LSHTM. Basic Statistics for PHP students: Exam Formulae, 2018-19

Context	Standard Error	95% CI	z, t, or chi-squared
Single mean (large sample)	$SE(\bar{x}) = \frac{s}{\sqrt{n}}$	$\bar{x} \pm 1.96 SE(\bar{x})$	$z = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$
Single mean (small sample)	as above	$\bar{x} \pm t\ SE(\bar{x})$ (1)	$t = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$ (1)
Single proportion (as percentage)	$SE(p) = \sqrt{\frac{p(100-p)}{n}}$	$p \pm 1.96 SE(p)$	$z = \frac{p - \pi_0}{SE(p)}$
Comparing two means (large sample)	$SE(diff) = \sqrt{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}$	$(\bar{x}_1 - \bar{x}_2) \pm 1.96 SE(diff)$	$z = \frac{\bar{x}_1 - \bar{x}_2}{SE(diff)}$
Comparing two means (small sample)	$SE(diff) = s_p \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}$ (2)	$(\bar{x}_1 - \bar{x}_2) \pm t SE(diff)$ (3)	$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(diff)}$ (3)
Comparing two proportions (as percentage)	$SE(diff) = \sqrt{\frac{\bar{p}(100-\bar{p})}{n_1} + \frac{\bar{p}(100-\bar{p})}{n_2}}$ *Where $\bar{p}$ = overall proportion*		$z = \frac{p_1 - p_2}{SE(diff)}$
	$SE(diff) = \sqrt{\frac{p_1(100-p_1)}{n_1} + \frac{p_2(100-p_2)}{n_2}}$	$(p_1 - p_2) \pm 1.96 SE(diff)$	
Testing association in an r X c contingency table			$X^2 = \sum \frac{(O-E)^2}{E}$ (4)

Sample size calculations	Estimating an effect with a certain degree of precision (5% significance level)	Testing a hypothesis – comparing 2 groups (5% significance level, 80% power)
Proportions (as percentage)	$n = \frac{3.84 \times p(100-p)}{\text{precision}^2}$	$n = \frac{\left(1.96\sqrt{2\bar{p}(100-\bar{p})} + 0.84\sqrt{p_1(100-p_1) + p_2(100-p_2)}\right)^2}{D_{min}^2}$
Means	$n = \frac{3.84 \times SD^2}{\text{precision}^2}$	$n = \frac{(0.84 + 1.96)^2 \times (\sigma_1^2 + \sigma_2^2)}{D_{min}^2}$

NOTATION	Single sample		For 2 groups			
	Population	**Sample**	**Population 1**	**Population 2**	**Sample 1**	**Sample 2**
Mean	μ	$\bar{x}$	μ_1	μ_2	$\bar{x}_1$	$\bar{x}_2$
SD	σ	s	σ_1	σ_2	s_1	s_2
Proportion	π	p	π_1	π_2	p_1	p_2
Sample size		n			n_1	n_2

μ_0 = A known reference mean; π_0 = A known reference proportion; $\bar{X}_i$ = an observed mean in group i

p_i = an observed proportion in group i, expressed as a percentage (if proportions are not expressed as percentages then replace 100 by 1 in above formulae).

(1) Student t distribution with degrees of freedom = n - 1; for 95% CI use 5% (two-tailed) point.

(2) $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$ (pooled estimate of common standard deviation)

(3) Degrees of freedom = $n_1 + n_2 - 2$; for 95% CI use 5% (two-tailed) point.

(4) O= observed cell count; E=expected count if null hypothesis were true; degrees of freedom = (r-1)X(c-1), where r=number of rows, c=number of columns.