

# Statistical Organizers

Ruth Wunderlich

Linden Hall (Now!)  
Lititz, Pennsylvania

Best Practices  
Thursday, June 13<sup>th</sup>  
AP Reading 2013

## Initial Descriptive Statistics Organizer (Chapter's 1-3)

#	Type of Variable(s)	Graphical Displays	Verbal Descriptions	Numerical Summaries
1	Categorical			
2	Categorical			
3	Categorical			
1	Quantitative (Discrete)			
1	Quantitative			
1	Categorical (2 values)			
1	Quantitative			
1	Categorical			
2	Quantitative			
1	Quantitative (Continuous)			

## Descriptive Statistics Sheet (with Tests)

#	Type of Variable(s)	Graphical Display	Verbal Descriptions	Numerical Summaries	Statistical Procedure(s)	Sampling Distribution(s)
	two categories only	Pie Chart Bar Graph				
	two categories only	Multiple Bar Graph Two Way Table				
		Histogram Stemplot Boxplot				
	two categories	Back to Back Stemplots				
		Parallel boxplots				
		Scatterplot				

## Descriptive Statistics Sheet (with Tests)

#	Type of Variable(s)	Graphical Display	Verbal Descriptions	Numerical Summaries	Statistical Procedure(s)	Sampling Distribution(s)
1	category	Pie Chart Bar Graph	more than less than twice as many almost the same - but few - make comparisons	%	Chi-Square Goodness of Fit	
	two categories only		same but describe conditional / marginal		frequency / count proportions	
2	category	Multiple Bar Graph	same but describe conditional / marginal		2 sample test for proportion	N: Distribution $Z$
		<u>Two Way Table</u>			Chi-Square test for relationship	Chi-Square distribution
1	quantitative	Histogram Stemplot Boxplot	<ul style="list-style-type: none"> <li>• shape</li> <li>• center</li> <li>• spread</li> <li>• outliers</li> <li>• context</li> </ul>	Mean standard deviation	1-sample test	t-distribution
1	two categories	Back to Back Stemplots	<ul style="list-style-type: none"> <li>• shape</li> <li>• center</li> <li>• spread</li> <li>• outliers</li> <li>• context</li> </ul>	several means several standard deviations	2-sample test	t-distribution
	category quantitative	Parallel boxplots	<ul style="list-style-type: none"> <li>• comparison</li> </ul>		ANOVA (3) or plugin	F-distribution
2	quantitative	Scatterplot	<ul style="list-style-type: none"> <li>• form</li> <li>• direction</li> <li>• strength</li> <li>• context</li> </ul>	correlation slope regression linear $r, r^2$	regression	t-distribution

$$\text{Residual} = \text{observed} - \text{predicted}$$

$$= y - \hat{y}$$

### Inference Formula Sheet

all give  
V or  
10

if either  
or sample  
or you will  
use to find  
0  
100%

variance

30  
70  
10  
10  
at least  
at least

Test	Conditions	Null Hypothesis	Test Statistic	Confidence Interval
One-sample z-test for inference of population means	1. SRS 2. Normal Dist 3. KNOWN pop SD	Ho: $\mu = \mu_0$ Ha: $\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$\bar{X} \pm z^* (\sigma/\sqrt{n})$
One-sample t-test for population means	1. SRS 2. $30 < 15$ roughly symmetric, single peak, no outliers 3. $30 = 15$ can be used except in presence of outliers/skewness 4. $30 > 15$ used w/ clearly skewed dist when sample is large	Ho: $\mu = \mu_0$ Ha: $\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$	$\bar{X} \pm T^* (\sigma/\sqrt{n})$ * n-1 degrees freedom
Paired t-test: One sample t-test on the differences of dependent sample means	1. SRS 2. N pop Dist 3. Do not know pop SD	Ho: $\mu = 0$	$T = \frac{\bar{X} - 0}{s/\sqrt{n}}$	$\bar{X} \pm T^* (s/\sqrt{n})$
Two-sample t-test for inference on population means	1. 2 SRS that are indep 2. Dist have similar shapes and no outliers in data 3. Do not know pop SD OR $\mu$	Ho: $\mu_1 = \mu_2$ Ha: $\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \neq \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$ * smaller DF $n_1 - 1$ or $n_2 - 1$	$(\bar{X}_1 - \bar{X}_2) \pm T^* \left( \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} \right)$
One-Proportion z-test	Use when sample size is so large that both $p_0$ & $n(1-p_0)$ are 10 or more	Ho: $p = p_0$ Ha: $p > p_0$ $p < p_0$ $p \neq p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$\hat{p} = \frac{\# \text{ of successes} / \# \text{ of failures}}{n}$ $\hat{p} \pm z^* \left( \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$ p. 3, 503 $\hat{p} = \frac{\# \text{ of successes in sample} + 2}{n + 4}$
One-Proportion interval plus four procedures for intervals	Use when C level = 90% and sample size is at least 10	$ZP(z)$		
Two-Proportion z-test	Use when the counts of successes & failures are 5 or more in both samples	Ho: $p_1 = p_2$ Ha: $p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$ $\hat{p} = \frac{\# \text{ success in both samples}}{\# \text{ of indivs in both samples}}$	$(\hat{p}_1 - \hat{p}_2) \pm z^* \left( \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right)$ $(\hat{p}_1 - \hat{p}_2) \pm z^* \left( \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 + 2}} \right)$ $\hat{p}_1 \neq \hat{p}_2$ add 1 $n_1 \neq n_2$ add 2
Two-Proportion interval plus four procedures for intervals	Use when sample size is at least 5 in each group with any counts of successes & failures	$ZP(z)$		

$\frac{\hat{p}(1-\hat{p})}{n}$   
\*  
+ 2

$$df = (r-1)(c-1)$$

input info  
stat  
table  
• chi-square goodness of fit  
• chi-square test  
• ok

regression - choose 1<sup>st</sup> opt  
response - y  
predict - x  
option  
Prediction Intervals conf int  o.k.

### Inference Formula Sheet (Page 2)

Test	Conditions	Null Hypothesis	Test Statistic	Confidence Interval
Chi Square test relationship of variables	All cell counts @ least 1 No more than 20% less than 5 SRS	H <sub>0</sub> : There is no relationship H <sub>a</sub> : There is some relationship	Expected count = $\frac{\text{row total} \times \text{column total}}{\text{table total}}$ Chi Sq test stat $\chi^2 = \sum \frac{\text{Observed} - \text{Expected}}{\text{Expected}}$	..... ..... .....
Chi Square Goodness of Fit n=1	All cell counts @ least 1 No more than 20% are less than 5 expected values SRS	H <sub>0</sub> : P <sub>1</sub> = P <sub>2</sub> = P <sub>3</sub> ... H <sub>a</sub> : not all P <sub>i</sub> =	$\chi^2 = \sum \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}}$	..... ..... .....
Lack of Correlation will not be on test	SRS Linear relationship Normal distribution of response with same standard deviation	H <sub>0</sub> : r = 0 H <sub>a</sub> : r ≠ 0	r = test statistic	..... ..... .....
Linear Regression Slope		H <sub>0</sub> : β = 0 $\pm \frac{b}{SE_b}$ H <sub>a</sub> : β > 0 <small>min to b - two sided if you want one sided not possible</small>	$t = \frac{b}{SE_b}$ - SE of slope df = n - 2	$b \pm t^* SE_b$ - next to slope b - slope SE - standard error of n - 2
Linear Regression mean response				$\hat{y} \pm t^* SE_{\hat{\mu}}$ - next to predicted df = n - 2
Linear Regression Prediction				$\hat{y} \pm t^* SE_{\hat{y}}$ - CI - mean PI - prediction df = n - 2
ANOVA will not be on test		Ind. SRS All pops have same SD Normal distribution	All populations have same mean	$F = \frac{MSG}{MSE}$ Mean square for groups Mean square for error

20  
20  
biggest we would expect

$$df = (I-1), (N-I)$$