Math-156 - Exam #1 - Practice (S22)

General Information

- You will have 50 minutes to complete Exam #1
- You are not allowed to use anything on the exam besides a pencil/pen, paper, and a basic calculator
- this practice exam includes a formula sheet. This will be the exact same set of formulas you'll receive on the actual exam.

Exam topics

1) Probability rules
2) Expected value and standard deviation
3) Probability models
4) Confidence intervals

Types of content to expect

- 1 question consisting of several true/false statements
- 2 questions resembling those on the homework (ie: a short prompt with several parts)
- 1 question resembling those on in-class labs (ie: StatKey output and real data)

Formulas

Basic Probability Rules:

- Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Multiplication rule: $P(A \cap B) = P(A) * P(B|A)$ or $P(A) * P(B)$ if A and B are independent.
- Compliment rule: $P(A) + P(A^C) = 1$

Expected Value and Variance of Random Variables:

- $E(X) = \sum_i P(X = x_i) * x_i$
- $Var(X) = \sum_i P(X = x_i) * (x_i - E(X))^2)$

Table of Statistical Notation:

	Population Parameter	Estimate (from sample)
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Proportion	p	$\hat{p}$
Correlation	ρ	r
Regression	β_0, β_1	b_0, b_1

Table of Standard Error Formulas (used for Confidence Intervals):

Estimate	Standard Error	CLT Conditions
$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$np \geq 10$ and $n(1-p) \geq 10$
$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	normal population or $n \geq 30$
$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$n_i p_i \geq 10$ and $n_i(1-p_i) \geq 10$ for $i \in \{1, 2\}$
$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	normal populations or $n_1 \geq 30$ and $n_2 \geq 30$
r	$\sqrt{\frac{1-\rho^2}{n-2}}$	normal population (both vars) or $n > 30$

Values of “c” for various confidence levels using the Normal Model:

Confidence Level	90	95	99
c	1.65	1.96	2.58

Question #1 (true/false)

For each statement (A - F), indicate whether the statement is true (T) or false (F):

A) If the standard deviation of a random variable is small, most outcomes tend to be close to the variable's expected value

B) Assuming independence is reasonable when sampling cases from a large population, but not from a small population

C) When drawing a single card from a deck, the outcomes of "heart" and "spade" are disjoint

D) A 95% confidence interval is expected to describe 95% of the cases in the population of interest

E) To properly calibrate a 95% confidence interval using the t-distribution, one would use a larger value of "c" than if the Normal distribution were used.

F) The binomial distribution can be reasonably well-approximated by a Normal distribution if a small sample of categorical data is drawn from a population.

G) True

H) True

I) True

J) False, a confidence interval is an estimate of a population parameter (ie: mean, proportion, etc.), not individual cases

K) True

L) False, this is only true for a large sample ($n * p \geq 10$ and $n * (1 - p) \geq 10$)

Question #2 (textbook)

It is the job of safety engineers to determine the ability of industrial workers to operate a machine's emergency shutoff device, a task requiring sufficient physical strength and coordination. Among a group of test subjects, 66% could successfully use the shutoff with their left hand, while 82% could use the shutoff with their right hand, and 51% could use it with both hands.

A) What percentage of workers in the sample could *not* successfully use the shutoff with either hand? Show any work necessary to arrive at your answer.

- This is easiest to approach with a Venn diagram. 15% can use only their left hand, 31% can use only their right, and 51% can use both. By the complement rule, this leaves 3% who cannot use either.

B) Are success with the right hand and success with the left hand *disjoint* events? Briefly explain.

- No, these are not disjoint because both can occur at the same time (ie: a randomly chosen person could be able to succeed with both hands).

C) Are success with the right hand and success with the left hand *independent* events? Briefly explain.

- No, $P(R \cap L) = 0.51$ according to the data, but this is $\neq P(R) * P(L) = 0.66 * 0.82 = 0.54$

D) Suppose a small factory has a team of three randomly chosen workers operating these machines. Based upon these test subjects, what is the probability that no one on the team can successfully operate the shutoff using their right hand?

- $P(R_1^C \cap R_2^C \cap R_2^C) = (1 - 0.82) * (1 - 0.82) * (1 - 0.82) = 0.0058$

Question #3 (textbook)

A consumer organization inspected a random sample of new cars looking for appearance defects (dents, scratches, paint chips, etc.) They found that none of the cars had more than 3 defects, but 7% had 3 defects, 11% had 2 defects, and 21% had 1 defect.

A) Using a table, write out a probability model for the number of defects present in a randomly selected new car.

X	0	1	2	3
$P(X)$	0.61	0.21	0.11	0.07

B) Find the expected number of appearance defects in randomly selected new car. Show your calculation.

$$E(X) = 0.61 * 0 + 0.21 * 1 + 0.11 * 2 + 0.07 * 3 = 0.64$$

C) Find the standard deviation of the number of appearance defects in randomly selected new car. Show your calculation.

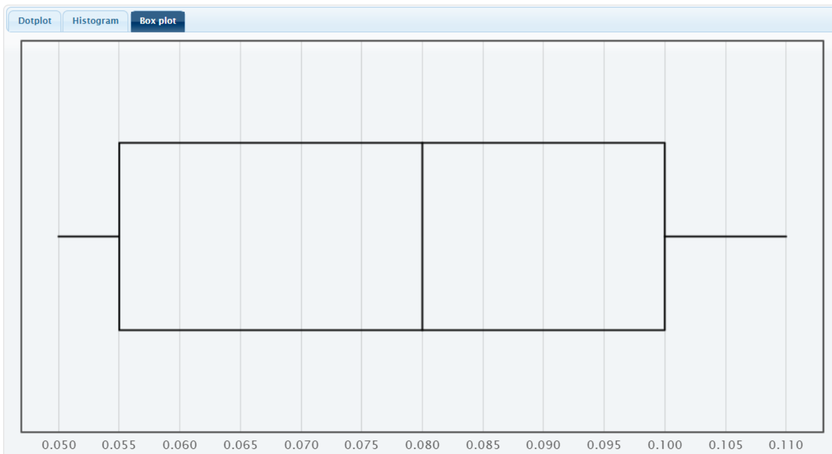
$$Var(X) = 0.61 * (0 - 0.64)^2 + 0.21 * (1 - 0.64)^2 + 0.11 * (2 - 0.64)^2 + 0.07 * (3 - 0.64)^2 = 0.87$$

$$SD(X) = \sqrt{0.87} = 0.93$$

Question #4 (lab)

Following a controversy at the 2008 Olympics, researchers conducted a study to determine whether new scientifically designed swimsuits could improve swim velocity relative to traditional swim suits. In the study, 12 professional swimmers were recruited to swim 1500m for time under two conditions: in one swim they wore a scientifically designed suit, and in the other they wore a traditional swim suit. The order of these swims was randomly determined, and they were performed on separate days.

Show below is StatKey output summarizing the difference in swim velocities for these 12 swimmers (new - traditional):



Summary Statistics

Statistic	Value
Sample Size	12
Mean	0.077
Standard Deviation	0.022
Minimum	0.05
Q_1	0.055
Median	0.080
Q_3	0.100
Maximum	0.11

A) Consider estimating the mean improvement in swim velocity for the population of all professional swimmers using a confidence interval. Should a Normal distribution or a t-distribution be used to calibrate the interval's margin of error? Briefly explain.

- T-distribution, since the outcome of interest is a single mean (quantitative) and we must estimate the population's standard deviation via the sample standard deviation.

B) Regardless of your answer to Part A, find a 95% confidence interval estimate for the population's mean improvement *using a Normal distribution*. Show your work.

$$\text{point est} \pm MOE$$

$$0.077 \pm 1.96 * \frac{0.022}{\sqrt{12}} = (0.065, 0.089)$$

C) Based upon the confidence interval you found in Part B, is it statistically plausible that the new suits do not actually improve swim velocity? Briefly explain.

- No, the 95% CI does not suggest a difference of zero is plausible. Only positive values are present in the interval, which means we can be confident that velocities are higher with the wetsuit

D) Had you used the t-distribution in the confidence interval calculation you performed in part B, would the resulting interval have been wider or narrower than the one you calculated using the Normal distribution? Briefly explain.

- Wider, as c would have been larger than 1.96 with the t-distribution (to properly account for the added certainty introduced by estimating σ as 0.022)

E) Suppose a colleague from a non-scientific discipline asks you explain the meaning of "95% confidence" as the term is used in your 95% confidence interval from Part B. In your own words, how would you explain the meaning of a "confidence level" or "95% confidence". Please limit your response to at most 3 sentences.

- A confidence level is the expected long-run success rate of a method used to form an interval estimate. Thus, a 95% confidence interval was built using a procedure that will succeed in capturing the parameter it seeks to estimate in 95% of samples/studies.