NAME: ______________________________ SECTION: ____________

INSTRUCTIONS:

- Write your name and section number at the top of every page.
- During this exam, you may have access to R/RStudio, Excel, your calculator, and 3 formula sheets handwritten by you prior to the exam.
- You may not communicate in any way with anyone for the duration of this exam, with the exception of your instructor or proctor.
- You are instructed to log off all email accounts and social media/chat/instant messaging (IM) sites *before* starting the exam. You must remain logged off for the duration of the exam.
- You are instructed to close all browsers *before* starting the exam. You may not open any browsers, nor log back on to email/social media/chat/IM sites until you have turned in your exam.

To indicate that you have read and understand these instructions, and to affirm your commitment to academic integrity, SIGN HERE:

____

1. Police report that 78% of drivers stopped on suspicion of drunk driving are given a breath test, 36% are given a blood test, and 22% are given both tests. (*DeVeaux, Velleman, & Bock 2010, p358-365*)

   Let A be the event that the suspect is given a breath test.
   Let B be the event that the suspect it given a blood test.
   $P(A) = 0.78$
   $P(B) = 0.36$
   $P(A \cap B) = 0.22$

   (a) What is the probability that a randomly selected DUI (DWI) suspect is given a test?

   $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.78 + 0.36 - 0.22 = 0.92$

   (b) What is the probability that a suspected drunk driver who is pulled over gets either a blood test or a breath test but <u>not</u> both?

   $P(A \cap B') + P(A' \cap B) = (0.78 - 0.22) + (0.36 - 0.22) = 0.56 + 0.14 = 0.70$

   (c) What is the probability that a suspected drunk driver who is stopped gets neither test?

   $P(A' \cap B') = 1 - P(A \cup B) = 1 - 0.92 = 0.08$

   (d) Are giving a suspected drunk driver a blood test and giving the suspect a breath test mutually exclusive (disjoint)? Justify your answer.

   No, they are not mutually exclusive because $P(A \cap B) \neq 0$.

   (e) Are the two tests independent? Show why or why not.

   No, they are not independent. $P(A \cap B) = 0.22$ and $P(A) \times P(B) = 0.78 \times 0.36 = 0.2808$. $P(A \cap B) \neq P(A) \times P(B)$.

NAME: ______________________________ SECTION: ____________

2. The communications monitoring company *Postini* has reported that 91% of e-mail messages are spam. Suppose your Inbox contains 25 messages. (*DeVeaux, Velleman, & Bock 2010, p410*)

Let X be the number of spam messages of the $n = 25$ email messages in the Inbox.

$$X \sim \text{Binomial } (n = 25, p = 0.91)$$
$$P(X = x) = \binom{25}{x}(0.91)^x(0.09)^{25-x} \quad \text{for } X \in \{0, 1, 2, ..., 25\}$$

(a) Find the probability that all of your 25 messages are spam.

$$P(X = 25) = \binom{25}{25}(0.91)^{25}(0.09)^0 = (0.91)^{25} = 0.0946$$

RStudio: dbinom(25,25,0.91)

(b) Find the probability that none of your 25 messages are spam.

$$P(X = 0) = \binom{25}{0}(0.91)^0(0.09)^{25} = (0.09)^{25} = 7.179 \times 10^{-27} \approx 0$$

RStudio: dbinom(0,25,0.91)

(c) Find the probability that 20 *or more* of your 25 messages are spam.

$$\begin{aligned} P(X \geq 20) &= P(X = 20) + P(X = 21) + P(X = 22) + P(X = 23) + P(X = 24) + P(X = 25) \\ &= \binom{25}{20}(0.91)^{20}(0.09)^5 + \binom{25}{21}(0.91)^{21}(0.09)^4 + \binom{25}{22}(0.91)^{22}(0.09)^3 + \\ &\binom{25}{23}(0.91)^{23}(0.09)^2 + \binom{25}{24}(0.91)^{24}(0.09)^1 + \binom{25}{25}(0.91)^{25}(0.09)^0 \\ &= 0.0476 + 0.1145 + 0.2106 + 0.2777 + 0.2340 + 0.0946 = 0.9790 \end{aligned}$$

RStudio: sum(dbinom(c(20:25),25,0.91)) or 1-pbinom(19,25,0.91)

(d) Find the probability that *at least* 3 of your 25 messages are <u>not</u> spam.

$$\begin{aligned} P(X \leq 22) &= 1 - P(X \geq 23) = 1 - [P(X = 23) + P(X = 24) + P(X = 25)] \\ &= 1 - (0.2777 + 0.2340 + 0.0946) = 1 - 0.6063 = 0.3937 \end{aligned}$$

RStudio: pbinom(22,25,0.91) or alternatively 1-pbinom(2,25,0.09)

(e) What is the expected number of messages that are spam? Do <u>not</u> round your answer.

$\mu = np = 25 \times 0.91 = 22.75$

3. Assume that the duration of human pregnancies can be described by a normal distribution with a mean of 266 days and a standard deviation of 16 days. (*DeVeaux, Velleman, & Bock 2010, p455*)

   Let X be the length (in days) of human pregnancy.

   $$X \sim \text{Normal}\ (\mu = 266, \sigma^2 = 16^2)$$

   (a) What proportion of human pregnancies last between 270 and 280 days?

   $$P(270 < X \leq 280) = P(X \leq 280) - P(X \leq 270) = 0.8092 - 0.5987 = 0.2105$$

   RStudio: pnorm(280,266,16)-pnorm(270,266,16)

   (b) The shortest 25% of human pregnancies last less than how many days?

   $$P(X \leq q) = 0.25 \quad \rightarrow \quad q = 255.2 \text{ days}$$

   RStudio: qnorm(0.25,266,16)

   (c) Suppose a certain obstetrician is currently providing prenatal care to 60 pregnant women. What is the distribution of the mean length of pregnancy for a sample of 60 women?

   $$\bar{X} \sim \text{Normal}\ \left(\mu = 266, \frac{\sigma^2}{n} = \frac{16^2}{60} = 4.2667\right) \quad \text{or} \quad \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{60}} = 2.0656$$

   (d) What is the probability that the mean duration of pregnancy for these 60 women will be less than 260 days?

   $$P(\bar{X} < 260) = 0.0018$$

   RStudio: pnorm(260,266,16/sqrt(60))

   (e) Suppose your assumption is incorrect and the distribution of pregnancy duration is actually skewed left. Your responses to parts (c) and (d) should not change considerably, due to what fundamental theorem of probability?

   The Central Limit Theorem

NAME: ______________________________ SECTION: ____________

4. Is there a difference between antibacterial soap (AB soap) and antibacterial spray (AB spray) in the effectiveness of eliminating bacteria? A student decided to investigate this. Using a random number generator to determine treatment - AB spray versus AB soap - she washed her hands accordingly and then placed her right hand on a sterile media plate designed to encourage bacteria growth. After a 2-day incubation period, she counted the bacteria colonies. The student replicated the procedure 8 times for each treatment. Her data on the bacterial counts are summarized in the table below. (*DeVeaux, Velleman, & Bock 2010, p713-724*)

AB spray	AB soap
51	70
5	164
19	88
18	111
58	73
50	119
82	20
17	95

(a) Is the student's investigation an experiment or an observational study? An experiment

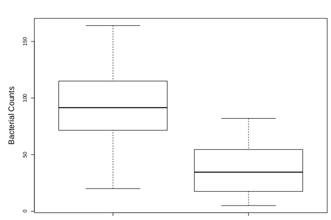
(b) Provide the five number summary (minimum, Q1, median, Q3, maximum) for the bacterial counts following treatment with AB spray.

Method	Min	Q1	Median	Q3	Max
By hand	5	17.5	34.5	54.5	82
RStudio	5	17.75	34.5	52.75	82

(c) Provide the five number summary for the bacterial counts following treatment with AB soap.

Method	Min	Q1	Median	Q3	Max
By hand	20	71.5	91.5	115	164
RStudio	20	72.25	91.5	113	164

(d) Sketch side-by-side boxplots (on the same set of axes) of the bacterial counts following the two treatments. Include tick marks and labels on the axes. Indicate the treatment corresponding to each boxplot. Provide a brief discussion of the distributions of bacterial counts comparing the two treatments.



The antibacterial spray appears to be more effective (lower bacteria counts following treatment) than the antibacterial soap.

5. Is there evidence in the student's data of a significant difference in the effectiveness in eliminating bacteria of soap versus spray? Conduct a formal statistical test to determine if the mean bacterial count following treatment with AB soap is significantly different from the mean bacterial count following treatment with AB spray. Be sure to state the hypotheses, set a reasonable significance level, give the test statistic (**not** assuming equal population variances), make a decision with respect to H_0, and provide a conclusion that is clear to anyone with *or without* a background in statistics.

   Let μ_1 be the true mean bacterial count following treatment with antibacterial spray.
   Let μ_2 be the true mean bacterial count following treatment with antibacterial soap.

   H_0: $\mu_1 = \mu_2$
   H_a: $\mu_1 \neq \mu_2$
   $\alpha = 0.05$

$$t_{stat} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{37.5 - 92.5}{\sqrt{\frac{(26.56)^2}{8} + \frac{(41.96)^2}{8}}} = -3.1326$$

   RStudio: pt(-3.1326,7) = 0.00828

   p-value $= 2 \times 0.00828 = 0.0165 < \alpha$, so reject H_0 and conclude that the mean bacterial count following treatment with AB soap *is* significantly different from the mean bacterial count following treatment with AB spray.

6. In a March 2006 Pew Research Center survey of 1286 cell phone users, 283 admitted that "When I'm on the cell phone, I'm not always truthful about exactly where I am." (*DeVeaux, Velleman, & Bock 2010, p457*)

   (a) Use this data to construct and interpret a 90% confidence interval for the proportion of cell phone users who are not always honest about their whereabouts when talking on the phone.

   Let p be the true proportion of cell phone users who are not always honest about their whereabouts when talking on the phone.

   $\hat{p} = \frac{283}{1286} = 0.220$
   $z^* = 1.645$ [RStudio: qnorm(0.95)]

   90% CI for p: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.220 \pm 1.645\sqrt{\frac{0.220(1-0.220)}{1286}} = (0.201, 0.239)$

   We are 90% confident that the true proportion of cell phone users who are not always honest about their whereabouts when talking on the phone is between 0.201 and 0.239.

   (b) Is this interval reliable? Verify the conditions for this method of inference.

   Yes, the interval is reliable. The sample size is large enough according to the conditions:

$$n\hat{p} = 1286 \times 0.220 = 282.92 > 5$$

$$n(1 - \hat{p}) = 1286 \times (1 - 0.220) = 1003.08 > 5$$

NAME: ________________________________ SECTION:____________

7. It is commonly believed that most students put on weight during their first year of college. Have you ever heard of the 'Freshmen 15'? Cornell University Professor of Nutrition, David Levitsky, decided to test this notion. He recruited freshmen from his introductory health course to participate in his study. Participants were weighed during the first week of the semester, and then again at 12 weeks. The data on weight (in pounds) for 14 of these students is provided below. (*DeVeaux, Velleman, & Bock 2010, p626-631*)

Week 1 Weight	Week 12 Weight	diff
148	150	2
164	165	1
137	138	1
198	201	3
122	124	2
146	146	0
150	151	1
187	192	5
94	96	2
105	105	0
127	130	3
142	144	2
140	143	3
107	107	0

(a) Construct and interpret a 95% confidence interval for the change in weight from week one to week 12 of the first semester freshman year.

Let μ_d be the true mean change in weight from week one to week 12 of freshman year.

$\bar{d} = 1.786, s_d = 1.424, n = 14,\ t^* = 2.160$ [RStudio: qt(0.975,13)]
95% CI for μ_d:

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}} \quad = \quad 1.786 \pm 2.160 \Big(\frac{1.424}{\sqrt{14}} \Big) \quad = \quad (0.964, 2.608)$$

We are 95% confident that the true mean change in weight from week one to week 12 of the first semester freshman year is between 0.964 and 2.608 pounds.

(b) Does your confidence interval support the notion of the 'Freshman 15' - that on average college students put on 15 pounds in their freshman year? Why or why not.
No, our CI does not support the notion of the 'Freshman 15' as the value 15 does not fall within it. The average weight gain freshman year (at least during the first semester) appears to be considerably less than 15 pounds.

(c) Is it reasonable to infer about all freshmen using the results from Dr. Levitsky's study? Explain why or why not.
We should be cautious about making such inference as Dr. Levitsy's students are not a random sample and likely are not representative of all college students.

8. Do storks bring babies? The table below summarizes data on the human population (in the thousands) of Oldenburg, Germany in the beginning of the 1930s and the number of storks nesting in the town. (*Ornithologische Monatsberichte, 44, no.2*)

Human Population	Number of Storks
55.5	140
55.5	148
64.9	175
67.5	195
69.0	245
72.0	250
75.5	250

(a) What is the explanatory variable and what is the response?

- Explanatory variable (x): Number of Storks

- Response variable (y): Human Population

(b) Create a scatter plot (you do not need to sketch the graph here) and comment on the form, direction, and strength of the relationship between the number of storks in town and the local human population.

There appears to be a strong, positive, linear relationship between the number of storks in town and the local human population.

(c) Find the correlation. $r = 0.941$

(d) Find the least-squares regression line, $\hat{y} = b_0 + b_1 x$, describing the relationship between the number of storks and the local human population (in the thousands). Write down the equation of the fitted regression line.

$$\hat{y} = 35.489 + 0.1507x$$

$$\widehat{\text{popln}} = 35.489 + 0.1507(\text{storks})$$

(e) Interpret both the slope, b_1, and the intercept, b_0, in terms of the number of storks in town and the local population (in the thousands).

We predict a local human population of 35,489 for a town that does not have any storks.

For each additional stork in town, we expect the local human population to increase by 151 people.

NAME: ______________________________ SECTION:____________

(f) What is the regression (residual) standard error, s_ϵ? $s_\epsilon = 2.864$

(g) Find the SSE (sum of squared errors).

$$\text{SSE} = s_\epsilon^2(n-2) = (2.864)^2(5) = 41.0125$$

(h) What proportion of the variation in the local human population is explained by the number of storks in town?

$$R^2 = 0.886$$

(i) Is there evidence in this sample of data of a significant relationship between the number of storks in town and the local human population? Formally test the slope, including the hypotheses, significance level, test statistic, decision and conclusion.

Let β_1 be the true slope describing the change in the local human population associated with changes in the number of storks in town.

H_0: $\beta_1 = 0$ (no relationship)
H_a: $\beta_1 \neq 0$ (there is a relationship)
$\alpha = 0.05$
$t_{stat} = 6.223$
p-value$= 0.00157 < \alpha$
Reject H_0 and conclude that there evidence in this sample of data of a significant relationship between the number of storks in town and the local human population.

(j) Based on your analysis, can you conclude that a greater number of storks in a town causes a larger local human population (in essence, that the idea that storks bring babies is not so far-fetched)? If not, provide a possible explanation for this phenomenon.

No, we cannot conclude that an increase in the number of storks in town *causes* an increase in the local human population. There may be factors (such as agreeable weather, available food, resources) in an area that attracts both humans and storks.

Do not write below this line.

Question	U	S	E
8(b)			
8(e)			
8(i)			
8(j)			