Formula Sheet - Midterm 1

Descriptive Measures (sample size n)

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

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

Sample Standard Deviation: $s = \sqrt{s^2}$

Coefficient of Variation: $\mathrm{CV} = \frac{s}{|\bar{x}|}$

Sample Covariance: $\mathrm{Cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$

Sample Correlation: $r = \frac{\mathrm{Cov}(x,y)}{s_x s_y}$

Properties of Estimators

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

Variance: $\mathrm{Var}(\hat{\theta}) = \mathbb{E}\Big[\big(\hat{\theta} - \mathbb{E}[\hat{\theta}]\big)^2\Big]$

Mean Squared Error: $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2$

Statistics and Their Distributions

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

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.