Summary Table for Statistical Techniques

ESTIMATES

Inference	Parameter	Statistic	Type of Data	Examples	Analysis	Conditions
Estimating a proportion	One population proportion p	sample proportion $\hat{p}$	categorical (binary)	What is the proportion of males in the world? What is the proportion of students that smoke?	1-proportion Z-interval $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	n $\hat{p} \geq 10$ and n (1- $\hat{p}$) ≥ 10
Estimating a mean	One population mean μ	sample mean $\bar{x}$	quantitative	What is the average weight of adults? What is the average cholesterol level of adult females?	1-sample t-interval $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$	data approximately normal or have a large sample size ($n \geq 30$)
Estimating the difference of two proportions	difference in two population proportions p_1-p_2	difference in two sample proportions $\hat{p}_1 - \hat{p}_2$	categorical (binary)	Are the percentages of male and female smokers different? Are the percentages of upper- and lower- class binge drinkers different?	two-proportions Z-interval $(\hat{p}_1 - \hat{p}_2) \pm$ $z^* \times s.e._{\hat{p}_1 - \hat{p}_2}$ See notes for s.e. formula	independent samples from the two populations n $\hat{p} \geq 10$ and (1-n) $\hat{p} \geq 10$ for each sample
Estimating the difference of two means	difference in two population means μ_1-μ_2	difference in two sample means $\bar{x}_1 - \bar{x}_2$	quantitative	Are the mean GPAs of males and females different? Do vitamin C takers get, on average, fewer colds than non vitamin C takers?	two-sample t-interval $(\bar{x}_1 - \bar{x}_2) \pm t^* \times s.e._{\bar{x}_1 - \bar{x}_2}$ See text, page 445, for the s.e. of the difference	independent samples from the two populations data in each sample are about normal or large samples ($n_i \geq 30$)
Estimating a mean with paired data	mean of paired difference μ_D	sample mean of difference $\bar{x}_d$	quantitative	Is there a difference in pulse rates, on the average, before and after exercise?	paired t-interval $\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n}}$	differences approximately normal or have a large number of pairs ($n \geq 30$)

TESTS

Inference	Parameter	Statistic	Type of Data	Examples	Analysis	Conditions
Test about a mean	One population mean μ	sample mean $\bar{x}$	quantitative	Is the average GPA of juniors at Penn State higher than 3.0? Is the average Winter temperature in State College less than 42˚F?	$H_o: \mu = \mu_o$ $H_a: \mu \neq \mu_o$ or $H_a: \mu > \mu_o$ or $H_a: \mu < \mu_o$ The one sample t test: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	data approximately normal or have a large sample size ($n \geq 30$)
Test about a proportion	One population proportion p	sample proportion $\hat{p}$	categorical (binary)	Is the proportion of females different from 0.5? Is the proportion of students who fail Stat200 less than 0.1?	$H_o: p = p_o$ $H_a: p \neq p_o$ or $H_a: p > p_o$ or $H_a: p < p_o$ The one proportion Z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$n\,p_o \geq 10$ and $n\,(1-p_o) \geq 10$
Test to compare two means	difference in two population means μ_1-μ_2	difference in two sample means $\bar{x}_1 - \bar{x}_2$	quantitative	Do the mean pulse rates of exercisers and non-exercisers differ? Is the mean EDS score for dropouts greater than the mean EDS score for graduates?	$H_o: \mu_1 = \mu_2$ $H_a: \mu_1 \neq \mu_2$ or $H_a: \mu_1 > \mu_2$ or $H_a: \mu_1 < \mu_2$ The two sample t test: $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s.e._{\bar{x}_1 - \bar{x}_2}}$ See text, page 445, for the s.e. of the difference	independent samples from the two populations data in each sample are about normal or large samples ($n_i \geq 30$)
Test about a mean with paired data	mean of paired difference μ_D	sample mean of difference $\bar{x}_d$	quantitative	Is the difference in IQ of pairs of twins zero? Are the pulse rates of people higher after exercise?	$H_o: \mu_D = 0$ $H_a: \mu_D \neq 0$ or $H_a: \mu_D > 0$ or $H_a: \mu_D < 0$ $t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$	differences approximately normal or have a large number of pairs ($n \geq 30$)
Test to compare two proportions	difference in two population proportions p_1-p_2	difference in two sample proportions $\hat{p}_1 - \hat{p}_2$	categorical (binary)	Is the percentage of males with lung cancer higher than the percentage of females with lung cancer? Are the percentages of upper- and lower-class binge drinkers different?	$H_o: p_1 = p_2$ $H_a: p_1 \neq p_2$ or $H_a: p_1 > p_2$ or $H_a: p_1 < p_2$ The two proportion z test: $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{s.e._{\hat{p}_1 - \hat{p}_2}}$ See notes for s.e. formula	independent samples from the two populations $n\,\hat{p} \geq 10$ and $(1-n)\,\hat{p} \geq 10$ for each sample (or at least 5 for a two-tailed test).

Test to compare several means	Population means of the k populations $\mu_1, \mu_2, \ldots, \mu_k$	Sample means of the k populations $x_1, x_2, \ldots, x_k$	quantitative	Is there a difference between the mean GPA of Freshman, Sophomore, Junior and Senior classes?	$H_o: \mu_1 = \mu_2 = \cdots = \mu_k$ H_a: not all the means are equal The F test for one-way ANOVA: $F = \frac{MS_{Between}}{MS_{Within}}$	each population is normally distributed independent samples from the k populations equal population standard deviations

RELATIONSHIPS

Inference	Parameter	Statistic	Type of Data	Examples	Analysis	Conditions
Test about a slope	slope of the population regression line β	sample estimate of the slope b	quantitative	Can height be used to predict weight?	$H_o: \beta = 0$ $H_a: \beta \neq 0$ or $H_a: \beta > 0$ or $H_a: \beta < 0$ The t test with n-2 degrees of freedom: $t = \frac{b - 0}{s.e.(b)}$	- relationship must be linear - the error terms are normally distributed - the errors terms have equal variances - the error terms are independent of each other
Association between categorical variables	relationship between two or more categorical variables	the observed counts in a two-way table	categorical	Is there a relationship between smoking and lung cancer?	H_o: The two variables are not related H_a: The two variables are related The chi-square statistic: $\chi^2 = \sum_{\substack{all \\ cells}} \frac{(Observed - Expected)^2}{Expected}$	- all expected counts must be at least 5
Association between two quantitative variables	ρ	Pearson's r	quantitative	Is there a linear relationship between height and weight of a person?	$H_o: \rho = 0$ $H_a: \rho \neq 0$ $t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$ n-2 degrees of freedom	- relationship must be linear