MATH10282 Introduction to Statistics

Formulas to remember for the final exam in May/June 2022

Given a data set, know how to compute the sample mean, sample variance and sample median.

Let $x_1, x_2, \ldots, x_n$ **denote a data set and let** $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ **denote the order statistics in ascending order. Then the** p**th quartile is**

$$Q(p) = x_{\left(r^{'}\right)} + \left\{p(n+1) - r^{'}\right\} \left\{x_{\left(r^{'}+1\right)} - x_{\left(r^{'}\right)}\right\},$$

where $r^{'} = [p(n+1)]$**.**

$\widehat{\theta}$ **is an unbiased estimator of** θ **if** $E\left(\widehat{\theta}\right) = \theta$**.**

The bias of $\widehat{\theta}$ **is** $E\left(\widehat{\theta}\right) - \theta$**.**

$\widehat{\theta}$ **is asymptotically unbiased estimator of** θ **if its bias approaches zero as** $n \to \infty$**.**

The mean squared error of $\widehat{\theta}$ **is** $E\left[\left(\widehat{\theta} - \theta\right)^2\right]$**.**

$\widehat{\theta}$ **is a consistent estimator of** θ **if** $\lim_{n \to \infty} E\left[\left(\widehat{\theta} - \theta\right)^2\right] = 0$**.**

If $X \sim$ **Poisson** (a) **then** $E(X) = a$ **and** $Var(X) = a$**.**

If $X \sim$ **Exp** (a) **then** $E(X) = 1/a$ **and** $Var(X) = 1/a^2$**.**

Suppose $X_1, X_2, \ldots, X_n$ **is a random sample from** $N\left(\mu, \sigma^2\right)$**, where** σ **is known. We reject** $H_0 : \mu = \mu_0$ **versus** $H_1 : \mu \neq \mu_0$ **if** $\sqrt{n}\left|\overline{X} - \mu_0\right| / \sigma > z_{1-\frac{\alpha}{2}}$**.**

Suppose $X_1, X_2, \ldots, X_n$ **is a random sample from** $N\left(\mu, \sigma^2\right)$**, where** σ **is known. We reject** $H_0 : \mu = \mu_0$ **versus** $H_1 : \mu < \mu_0$ **if** $\sqrt{n}\left(\overline{X} - \mu_0\right) / \sigma < z_{\alpha}$**.**

If $X_1, X_2, \ldots, X_n$ **is a random sample from** $N\left(\mu, \sigma^2\right)$ **and** $\overline{X}$ **denotes the sample mean then** $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$**.**

Let $\mathbf{X} = (X_1, \ldots, X_n)$**, with** $X_1, \ldots, X_n$ **an independent random sample from a distribution** F_X **with unknown parameter** θ**. Let** $I(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$ **denote an interval estimator for** θ**.**

(i) $I(\mathbf{X})$ **is a** $100(1-\alpha)\%$ **confidence interval if**

$$\Pr\left(a(\mathbf{X}) < \theta < b(\mathbf{X})\right) = 1 - \alpha;$$

(ii) the coverage probability of $I(\mathbf{X})$ **is**

$$\Pr\left(a(\mathbf{X}) < \theta < b(\mathbf{X})\right);$$

(iii) the coverage length of $I(\mathbf{X})$ **is** $b(\mathbf{X}) - a(\mathbf{X})$**.**

The gamma function defined by $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$**.**

The property that $\Gamma(n) = (n-1)!$ **where** n **is a positive integer.**

When testing $H_0 : \mu = \mu_0$ **versus** $H_1 : \mu \neq \mu_0$ **the Type I error occurs if** H_0 **is rejected when in fact** $\mu = \mu_0$**.**

When testing $H_0 : \mu = \mu_0$ **versus** $H_1 : \mu \neq \mu_0$ **the Type II error occurs if** H_0 **is accepted when in fact** $\mu \neq \mu_0$**.**

The significance level is the probability of type I error.