS_{Tot} - SS_A - SS_B - SS_{AB}.$

Chi-Square Tests

Chi-Square Test for Goodness of Fit:

$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \sim \chi^2_{k-1}$, k = number of categories or groups, $E_j = n\,p_j$ with p_j being the proportion of occurrence of the jth group.

Chi-Square Test for Independence and Homogeneity:

$\chi^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$, where

where r = # of rows, c = # of columns

Non-Parametric Statistics

Wilcoxon Rank Sum Test

$W = \sum_{i=1}^{n} R_i$ where $R_i = rank\ \ of\ \ Y_i$

$$mean(W) = \frac{n(m+n+1)}{2},\ and\ \ Var(W) = \frac{mn(m+n+1)}{12}.$$

where m =sample size of X-sample and n = sample size of Y-sample.

Wilcoxon Signed Rank Test

$T^+ = \sum_{i=1}^{n} \psi^+ R_i$, $where\ R_i = rank\ \ of\ \ the\ \ absolute\ \ difference\ (d = X - Y)$.

and

$$\psi_i = \begin{cases} 1 & if\ \ the\ \ difference\ is\ \ positive \\ 0 & if\ \ the\ \ difference\ is\ negative. \end{cases}$$

$$mean(T^+) = \frac{n(n+1)}{4}\ \ and\ \ Var(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

Kruskal-Wallis Test

$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{w_i^2}{n_i} - 3(n+1) \sim \chi^2_{k-1}$, where k = degrees of freedom