STA 103 (SS2 2025): Formula Sheet Final
Instructor: Wookyeong Song
Institution: UC Davis

- Summation notation: $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$
- Linear properties of summation:
  1. $\sum_{i=1}^{n}(ax_i) = a(\sum_{i=1}^{n} x_i)$
  2. $\sum_{i=1}^{n}(x_i + b) = (\sum_{i=1}^{n} x_i) + nb$
  3. $\sum_{i=1}^{n}(x_i + y_i) = (\sum_{i=1}^{n} x_i) + (\sum_{k=1}^{n} y_i)$
- Suppose $y_i = ax_i + b$ and $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = v$, then:
  1. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = (\frac{1}{n}\sum_{i=1}^{n} x_i^2) - (\overline{x})^2$
  2. $\sum_{i=1}^{n} y_i = a(\sum_{i=1}^{n} x_i) + nb$
  3. $\overline{y} = a\overline{x} + b$
  4. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = a^2 v$
- PMFs (Probability Mass Functions)
  1. Valid PMF Requirements
  2. Table format
  3. Function format
  4. Modifying PMFs for functions of a random variable.
  5. Modifying PMFs for conditional distributions.
- Expected value
  1. $E(X) = \sum_{i=1}^{n} x_i p(x_i)$
  2. $E(f(X)) = \sum_{i=1}^{n} f(x_i)p(x_i)$
- Variance and Standard Deviation
  1. $\mathrm{Var}(X) = E[(X - E(X))^2]$
  2. $\mathrm{Var}(X) = E(X^2) - (E(X))^2$
  3. $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$
- Bernoulli Distribution
  1. $X \sim \mathrm{Bernoulli}(p)$, Parameters: p.
  2. PMF:
  $$p(x) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x = 0 \text{ or } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$
  3. Mean $E(X) = p$, Variance $\mathrm{Var}(X) = p(1-p)$.
- Binomial Distribution
  1. $X \sim \mathrm{Binomial}(n, p)$, Parameters: n, p.
  2. PMF:
  $$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$
  3. $\binom{n}{x} = \frac{n!}{(n-x)!x!}$
  4. Mean $E(X) = np$,
     Variance $\mathrm{Var}(X) = np(1-p)$.
- Poisson Distribution
  1. $X \sim \mathrm{Poisson}(\lambda)$, Parameters: λ.
  2. PMF:
  $$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \ldots$$
  3. $\binom{n}{x} = \frac{n!}{(n-x)!x!}$
  4. Mean $E(X) = \lambda$, Variance $\mathrm{Var}(X) = \lambda$.
- JPMFs (Joint Probability Mass Functions)
  1. Valid JPMF Requirements
  2. Table format
  3. Formula format
  4. Computing marginal distributions from JPMFs.
  5. Modifying JPMFs for functions of random variables.
  6. Modifying JPMFs for conditional distributions.
- Expected value of a function of X and Y:
  1. $E(f(X, Y)) = \sum_{x,y} f(x, y)p(x, y)$.
  2. $E(X \mid Y = y) = \sum_x x p_X(x \mid Y = y)$.
  3. $E(Y \mid X = x) = \sum_y y p_Y(y \mid X = x)$.
- Independence: two random variables X and Y are independent if and only if:
  1. $p(x, y) = p_X(x) \cdot p_Y(y)$ for all possible x, y **or**
  2. $p_X(x|Y = y) = p_X(x)$ for all possible x, y **or**
  3. $p_Y(y|X = x) = p_Y(y)$ for all possible x, y
- Covariance of X and Y:
  $$\begin{aligned} \mathrm{cov}(X, Y) &= E[\,(X - E(X))(Y - E(Y))\,] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$
- Covariance Properties
  1. $\mathrm{cov}(aX, bY) = a \cdot b \cdot \mathrm{cov}(X, Y)$
  2. $\mathrm{cov}(X, c) = \mathrm{cov}(c, X) = 0$
  3. $\mathrm{cov}(X + b, Y + c) = \mathrm{cov}(X, Y)$
  4. $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
  5. $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$
  6. If X & Y are independent then $\mathrm{cov}(X, Y) = 0$

- Expectation formula:

$$E(aX+b) = aE(X)+b$$
$$E(aX+bY) = aE(X)+bE(Y)$$

- Variance formula:

$$\mathrm{var}(aX+b) = \mathrm{var}(aX) = a^2\,\mathrm{var}(X)$$
$$\mathrm{var}(aX+bY) = a^2\mathrm{var}(X)+b^2\mathrm{var}(Y)+2ab\,\mathrm{cov}(X,Y)$$
$$\mathrm{var}(aX-bY) = a^2\mathrm{var}(X)+b^2\mathrm{var}(Y)-2ab\,\mathrm{cov}(X,Y)$$

- Continuous probability density functions (PDF): If X has PDF $p(x)$ then

$$Pr(a \le X \le b) = \int_a^b p(x)\,dx$$
$$E(X) = \int_a^b x\,p(x)\,dx$$
$$E(f(X)) = \int_a^b f(x)\,p(x)\,dx$$

- Uniform Random Variable
  - The PDF for $Y \sim \mathrm{Unif}[a,b]$ is:

$$f(y) = \begin{cases} \frac{1}{(b-a)}, & a \le y \le b, \\ 0, & \text{otherwise.} \end{cases}$$

  - Mean and Variance:

$$E(Y) = \frac{a+b}{2}, \quad \mathrm{Var}(Y) = \frac{(b-a)^2}{12}.$$

- Normal Random Variables
  - The PDF for $X \sim N(\mu, \sigma^2)$ is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

  - Mean and Variance:

$$E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

- z-scores: For any random variable X we can change units to W via

$$W = \frac{X-a}{b}$$
$$X = a + bW.$$

  If we use $a = E(X)$ and $b = sd(X)$ then W is called a z-score transformation of X and satisfies

$$E(W) = 0 \text{ and } \mathrm{sd}(W) = 1.$$

- Special Normal/Gaussian Cases: If $X \sim \mathcal{N}(a,b)$ then

$$E(X) = a$$
$$sd(X) = \sqrt{b}$$
$$W = \frac{X-a}{\sqrt{b}} \sim N(0,1)$$

- The central limit theorem (CLT): If $X_1, \ldots, X_n$ are independent random variables, all with the same PMF or PDF, then

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \approx N(a,b)$$

  for large n (usually $n > 30$ is good), where $a = E(\overline{X})$ and $b = \mathrm{Var}(\overline{X})$.

- For i.i.d. RVs $X_1, \ldots, X_n$, with $E(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$.

$$E(\bar{X}) = E(X_1) = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X_1)}{n} = \frac{\sigma^2}{n}.$$

In each case you will base your inference on a test statistic $\hat{\theta}$ which an estimator of a population parameter and you will be asked to compute the following:

- Compute the Z-score for the observed value of your test statistic $\hat{\theta}$:

$$Z = \frac{\hat{\theta} - \theta}{\mathrm{S}E(\hat{\theta})} \sim N(0,1),$$

  assuming the given probability model for the data is correct, i.e. values assumed under the null hypothesis H_0.

- Determine if your observed Z-score is consistent with a typical random fluctuations in Z-scores. Use the CLT to quantify how rare the observed Z-score is, i.e. the p-value.

- Produce $100(1-\alpha)\%$ approximate **Confidence Intervals** for the unknown population parameter:

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}}\,\mathrm{S}E(\hat{\theta}),$$

  where formula for $\mathrm{S}E(\hat{\theta})$ can be computed but values for $\mathrm{S}E(\hat{\theta})$ are the "plug-in" estimates from your data or "conservative" estimates for one-sample proportion estimate $\hat{p}$.

- Two-sided Hypothesis Testing:

$$H_0 : \theta = \theta_0 \quad H_a : \theta \ne \theta_0.$$

- One-sided Hypothesis Testing "$>$":

$$H_0 : \theta \le \theta_0 \text{ (equivalently, } \theta = \theta_0) \quad H_a : \theta > \theta_0.$$

- One-sided Hypothesis Testing "<":

$$H_0 : \theta \geq \theta_0 \text{ (equivalently, } \theta = \theta_0) \quad H_a : \theta < \theta_0.$$

- The null H_0 must have the equality =. If the statement we want to prove is inequality > or <, then we put it on alternative H_a.
- Perform hypothesis test at significance level α for the unknown population parameter θ,
  - Step 1, find the Z-score.
  - Step 2, find the p-value following the direction of H_a.
  - Step 3, compare p-value with significance level, decide whether we reject the null H_0 or not.
- The population correlation coefficient ρ (or ρ_{XY}), defined as

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- Given paired data $(X_1, Y_1), \ldots, (X_n, Y_n)$, the sample correlation coefficient is defined as:

$$\hat{\rho} = r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

- Simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2), \text{ i.i.d.}$$

- The fitted (predicted) response is given by

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,$$

  where

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

- The variance σ^2 is estimated by

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2.$$

- Sampling variability of $\hat{\beta}_1$ given $X_1, \ldots, X_n$ are fixed. We expect it to be

$$\hat{\beta}_1 \sim N(E(\hat{\beta}_1), \text{Var}(\hat{\beta}_1)),$$

  where $E(\hat{\beta}_1) = \beta_1$, and

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \approx \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- The Z-score for $\hat{\beta}_1$ is given by

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\text{sd}(\hat{\beta}_1)} \sim N(0, 1).$$