Midterm 2 - STAT 301
Fall 2019

Name:

UIN:

Signature:

:

1. Do not open this test until told to do so.
2. This is a closed book examination, However you may use the cheat sheet provided to you, the t and z tables. You should have no other printed or written material with you on the exam. But scrap paper is allowed.
3. You have 60 minutes to work on this exam. There are **16** multiple choice questions.
4. On the scantron please state the version of exam that you have.
5. You may use a calculator in the exam.
6. If there is no correct answer or if multiple answers are correct, select the **best** answer.
7. Unless stated otherwise, do all tests at the 5% level.
8. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification (however we are limited in the amount of help we can offer).
9. Please only give one answer per question (the one that is closest to the solution).
10. Good Luck!!!

(1-3) Hypokalemia is diagnosed when the mean μ blood potassium level is **below** 4.0mEq/dl. A few blood samples from a patient are drawn. If the potassium level in the sample mean is **less** than 4.0, a diagnoses of low potassium is made.

A patient goes to a clinic. Their potassium levels varies according to a Normal($\mu = 4.3, \sigma = 0.2$) distribution (this patient, by definition, does not have low potassium since $\mu = 4.3 > 4.0$).

(1) **One** blood sample is drawn from the healthy patient described above. If the blood sample is less than 4.0 she is (mistakenly) diagnosed with low potassium.

What is the chance of a misdiagnoses?

(A) 93.3% (B) 6.6% (C) 1.5% (D) 0.015% (E) -1.5%

(2) **5** blood samples are drawn from the healthy patient described above. If the sample mean is less than 4.0 she is (mistakenly) diagnosed with low potassium.

What is the chance of a misdiagnoses?

(A) 3.35% (B) 96.3% (C) 0.04% (D) 100% (E) 6.6%

(3) In general, what is the advantage of using the sample mean based on **5** blood samples rather than **one** blood sample?

A false positive is making a diagnoses when the patient is healthy.

(A) The probability of a false positive **increases** with a larger sample size.

(B) The probability of a false positive **decreases** with a larger sample size.

(C) The probability of a false positive remains the same regardless of sample size.

(D) There is no clear advantage.

(4) **5** blood samples are drawn from another patient (with unknown population potassium level). The sample mean of these blood samples is $\bar{x} = 4.1$mEq/dl.

The standard deviation of a blood sample is known to the $\sigma = 0.2$mEq/dl. Using the normal distribution, construct a **99**% confidence interval for their true mean potassium level.

(A) [3.92,4.28] [B] [4.62,3.59] [C] [3.71,4.49] [D] [3.77,4.30] [E] [3.87,4.32]

(5) Given the 95% confidence interval $[-3, 5]$ what is the sample mean and the margin of error?

(A) $\bar{x} = 2.5$, MoE $= 5$ (B) $\bar{x} = 0$, MoE $= 5$ (C) $\bar{x} = -1.5$, MoE$= 4$

(D) $\bar{x} = 1$, MoE$= 4$ (E) $\bar{x} = 1$, MoE $= 8$.

(6) The **maximum** weight a delivery truck can carry is **1000kg**.

The weight of an electrical appliance is normally distributed with mean $\mu = 52.5$kg and standard deviation $\sigma = 20$kg.

16 appliances are loaded into the truck, what is the probability that the truck will **exceed** the maximum weight?

(A) 0.0001% (B) 2.27% (C) 30.8% (D) 50% (E) 5%

(7) Compare the margin of error of a 99% confidence interval with the margin of error of a 80% confidence interval (both constructed using the same data). Which statement(s) is correct?

(A) The 80% confidence interval is **wider** than the 99% confidence interval.

(B) The 80% confidence interval is about **half** the length of a 99% confidence interval.

(C) The 80% confidence interval is about **80%** the length of the 99% confidence interval.

(D) The 80% confidence interval is about **two thirds** the length of the 99% confidence interval.

(E) (A) and (C).

(8) The margin of error of a 95% confidence interval is **10**. By how much should one change the sample size such that the margin of error reduces to **2**?

(A) The sample size should be **5** times larger.

(B) The sample size should be **1/5** smaller.

(C) The sample size should be **10** times larger.

(D) The sample size should be **25** times larger.

(E) The sample size should be **1/10** smaller.

(9) Fill in the blanks:

We use the t-distribution instead of the normal distribution when we ____(1)____.

Using a t-distribution instead of a normal distribution leads to a ____(2)____ confidence interval.

The t-disribution ____(3)____ correct for the lack of normality of the original data.

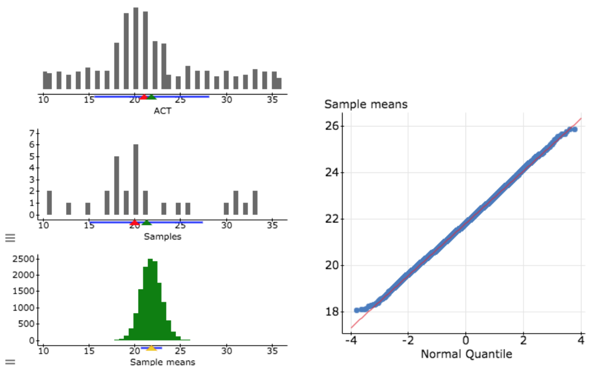
	(1)	(2)	(3)
A	if the data is skewed	narrower	can
B	if the data has heavy tails	narrower	cannot
C	estimate the standard deviation	narrower	can
D	if the data is skewed	wider	can
E	estimate the standard deviation	wider	cannot

(10) **30** students who did the ACT exams are randomly sampled. The sample mean of their grades is $\bar{x} = 21.3$ and the sample standard deviation is $s = 6$. Using the t-distribution construct a **98**% confidence interval for the mean ACT grade.

(A) $[21.3 \pm 2.462 \times 6]$ (B) $[21.3 \pm 2.462 \times 1.1]$ (C) $[21.3 \pm 2.462 \times 0.2]$

(D) $[21.3 \pm 1.311 \times 0.2]$ (E) $[21.3 \pm 1.311 \times 0.2]$.

(11) The distributions of the ACT scores and the sample mean are given below. Together with the QQplot of the sample mean.



Based on these plots what can we say about the 98% confidence interval for the (population) mean constructed in (Q10)?

(A) The mean **must** lie inside this confidence interval.

(B) The mean **does not** lie inside the interval.

(C) We have close to **98**% confidence that the mean lies in the interval.

(D) We have substantially less than **98**% confidence the mean lies in the interval. The true confidence level is close to 50%.

(E) The distribution of the sample mean is highly right skewed.

(12) It is well documented that the amount of CO_2 entering our atmosphere has substantially increased over the past 50 years. Scientists are studying the impact this has on the size of Arctic ice (which varies each year).

The size of Arctic ice is measured each year (in $10^6 m^2$). The mean size of the ice prior to 1999 was known to be **7** ($10^6 m^2$). Scientists hypothesis that the size of Arctic ice has **decreased**. They measure the size of Arctic ice yearly between 1999-2016. The data is summarized below. Let μ denote the population mean. Based on the data what can we conclude about the size of ice in the Arctic?

Data can be found at `https://www.epa.gov/climate-indicators/`.

One sample T hypothesis test:
μ : Mean of variable
$H_0 : \mu = 7$
$H_A : \mu < 7$

Hypothesis test results:

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
IceShelf	5.355	0.19467461	17	-8.4499978	<0.0001

(A) $H_0 : \mu \leq 7$ vs $H_A : \mu > 7$. The p-value is less than 1% there **is** evidence that the size of the ice **has increased**.

(B) $H_0 : \mu \leq 7$ vs $H_A : \mu > 7$. The p-value is less than 1% there is **no** evidence that the size of the ice has increased.

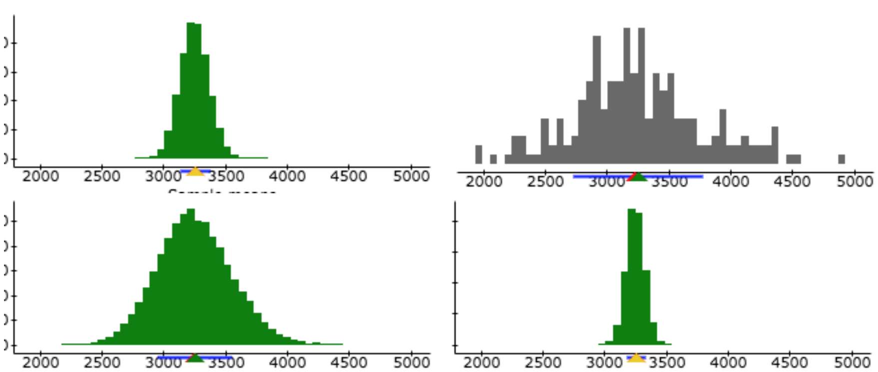
(C) $H_0 : \mu \geq 7$ vs $H_A : \mu < 7$. The p-value is less than 1% there is **no** evidence that the size of the ice has decreased.

(D) $H_0 : \mu \geq 7$ vs $H_A : \mu < 7$. The p-value is less than 1% there **is** strong evidence that the size of the ice shelf **has decreased**.

(E) $H_0 : \mu \geq 7$ vs $H_A : \mu < 7$. The p-value is so small there is **nothing** to suggest that the size of ice has decreased.

(13) Below are four histograms. One is the histogram of Shell strengths of eggs population of heights. The other three are the histograms of the distribution of sample means (taken from this population). The sample means are evaluated using different sample sizes. Match the histogram to the sample size.

Plot numbering: Top Left: (1). Top Right: (2). Bottom Left: (3) and Bottom Right: (4).



	(1)	(2)	(3)	(4)
A	10	60	Original	5
B	Original	60	5	10
C	20	Original	60	10
D	20	Original	10	60
E	10	5	Original	60

(14) A professor suspects that students who take GENE301 at 8am tend to do **better** than the students who take GENE301 at other times in the day. The (population) **mean** score for students who take GENE301 at any time of the day is 72%. In a sample of 30 students who took the class at 8am, the sample mean was 75%.

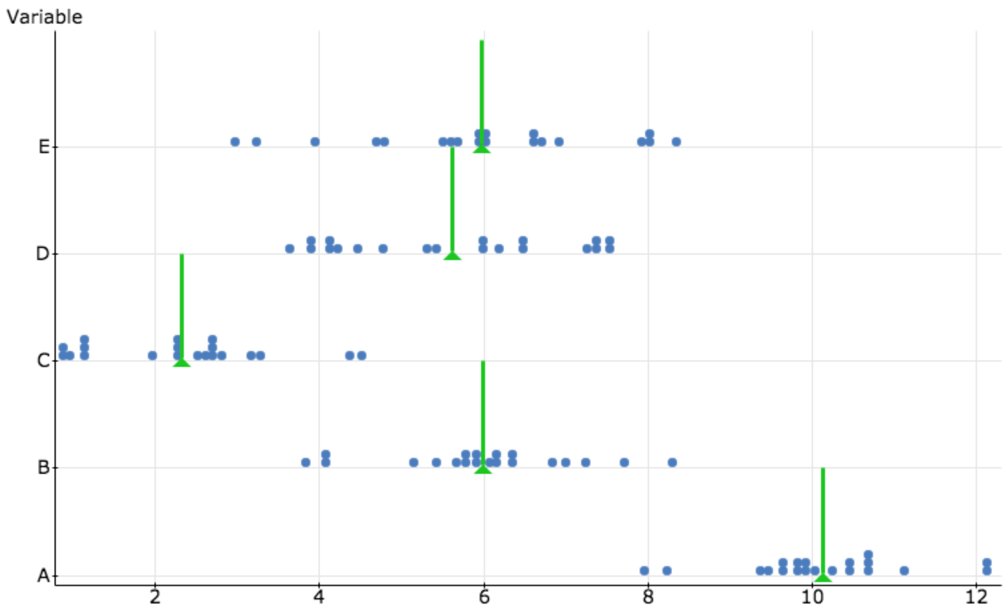
What is the professor's hypothesis of interest?

(A) $H_0 : \mu \geq 75\%$ vs $H_A : \mu < 75\%$ (B) $H_0 : \mu \leq 75\%$ vs $H_A : \mu > 75\%$

(C) $H_0 : \mu \geq 72\%$ vs $H_A : \mu < 72\%$ (D) $H_0 : \mu \leq 72\%$ vs $H_A : \mu > 72\%$

(E) $H_0 : \mu = 72\%$ vs $H_A : \mu \neq 72\%$.

(15-16) The dot plots of **5** data sets are plotted below (all with sample size 20). For each data set the sample mean is evaluated. This is denoted by the thick vertical line for each data set.



(15) You test the hypothesis $H_0 : \mu \leq 4$ vs $H_A : \mu > 4$. Which gives the **largest p-value**?

(A) (B) (C) (D) (E).

(16) You test the hypothesis $H_0 : \mu \leq 4$ vs $H_A : \mu > 4$. Which gives the **smallest p-value**?

(A) (B) (C) (D) (E).