

Formula Sheet - Midterm 2

Descriptive Measures (sample size n)

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

$$s = \sqrt{s^2} \quad CV = \frac{s}{|\bar{x}|}$$

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad r = \frac{\text{Cov}(x, y)}{s_x s_y}$$

Properties of Estimators

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

$$\text{Var}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right]$$

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$$

Statistics and Their Distributions

Statistic	Distribution
$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$Z \sim \mathcal{N}(0, 1)$
$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	$T \sim t_{n-1}$
$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	$Z \sim \mathcal{N}(0, 1)$
$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$ ($\nu \approx \min(n_1 - 1, n_2 - 1)$)	$T \sim t_\nu$
$J = \frac{(n-1)S^2}{\sigma^2}$	χ_{n-1}^2
$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	$F_{(n_1-1, n_2-1)}$
$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$ (paired differences D_i)	t_{n-1}
$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ (correlation test)	t_{n-2}

Simple Linear Regression

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right)\right)$$

$$s^2 = \frac{\sum u_i^2}{n-2}, \quad \frac{(n-2)s^2}{\sigma^2} \sim \chi_{n-2}^2$$

Prediction

$$\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Average (mean) prediction at x_0 :

$$\hat{y}(x_0) \sim N\left(\beta_0 + \beta_1 x_0, \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)\right)$$

New observation at x_0 :

$$y(x_0) \sim N\left(\beta_0 + \beta_1 x_0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)\right)$$

ANOVA Decomposition

$$\text{Total SS: SST} = \sum (y_i - \bar{y})^2$$

$$\text{Regression SS: SSR} = \sum (\hat{y}_i - \bar{y})^2, \quad \text{Error SS: SSE} = \sum (y_i - \hat{y}_i)^2$$

$$\text{SST} = \text{SSR} + \text{SSE}$$

Goodness of Fit

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

Normality Test

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} K^2 \right)$$

where S = sample skewness, K = sample excess kurtosis

Notes: (i) S^2, S_1^2, S_2^2 are sample variances; S_D is the sample sd of differences. (ii) Welch df shown as a simple conservative approximation.