Intermediate Statistics Formulas in LaTeX

Dan Liu

2025-01-02

I think students should be able to use this class as an opportunity to learn how to typeset with LaTeX. Put in some time to learn it! To help you, I've provided most of the formulas you'll need for this course below. Ideally, you should be taking good notes of these formulas yourselves with whatever method of note taking you currently use, but feel free to use these to make your submissions better.

At the end, I've also supplied a snippet of LaTeX to show how to create a matrix and vector.

Lecture 1 - Descriptive Statistics

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

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

Lecture 2 - Sampling Distributions

$$P(N = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu = E[N] = np$$

$$\sigma^2 = Var[N] = np(1-p)$$

$$\widehat{p} = N/n$$

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

$$\frac{N}{n} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

$$\frac{N/n - p}{\sqrt{p(1-p)/n}} \sim \mathcal{N}(0, 1)$$

$$\widehat{p_1} - \widehat{p_2} \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

$$X = \begin{cases} 1, & \text{Success} \\ 0, & \text{Failure} \end{cases}$$

$$X = \begin{cases} 1, & P(\text{Success}) \\ 0, & P(\text{Failure}) \end{cases}$$

$$\mu = E[X]$$

$$\widehat{p} = \frac{N}{n}$$

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0,1)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$r(x,y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim \mathcal{T}_{n-1}$$

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_1+n_2-2}$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathcal{F}_{n_1-1,n_2-1}$$

$$\frac{[(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)] \Big/ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}{\sqrt{\frac{(n_1-1)S_1^2/\sigma_1^2 + (n_2-1)S_2^2/\sigma_2^2}{n_1+n_2-2}}} \sim \mathcal{T}_{n_1+n_2-2}$$

Lecture 3 - Interval Estimation

$$\left[\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim \mathcal{T}_{n-1}$$

$$\left[\bar{X} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]$$

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)}\right]$$

$$\left[\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

Lecture 4 - Inference on Population Mean & Proportion

$H_a : \mu \neq \mu_0$ (two-sided)

$H_a : \mu < \mu_0$ (left-tailed)

$H_a : \mu > \mu_0$ (right-tailed)

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim \mathcal{N}(0, 1)$$

$$\begin{aligned}
&P_{H_0}(|Z| \geq |z|) = 2\Phi(-|z|) < \alpha && \text{for } H_a : \mu \neq \mu_0 \\
&P_{H_0}(Z > z) = 1 - \Phi(z) < \alpha && \text{for } H_a : \mu > \mu_0 \\
&P_{H_0}(Z < z) = \Phi(z) < \alpha && \text{for } H_a : \mu < \mu_0
\end{aligned}$$

$$T = \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim \mathcal{T}_{n-1}$$

$$\begin{aligned}
&P_{H_0}(|T| \geq |t|) = 2\mathbf{pt}(-|t|, n-1) < \alpha && \text{for } H_a : \mu \neq \mu_0 \\
&P_{H_0}(T > t) = 1 - \mathbf{pt}(t, n-1) < \alpha && \text{for } H_a : \mu > \mu_0 \\
&P_{H_0}(T < t) = \mathbf{pt}(t, n-1) < \alpha && \text{for } H_a : \mu < \mu_0
\end{aligned}$$

$$D = Y - X \sim \mathcal{N}(\mu_Y - \mu_X, \sigma^2)$$

$$Z = \frac{\sqrt{n}(\widehat{p} - p_0)}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0, 1)$$

$$\begin{aligned}
&P_{H_0}(|Z| \geq |z|) = 2\Phi(-|z|) < \alpha && \text{for } H_a : p \neq p_0 \\
&P_{H_0}(Z > z) = 1 - \Phi(z) < \alpha && \text{for } H_a : p > p_0 \\
&P_{H_0}(Z < z) = \Phi(z) < \alpha && \text{for } H_a : p < p_0
\end{aligned}$$

$$Z = \mathcal{N}(0, 1) \text{ is equivalent to } \chi^2 = Z^2 \sim \chi^2_1$$

Lecture 5 - Inference on Two Population Means and Proportions

Note: statistic formulas are written without assuming H_0!

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim \mathcal{N}(0, 1)$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_p^2/n_1 + S_p^2/n_2}} \sim \mathcal{T}_{n_1+n_2-2}$$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim \mathcal{T}_k$$

$$k = \left\lceil \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{1}{n_1-1}(s_1^2/n_1)^2 + \frac{1}{n_2-1}(s_2^2/n_2)^2} \right\rceil$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathcal{F}_{n_1-1,n_2-1}$$

$$\widehat{p_1} - \widehat{p_2} \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

$$SE = \sqrt{\widehat{p_1}(1-\widehat{p_1})/n_1 + \widehat{p_2}(1-\widehat{p_2})/n_2}$$

$$\widehat{p} = \frac{N_1 + N_2}{n_1 + n_2}$$

$$Z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \sim \mathcal{N}(0,1)$$

Lecture 6 - Inference on Two-Way Tables

$$E_{i,j} = \frac{O_{i,+}O_{+,j}}{O_{+,+}}$$

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi_k^2 = \chi_{(r-1)(c-1)}^2$$

Lecture 7 - Analysis of Variance (ANOVA)

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \bar{X}_{\cdot,\cdot})^2$$

$$\bar{X}_{\cdot,\cdot} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{X}_{i,\cdot}$$

$$SSB = \sum_{i=1}^{k} n_i(\bar{X}_{i,\cdot} - \bar{X}_{\cdot,\cdot})^2$$

$$\bar{X}_{i,\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}$$

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{i,\cdot})^2 = \sum_{i=1}^{k} (n_i - 1) S_i^2$$

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{i,\cdot})^2$$

$$MSB = \frac{SSB}{k-1}$$

$$MSE = \frac{SSE}{n-k}$$

$$F = \frac{MSB}{MSE} = \frac{SSB/(k-1)}{SSE/(n-k)} \sim \mathcal{F}_{k-1,n-k}$$

$$R^2 = \frac{SSB}{SST} \in (0,1)$$

$$\psi = a_1\mu_1 + \cdots + a_k\mu_k$$

$$C_\psi = \sum_{i=1}^{k} a_i \bar{X}_{i,\cdot}$$

$$C_\psi \sim \mathcal{N}(\psi, \sigma_\psi^2)$$

$$S_p^2 = \frac{1}{\sum_{i=1}^{k}(n_i - 1)} \sum_{i=1}^{k} (n_i - 1) S_i^2 = \frac{SSE}{n-k} = MSE$$

$$SE = \sqrt{S_p^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}}$$

$$\frac{SSE}{\sigma^2} = \sum_{i=1}^{k} \frac{(n_i - 1) S_i^2}{\sigma^2} \sim \chi_{n-k}^2$$

$$T = \frac{C_\psi}{\sqrt{S_p^2 \sum_{i=1}^{k} \frac{1}{n_i} a_i^2}} = \frac{C_\psi / \sqrt{\sigma^2 \sum_{i=1}^{k} \frac{1}{n_i} a_i^2}}{\sqrt{\frac{SSE}{\sigma^2(n-k)}}} \sim \mathcal{T}_{n-k}$$

$$P_{H_0}(|T| > |t|) = 2(1 - pt(|t|, n-k)) < \alpha$$

$$C_\psi \pm t_{1-\alpha/2}(n-k) \sqrt{S_p^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}}$$

$$T_{i,j} = \frac{\bar{X}_{i,\cdot} - \bar{X}_{j,\cdot}}{\sqrt{S_p^2(\frac{1}{n_i} + \frac{1}{n_j})}} \sim \mathcal{T}_{n-k}$$

$$P_{H_0}(|T_{i,j}| > |t_{i,j}|) = 2[1 - \mathrm{pt}(|t_{i,j}|, n-k)] < \alpha$$

$$|\bar{x}_{i,\cdot} - \bar{x}_{j,\cdot}| > t_{1-\alpha/2}(n-k)\sqrt{2s_p^2/n_1}$$

Lectures 8 & 9 - Regressions

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$$

$$\widehat{\beta} = \begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \end{bmatrix} [(X'X)^{-1}X']Y$$

$$S^2 = \frac{SSE}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

$$SE_{\widehat{\beta_0}} = \sqrt{(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2})S^2}$$

$$SE_{\widehat{\beta_1}} = \sqrt{(\frac{1}{\sum(x_i - \bar{x})^2})S^2}$$

$$\widehat{\beta_0} \pm t_{1-\alpha/2}(n-2) \cdot SE_{\widehat{\beta_0}}$$

$$\widehat{\beta_1} \pm t_{1-\alpha/2}(n-2) \cdot SE_{\widehat{\beta_1}}$$

$$SE_{\widehat{\mu_{Y^*}}} = \sqrt{(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2})S^2}$$

$$\widehat{\beta_0} + \widehat{\beta_1}x^* \pm t_{1-\alpha/2}(n-2) \cdot SE_{\widehat{\mu_{Y^*}}}$$

$$SE_{\widehat{Y^*}} = \sqrt{(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2})S^2}$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$SSM = \sum_{i=1}^{n}(\widehat{y_i} - \bar{y})^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

$$SST = SSM + SSE$$

$$MSE = \frac{SSE}{n-2} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = S^2$$

$$F = \frac{MSM}{MSE} = \frac{SSM/1}{SSE/(n-2)} \sim \mathcal{F}_{1,n-2}$$

$$R^2 = \frac{SSM}{SST}$$

$$\widehat{\beta_j} \sim \mathcal{N}(\beta_j, \sigma_j^2)$$

$$\widehat{\beta_j} \pm t_{1-\alpha/2}(n-k-1) \cdot SE_{\widehat{\beta_j}}$$

$$T_j = \frac{\widehat{\beta_j}}{SE_{\widehat{\beta_j}}}$$

$$MSE = \frac{1}{n-k-1} SSE = S^2$$

$$F = \frac{MSM}{MSE} \sim \mathcal{F}_{k,n-k-1}$$

$$R^2 = \frac{SSM}{SST}$$

$$R^2_{adj} = \frac{n-1}{n-k-1}R^2 - \frac{k}{n-k-1}$$

Matrices in LaTeX

$$\text{matrix} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

$$\text{vector} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$