Math 163 Objective by Chapter

Displaying Distributions, Chapter 1

1. Identify the individuals and variables in a set of data.
2. Identify each variable as categorical or quantitative. Identify the units in which each quantitative variable is measured.
3. Recognize when a pie chart can and cannot be used.
4. Make a bar graph of the distribution of a categorical variable, or in general to compare related quantities.
5. Interpret pie charts and bar graphs.
6. Make a histogram of the distribution of a quantitative variable.
7. Make a stemplot of the distribution of a small set of observations.
8. Look for the overall pattern and for major deviations from the pattern.
9. Assess from a histogram or stemplot whether the shape of a distribution is roughly symmetric, distinctly skewed, or neither.
10. Recognize outliers and give plausible explanation for them.

Describing Distributions (Quantitative Variable), Chapter 2

1. Describe the overall pattern by giving numerical measures of center and spread in addition to a verbal description of shape.
2. Decide which measures of center and spread are more appropriate: the mean and standard deviation (especially for symmetric distributions) or the five-number summary (especially for skewed distributions).
3. Find the median $M$ and the quartiles $Q_1$ and $Q_3$ for a set of observations.
4. Find the five-number summary and draw a boxplot; assess center, spread, symmetry, and skewness from a boxplot.
5. Find the mean $\bar{x}$ and the standard deviation $s$ for a set of observations.
6. Understand that the median is more resistant than the mean. Recognize that skewness in a distribution moves the mean away from the median toward the long tail.
7. Know the basic properties of the standard deviation: $s \geq 0$ always; $s = 0$ only when all observations are identical and increases as the spread increases; $s$ has the same units as the original measurements; $s$ is pulled strongly up by outliers or skewness.

Curves and Normal Distributions, Chapter 3

1. Know that areas under a density curve represent proportions of all observations and that the total area under a density curve is 1.
2. Approximately locate the median (equal-areas point) and the mean (balance point) on a density curve.
3. Know that the mean and median both lie at the center of a symmetric density curve and that the mean moves farther toward the long tail of a skewed curve.
4. Recognize the shape of Normal curves and estimate by eye both the mean and standard deviation from such a curve.
5. Use the 68–95–99.7 rule and symmetry to state what percent of the observations from a Normal distribution fall between two points when both points lie at the mean or one, two, or three standard deviations on either side of the mean.

6. Find the standardized value (z-score) of an observation. Interpret z-scores and understand that any Normal distribution becomes standard Normal $N(0, 1)$ when standardized.

7. Given that a variable has a Normal distribution with a stated mean $\mu$ and standard deviation $\sigma$, calculate the proportion of values above a stated number, below a stated number, or between two stated numbers.

8. Given that a variable has a Normal distribution with a stated mean $\mu$ and standard deviation $\sigma$, calculate the point having a stated proportion of all values above it or below it.

**Scatterplots and Correlation, Chapter 4**

1. Identify the explanatory and response variables in situations where one variable explains or influences another.

2. Make a scatterplot to display the relationship between two quantitative variables measured on the same subjects. Place the explanatory variable (if any) on the horizontal scale of the plot.

3. Add a categorical variable to a scatterplot by using a different plotting symbol or color.

4. Describe the direction, form, and strength of the overall pattern of a scatterplot. In particular, recognize positive or negative association and linear (straight-line) patterns. Recognize outliers in a scatterplot.

5. Judge whether it is appropriate to use correlation to describe the relationship between two quantitative variables. Find the correlation $r$.

6. Know the basic properties of correlation: $r$ measures the direction and strength of only straight-line relationships; $r$ is always a number between $-1$ and $1$; $r = \pm 1$ only for perfect straight-line relationships; $r$ moves away from 0 toward $\pm 1$ as the straight-line relationship gets stronger.

Note: $r^2$ does not determine the association between variables. Students have a difficult time understanding what the residual plot is and what it tells you. How do you know if you have a “good” linear model? Look at the scatterplot, $r$, $r^2$, and residual plot.

Students have a difficult time interpreting specifically what the slope means and the units associated with the slope.

**Regression, Chapter 5**
1. Understand that regression requires an explanatory variable and a response variable. Use a calculator or software to find the least-squares regression line of a response variable $y$ on an explanatory variable $x$ from data.
2. Explain what the slope $b$ and the intercept $a$ mean in the equation $\hat{y} = a + bx$ of a regression line.
3. Draw a graph of a regression line when you are given its equation.
4. Use a regression line to predict $y$ for a given $x$. Recognize extrapolation and be aware of its dangers.
5. Find the slope and intercept of the least-squares regression line from the means and standard deviations of $x$ and $y$ and their correlation.
6. Use $r^2$, the square of the correlation, to describe how much of the variation in one variable can be accounted for by a straight-line relationship with another variable.
7. Recognize outliers and potentially influential observations from a scatterplot with the regression line drawn on it.
8. Calculate the residuals and plot them against the explanatory variable $x$. Recognize that a residual plot magnifies the pattern of the scatterplot of $y$ versus $x$.
9. Understand that both $r$ and the least-squares regression line can be strongly influenced by a few extreme observations.
10. Recognize possible lurking variables that may explain the observed association between two variable $x$ and $y$.
11. Understand that even a strong correlation does not mean that there is a cause-and-effect relationship between $x$ and $y$.
12. Give plausible explanations for an observed association between two variables: direct cause and effect, the influence of lurking variables, or both.

**Two-Way Tables, Chapter 6**

1. From a two-way table of counts, find the marginal distributions of both variables by obtaining the row sums and column sums.
2. Express any distribution in percents by dividing the category counts by their total.
3. Describe the relationship between two categorical variables by computing and comparing percents. Often this involves comparing the conditional distributions of one variable for the different categories of the other variable.
4. Recognize Simpson’s paradox and be able to explain it.

**Producing Data: Sampling, Chapter 8**

1. Identify the population in a sampling situation.
2. Recognize bias due to voluntary response samples and other inferior sampling methods.
3. Use software or Table B of random digits to select a simple random sample (SRS) from a population.
4. Recognize the presence of undercoverage and nonresponse as sources of error in a sample survey. Recognize the effect of the wording of questions on the response.
5. Use random digits to select a stratified random sample from a population when the strata are identified.
6. Recognize when stratified random sampling would provide an advantage over simple random sampling.

**Producing Data: Experiments, Chapter 9**

1. Distinguish between an observational study and an experiment.
2. Recognize the effect of confounding of an explanatory variable with a lurking variable.
3. Recognize that experiments are superior to observational studies in establishing causal relationships between the response and explanatory variables because assignment to groups (randomly) maximizes the chance that the groups being compared are alike in all ways except for the imposed difference in explanatory variable level.
4. Identify the factors (explanatory variables), treatments, response variables, and individuals or subjects in an experiment.
5. Outline the design of a completely randomized experiment using a diagram like that in Figure 9.4. The diagram in a specific case should show the sizes of the groups, the specific treatments, and the response variable.
6. Use software or Table B of random digits to carry out the random assignment of subjects to groups in a completely randomized experiment.
7. Recognize the placebo effect and when blinding or double-blinding should be used.
8. Explain why randomized comparative experiments can give good evidence for cause-and-effect relationships.
10. Understand the role of blocking in reducing the effect of an extraneous source of variation in a problem. If possible, demonstrate that blocking in an experiment is analogous to “controlling” for a known lurking variable in an observational study.
11. Understand that a matched pairs design is a kind of randomized block experiment in which we block over subjects to reduce the impact of subject variation.
12. Be able to outline a block design and a matched pairs experiment.

**Introducing Probability, Chapter 10**

1. Recognize that some phenomena are random. Probability describes the long-run regularity of random phenomena.
2. Understand that the probability of an event is the proportion of times the event occurs in very many repetitions of a random phenomenon. Use the idea of probability as long-run proportion to think about probability.
3. Use basic probability rules to detect illegitimate assignments of probability: any probability must be a number between 0 and 1, and the total probability assigned to all possible outcomes must be 1.
4. Use basic probability rules to find the probabilities of events that are formed from other events. The probability that an event does not occur is 1 minus its probability of occurring. If two events are disjoint, the probability that one or the other occurs is the sum of their individual probabilities.

5. Find probabilities in a finite probability model by adding the probabilities of their outcomes. Find probabilities in a continuous probability model as areas under a density curve.

6. Use the notation of random variables to make compact statements about random outcomes, such as \( P(\bar{x} \leq 4) = 0.3 \). Be able to interpret such statements.

**Sampling Distributions, Chapter 11**

1. Identify parameters and statistics in a sample or experiment.
2. Recognize the fact of sampling variability: a statistic will take different values when you repeat a sample or experiment.
3. Interpret a sampling distribution as describing the values taken by a statistic in all possible repetitions of a sample or experiment under the same conditions.
4. Interpret the sampling distribution of a statistic as describing the probabilities of its possible values.
5. Recognize when a problem involves the mean \( \bar{x} \) of a sample. Understand that \( \bar{x} \) estimates the mean \( \mu \) of the population from which the sample is drawn.
6. Use the law of large numbers to describe the behavior of \( \bar{x} \) as the size of the sample increases.
7. Find the mean and standard deviation of a sample mean \( \bar{x} \) from an SRS of size \( n \) when the mean \( \mu \) and standard deviation \( \sigma \) of the population are known.
8. Understand that \( \bar{x} \) is an unbiased estimator of \( \mu \) and that the variability of \( \bar{x} \) about its mean \( \mu \) gets smaller as the sample size increases.
9. Understand that \( \bar{x} \) has approximately a Normal distribution when the sample is large (central limit theorem). Use this Normal distribution to calculate probabilities that concern \( \bar{x} \).

**General Rules of Probability, Chapter 12**

1. Use Venn diagrams to picture relationships among several events.
2. Use the general addition rule to find probabilities that involve overlapping events.
3. Understand the idea of independence. Judge when it is reasonable to assume independence as part of a probability model.
4. Use the multiplication rule for independent events to find the probability that all of several independent events occur.
5. Use the multiplication rule for independent events in combination with other probability rules to find the probabilities of complex events.
6. Understand the idea of conditional probability. Find conditional probabilities for individuals chosen at random from a table of counts of possible outcomes.

7. Use the general multiplication rule to find \( P(A \text{ and } B) \) from \( P(A) \) and the conditional probability \( P(B \mid A) \).

8. Use tree diagrams to organize several-stage probability models.

**Binomial Distributions, Chapter 13**

1. Recognize the binomial setting: a fixed number \( n \) of independent success-failure trials with the same probability \( p \) of success on each trial.
2. Recognize and use the binomial distribution of the count of successes in a binomial setting.
3. Use the binomial probability formula to find the probabilities of events involving the count \( X \) of successes in a binomial setting for small values of \( n \).
4. Find the mean and standard deviation of a binomial count \( X \).
5. Recognize when you can use the Normal approximation to a binomial distribution. Use the Normal approximation to calculate probabilities that concern a binomial count \( X \).

**Confidence Intervals: The Basics, Chapter 14**

1. State in nontechnical language what is meant by “95% confidence” or other statements of confidence in statistical reports.
2. Know the four-step process for any confidence interval.
3. Calculate a confidence interval for the mean \( \mu \) of a Normal population with known standard deviation \( \sigma \), using the formula, \( \bar{x} \pm z\sigma/\sqrt{n} \).
4. Understand how the margin of error of a confidence interval changes with the sample size and the level of confidence \( C \).
5. Find the sample size required to obtain a confidence interval of specified margin of error \( m \) when the confidence level and other information are given.

**Tests of Significance: The Basics, Chapter 15**

1. State the null and alternative hypotheses in a testing situation when the parameter in question is a population mean \( \mu \).
2. Explain in nontechnical language the meaning of the \( P \)-value when you are given the numerical value of \( P \) for a test.
4. Calculate the one-sample \( z \) test statistic and the \( P \)-value for both one-sided and two-sided tests about the mean \( \mu \) of a Normal population.
5. Assess statistical significance at standard levels \( \alpha \), either by comparing \( P \) with \( \alpha \) or by comparing \( z \) with standard Normal critical values.

**Inference in Practice- Chapter 16**
1. Identify sources of error in a study that are not included in the margin of error of a confidence interval, such as undercoverage or nonresponse.
2. Recognize that significance testing does not measure the size or importance of an effect. Explain why a small effect can be significant in a large sample and why a large effect can fail to be significant in a small sample.
3. Recognize that any inference procedure acts as if the data were properly produced. The \( z \) confidence interval and test require that the data be an SRS from the population.

Students should be able to:

1. Recognize when a problem requires inference about population means (quantitative response variable) or population proportions (usually categorical response variable).
2. Recognize from the design of a study whether one-sample, matched pairs, or two-sample procedures are needed.
3. Based on recognizing the problem setting, choose among the one- and two-sample \( t \) procedures for means and the one-and two-sample \( z \) procedures for proportions.

**Inference about a Population Mean - Chapter 18**

1. Verify that the \( t \) procedures are appropriate in a particular setting. Check the study design and the distribution of the data and take advantage of robustness against lack of Normality.
2. Recognize when poor study design, outliers, or a small sample from a skewed distribution make the \( t \) procedures risky.
3. Use the one-sample \( t \) procedure to obtain a confidence interval at a stated level of confidence for the mean \( \mu \) of a population.
4. Carry out a one-sample \( t \) test for the hypothesis that a population mean \( \mu \) has a specified value against either a one-sided or a two-sided alternative. Use software to find the \( P \)-value or Table C to get an approximate value.
5. Recognize matched pairs data and use the \( t \) procedures to obtain confidence intervals and to perform tests of significance for such data.

**Two-Sample Problems - Chapter 19**

1. Verify that the two-sample \( t \) procedures are appropriate in a particular setting. Check the study design and the distribution of the data and take advantage of robustness against lack of Normality.
2. Give a confidence interval for the difference between two means. Use software if you have it. Use the two-sample \( t \) statistic with conservative degrees of freedom and Table C if you do not have statistical software.
3. Test the hypothesis that two populations have equal means against either a one-sided or a two-sided alternative. Use software if you have it. Use the two-sample $t$ test with conservative degrees of freedom and Table C if you do not have statistical software.

4. Know that procedures for comparing the standard deviations of two Normal populations are available, but that these procedures are risky because they are not at all robust against non-Normal distributions.

**Inference about a Population Proportion- chapter 20**

1. Verify that you can safely use either the large-sample or the plus four $z$ procedure in a particular setting. Check the study design and the guidelines for sample size.

2. Use the large-sample $z$ procedure to give a confidence interval for a population proportion $p$. Understand that the true confidence level may be substantially less than you ask for unless the sample is very large and the true $p$ is not close to 0 or 1.

3. Use the plus four modification of the $z$ procedure to give a confidence interval for $p$ that is accurate even for small samples and for any value of $p$.

4. Use the $z$ statistic to carry out a test of significance for the hypothesis $H_0: p = p_0$ about a population proportion $p$ against either a one-sided or a two-sided alternative. Use software or Table A to find the $P$-value, or Table C to get an approximate value.

**Comparing Two Proportions, Chapter 21**

1. Verify that you can safely use either the large-sample or the plus four $z$ procedures in a particular setting. Check the study design and the guidelines for sample size.

2. Use the large-sample $z$ procedure to give a confidence interval for the difference $p_1 - p_2$ between proportions in two populations based on independent samples from the populations. Understand that the true confidence level may be less than you ask for unless the samples are quite large.

3. Use the plus four modification of the $z$ procedure to give a confidence interval for $p_1 - p_2$ that is accurate even for very small samples and for any values of $p_1$ and $p_2$.

4. Use a $z$ statistic to test the hypothesis $H_0: p_1 = p_2$ that the proportions in two distinct populations are equal. Use software or Table A to find the $P$-value, or Table C to get an approximate value.

**Two Categorical variables: The Chi-Square Test, Chapter 22**

1. Understand that the data for a chi-square test must be presented as a two-way table of counts of outcomes.

2. Use percents to describe the relationship between any two categorical variables, starting from the counts in a two-way table.
3. Locate the chi-square statistic, its $P$-value, and other useful facts (row or column percents, expected counts, terms of chi-square) in output from your software or calculator.
4. Use the expected counts to check whether you can safely use the chi-square test.
5. Explain what null hypothesis the chi-square statistic tests in a specific two-way table.
6. If the test is significant, compare percents, compare observed with expected cell counts, or look for the largest terms of the chi-square statistic to see what deviations from the null hypothesis are most important.
7. Calculate the expected count for any cell from the observed counts in a two-way table. Check whether you can safely use the chi-square test.
8. Calculate the term of the chi-square statistic for any cell, as well as the overall statistic.
9. Give the degrees of freedom of a chi-square statistic. Make a quick assessment of the significance of the statistic by comparing the observed value with the degrees of freedom.
10. Use the chi-square critical values in Table D to approximate the $P$-value of a chi-square test.