Assumptions for Inference	And the Conditions That Support or Override Them
Proportions (z)	
• **One sample**	
1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Sample is sufficiently large.	2. Successes and failures each ≥ 10.
• **Two groups**	
1. Groups are independent.	1. (Think about how the data were collected.)
2. Data in each group are independent.	2. Both are SRSs and $n < 10\%$ of populations OR random allocation.
3. Both groups are sufficiently large.	3. Successes and failures each ≥ 10 for both groups.
Means (t)	
• **One Sample** (df $= n - 1$)	
1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Population has a Normal model.	2. Histogram is unimodal and symmetric.*
• **Matched pairs** (df $= n - 1$)	
1. Data are matched.	1. (Think about the design.)
2. Individuals are independent.	2. SRS and $n < 10\%$ OR random allocation.
3. Population of differences is Normal.	3. Histogram of differences is unimodal and symmetric.*
• **Two independent groups** (df from technology)	
1. Groups are independent.	1. (Think about the design.)
2. Data in each group are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Both populations are Normal.	3. Both histograms are unimodal and symmetric.*
Distributions/Association (χ^2)	
• **Goodness of fit** (df = # of cells $- 1$; one variable, one sample compared with population model)	
1. Data are counts.	1. (Are they?)
2. Data in sample are independent.	2. SRS and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.
• **Homogeneity** [df $= (r - 1)(c - 1)$; many groups compared on one variable]	
1. Data are counts.	1. (Are they?)
2. Data in groups are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Groups are sufficiently large.	3. All expected counts ≥ 5.
• **Independence** [df $= (r - 1)(c - 1)$; sample from one population classified on two variables]	
1. Data are counts.	1. (Are they?)
2. Data are independent.	2. SRSs and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.
Regression (t, df $= n - 2$)	
• **Association** between two quantitative variables ($\beta = 0$?)	
1. Form of relationship is linear.	1. Scatterplot looks approximately linear.
2. Errors are independent.	2. No apparent pattern in residuals plot.
3. Variability of errors is constant.	3. Residuals plot has consistent spread.
4. Errors have a Normal model.	4. Histogram of residuals is approximately unimodal and symmetric, or normal probability plot reasonably straight.*
	(*less critical as n increases)

Quick Guide to Inference

Think			Show				Tell?
Inference about?	**One group or two?**	**Procedure**	**Model**	**Parameter**	**Estimate**	**SE**	**Chapter**
Proportions	One sample	1-Proportion z-Interval	z	p	$\hat{p}$	$\sqrt{\frac{\hat{p}\hat{q}}{n}}$	19
		1-Proportion z-Test				$\sqrt{\frac{p_0 q_0}{n}}$	20, 21
	Two independent groups	2-Proportion z-Interval	z	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	22
		2-Proportion z-Test				$\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}, \ \hat{p} = \frac{y_1 + y_2}{n_1 + n_2}$	22
Means	One sample	t-Interval t-Test	t df $= n - 1$	μ	$\bar{y}$	$\frac{s}{\sqrt{n}}$	23
	Two independent groups	2-Sample t-Test 2-Sample t-Interval	t df from technology	$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	24
	Matched pairs	Paired t-Test Paired t-Interval	t df $= n - 1$	μ_d	$\bar{d}$	$\frac{s_d}{\sqrt{n}}$	25
Distributions (one categorical variable)	One sample	Goodness-of-Fit	χ^2 df $= cells - 1$			$\sum \frac{(Obs - Exp)^2}{Exp}$	26
	Many independent groups	Homogeneity χ^2 Test	χ^2 df $= (r - 1)(c - 1)$				
Independence (two categorical variables)	One sample	Independence χ^2 Test					
Association (two quantitative variables)	One sample	Linear Regression t-Test or Confidence Interval for β	t df $= n - 2$	β_1	b_1	$\frac{s_e}{s_x \sqrt{n - 1}}$ (compute with technology)	27
		*Confidence Interval for μ_ν		μ_ν	$\hat{y}_\nu$	$\sqrt{SE^2(b_1) \cdot (x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$	
		*Prediction Interval for y_ν		y_ν	$\hat{y}_\nu$	$\sqrt{SE^2(b_1) \cdot (x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$	
Inference about?	**One group or two?**	**Procedure**	**Model**	**Parameter**	**Estimate**	**SE**	**Chapter**