1. A businessman interested in purchasing a laundry in a small town wants to know what percentage of households there would use a pickup and delivery service. He would need 40% participation to go ahead with his plan. He mails out questionnaires to all 2000 homes in the town and gets 310 responses. Can he use the percentage of the 310 indicating interest as a reasonable estimate of likely participation? What specific concerns would you have with this procedure? Keep your answer brief and specific.

2. Consider the following sample data: 20.4, 15.4, 9.1, 17.7, 23.3, 14.6, 20.6, 10.1

   a) Find the sample mean $\bar{x}$

   b) Find the sample standard deviation $s$.

   c) Find the median.
3. Consider the following frequency table for a sample of quiz scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

a) Find the mean $\bar{x}$ of the sample of quiz scores.

b) Find the sample standard deviation $s$ of the sample of quiz scores.

c) Find the median of the sample of quiz scores.
4. Consider this density curve for a population of data, and the points A,B,C on the horizontal axis. (Think of this curve as a very fine histogram of the data).

Of the points A,B,C one is the population mean, one is the population median and one is the first quartile $Q_1$. Tell which of A,B,C is the mean, which is the median, and which is the first quartile and (important!) explain your reasoning.
5. Below are scatterplots for four datasets, each involving variables $x, y$. The correlations for the four datasets—NOT IN ORDER—are $r = -0.953, r = -0.668, r = -0.078, r = 0.673$. Match the correlations to the scatterplots—i.e. tell which scatterplot has which of these correlations. Briefly explain each choice.
6. Suppose have some mound shaped, or roughly normal data and we find $\bar{x}, s$.
   
a) Roughly what percent of the observations do we expect to find in the interval $\bar{x} \pm 2s$? Why?

   b) Roughly what percent of the observations do we expect to find in the interval $\bar{x} \pm s$? Why?

7. We have two events $A, B$ with $P(A) = .5, P(B) = .3$. In each case below find $P(A \cup B)$.
   
   (a) $A, B$ are disjoint i.e. mutually exclusive i.e. $A \cap B = \phi$.

   (b) $P(A \cap B) = .2$.

   (c) $A, B$ are independent.

   (d) $P(A | B) = 1/3$. 
8. Here is a breakdown of a group by gender and smoking status.

<table>
<thead>
<tr>
<th></th>
<th>Smoker</th>
<th>Non-smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>75</td>
</tr>
</tbody>
</table>

We choose a group member at random. In answering the following questions, leave the arithmetic undone--leave all answers as unsimplified fractions.

(a) What is the probability we get a female smoker?

(b) What is the conditional probability we get a smoker, given that we get a female?

(c) What is the conditional probability we get a female given that we get a smoker?

9. Historical experience shows that 10% of all class C loans wind up in default. Assume that with respect to default, all class C loans are independent of one another.

(a) A small savings and loan bank makes 20 class C loans. What is the probability that no more than 3 wind up in default?

(b) A mammoth financial institution makes 100,000 class C loans. Let $X$ the number of these loans that wind up in default. What are the expected value and standard deviation of $X$?

(c) Consider the loans in (b). What is the probability that no more than 10,100 of these class C loans wind up in default?
10. A child’s deck of cards contains 8 tiger cards and 10 elephant cards. If we deal out 5 of these cards at random, what is the probability we get exactly 3 tiger cards?

11. A normal population of measurements has mean $\mu = 75$ and standard deviation $\sigma = 12$.
   (a) What fraction of the population falls between 70 and 83?
   (b) What is the 60th percentile of this population?

12. A factory has two independent production lines. Daily production for line 1 is a random variable $X_1$ with mean 1000 and standard deviation 100. Daily production for line 2 is a random variable $X_2$ with mean 1500 and standard deviation 120. What are the mean and standard deviation of the total daily production $T = X_1 + X_2$?
13. A normal population has mean $\mu = 200$ and standard deviation $\sigma = 45$. We plan to take a sample of 15 observations from this population. What is the probability the sample mean $\bar{x}$ will exceed 208?

14. Two normal populations have means $\mu_1 = 95, \mu_2 = 80$ and standard deviations $\sigma_1 = 22, \sigma_2 = 30$ respectively. We plan to take random samples of sizes $n_1 = 40, n_2 = 25$ respectively. What is the probability the difference $\bar{x}_1 - \bar{x}_2$ of the two sample means will be less than 20?

15. Here is a sample of 7 observations from a normal population:

40.7, 30.3, 45.8, 31.9, 35.9, 40.5, 44.4

Give a 95% confidence interval for the population mean $\mu$. 
16. We take a random sample of 800 Louisiana adults and find that 475 of them give the governor a good or excellent job rating. Give a 95% confidence interval for the proportion \( p \) of all Louisiana adults who give the governor a good or excellent job rating.

17. We want a 99% confidence interval for the mean \( \mu \) of a large population. We believe that the standard deviation \( \sigma = 25 \). If we want our margin of error (the plus and minus term in the confidence interval) to be no larger than 4, what is the minimum sample size we need?

18. We want a 90% confidence interval for the proportion of adults in a Midwestern state who prefer American cars. We want our margin of error (the plus and minus term in the confidence interval) to be no larger than .05. What is the minimum sample size we need?
19. The mathematics SAT test was designed to have a mean $\mu = 500$. A random sample of 150 scores from one state yielded $\bar{x} = 515, s = 104$. At significance level $\alpha = .05$, is this sample mean significantly higher than the intended mean? Answer in steps.
   a) Define an appropriate mean and give the appropriate null hypothesis and alternative hypothesis.

   b) Give the formula for the test statistic and the rejection region.

   c) Calculate the test statistic and state whether we reject the null hypothesis.

   d) Are these results significantly above the intended average?

   e) Give the P-value for this result.
20. A health investigator takes a random sample of 200 women and another independent random sample of 140 men in a population and takes their blood pressure. She obtains 
\( \bar{x}_w = 77, s_w = 14, \bar{x}_m = 83, s_m = 17 \). Give a 95% confidence interval for the difference 
\( \mu_m - \mu_w \) in mean blood pressures between men and women in this population.

21. If 340 of 500 randomly selected Democrats favor Proposition A and 220 of 500 randomly selected Republicans favor Proposition A, give a 92% confidence interval for the difference 
\( p_D - p_R \) in the proportions of all Democrats and all Republicans who favor Proposition A.
22. Acme Tutors promises to improve their student LSAT scores. A neutral investigator obtains a random sample of 10 prelaw students at State U who have not taken the course and gives them an LSAT. He then sends them through Acme Tutors services and finally gives them another LSAT. Here are the results. Do these results support Acme’s claim, at least on average? ($\alpha = .05$) Answer in steps below.

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>163</td>
<td>156</td>
<td>141</td>
<td>139</td>
<td>165</td>
<td>139</td>
<td>153</td>
<td>154</td>
<td>145</td>
<td>161</td>
</tr>
<tr>
<td>After</td>
<td>161</td>
<td>162</td>
<td>142</td>
<td>136</td>
<td>167</td>
<td>147</td>
<td>162</td>
<td>154</td>
<td>157</td>
<td>163</td>
</tr>
</tbody>
</table>

a) Give an appropriate null hypothesis and an appropriate alternative hypothesis.

b) Give the formula for the test statistic and rejection criterion.

c) Calculate the test statistic and decide: do we reject the null hypothesis?

d) Do we have evidence to refute the company’s claim?

e) Give the P-value.
23. We take independent samples from two normal populations and get

\[ n_1 = 8, \bar{x}_1 = 65, s_1 = 18, n_2 = 10, \bar{x}_2 = 55, s_2 = 10. \]

Do we have evidence (\( \alpha = .05 \)) that the two population means differ? Answer in steps:

a) Give the null hypothesis and alternative hypothesis.

b) Give the formula for the test statistic and the rejection criterion. Be sure to specify the degrees of freedom you should use.

c) Calculate the test statistic and tell whether or not you should reject the null hypothesis.

d) Do we have evidence the population means differ?

e) Give the P-value.
24. Here are data from 5 cases for the related variables $x, y$. We seek to estimate the coefficients for the simple linear regression model $y = \beta_0 + \beta_1 x + e$.

\[
\begin{array}{cccccc}
  x & 1 & 3 & 4 & 6 & 9 \\
  y & 6 & 10 & 10 & 15 & 24 \\
\end{array}
\]

Using this data
a) Estimate $\beta_1$.

b) Find the standard error (estimated standard deviation) of the estimator of $\beta_1$. 

25. We spilled coffee on an Excel regression printout and so several numbers are missing. The remaining portion of the printout is below. From this answer the questions further below.

**SUMMARY OUTPUT**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>1081.190201</td>
<td>360.3967</td>
<td>4.503953436</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>2080.464465</td>
<td>80.01786</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>3161.654667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.914610785</td>
<td>7.268157584</td>
<td>0.538597</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>0.790139877</td>
<td>0.242537121</td>
<td>3.25781</td>
</tr>
<tr>
<td>X Variable 2</td>
<td>0.274687732</td>
<td>0.241660687</td>
<td>1.136667</td>
</tr>
<tr>
<td>X Variable 3</td>
<td>-0.542100479</td>
<td>0.258441741</td>
<td>-2.09757</td>
</tr>
</tbody>
</table>

a) Find “R square”

b) Find “Standard Error” = \( \hat{\sigma} \) = the estimate of the standard deviation of the error terms.

c) Which of the independent variables appear to contribute to our model? Which one(s) seem dubious and are candidates to be dropped from the model? Explain your answers.

d) Give a 95% confidence interval for the true coefficient of x1.