# Summary of Anova and Regression

Tuesday, 8 November 2022 21:12



| | $H_0$ | The statistical value/test | $H_1$ | Rejection region |
|---|---|---|---|---|
| 1. Goodness of Fit | The data belongs to a certain distribution | $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$<br>$= \sum_{i=1}^{k} \frac{O_i^2}{E_i} - n$<br>k - groups<br>n - observations<br>$E_i$ - expected value according to $H_0$ in group<br>$O_i$ - observed value in group<br>$E_i \geq 5$ | The data does not belongs to the certain distribution | $\chi^2 > \chi^2_\alpha \overbrace{(k-t)}^{DF}$<br>t – the number of values necessary to calculate the $E_i$ |
| 2. Contingency Tables | The two qualitative variables are independent | $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$<br>$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$<br>$n_i$ – sum of row i<br>$n_j$ – sum of column j<br>n – total sum | The two variables are dependent | $\chi^2 > \chi^2_{\alpha, \overbrace{(r-1)*(c-1)}^{DF}}$<br>r – number of rows<br>c – number of columns |
| 3. Analysis of Variance<br>ANOVA | $\mu_1 = \mu_2 = \ldots = \mu_k$<br>n – number of observations in each group<br>k – number of groups | $T_i = \sum_j Y_{ij}$<br>i – sum of observations in group i DF = k−1<br>$T = \sum_{i,j} Y_{ij}$<br>sum of all observations<br>$SST = \sum_{i,j} Y_{ij}^2 - \frac{T^2}{kn}$<br>$SSB = \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{kn}$<br>SSW=SST – SSB<br>$F = \frac{SSB/(k-1)}{SSW/k(n-1)}$ | $\exists (i,j) \quad \mu_i \neq \mu_j$ | $F > F_\alpha(k-1, k(n-1))$ |
| 4. Analysis of Variance with uneven number of observations in each group | $\mu_1 = \mu_2 = \ldots = \mu_k$ | N instead of nk<br>$n_i$ instead of n<br>$n_i$ = number of observations in group i<br>$N = \sum_{i=1}^{k} n_i$ | $\exists (i,j) \quad \mu_i \neq \mu_j$ | $F > F_\alpha(k-1, \underbrace{N-k}_{DF})$ |

Chisq.test(x = x_data , P= <vector of probabilities>

•Calculate chi$^2$ value: qchisq(DF=df, P=0.95)

Chisq.test(x= Matrix)



| | $H_0$ | The statistical value/test | $H_1$ | Rejection region |
|---|---|---|---|---|
| 5. correlation | ρ=0 | $R^2 = \frac{SSR}{SST} = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSR}{SST}$<br>$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\hat{b}^2 \cdot S_{xx}}{S_{yy}} = \frac{\hat{b} \cdot S_{xy}}{S_{yy}}$<br>$R^2 = \frac{\hat{b} \cdot S_{xy}}{S_{yy}} = \frac{S_{xy}}{S_{yy}/b} = \frac{(S_{xy})^2}{S_{yy}(S_{XX})} =$<br>$\frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} =$<br>$\frac{\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}$ | ρ≠0 | According to table |
| 6. regression<br>$\hat{Y} = \hat{a} + \hat{b}X$ | b=0 | $\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$<br>$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}$<br>$SST = S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$<br>$SSR = \hat{b}^2 \cdot S_{xx} = \hat{b} \cdot S_{xy}$<br>$SSE = SST - SSR$<br>$F = \frac{(n-2) \cdot SSR}{SSE} = \frac{(n-2) \cdot R^2}{1 - R^2}$ | b≠0 | $F > F_\alpha(1, n-2)$ |