Chapter 3 Formulas

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

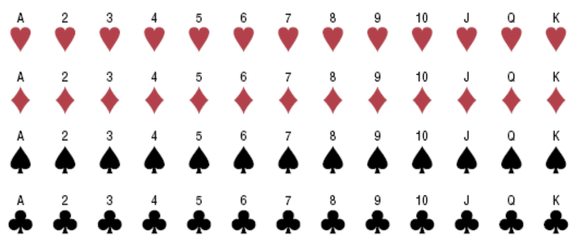
Chapter 4 Formulas

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: $P(A \cap B) = P(A) \cdot P(B \mid A)$	Conditional Probability Rule: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

Sum of Two Dice

		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

Standard Deck of Cards



Face Cards are Jack (J), Queen (Q) and King (K)

Chapter 5 Formulas

Discrete Distribution Table: $0 \le P(x) \le 1$ $\sum P(x) = 1$	Discrete Distribution Mean: $\mu = \sum(x \cdot P(x))$
Discrete Distribution Variance: $\sigma^2 = \sum(x^2 \cdot P(x)) - \mu^2$	Discrete Distribution Standard Deviation: $\sigma = \sqrt{\sigma^2}$
Binomial Distribution: $P(X = x) = {}_nC_x \cdot p^x \cdot q^{n-x}$	Binomial Distribution Mean: $\mu = n \cdot p$ Variance: $\sigma^2 = n \cdot p \cdot q$ Standard Deviation: $\sigma = \sqrt{n \cdot p \cdot q}$
Poisson Distribution: $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$	Unit Change for Poisson Distribution: New μ = old $\mu \left(\frac{\text{new units}}{\text{old units}}\right)$

Signs Are Important For Discrete Distributions!		
P(X=x)	**P(X≤x)**	**P(X≥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 the same as		
=BINOM.DIST(x,n,p,false)	=BINOM.DIST(x,n,p,true)	=1-BINOM.DIST(x–1,n,p,true)
=POISSON.DIST(x,mean,false)	=POISSON.DIST(x,mean,true)	=1–POISSON.DIST(x–1,mean,true)
Where:	**P(X>x)**	**P(X<x)**
x is the value in the	More than	Less than
question that you are	Greater than	Below
finding the probability for.	Above	Lower than
p is the proportion of a	Higher than	Shorter than
success expressed as a	Longer than	Smaller than
decimal between 0 and 1	Bigger than	Decreased
n is the sample size	Increased	Reduced
	Smaller	Larger
The mean has to be rescaled to	=1–BINOM.DIST(x,n,p,true)	=BINOM.DIST(x–1,n,p,true)
the units of the question.	=1–POISSON.DIST(x,mean,true)	=POISSON.DIST(x–1,mean,true)

Common Symbols

n = Sample Size
s^2 = Sample Variance
s = Sample Standard Deviation
p̂ = Sample Proportion

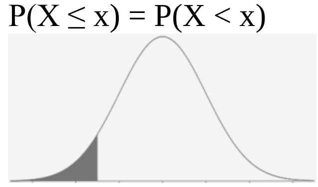
N = Population Size
σ^2 = Population Variance
σ = Population Standard Deviation
p = Population Proportion

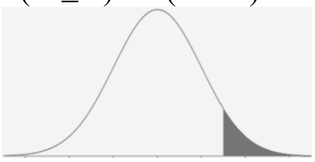
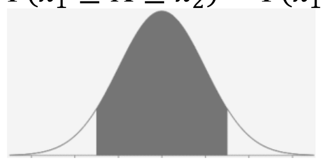
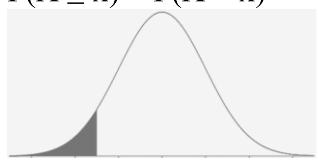
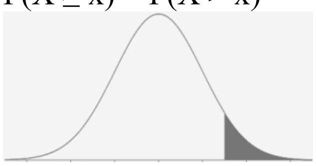
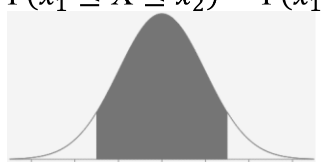
$\bar{x}$ = Sample Mean
μ = Population Mean

Chapter 6 Formulas

Note that for a continuous distribution there is no area at a line under the curve, so ≥ and > will have the same probability and use the same Excel commands.

Note that the NORM.S.DIST and NORM.S.INV functions are for a standard normal when μ=0 and σ=1.

Standard Normal Distribution: μ = 0, σ = 1 Z-score: $z = \frac{x-\mu}{\sigma}$	Central Limit Theorem: Z-score: $z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$
Normal Distribution Probabilities:	P(X ≤ x) = P(X < x)  Excel: =NORM.DIST(x,μ ,σ,true)

$P(X \geq x) = P(X > x)$  Excel: = 1–NORM.DIST(x,μ,σ,true)	$P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)$  Excel: =NORM.DIST(x2,μ,σ,true)-NORM.DIST(x1,μ,σ,true)
Inverse Normal Distribution:	$P(X \leq x) = P(X < x)$  Excel: =NORM.INV(area,μ,σ)
$P(X \geq x) = P(X > x)$  Excel: =NORM.INV(1–area,μ,σ)	$P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)$  Excel: x_1 =NORM.INV(area/2,μ,σ) x_2 =NORM.INV(1–area/2,μ,σ)

Chapter 7 Formulas

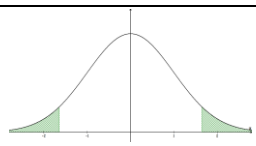
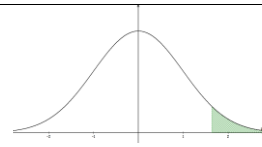
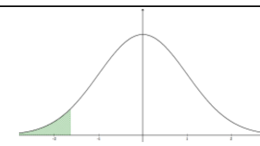
Confidence Interval for One Proportion $\hat{p} \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}\hat{q}}{n}\right)}$ $\hat{p} = \frac{x}{n}$ $\hat{q} = 1 - \hat{p}$	Sample Size for Proportion: $n = p^* \cdot q^* \left(\frac{z_{\alpha/2}}{E}\right)^2$ Always round up to whole number. If p is not given use p*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

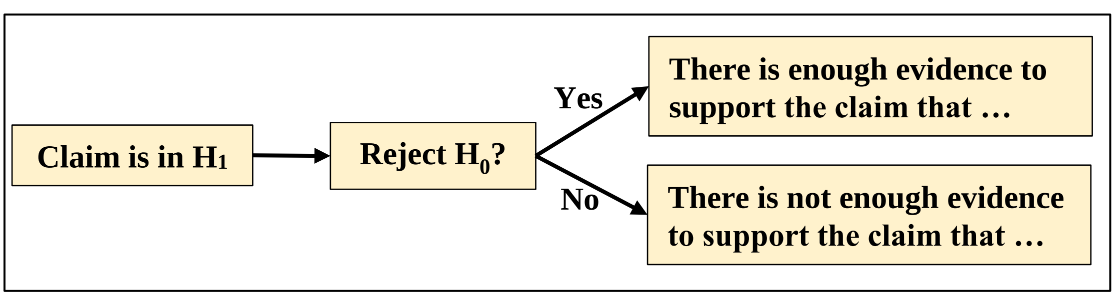
Data > Data Analysis > Descriptive Statistics. Check Confidence Level for Mean to get the margin of error.

Chapter 8 Formulas-Hypothesis Testing

Hypothesis Test for One Mean Use z-test when σ is given. Use t-test when s is given. If n < 30, population needs to be normal.	Type I Error-Reject H_0 when H_0 is true. α = P(Type I error) Type II Error-Fail to reject H_0 when H_0 is false.
Z-Test: $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ $z = \frac{\bar{x}-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}$	T-Test: $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ $t = \frac{\bar{x}-\mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$
Hypothesis Test for One Proportion $H_0: p = p_0$ $H_1: p \neq p_0$ $z = \frac{\hat{p}-p_0}{\sqrt{\left(\frac{p_0 q_0}{n}\right)}}$	The Rejection Rule: • p-value method: reject H_0 when the p-value ≤ α. • Critical value method: reject H_0 when the test statistic is in the critical tail(s).

Look for these key words to help set up your hypotheses:

Two-tailed Test	Right-tailed Test	Left-tailed Test
$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0$	$H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0$
		
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	
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



Finish conclusion with context and units from question.

Chapter 9 Formulas

Hypothesis Test for Two Dependent Means $H_0: \mu_D = 0$ $H_1: \mu_D \neq 0$ $t = \frac{\bar{D} - \mu_D}{\left(\frac{s_D}{\sqrt{n}}\right)}$	Confidence Interval for Two Dependent Means $\bar{D} \pm t_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right)$
Hypothesis Test for Two Independent Means T-Test: Assume variances are unequal $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$	$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\left(\frac{s_1^2}{n_1}\right)^2\left(\frac{1}{n_1 - 1}\right) + \left(\frac{s_2^2}{n_2}\right)^2\left(\frac{1}{n_2 - 1}\right)\right)}$

Chapter 10 Formulas

ANOVA Single Factor k = #of groups, N = total of all n's

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

H_1: At least one mean is different. CV: Always a right-tailed F, use Excel =F.INV.RT(α, df_B, df_W)

ANOVA Table

Source	SS	df	MS	F (Test Statistic)
Between (Treatment or Factor)	$SS_B = \Sigma n(\bar{x} - \bar{x}_{GM})^2$	k–1	$MS_B = SS_B / df_B$	$F = MS_B / MS_W$
Within (Error)	$SS_W = \Sigma(n-1)s^2$	N–k	$MS_W = SS_W / df_W$	
Total	SS_T	N–1		

Chapter 11 Formulas-Correlation and Regression

$SS_{xx} = (n-1)s_x^2$ $SS_{yy} = (n-1)s_y^2$ $SS_{xy} = \sum(xy) - n \cdot \bar{x} \cdot \bar{y}$	Sample Correlation Coefficient $r = \frac{SS_{xy}}{\sqrt{(SS_{xx} \cdot SS_{yy})}}$
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	Regression Equation (Line of Best Fit) $\hat{y} = b_0 + b_1 x$ Slope = $b_1 = \frac{SS_{xy}}{SS_{xx}}$ y-Intercept = $b_0 = \bar{y} - b_1\bar{x}$
Standard Error of Estimate $s = s_{est} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{MSE}$	Residual $e_i = y_i - \hat{y}_i$
Prediction Interval $\hat{y} \pm t_{\alpha/2} \cdot s\sqrt{\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{SS_{xx}}\right)}$	Coefficient of Variation $R^2 = (r)^2 = \frac{SSR}{SST}$