QTM 100 Formula Sheet

Mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$; Standard deviation: $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$;

Boxplot whiskers: $Q_1 - 1.5IQR,\ Q_3 + 1.5IQR$;

Complement rule: $P(A^c) = 1 - P(A)$; General addition rule: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$;

Multiplication rule: $P(A\ and\ B) = P(A) \times P(B)$; Conditional probability rule: $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$;

z-score: $z = \frac{x-\mu}{\sigma}$; *x* from *z*-score: $x = \mu + z\sigma$;

Binomial probability: $P(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, x = 0, 1, 2, \dots, n$

Binomial distribution mean: $\mu = np$; Binomial distribution standard deviation: $\sigma = \sqrt{np(1-p)}$;

Sampling distributions standard deviations: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$; $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Confidence interval for proportion: $\hat{p} \pm z_\alpha \times se$; Standard error in CI for proportion: $se_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

z-score of sample proportion: $z = \frac{\hat{p}-p_0}{se_0}$; Standard error for HT for p_0: $se_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$

CI for difference of two proportions: $(\hat{p}_1 - \hat{p}_2) \pm z_\alpha \times se$; $se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$;

Two-sample *z* test: $z = \frac{(\hat{p}_1-\hat{p}_2)-0}{se_0}$; $se_0 = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$; $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$;

Chi squared: $\chi^2 = \sum \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$; expected $= \frac{\text{row}\times\text{column}}{\text{total}}$; $df = (r-1) \times (c-1)$;

Confidence interval for mean: $\bar{x} \pm t_\alpha \times se$; *t*-score: $t = \frac{\bar{x}-\mu}{se_{\bar{x}}}$;

Standard error in CI for mean: $se_{\bar{x}} = \frac{s}{\sqrt{n}}$; Degrees of freedom in *t* distribution: $df = n - 1$

CI for difference of two means: $\bar{x}_1 - \bar{x}_2 \pm t_\alpha \times se$; Two-sample t test: $t = \frac{(\bar{x}_1-\bar{x}_2)-0}{se}$;

Unequal variance: $se = \sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}$; $df = \min\{n_1 - 1, n_2 - 1\}$

Equal variance: $s_p = \sqrt{\frac{(n_1-1)s_1{}^2+(n_2-1)s_2{}^2}{n_1+n_2-2}}$; $se = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$; $df = n_1 + n_2 - 2$;

ANOVA: $F = \frac{MS_{group}}{MS_{error}}$; Between-group variability: $MS_{group} = \frac{SS_{group}}{df1}$; $df1 = g - 1$;

Within-group variability: $MS_{error} = \frac{SS_{error}}{df2}$; $df2 = N - g$

Regression line: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$; Slope: $\hat{\beta}_1 = r\frac{s_y}{s_x}$; residual $= y - \hat{y}$; $s = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$;

Regression slope test statistic: $t = \frac{\hat{\beta}_1-0}{se_{\hat{\beta}_1}}$; CI for regression line slope: $\hat{\beta}_1 \pm t_\alpha \times se_{\hat{\beta}_1}$; $df = n - 2$;

$r^2 = \frac{SS_{regression}}{SS_{total}}$;