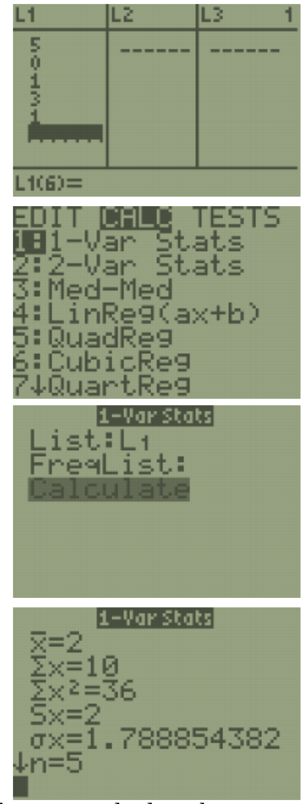
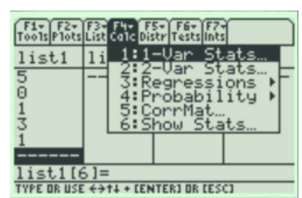
Mostly Harmless Statistics Formula Packet

Chapter 3 Formulas

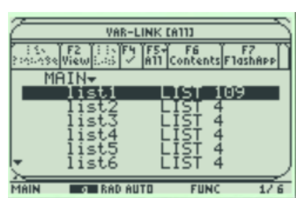


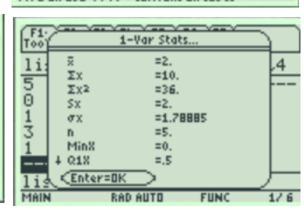
Sample Mean: $\bar{x} = \frac{\sum x}{n}$	**Population Mean:** $\mu = \frac{\sum x}{N}$
Weighted Mean: $\bar{x} = \frac{\sum(xw)}{\sum w}$	**Range** = Max – Min
Sample Standard Deviation: $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$	**Population Standard Deviation** = σ
Sample Variance: $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$	**Population Variance** = σ^2
Coefficient of Variation: CVar = $\left(\frac{s}{\bar{x}} \cdot 100\right)\%$	**Z-Score:** $z = \frac{x-\bar{x}}{s}$
Percentile Index: $i = \frac{(n+1)\cdot p}{100}$	**Interquartile Range:** IQR = $Q_3 - Q_1$
Empirical Rule: z = 1, 2, 3 ⇒ 68%, 95%, 99.7%	**Outlier Lower Limit:** $Q_1 - (1.5 \cdot IQR)$
Chebyshev's Inequality: $\left(\left(1 - \frac{1}{(z)^2}\right) \cdot 100\right)\%$	**Outlier Upper Limit:** $Q_3 + (1.5 \cdot IQR)$

TI-84: Enter the data in a list and then press [STAT]. Use cursor keys to highlight CALC. Press 1 or [ENTER] to select **1:1-Var Stats**. Press [2nd], then press the number key corresponding to your data list. Press [Enter] to calculate the statistics. Note: the calculator always defaults to L_1 if you do not specify a data list. s_x is the sample standard deviation. You can arrow down and find more statistics. Use the min and max to calculate the range by hand. To find the variance simply square the standard deviation.



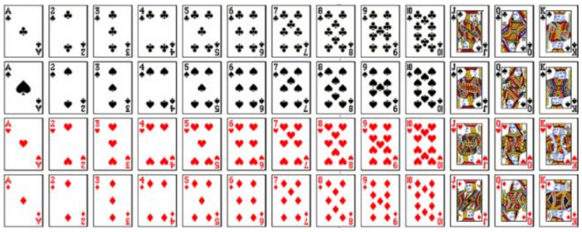
TI-89: Press [APPS], select **FlashApps** then press [ENTER]. Highlight **Stats/List Editor** then press [ENTER]. Press [ENTER] again to select the main folder. To clear a previously stored list of data values, arrow up to the list name you want to clear, press [CLEAR], then press enter. Press [F4], select 1: 1-Var Stats. To get the list name to the List box, press [2^{nd}] [Var-Link], arrow down to list1 and press [Enter]. This will bring list1 to the List box. Press [Enter] to enter the list name and then enter again to calculate. Use the down arrow key to see all the statistics. Sx is the sample standard deviation. You can arrow down and find more statistics. Use the min and max to calculate the range by hand. To find the variance simply square the standard deviation or take the last sum of squares divided by $n - 1$.





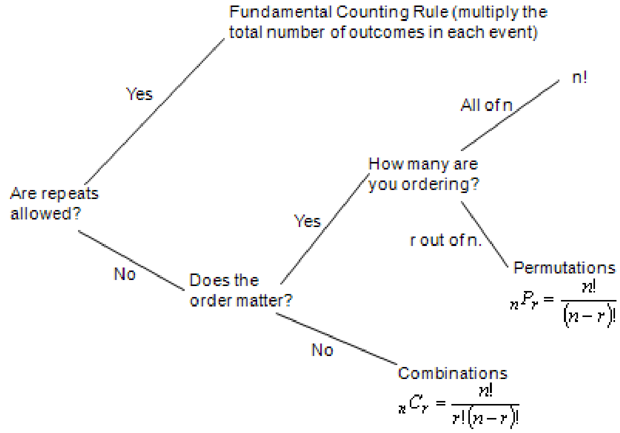
Chapter 4 Formulas

Complement Rules: $P(A) + P(A^C) = 1$ $P(A) = 1 - P(A^C)$ $P(A^C) = 1 - P(A)$	Mutually Exclusive Events: $P(A \cap B) = 0$
Union Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Independent Events: $P(A \cap B) = P(A) \cdot P(B)$
Intersection Rule: $P(A \cap B) = P(A) \cdot P(B\|A)$	Conditional Probability Rule: $P(A\|B) = \frac{P(A \cap B)}{P(B)}$
Fundamental Counting Rule: $m_1 \cdot m_2 \cdots m_n$	Factorial Rule: $n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$
Combination Rule: ${}_nC_r = \frac{n!}{(r!(n-r)!)}$	Permutation Rule: ${}_nP_r = \frac{n!}{(n-r)!}$



clubs = ♣, spades = ♠, hearts = ♥, diamonds = ♦

+		Second Die					
		1	2	3	4	5	6
First Die	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12



Chapter 5 Formulas

Discrete Distribution Table: $0 \le P(x_i) \le 1$ $\sum P(x_i) = 1$	Discrete Distribution Mean: $\mu = \Sigma(x_i \cdot P(x_i))$
Discrete Distribution Variance: $\sigma^2 = \sum(x_i^2 \cdot P(x_i)) - \mu^2$	Discrete Distribution Standard Deviation: $\sigma = \sqrt{\sigma^2}$
Geometric Distribution: $P(X = x) = p \cdot q^{(x-1)}$, $x = 1, 2, 3, \ldots$ Mean: $\mu = \frac{1}{p}$ Variance: $\sigma^2 = \frac{1-p}{p^2}$ Standard Deviation: $\sigma = \sqrt{\frac{1-p}{p^2}}$	Binomial Distribution: $P(X = x) = {}_nC_x \cdot p^x \cdot q^{(n-x)}$, $x = 0, 1, 2, \ldots, n$ Binomial Distribution Mean: $\mu = n \cdot p$ Variance: $\sigma^2 = n \cdot p \cdot q$ Standard Deviation: $\sigma = \sqrt{n \cdot p \cdot q}$
Hypergeometric Distribution: $P(X = x) = \frac{{}_aC_x \cdot {}_bC_{n-x}}{{}_NC_n}$	p = P(success) q = P(failure) = 1 – p n = sample size N = population size
Unit Change for Poisson Distribution: New μ = old $\mu\left(\frac{\text{new units}}{\text{old units}}\right)$	Poisson Distribution: $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$

P($X = x$)	**P($X \le x$)**	**P($X \ge x$)**
Is	Is less than or equal to	Is greater than or equal to
Is equal to	Is at most	Is at least
Is exactly the same as	Is not greater than	Is not less than
Has not changed from	Within	Is more than or equal to
Is the same as		
Excel =binom.dist(x,n,p,0) =HYPGEOM.DIST(x,n,a,N,0) =POISSON.DIST(x,μ,0)	**Excel** =binom.dist(x,n,p,1) = HYPGEOM.DIST(x,n,a,N,1) =POISSON.DIST(x,μ,1)	**Excel** =1-binom.dist(x-1,n,p,1) =1- HYPGEOM.DIST(x-1,n,a,N,1) =1-POISSON.DIST(x–1,μ,1)
TI Calculator geometpdf(p,x) binompdf(n,p,x) poissonpdf(μ,x)	**TI Calculator** binomcdf(n,p,x) poissoncdf(μ,x)	**TI Calculator** 1-binomcdf(n,p,x-1) 1-poissoncdf(μ,x-1)

P($X > x$)	**P($X < x$)**
More than	Less than
Greater than	Below
Above	Lower than
Higher than	Shorter than
Longer than	Smaller than
Bigger than	Decreased
Increased	Reduced
Excel =1-binom.dist(x,n,p,1) =1- HYPGEOM.DIST(x,n,a,N,1) =1-POISSON.DIST(x,μ,1)	**Excel** =binom.dist(x-1,n,p,1) = HYPGEOM.DIST(x-1,n,a,N,1) =POISSON.DIST(x-1,μ,1)
TI Calculator 1-binomcdf(n,p,x) 1-poissoncdf(μ,x)	**TI Calculator** binomcdf(n,p,x-1) poissoncdf(μ,x-1)

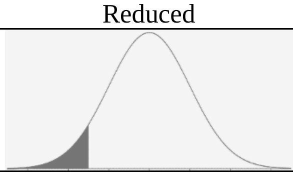
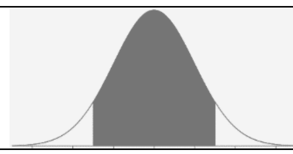
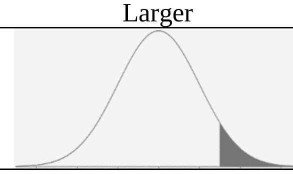
How do you tell them apart?

- Geometric – A percent or proportion is given. There is no set sample size until a success is achieved.
- Binomial – A percent or proportion is given. A sample size is given.
- Hypergeometric – Usually frequencies of successes are given instead of percentages. A sample size is given.
- Poisson – An average or mean is given. There is no set sample size until a success is achieved.

Chapter 6 Formulas

Uniform Distribution: $f(x) = \frac{1}{b-a}$, for $a \le x \le b$ $P(X \ge x) = P(X > x) = \left(\frac{1}{b-a}\right) \cdot (b - x)$ $P(X \le x) = P(X < x) = \left(\frac{1}{b-a}\right) \cdot (x - a)$ $P(x_1 \le X \le x_2) = P(x_1 < X < x_2) = \left(\frac{1}{b-a}\right) \cdot (x_2 - x_1)$	Exponential Distribution: $f(x) = \frac{1}{\mu} e^{(-x/\mu)}$, for $x \ge 0$ $P(X \ge x) = P(X > x) = e^{-x/\mu}$ $P(X \le x) = P(X < x) = 1 - e^{-x/\mu}$ $P(x_1 \le X \le x_2) = P(x_1 < X < x_2) = e^{(-x_1/\mu)} - e^{(-x_2/\mu)}$
Standard Normal Distribution: $\mu = 0$, $\sigma = 1$ z-score: $z = \frac{x-\mu}{\sigma}$ $x = z\sigma + \mu$	Central Limit Theorem: Z-score: $z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

When $\mu = 0$ and $\sigma = 1$ use the NORM.S. DIST or NORM.S.INV function in Excel for a standard normal distribution.

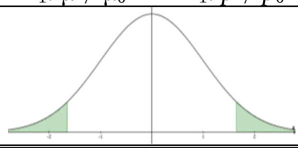
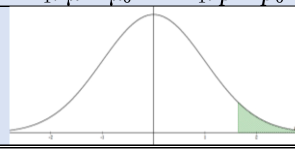
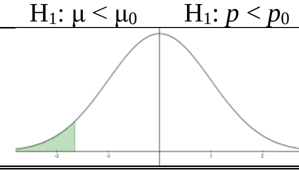
$P(X \le x)$ or $P(X < x)$	**$P(x_1 < X < x_2)$ or $P(x_1 \le X \le x_2)$**	**$P(X \ge x)$ or $P(X > x)$**
Is less than or equal to	Between	Is greater than or equal to
Is at most		Is at least
Is not greater than		Is not less than
Within		More than
Less than		Greater than
Below		Above
Lower than		Higher than
Shorter than		Longer than
Smaller than		Bigger than
Decreased		Increased
Reduced		Larger
		
Excel **Finding a Probability** =NORM.DIST(x,μ,σ,true) **Finding a Percentile** =NORM.INV(area,μ,σ)	**Excel** **Finding a Probability** =NORM.DIST(x_2,μ,σ,true) – NORM.DIST(x_1,μ,σ,true) **Finding a Percentile** x_1 =NORM.INV((1–area)/2,μ,σ) x_2 =NORM.INV(1–((1–area)/2),μ,σ)	**Excel** **Finding a Probability** =1–NORM.DIST(x,μ,σ,true) **Finding a Percentile** =NORM.INV(1–area,μ,σ)
TI Calculator **Finding a Probability** =normalcdf(-1E99,x,μ,σ) **Finding a Percentile** =invNorm(area,μ,σ)	**TI Calculator** **Finding a Probability** =normalcdf(x_1,x_2,μ,σ) **Finding a Percentile** x_1 =invNorm((1–area)/2,μ,σ); x_2 =invNorm(1–((1–area)/2),μ,σ)	**TI Calculator** **Finding a Probability** =normalcdf(x,1E99,μ,σ) **Finding a Percentile** =invNorm(1–area,μ,σ)

Chapter 7 Formulas

Confidence Interval for One Proportion $\hat{p} \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}\hat{q}}{n}\right)}$ $\hat{p} = \frac{x}{n}$ $\hat{q} = 1 - \hat{p}$ TI-84: 1-PropZInt	**Sample Size** Always round up to whole number. **Proportion** **Mean** $n = p^* \cdot q^* \left(\frac{z_{\alpha/2}}{E}\right)^2$ $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$ If p is not given use p* = 0.5. E = Margin of Error
***z*- Confidence Interval for One Mean** Use z-interval when σ is given. TI-84: ZInterval $\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	**t-Confidence Interval for One Mean**, $df = n - 1$; Use t-interval when s is given. If $n < 30$, population needs to be normal. TI-84: TInterval $\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
***Z*-Critical Values** Excel: $z_{\alpha/2}$ =NORM.INV(1–area/2,0,1) TI-84: $z_{\alpha/2}$ = invNorm(1–area/2,0,1)	**t-Critical Values** Excel: $t_{\alpha/2}$ =T.INV(1–area/2,*df*) TI-84: $t_{\alpha/2}$ = invT(1–area/2,*df*)

Chapter 8 Formulas

Z-Test: $z = \frac{\bar{x}-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ TI-84: Z-Test H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$ Use z-test when σ is given.	**t-Test:** $t = \frac{\bar{x}-\mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$ TI-84: T-Test H_0: $\mu = \mu_0$ If $n < 30$, population needs to be normal. H_1: $\mu \neq \mu_0$ Use t-test when s is given.
z-Critical Values Excel: $z_{\alpha/2}$ = NORM.INV(1–α/2,0,1) $z_{1-\alpha}$ = NORM.INV(1–α,0,1) z_{α} = NORM.INV(α,0,1) TI-84: $z_{\alpha/2}$ = invNorm(1–α/2,0,1) $z_{1-\alpha}$ = invNorm(1–α,0,1) z_{α} = invNorm(α,0,1)	**t-Critical Values** Excel: $t_{\alpha/2}$ =T.INV(1–α/2,*df*) $t_{1-\alpha}$ = T.INV(1–α,*df*) t_{α} = T.INV(α,*df*) TI-84: $t_{\alpha/2}$ = invT(1–α/2,*df*) $t_{1-\alpha}$ = invT(1–α,*df*) t_{α} = invT(α,*df*)
Hypothesis Test for One Proportion $z = \frac{\hat{p}-p_0}{\sqrt{\left(\frac{p_0 q_0}{n}\right)}}$ TI-84: 1-PropZTest **Type I Error**- Reject H_0 when H_0 is true. **Type II Error**- Fail to reject H_0 when H_0 is false.	**Rejection Rules:** P-value method: reject H_0 when the p-value $\leq \alpha$. Critical value method: reject H_0 when the test statistic is in the critical region (shaded tails). Confidence Interval method for mean, reject H_0 when the hypothesized value found in H_0 is outside the bounds of the confidence interval.

Two-tailed Test	Right-tailed Test	Left-tailed Test
H_0: $\mu = \mu_0$ or H_0: $p = p_0$ H_1: $\mu \neq \mu_0$ H_1: $p \neq p_0$	H_0: $\mu = \mu_0$ or H_1: $p = p_0$ H_1: $\mu > \mu_0$ H_1: $p > p_0$	H_0: $\mu = \mu_0$ or H_1: $p = p_0$ H_1: $\mu < \mu_0$ H_1: $p < p_0$
		
Claim is in the Null Hypothesis		
=	≤	≥
Is equal to	Is less than or equal to	Is greater than or equal to
Is exactly the same as	Is at most	Is at least
Has not changed from	Is not more than	Is not less than
Is the same as	Within	Is more than or equal to
Claim is in the Alternative Hypothesis		
≠	>	<
Is not	More than	Less than
Is not equal to	Greater than	Below
Is different from	Above	Lower than
Has changed from	Higher than	Shorter than
Is not the same as	Longer than	Smaller than
	Bigger than	Decreased
	Increased	Reduced

Chapter 9 Formulas

Hypothesis Test for Two Dependent Means H_0: $\mu_D = 0$ H_1: $\mu_D \neq 0$ $t = \frac{\bar{D}-\mu_D}{\left(\frac{s_D}{\sqrt{n}}\right)}$ TI-84: T-Test	**Confidence Interval for Two Dependent Means** $\bar{D} \pm t_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right)$ TI-84: TInterval
Hypothesis Test for Two Independent Means Z-Test: H_0: $\mu_1 = \mu_2$ TI-84: 2-SampZTest H_1: $\mu_1 \neq \mu_2$ $z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}}$	**Confidence Interval for Two Independent Means Z-Interval** $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$ TI-84: 2-SampZInt

Hypothesis Test for Two Independent Means

H$_0$: $\mu_1 = \mu_2$ TI-84: 2-SampTTest

H$_1$: $\mu_1 \neq \mu_2$

T-Test: Assume variances are unequal

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\left(\frac{s_1^2}{n_1}\right)^2\left(\frac{1}{n_1-1}\right) + \left(\frac{s_2^2}{n_2}\right)^2\left(\frac{1}{n_2-1}\right)\right)}$$

T-Test: Assume variances are equal

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad df = n_1 - n_2 - 2$$

Confidence Interval for Two Independent Means

TI-84: 2-SampTInt

T-Interval: Assume variances are unequal

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\left(\frac{s_1^2}{n_1}\right)^2\left(\frac{1}{n_1-1}\right) + \left(\frac{s_2^2}{n_2}\right)^2\left(\frac{1}{n_2-1}\right)\right)}$$

T-Interval: Assume variances are unequal

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\left(\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)}$$

$df = n_1 - n_2 - 2$

Hypothesis Test for Two Proportions

H$_0$: $p_1 = p_2$ TI-84: 2-PropZTest

H$_1$: $p_1 \neq p_2$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\left(\hat{p}\cdot\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)}} \qquad \hat{p}_1 = \frac{x_1}{n_1} \qquad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\hat{p} = \frac{(x_1 + x_2)}{(n_1 + n_2)} = \frac{(\hat{p}_1 \cdot n_1 + \hat{p}_2 \cdot n_2)}{(n_1 + n_2)} \qquad \hat{q} = 1 - \hat{p}$$

Confidence Interval for Two Proportions

TI-84: 2-PropZInt

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}\right)}$$

$$\hat{p}_1 = \frac{x_1}{n_1} \qquad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\hat{q}_1 = 1 - \hat{p}_1 \qquad \hat{q}_2 = 1 - \hat{p}_2$$

Hypothesis Test for Two Variances

$H_0: \sigma_1^2 = \sigma_2^2$ TI-84: 2-SampFTest

$H_1: \sigma_1^2 \neq \sigma_2^2$

$F = \frac{s_1^2}{s_2^2}$ dfN = n$_1$ – 1, dfD = n$_2$ – 1

Hypothesis Test for Two Standard Deviations

$H_0: \sigma_1 = \sigma_2$ TI-84: 2-SampFTest

$H_1: \sigma_1 \neq \sigma_2$

$F = \frac{s_1^2}{s_2^2}$ dfN = n$_1$ – 1, dfD = n$_2$ – 1

- Proportions
  - 1 sample →
    $TS: z = \dfrac{\hat{p} - p}{\sqrt{\left(\dfrac{p \cdot q}{n}\right)}}$ H$_0: p = p_0$ 1-PropZTest
    where $\hat{p} = \dfrac{x}{n}, q = 1 - p$
  - 2 samples →
    $TS: z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\left(\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\right)}}$ H$_0: p_1 = p_2$ 2-PropZTest
    2-PropZInt
    $CI: (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\left(\left(\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right) + \left(\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)\right)}$
    $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$
- Means
  - 1 sample
    - σ known → H$_0: \mu = \mu_0$ Z-test $TS: z = \dfrac{\bar{x} - \mu_0}{(\sigma/\sqrt{n})}$
    - σ unknown → $TS: t = \dfrac{\bar{x} - \mu_0}{(s/\sqrt{n})}$, $df = n - 1$ H$_0: \mu = \mu_0$ t-test
  - 2 samples
    - Dependent samples (paired) → H$_0: \mu_D = 0$ t-test $TS: t = \dfrac{\bar{d} - \mu_d}{(s_d/\sqrt{n})}$, $CI: \bar{d} \pm t_{\alpha/2}\left(\dfrac{s_d}{\sqrt{n}}\right)$, $df = n - 1$
    - Independent samples
      - σ's known → H$_0: \mu_1 = \mu_2$ $TS: z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\dfrac{\sigma_1^2}{n_1}\right) + \left(\dfrac{\sigma_2^2}{n_2}\right)}}$ 2-SampZTest $CI: (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)}$ 2-SampZInt
      - σ's unknown
        - Pooled: No → H$_0: \mu_1 = \mu_2$ 2-SampTTest, Pooled: No
          $TS: t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\dfrac{s_1^2}{n_1}\right) + \left(\dfrac{s_2^2}{n_2}\right)}}, df = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\left(\left(\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right)^2\right) + \left(\dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)^2\right)\right)}$
          $CI: (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ 2-SampTInt , Pooled: No
        - Pooled: Yes → TS: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ 2-SampTTest, Pooled: Yes
          2-SampTInt , Pooled: Yes
          H$_0: \mu_1 = \mu_2$ $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\left(\left(\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\right)}$
          $df = n_1 + n_2 - 2$
- 2 Variances or St. Dev → H$_0: \sigma_1^2 = \sigma_2^2$ $TS: F = \dfrac{s_1^2}{s_2^2}$ Ndf = n1 – 1 , Ddf = n2 – 1 2-SampFTest

Chapter 10 Formulas

Goodness of Fit Test	**Test for Independence**
H_0: $p_1 = \hat{p}_1$, $p_2 = \hat{p}_2$, ⋯ $p_k = \hat{p}_k$ H_1: At least one proportion is different. $\chi^2 = \sum \frac{(O-E)^2}{E}$ df = k – 1 TI-84: χ^2 GOF-Test	H_0: Variable 1 and Variable 2 are independent. H_1: Variable 1 and Variable 2 are dependent. $\chi^2 = \sum \frac{(O-E)^2}{E}$ df = (R – 1)(C – 1) TI-84: χ^2-Test

Chapter 11 Formulas

One-Way ANOVA: H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
H_1: At least one mean is different.

Source	SS = Sum of Squares	*df*	MS = Mean Square	F
Between (Factor)	$\sum n_i(\bar{x}_i - \bar{x}_{GM})^2$	k – 1	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
Within (Error)	$\sum(n_i - 1)s_i^2$	N – k	$MSW = \frac{SSW}{N-k}$	
Total	SST	N – 1		

$\bar{x}_i$ = sample mean from the i^{th} group s_i^2 = sample variance from the i^{th} group
n_i = sample size of the i^{th} group $N = n_1 + n_2 + \cdots + n_k$ k = number of groups

Bonferroni test statistic: $t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\left(MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right)}}$ $H_0: \mu_i = \mu_j$ $H_1: \mu_i \neq \mu_j$ Multiply p-value by m = ${}_kC_2$, divide area for critical value by m = ${}_kC_2$.

Two-Way ANOVA:
Row Effect (Factor A): H_0: The row variable has no effect on the average ___________.
H_1: The row variable has an effect on the average ___________.

Column Effect (Factor B): H_0: The column variable has no effect on the average ________.
H_1: The column variable has an effect on the average ________.

Interaction Effect (A×B):
H_0: There is no interaction effect between row variable and column variable on the average ______.
H_1: There is an interaction effect between row variable and column variable on the average ______.

Source	**SS**	***df***	**MS**	**F**
A (row factor)	SS_A	a – 1	$MS_A = \frac{SS_A}{df_A}$	$F_A = \frac{MS_A}{MSE}$
B (column factor)	SS_B	b – 1	$MS_B = \frac{SS_B}{df_B}$	$F_B = \frac{MS_B}{MSE}$
A×B (interaction)	$SS_{A\times B}$	(a – 1) (b – 1)	$MS_{A\times B} = \frac{SS_{A\times B}}{df_{A\times B}}$	$F_{A\times B} = \frac{MS_{A\times B}}{MSE}$
Error (within)	SSE	ab(n – 1)	$MSE = \frac{SSE}{df_E}$	
Total	SST	N – 1		

Chapter 12 Formulas

$SS_{xx} = (n-1)s_x^2$ $SS_{yy} = (n-1)s_y^2$ $SS_{xy} = \sum(xy) - n \cdot \bar{x} \cdot \bar{y}$	**Correlation Coefficient** $r = \frac{SS_{xy}}{\sqrt{(SS_{xx} \cdot SS_{yy})}}$
Slope = $b_1 = \frac{SS_{xy}}{SS_{xx}}$ **y-intercept** = $b_0 = \bar{y} - b_1\bar{x}$ **Regression Equation** (Line of Best Fit): $\hat{y} = b_0 + b_1x$	**Correlation t-test** H_0: $\rho = 0$; H_1: $\rho \neq 0$ $t = r\sqrt{\left(\frac{n-2}{1-r^2}\right)}$ $df = n - 2$ **Slope t-test** H_0: $\beta_1 = 0$; H_1: $\beta_1 \neq 0$ $t = \frac{b_1}{\sqrt{\left(\frac{MSE}{SS_{xx}}\right)}}$ $df = n - p - 1 = n - 2$

Residual

$e_i = y_i - \hat{y}_i$

(Residual plots should have no patterns.)

Standard Error of Estimate

$s_{est} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{MSE}$

Prediction Interval

$\hat{y} \pm t_{\alpha/2} \cdot s_{est}\sqrt{\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{SS_{xx}}\right)}$

Slope/Model F-test

H_0: $\beta_1 = 0$; H_1: $\beta_1 \neq 0$

Source	SS = Sum of Squares	*df*	MS = Mean Square	F
Regression	SSR = $\frac{(SS_{xy})^2}{SS_{xx}}$	p	$MSR = \frac{SSR}{p}$	$F = \frac{MSR}{MSE}$
Error	SSE = SS_{yy} – SSR	*n* – p – 1	$MSE = \frac{SSE}{n-p-1}$	
Total	SST = SS_{yy}	*n* – 1		

Multiple Linear Regression Equation

$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p$

Coefficient of Determination

$R^2 = (r)^2 = \frac{SSR}{SST}$

Model F-Test for Multiple Regression

H_0: $\beta_1 = \beta_2 = \cdots \beta_p = 0$

H_1: At least one slope is not zero.

Adjusted Coefficient of Determination

$R^2_{adj} = 1 - \left(\frac{(1-R^2)(n-1)}{(n-p-1)}\right)$

Chapter 13 Formulas

Ranking Data

- Order the data from smallest to largest.
- The smallest value gets a rank of 1.
- The next smallest gets a rank of 2, etc.
- If there are any values that tie, then each of the tied values gets the average of the corresponding ranks.

Sign Test

H_0: Median = MD_0

H_1: Median ≠ MD_0

p-value uses binomial distribution with p = 0.5 and *n* is the sample size not including ties with the median or differences of 0.

- For a 2-tailed test, the test statistic, *x*, is the smaller of the plus or minus signs. If *x* is the test statistic, the p-value for a two-tailed test is the 2*P($X \leq x$).
- For a right-tailed test, the test statistic, *x*, is the number of plus signs. For a left-tailed test, the test statistic, *x*, is the number of minus signs. The p-value for a one-tailed test is the P($X \geq x$).

Wilcoxon Signed-Rank Critical Values

	1-Tailed α			2-Tailed α		
n	0.01	0.05	0.10	0.01	0.05	0.10
5	-	0	2	-	-	0
6	-	2	3	-	0	2
7	0	3	5	-	2	3
8	1	5	8	0	3	5
9	3	8	10	1	5	8
10	5	10	14	3	8	10
11	7	13	17	5	10	13
12	9	17	21	7	13	17
13	12	21	26	9	17	21
14	15	25	31	12	21	25
15	19	30	36	15	25	30
16	23	35	42	19	29	35
17	27	41	48	23	34	41
18	32	47	55	27	40	47
19	37	53	62	32	46	53
20	43	60	69	37	52	60
21	49	67	77	42	58	67
22	55	75	86	48	65	75
23	62	83	94	54	73	83
24	69	91	104	61	81	91
25	76	100	113	68	89	100
26	84	110	124	75	98	110
27	92	119	134	83	107	119
28	101	130	145	91	116	130
29	110	140	157	100	126	140

Wilcoxon Signed-Rank Test

n is the sample size not including a difference of 0.

When *n* < 30, use test statistic w_s is the absolute value of the smaller of the sum of ranks. CV uses table on next page.

If critical value is not in table then use an online calculator:
http://www.socscistatistics.com/tests/signedranks/

When $n \geq 30$, use z-test statistic: $z = \frac{\left(w_s - \left(\frac{n(n+1)}{4}\right)\right)}{\sqrt{\left(\frac{n(n+1)(2n+1)}{24}\right)}}$

Mann-Whitney U Test

When $n_1 \leq 20$ and $n_2 \leq 20$

$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$, $U_2 = R_2 - \frac{n_2(n_2+1)}{2}$.

$U = \text{Min}(U_1, U_2)$.

CV uses tables below.

If critical value is not in tables then use an online calculator:
https://www.socscistatistics.com/tests/mannwhitney/default.aspx

When $n_1 > 20$ and $n_2 > 20$ use z-test statistic:

$z = \frac{\left(U - \left(\frac{n_1 * n_2}{2}\right)\right)}{\sqrt{\left(\frac{n_1 \cdot n_2(n_1+n_2+1)}{12}\right)}}$

Mann-Whitney U Critical Values

Critical Values for 2-Tailed Mann-Whitney U Test for α = 0.05																			
	n_2																		
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
3	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Critical Values for 2-Tailed Mann-Whitney U Test for α = 0.01																			
	n_2																		
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0
3	-	-	-	-	-	-	-	0	0	0	1	1	1	2	2	2	2	3	3
4	-	-	-	-	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	-	-	-	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	-	-	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	-	-	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	-	-	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	-	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	-	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	-	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	46
12	-	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	-	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	-	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	-	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	-	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	-	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	-	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	0	3	8	13	18	24	30	36	42	46	54	60	67	73	79	86	92	99	105