SM-4331 Formula Sheet

1. **Distributions**

Distribution	Distribution function	Mean	Variance
Bernoulli	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	${}^nC_x\, p^x(1-p)^{n-x}$	np	$np(1-p)$
Poisson	$e^{-\lambda}\lambda^x/x!$	λ	λ
Normal	$(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Exponential	$\mu^{-1}e^{-x/\mu}$	μ	μ^2

2. **Sampling variances for estimators of the mean**

   Let $y_1, \dots, y_N$ be values of a population whose mean is $\mu := N^{-1}\sum_{i=1}^N y_i$, and (corrected) variance is $(N-1)\sum_{i=1}^N (y_i - \mu)^2$.

   - **Simple random sampling:** Let $\hat{\mu}$ be an estimator for μ from a sample of SRS of size n. Then
     $$\mathrm{Var}_{\mathrm{srs}}(\hat{\mu}) = \frac{N-n}{Nn}S^2$$
   - **Cluster sampling:** Let $\hat{\mu}$ be an estimator for μ using data from m out of a total of M clusters. Then
     $$\mathrm{Var}_{\mathrm{cl}}(\hat{\mu}) = \frac{M-m}{Mm}S_{\mathrm{cl}}^2$$
     where $S_{\mathrm{cl}}^2 = (M-1)^{-1}\sum_{j=1}^M (\mu_j - \mu)^2$ and μ_j are the cluster means.
   - **Stratified sampling:** Let $\hat{\mu}$ be the stratified sample estimator, where n_h elements out of a possible N_h elements from each strata $h = 1, \dots, H$ were taken. Then
     $$\mathrm{Var}(\hat{\mu}) = \sum_{h=1}^H \frac{N_h - n_h}{N_h n_h} S_h^2$$
     where S_h^2 is the (corrected) variance of the stratum h.

3. **Two-sample t-test (equal variances)**

   $$T = \frac{n_x + n_y - 2}{1/n_x + 1/n_y}\sqrt{\frac{\bar{X} - \bar{Y} - \delta}{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}}.$$

   Under $H_0 : \mu_x - \mu_y = \delta$, $T \sim t_{n_x+n_y-2}$.

4. **Simple linear regression**

   - **Model:** $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ for $i = 1, \dots, n$.

- **LSE:**

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \quad \hat{\beta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{j=1}^{n}(x_i - \bar{x})^2, \text{ and}$$

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{j=1}^n (x_i - \bar{x})^2}, \quad \mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{j=1}^n (x_i - \bar{x})^2}, \quad \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2\bar{x}}{\sum_{j=1}^n (x_i - \bar{x})^2}$$

- **Estimator for the variance of ϵ_i:** $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$.
- **Regression ANOVA:**

$$\text{Total SS} = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad \text{Reg SS} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2, \quad \text{Resid SS} = \sum_{i=1}^{n}(\hat{y}_i - \hat{y}_i)^2.$$

- **Regression correlation coefficients:**

$$R^2 = \frac{\text{Reg SS}}{\text{Total SS}}, \qquad \tilde{R}^2 = 1 - \frac{\text{Resid SS}/(n-2)}{\text{Total SS}/(n-1)}$$

- **Confidence interval:** A $(1-\alpha)\%$ confidence interval for $\mu(x)$ is

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{n-2}(\alpha/2) \cdot \hat{\sigma}\sqrt{\frac{\sum_{i=1}^n (x_i - x)^2}{n\sum_{j=1}^n (x_j - \bar{x})^2}}$$

- **Predictive interval:** A predictive interval which covers y with probability $(1-\alpha)$ is

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{n-2}(\alpha/2) \cdot \hat{\sigma}\sqrt{1 + \frac{\sum_{i=1}^n (x_i - x)^2}{n\sum_{j=1}^n (x_j - \bar{x})^2}}$$