Hypothesis Testing:

Means and proportions:
T-Test
2-sample T-test
1 prop-z-test
2-prop T-test
Matched pairs T-test

Chi^2 Test:
Goodness of Fit
Test of Independence
Test of Homogeneity

Confidence Intervals
compute using calculator
Margin of Error
Standard Error
Understanding changes with changes in sample size

Sampling Distribution:
\bar{X} is approximately normal
mean of the means is the mean
standard error is the standard deviation of the sampling distribution

X ~ B(n,p)
binomial random variables
using binompdf(n,p,k), binomcdf(n,p,k), 1-binomcdf(n,p,k)

X~U(a,b)

X~N(\mu,\sigma)

X~Exp(m)

Chapter 10: Problems 124-133.

Chapter 1: 1.1, 1.2, 1.3
variables: categorical and quantitative,
quant: variables: discrete and continuous
scale levels: nominal v. ordinal and ratio v. interval
frequency table, relative freq, cumulative relative frequency

Chapter 2: 2.1-2.7
Visualizations: bar graphs, pie charts, histograms*, line graphs, stem-leaf plots, box-whisker^, time series grpahs
* -- arguably best for quant. Data, ^ -- biological sciences use this frquently
skew left (right), symmetric,
percentiles, quartiles,
5 number summary, range, min, max, mean, median ,mode,
spread versus center. standard deviation and variance
IQR, outliers
dilations, translations

Ch. 3: 3.1-3.7
Definition of experiment and its outcomes and sample space S, fair experiment means equally likely outcomes. A discrete experiment if there are finite number of outcomes. The size of the sample space is given by $|S|$.
An event is a collection of outcomes. We usually use letters like A, B E for events.
The ***probability*** of an event is $|E|/|S|$… this is a ratio and we write P(E) for probability of an event.
$0 \leq P(E) \leq 1$.
Complex/Compound events: A AND B, A OR B, NOT A = A'.
$P(A')=1-P(A)$, equivalently $P(A)+P(A')=1$.
P(A OR B) =P(A)+P(B)-P(A AND B).
Tree diagrams, contingency tables, and Venn diagrams.
Examples of experiments: i) flipping a coin S={H,T}, ii) rolling a 6-sided die S={1,2,3,4,5,6}
Conditional Probabilities $P(A|B)$ is read as "the probality of A given that B has occurred".
$P(A|B)$ = P(A AND B) / P(B) = |A AND B| / |B|.
In general, $P(A|B) \neq P(B|A)$.
Two events A,B are ***mutually exclusive*** or disjoint if A AND B = Ø (the empty set).
Two events are ***independent*** if $P(A|B)=P(A)$. (Notice that this implies that $P(B|A)=P(A|B)$…. But not conversely.)

Ch. 4: 4.1-4.3
Discussed Discrete random variable. A random variable (r.v.) is a function whose domain is the sample space of an experiment. We denote the r.v. by X. If the output of X is a collection of integers, then X is said to be ***discrete***.

PDF stands for probability distribution function. For discrete r.v. X, it is a table with a row/column of outputs with a separate row/column for the probabilities.

E.g. of a PDF

x	*P*(*x*)
0	0.01
1	0.04
2	0.15
3	0.80

We denote the mean of a r.v. by E(X) and call this the ***expected value*** of X.
E(X)=Σ x p(x) and \sigma = Σ $(x-\mu)^2$ p(x).

The nicest example of a discrete random variable is the binomial distribution: X ~ B(*n*,p)
whereX counts the number of successes in *n* trials and the probability of a success on a given trial is p.

If X~B(n,p), then E(X) = np and \sigma = $\sqrt{np(1-p)}$

P(X=k) = binompdf(n,p,k)
P(X≤k) = binomcdf(n,p,k)
P(x≥k) = 1- binomcdf(n,p,k-1)

Ch. 5: 5.1-5.2
Continuous random variables. Main difference is that we calculate probabilities in terms of area.
P(E)= area of event E / total area.

Areas of circles and rectangles. The latter is said to be modeled by a uniform distribution. X~U(a,b). This means that all of the values between a and b are equally likely.
The graph of the function is f(x)=1/(b-a).
P(c<x<d) = (d-c)/ (b-a).
E(X) = (a+b)/2.

Ch. 6: 6.1-6.4
Normal Distribution. A bell-shaped curve is unimodal and symmetric centered at the mean μ and has a standard deviation of σ.
X~N(μ,σ)
Emperical rule is 68%,95%,99.7%: 65% of the values are within 1 standard deviation, 95% are within 2 standard deviations.
normalcdf(lower,upper, μ,σ) = probability that the value is between the lower and upper given values.
normalcdf(lower,x, μ,σ) = the percentile of the value x.
E.g. SAT Math scores are approximately normally distributed: X~N(500,100). So about 68% of the scores are between 400 and 600. A score of 600 is at about the 84%-tile.

Ch. 7: 7.1, 7.3
Central Limit Theorem
Given a r.v. X on a population not necessarily normal, the sampling distribution is $\bar{X}$ is normally distributed and $\bar{X}$ ~ N(μ, σ/$\sqrt{n}$).

The mean of the mean is the mean.

The standard error is SE = σ/$\sqrt{n}$.

Most of the time you will not be given σ so use the sample standard deviation s_X.

Ch. 8: 8.1, 8.2, 8.3

We create confidence intervals CI around our sample mean based on a confidence level CL. They have the form $\bar{x}$ ± ME or ($\bar{x}$ - ME, $\bar{x}$ + ME) where ME is the margin of error.

For population means ME = z_* (σ /$\sqrt{n}$) or better use ME = t_* (s_X/$\sqrt{n}$)
Here z_* and $_{t^*}$ are called the critical value and they depend on the CL.
z_* = invnovrm((1+CL)/2, 0, 1)

For t_* use the provided t-table knowing that the degrees of freedom df = n-1.

The larger the CL the wider the CI.
Also, increasing sample size while keeping CL fixed makes the ME smaller.
(H) Be able to compute minimum sample sizes to ensure certain ME.

For population proportion use that the standard error is SE = $\sqrt{\frac{p\prime(1-p\prime)}{n}}$ and ME = z_* SE, where p' is the sample proportion.