GSBA 545 Formula Sheet 2021

Descriptive Statistics

Population: $\mu = \frac{1}{N}\sum X_i$ $\sigma^2 = \frac{1}{N}\sum(X_i - \mu)^2$ $\sigma = \sqrt{\sigma^2}$

Sample: $\bar{x} = \frac{1}{N}\sum X_i$ $s^2 = \frac{1}{n-1}\sum(X_i - \bar{x})^2$ $s = \sqrt{s^2}$

Coefficient of Variation: $c_v = \frac{\sigma}{\mu}$ or $\frac{s}{\bar{x}}$

Normal Distribution/Sampling

If X is normal with mean μ and std dev σ, then $Z = \frac{X-\mu}{\sigma}$.

If $n \geq 30$, then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Other Distributions

	Mean	Variance
Uniform:	$\mu = \frac{a+b}{2}$	$\sigma^2 = \frac{(b-a)^2}{12}$
Exponential:	$\mu = \frac{1}{\lambda}$	$\sigma^2 = \frac{1}{\lambda^2}$
Poisson:	$\mu = \lambda$	$\sigma^2 = \lambda$

Confidence Intervals/Sample Size Determination

Means: $n \geq 30$: $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm Z_{\alpha/2}\frac{s}{\sqrt{n}}$

$n < 30$: $\bar{X} \pm t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}, df = n - 1$

$$n = \left(\frac{Z_{\alpha/2} \times \sigma}{E}\right)^2$$

Proportions: $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$,

$$n = p(1-p)\left(\frac{Z_{\alpha/2}}{E}\right)^2$$

Confidence Interval Z-values:

90% $Z = 1.645$
95% $Z = 1.96$
99% $Z = 2.575$

Linear Regression

$$\hat{Y} = b_0 + b_1 X \quad \text{Error: } e_i = Y_i - \hat{Y}_i$$

PI (estimate): $\hat{Y} \pm t_{0.025,df} RMSE$

CI (estimate): $\hat{Y} \pm t_{0.025,df}\frac{RMSE}{\sqrt{n}}$

Confidence Interval (β_j): $b_j \pm t_{0.025,dfError} s_{b_j}$

Discrete Distributions

$$\mu = E(X) = \sum X \cdot P(X)$$

$$\sigma^2 = Var(X) = \sum (X - \mu)^2 \cdot P(X)$$

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Hypothesis Testing

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ or } Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \text{ with } df = n - 1$$

	Z-values	
α	One-sided	Two-sided
0.10	1.28	1.645
0.05	1.645	1.96
0.01	2.33	2.575

Two-Sample Hypothesis Testing

Independent: $df = n_1 + n_2 - 2$

$t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{se(\bar{x}_1 - \bar{x}_2)}$, where

$$se(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Proportions:

$Z = \frac{\hat{p}_1 - \hat{p}_2}{se(\hat{p}_1 - \hat{p}_2)}$, where

$$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Paired: $df = n - 1$, $t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}}$