Regression I Midterm

Formulas

Sample Statistics

Mean ($\mu_x; \bar{x}$)

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

Variance ($Var(X); s_X^2; \sigma_X^2$)

$$Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2 \tag{2}$$

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_X)^2 \tag{3}$$

Covariance ($Cov(X,Y); s_{XY}; \sigma_{XY}$)

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y \tag{4}$$

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_X)(y_i-\mu_Y) \tag{5}$$

Correlation ($Corr(X,Y); r_{XY}; \rho_{XY}$)

$$Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)*sd(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{6}$$

Bivariate Regression Statistics

Slope ($\hat{B}_1$)

$$\hat{B}_1 = \frac{\sum(x_i-\mu_x)(y_i-\mu_y)}{\sum(x_i-\mu_x)^2} = \frac{Cov(X,Y)}{Var(X)} \tag{7}$$

Constant ($\hat{B}_0$)

$$\hat{B}_0 = \bar{y} - \hat{B}_1\bar{x} \tag{8}$$

Standard Error of the Regression ($\hat{\sigma}$)

$$\hat{\sigma}^2 = \frac{1}{(n-k-1)}\sum_{i=1}^{n}\hat{u}_i^2 \tag{9}$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \tag{10}$$

Standard Error (B_1)

$$s.e.(\hat{B}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum(x_i-\mu_x)^2}} \tag{11}$$

Goodness of Fit

Explained ("Model") Sum of Squares (SSE)

$$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{12}$$

Residual Sum of Squares (SSR)

$$SSR = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}\hat{u}_i^2 \tag{13}$$

Total Sum of Squares (SST)

$$SST = \sum_{i=1}^{n}(y_i-\bar{y})^2 = SSE + SSR \tag{14}$$

Coefficient of Determination (R^2)

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = Corr(\hat{y}, y) \tag{15}$$

$$Adj.R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} \tag{16}$$

Mean Squared Error (MSE)

$$MSE = \frac{SSR}{n-k-1} \tag{17}$$

$$RootMSE = \sqrt{MSE} \tag{18}$$

Multiple Regression Statistics

Slope ($\hat{B}_1$)

$$\hat{B}_1 = \left(\sum_{i=1}^{n}\hat{ri1}y_i\right) / \left(\sum_{i=1}^{n}\hat{ri1}^2\right) \tag{19}$$

$$where$$
$$\hat{ri1} = residuals(x_1 = B_0 + B_2x_2 + B_3x_3 + \cdots + B_kx_k)$$

Standard Error ($\hat{B}_1$)

$$s.e.(\hat{B}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum(x_{i1}-\mu_{x1})^2(1-R_1^2)}} \tag{20}$$

$$where$$
$$R_1^2 = R^2(x_1 = B_0 + B_2x_2 + B_3x_3 + \cdots + B_kx_k)$$

Hypothesis Testing Statistics

Test (t) statistic for $\hat{B}_1$ when $H_0: B_1 = 0$

$$t_{\hat{B}_1} = \hat{B}_1/s.e.(\hat{B}_1) \tag{21}$$

Confidence interval around $\hat{B}_1$

$$\hat{B}_1 \pm t_{crit} * s.e.(\hat{B}_1) \tag{22}$$

F Statistic, e.g. $H_0 := \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$

$$F = \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)} \tag{23}$$

$$where$$
$$q = df_r - df_{ur}$$

Statistical Properties

Linear Combination Algebra

$$E(aX+b) = aE(x) + b \tag{24}$$

$$E(aX+bY) = aE(X) + bE(Y) \tag{25}$$

$$Var(aX+b) = a^2Var(x) \tag{26}$$

$$Var(aX+bY) = a^2Var(X) + b^2Var(Y) + 2abCov(XY) \tag{27}$$

OLS Estimates

$$\sum_{i=1}^{n}\hat{u}_i^2 = 0 \tag{28}$$

$$\sum_{i=1}^{n}x_i\hat{u}_i = Cov(x_i\hat{u}_i) = 0 \tag{29}$$

Misc.

OLS Assumptions

Gauss-Markov
1. Model is linear in its parameters; Y is a linear function of X and u
2. We have a random sample.
3. We have variation in X
4. Zero Conditional Mean; $E(u) = 0; E(u|x) = 0$
5. Homoskedasticity (Constant error variance for all Xs); $Var(u|x) = \sigma^2$

Note: If OLS meets GM, it is (B)est (L)inear (U)nbiased (E)stimator.

Classic Linear Model
6. No perfect collinearity between Xs
7. Normally distributed u

Extra Assumptions
8. No correlation between errors, i.e. $Cov(u_iu_j) = 0$
9. Fixed Xs across repeated samples

Properties of Estimators

1. Unbiasedness, $E(W) = \theta$
2. Efficiency, W_1 is more efficient than W_2 if $Var(W_1) \leq Var(W_2)$

Large Sample Properties

3. Consistency, as n gets larger, distribution of W_n becomes more concentrated around θ
4. Asymptotic Normality, as n gets larger, distribution Z_n becomes more normal.
5. Central Limit Theory, the sampling distribution of standardized means drawn from any shape distribution will be asymptotically normal.

Hypothesis Testing Terms

Type 1 Error (α), probability of false rejection of H_0
Type 2 Error (β), probability of false acceptance of H_0
Correct Acceptance of $H_0 = 1 - \alpha$
Correct Acceptance of H_0, *(Power)* $= 1 - \beta$