Empirical Research in Management and Economics
Formula Sheet

January 25, 2022

1 Descriptive Statistics

1.1 Measure of Location

1.1.1 Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

1.1.2 Median

For odd number of elements in a dataset:

$$\tilde{x} = x_{\frac{n+1}{2}} \tag{2}$$

For even number of elements in a dataset:

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2} \tag{3}$$

1.1.3 Mode

$$Mo(x) = \max(f(x_i)) \tag{4}$$

1.1.4 Quartile

Measure of percentage of elements less than or equal to a term

1.2 Measure of Spread

1.2.1 Variance

Variance measured on the whole population

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{5}$$

1.2.2 Sample Variance

Variance measured on a sample population

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{6}$$

1.2.3 Standard Deviation and Sample Standard

$$\sigma = \sqrt{\sigma^2} \tag{7}$$

$$s = \sqrt{s^2} \tag{8}$$

1.2.4 Co-efficient of Variance

$$v = \frac{s}{\bar{x}} \tag{9}$$

1.3 Skewness

1.3.1 Types of Skewness

Name	Other Name	Characteristic
Right Skew	Positive Skew	Data concentrated on the lower side
Symmetric Distribution	Normal Distribution	Data distributed evenly
Left Skew	Negative Skew	Data concentrated on the higher side

1.3.2 Measure of Skewness

Skewness is measured by the Moment Co-efficient of Skewness.

$$g_m = \frac{m_3}{s^3}, \text{ where} \tag{10}$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \tag{11}$$

Type of Skewness The type of skewness from the value is g_m is:

Value of g_m	Type
$g_m = 0$	Symmetric
$g_m > 0$	Positive Skew
$g_m < 0$	Negative Skew

Degree of Skewness The degree of skewness from the value is g_m is:

Value of g_m	Degree
$\lvert g_m \rvert > 1$	High Skewness
$0.5 < \lvert g_m \rvert \geq 1$	Moderate Skewness
$\lvert g_m \rvert \leq 0.5$	Low Skewness

1.4 Kurtosis

Kurtosis is the measure of peakedness of data. Fisher's kurtosis measure is defined as:

$$\gamma = \frac{m_4}{s^4}, \text{ where} \tag{12}$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 \tag{13}$$

1.4.1 Type of Kurtosis

The types of kurtosis from the value of γ are:

Value of γ	Type
$\gamma = 0$	Normal Distribution or Mesokurtic
$\gamma < 0$	Flattened or Platykurtic
$\gamma > 0$	Peaked or Lepokurtic

2 Hypothesis Testing

2.1 T-Test

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \tag{14}$$

where:

$$\begin{aligned} \bar{X} &= \text{Sample Mean} \\ \mu &= \text{Assumed Mean} \\ s &= \text{Number of Samples} \\ n &= \text{Number of observations} \end{aligned}$$

If $T < t_c$ the H_0 is not rejected. t_c is a functions of level of significance (α) and degrees of freedom ($v = n - 1$).

2.2 χ^2 Test

$$\chi^2 = \sum_i \sum_j \frac{(h_{ij}^o - h_{ij}^e)^2}{h_{ij}^e} \tag{15}$$

where:

$$\begin{aligned} h_e &= \text{Expected Value} \\ h_o &= \text{Actual Value} \end{aligned}$$

If $\chi^2 < \chi_c^2$ then H_0 is not rejected. χ_c is a functions of level of significance (α) and degrees of freedom ($v = (i-1)(j-1)$).

3 Research and Survey Design

3.1 Population Covariance

$$\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) \tag{16}$$

3.2 Sample Covariance

$$\text{Cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{17}$$

3.3 Bravais-Pearson Correlation Co-efficient

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{18}$$

$$= \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x) \cdot \text{Var}(y)}} \tag{19}$$

$$= \frac{\text{Cov}(x,y)}{\sigma_x \cdot \sigma_y} \tag{20}$$

4 Estimation of Regression Function

For the regression functions:

$$Y_i = \beta_0 + \beta_1 X_1 \tag{21}$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 \tag{22}$$

$$\tag{23}$$

where Y_i is the observed dependent variable (DV), $\hat{Y}_i$ is the estimated DV, and X_i is the independent variable (IV).

$$u_i = Y_i - \hat{Y}_i \tag{24}$$

$$\Rightarrow Y_i = \hat{Y}_i + u_i \tag{25}$$

$$\Rightarrow Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + u_i \tag{26}$$

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{27}$$

The objective function is:

$$\min_{u_i} \sum u_i = \min \sum_i \left[Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right]^2$$

Since the regression function passes through: $\left(\bar{X}, \bar{Y}\right)$

$$\beta_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\begin{aligned}
\min_{u_i} \sum u_i &= \min \sum_i \left[Y_i - \bar{Y} + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 X_i \right]^2 \\
&= \min \sum_i \left[\left(Y_i - \bar{Y}\right) - \hat{\beta}_1 \left(X_i - \bar{X}\right) \right]^2 \\
&= \min \sum_i \left[\left(Y_i - \bar{Y}\right)^2 - 2 \cdot \left(Y_i - \bar{Y}\right) \cdot \hat{\beta}_1 \left(X_i - \bar{X}\right) + \hat{\beta}_1{}^2 \left(X_i - \bar{X}\right)^2 \right] \\
&= \min \left[\sum_i \left(Y_i - \bar{Y}\right)^2 - 2 \cdot \hat{\beta}_1 \sum_i \left(Y_i - \bar{Y}\right) \cdot \left(X_i - \bar{X}\right) + \hat{\beta}_1{}^2 \sum_i \left(X_i - \bar{X}\right)^2 \right] \\
\Rightarrow u_i^{\beta_1} &= -2 \sum_i \left(Y_i - \bar{Y}\right) + 2\hat{\beta}_1 \left(X_i - \bar{X}\right)^2 = 0 \qquad \text{(For optima Conditions)} \\
\Rightarrow \hat{\beta}_1 &= \boxed{\frac{\sum_i (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}} \\
\Rightarrow \hat{\beta}_0 &= \boxed{\bar{Y} - \hat{\beta}_1 \bar{X}}
\end{aligned}$$

4.1 Sum of Squares Error

$$TSS = \sum_i (Y_i - \bar{Y})^2 \tag{28}$$

$$= \underbrace{\sum_i (\hat{Y}_i - \bar{Y})}_{\text{Explained Sum of Square Error (ESS)}} + \underbrace{\sum_i u_i^2}_{\text{Residual Sum of Squares Error (RSS)}} \tag{29}$$

4.1.1 R^2: Coefficient of Determination

$$
\begin{aligned}
R^2 &= \frac{\text{ESS}}{\text{TSS}} && (30)\\
&= 1 - \frac{\text{RSS}}{\text{TSS}} && (31)\\
&= 1 - \frac{\sum_i u_i^2}{\sum_i (Y_i - \bar{Y})^2} && (32)
\end{aligned}
$$

For a regression analysis with single IV:

$$\sqrt{R^2} = v$$

4.1.2 $\bar{R^2}$: Coefficient of Determination

$$\bar{R^2} = 1 - \frac{\frac{\sum_i u_i^2}{(N-K-1)}}{\frac{\sum_i (Y_i - \bar{Y})^2}{(N-1)}} \tag{33}$$

where, N is the number of observations and K is the number of independent variables.

4.2 T-Test

Test for statistical significance of a single IV.

$$T = \frac{\hat{\beta}_1 - 0}{S_e(\hat{\beta}_1)} \tag{34}$$

4.3 F-Test

Test for statistical significance of all IVs together.

$$F = \frac{\frac{\text{ESS}}{(K-1)}}{\frac{\text{RSS}}{(N-K)}} \qquad (F \geq F_c, H_0 \text{ is rejected})$$

4.4 Test for Heteroskedasticity

Definition $\sigma_{\epsilon_i} \forall \epsilon_i \in [X_a, X_b] = \sigma_{\epsilon_i} \forall \epsilon_i \in [X_{b+1}, X_c]$

4.4.1 Durbin-Watson Test

$$d_e = \frac{\sum_{t=2}^{n} (\hat{u_t} - \hat{u_{t-1}})^2}{\sum_{t=2}^{n} \hat{u_t}^2} \tag{35}$$

For the H_0: No autocorrelation:

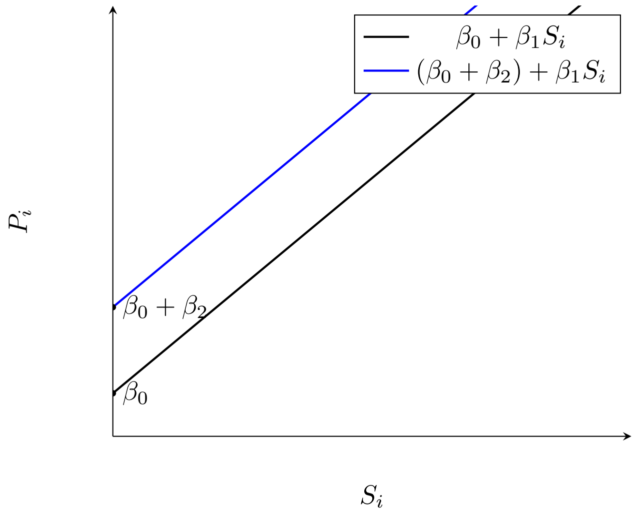
d	H_0
$0 \leq d_e \leq d_L$ & $(4 - d_L) \leq d_e \leq 4$	Rejected
$d_L < d_e \leq d_U$ & $(4 - d_U) < d_e \leq (4 - d_L)$	Decision Free Zone
$d_L < d_e < D_U$	Not rejected

5 Dummy Variables

5.1 Dummy Variable

$$P_i = \beta_0 + \beta_1 S_1 + \beta_2 D_i + \epsilon_i \tag{36}$$

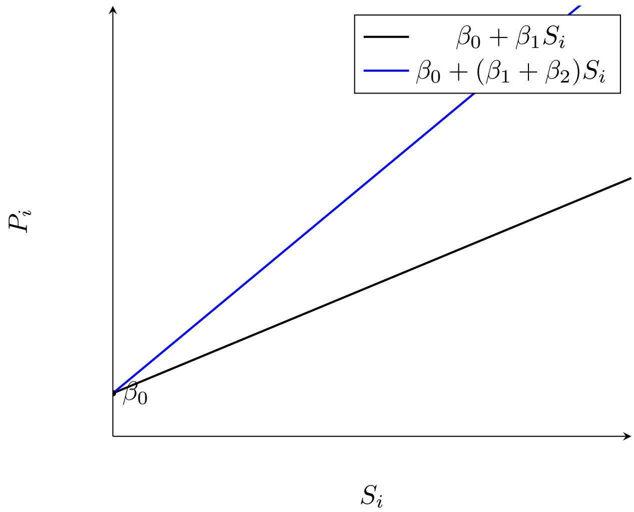
$$E(P_i) = \begin{cases} (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 S_i, & D_i = 1 \\ \hat{\beta}_0 + \hat{\beta}_1 S_i, & D_i = 0 \end{cases} \tag{37}$$



5.2 Slope Dummy Variable

$$P_i = \beta_0 + \beta_1 S_1 + \beta_2 (S_i \cdot D_i) + \epsilon_i \tag{38}$$

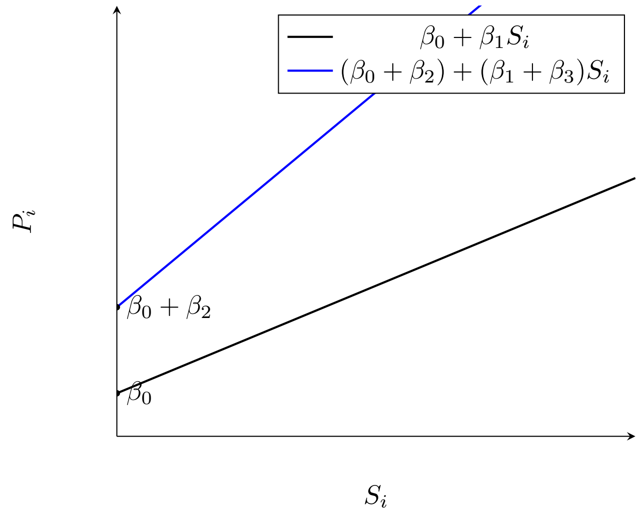
$$E(P_i) = \begin{cases} \hat{\beta}_0 + \left(\hat{\beta}_1 + \hat{\beta}_2\right) S_i, & D_i = 1 \\ \hat{\beta}_0 + \hat{\beta}_1 S_i, & D_i = 0 \end{cases} \tag{39}$$



5.3 Slope & Dummy Variable

$$P_i = \beta_0 + \beta_1 S_1 + \beta_2 D_i + \beta_3 S_i D_i + \epsilon_i \tag{40}$$

$$E(P_i) = \begin{cases} (\hat{\beta}_0 + \hat{\beta}_2) + \left(\hat{\beta}_1 + \hat{\beta}_3\right) S_i, & D_i = 1 \\ \hat{\beta}_0 + \hat{\beta}_1 S_i, & D_i = 0 \end{cases} \tag{41}$$



5.4 Multi-Categories Dummy Variable

$$P_0 = b_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \underbrace{b_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + b_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{\text{Leads to Perfect Multicollinearlity}} \tag{42}$$

5.4.1 Alternatives

- B_n captures the mean of each category, but F-Test is impossible

$$y = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} \tag{43}$$

- Computer drops automatically drops a variable

$$y = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} \tag{44}$$

- Manually dropping a variable

$$y = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} \tag{45}$$

6 Logistic Regression

For $Y_i \in \{0, 1\}$:

$$z_k = \beta_0 + \sum_{j=1}^{n} \beta_{jk} x_j + \epsilon_k, \beta_j \to \text{Logit Coefficient} \tag{46}$$

$$p = \frac{\exp^k}{1 + \exp^k} = \frac{1}{1 + \exp^{-k}} \tag{47}$$

where p is the probability of $y = 1$.