Statistics 401D
Examination 1 (100 points)

Name ____________________________  Key ____________________________

NOTE: I will allow partial credit only if you show all your work.

1. (14) Daily ozone concentrations (in ppb.) for an Eastern city in the U.S. for a period of 150 consecutive days are graphed in a box plot. Answer the questions given below:

\[ \text{Median} = 44 \]

(a) Give a numerical measure of the center of the distribution of these data. \( \text{Median} = 44 \)

(b) What is the shape of the distribution of the ozone measurements suggested by the boxplot?

\( \text{Right-skewed} \)

(c) Calculate the interquartile range (IQR) of this dataset.

\[ 68 - 30 = 38 \]

(d) Calculate the number of days for which the ozone concentration lie between 44 ppb. and 90 ppb.

\[ \frac{150}{2} - 1 = 74 \]

(e) What is the LAV (Lower Adjacent Value) for this data?

\[ 20 \]

(f) Find the 1/3\(^{rd}\) percentile (i.e., the .00333 ... quantile) of the data.

\[ \frac{\text{IQR}}{15} = \frac{38}{15} = 2.5333 \]

(g) Calculate the Q(.99) point of the data.

\[ \frac{\text{IQR}}{15} = \frac{38}{15} = 2.5333 \]

2. (14) In Los Angeles, it is found that 75% of all automobiles would fail an inspection because they had emissions that did not meet EPA regulations.

(a) A random sample of 5 cars are inspected. Calculate the probability that at most one car will fail the inspection.

\[ X \sim \text{B}(5; 0.75) \]

\[ P(X \leq 1) = P(X = 0) + P(X = 1) = \binom{5}{0}(0.75)^0(0.25)^5 + \binom{5}{1}(0.75)^1(0.25)^4 \]

\[ = 0.000977 + 5 \times 0.75 \times 0.03906 = 0.015625 \]

(b) Use the Normal approximation to the Binomial to find the probability that at most 85 out of 100 cars inspected will fail the inspection.

\[ X \sim \text{B}(100; 0.75) \]

\[ \mu = 75 \quad n(1-x) = 25 \]

\[ P(X \leq 85) \approx P \left( Z \leq \frac{85 - 75}{\sqrt{75 \times 0.25}} \right) \Rightarrow \text{No Continuity Correction} \]

\[ = P(Z \leq \frac{10}{4.33}) = P(Z \leq 2.31) \]

\[ = 0.9896 \]
3. (24) The lifetime of a mass produced electronic component is assumed to have a Normal distribution with mean $\mu = 1400$ hours with standard deviation $\sigma = 50$ hours. 

$$X = \text{Life Time} \sim N(1400, 50^2)$$

(a) What is the probability that a component, chosen randomly from a lot, will fail in less than 1300 hours. (Show your work).

$$P(X < 1300) = P \left( Z < \frac{1300 - 1400}{50} \right)$$

$$= P(Z < -2)$$

$$= .0228$$

(b) Compute the probability that at least 1 of these components selected randomly from a package containing 5, will fail in less than 1300 hours.

$$Y = \# \text{fail out of } 5 \sim Bin(0.0228, 5)$$

$$P(Y \geq 1) = 1 - P(Y = 0)$$

$$= 1 - \binom{5}{0}(0.0228)^0(1-0.0228)^5$$

$$= 1 - .8911 = .1089$$

(c) It is observed that 10% of these components will last for C hours before failing. Find C showing your work.

$$P(X > C) = .1$$

$$\frac{C - 1400}{50} = 1.28$$

$$C = 1400 + 1.28 \times 50$$

$$= 1464$$

(d) Compute the probability that the average lifetime of 100 of these components will be in the range of 1390 to 1405 hours.

$$\bar{X} \sim N(1400, \frac{50^2}{100})$$

$$P \left( \frac{1390 - 1400}{5} < Z < \frac{1405 - 1400}{5} \right)$$

$$= P \left( -2 < Z < 1 \right) = .8413 - .0228$$

$$= .8185$$
4. (24) A precision parts manufacturer produces heat-treated steel alloy bolts for use in rockets. Let $\mu$ and $\sigma$ denote the mean and standard deviation for the length of the bolts produced by this process, respectively. For the next three parts assume that a value of $\sigma = 0.2$ mm. is available from past data but $\mu$ is not known.

(a) What is the minimum sample size that should be used if it is desired to estimate $\mu$ to within $0.05$ mm. with 90% confidence? (Show your work)

$$ E = 0.05 $$

$$ n \geq \frac{3.05 \cdot (0.2)^2}{(0.05)^2} = \frac{1.645 \cdot 0.04}{0.0025} = 4.3 $$

$$ \Rightarrow \quad n \geq 43 $$

The company statistician obtains a random sample of 40 bolts from the production line and measures the length of each. The sample mean of these 40 measurements is $\bar{y} = 32.46$ mm.

(b) Provide a 95% confidence interval for $\mu$. Explain to the production manager in non-technical terms what you mean by saying that you have 95% confidence in this interval:

$$ \bar{y} \pm 2.025 \cdot \frac{S}{\sqrt{n}} $$

$$ 32.46 \pm 1.96 \cdot \frac{2}{\sqrt{40}} \Rightarrow (32.398, 32.522) $$

If similar intervals are constructed from many different random samples of bolts the true mean $\mu$ will be contained in 95% of the intervals.

(c) Perform a test of hypotheses $H_0 : \mu = 32.5$ versus $H_a : \mu \neq 32.5$ using $\alpha = 0.05$. Show the rejection region clearly and state how you made your decision.

$$ t_c = \frac{32.46 - 32.5}{2/\sqrt{40}} = -1.265 $$

$$ t_{0.025} = 1.96 \Rightarrow R.R. \left| t \right| > 1.96 $$

Since $\left| t_c \right| \notin R.R.,$ fail to Reject $H_0$ at $\alpha = 0.05$

(d) In the above study identify clearly (i) the variable of interest,(ii) the population under study (iii) the parameter of interest in the population (iv) the point estimate of the parameter you identified (vi) an estimate of this parameter that incorporates the uncertainty in the estimate.

(i) length of bolts
(ii) all bolts of this type produced by this process
(iii) 40 bolts sampled by the statistician
(iv) $\mu$, the population mean length of these bolts
(v) the sample mean $\bar{y} = 32.46$
(vi) the 99% CI for $\mu$ obtained in part (c).
5. (24) A research engineer for a tire manufacturer conjectures that the mean life of a tyre built with a new rubber compound exceeds 60,000 kilometers. To investigate this conjecture, he built \( n = 16 \) tires and tested them to end-of-life on a test wheel that simulated normal road conditions. The sample mean and sample standard deviation of the 16 tire life times measured were \( \bar{y} = 60,137.14 \) and \( s = 372.8 \) kilometers, respectively. We may assume that the life times of the new tyres are approximately Normally distributed with mean \( \mu \) miles.

(a) Formulate appropriate null and alternative hypotheses to determine if the data supports the engineer's conjecture. Compute a test statistic to perform a test of this hypotheses and state your conclusion using \( \alpha = .05 \). Illustrate the rejection region clearly.

\[
\begin{align*}
H_0: \mu &\leq 60,000 & H_a: \mu > 60,000 \\
H_0 &\text{ Rejected}\end{align*}
\]

\[
t_c = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{60,137.14 - 60,000}{372.8/\sqrt{16}} = 1.47
\]

\[
R.R. \quad t > t_{.05,15} = t > 1.753
\]

Since \( t_c \) is not in R.R., fail to reject \( H_0 \) at \( \alpha = .05 \).

(b) i. Explain in words (or define by a formula) what is meant by the p-value for the test in part (a) above.

\[
p = P_{H_0}(T_{15} \geq 1.47)
\]

Prob. of getting a \( t \)-statistic value as large as 1.47 if the null hypothesis is true.

ii. Approximately, compute (or bound) the p-value using an appropriate table.

Clearly, 1.47 falls in the range \((1.241, 1.753)\)

\[
\Rightarrow \quad .05 < p < .1
\]

iii. Use the p-value you computed to test the hypotheses you stated in part (a).

Since the p-value is not less than .05 we fail to reject \( H_0 \) at \( \alpha = .05 \).

(c) Calculate a 90% confidence interval for \( \mu \). Use this interval to test the hypothesis you stated in part (a) and state the result. What is the \( \alpha \)-level of this test?

\[
\bar{y} \pm t_{.05,15} \times \frac{s}{\sqrt{n}} = 60,137.14 \pm 1.753 \times \frac{372.8}{\sqrt{16}}
\]

\[
= 60,137.14 \pm 1.753 \times \frac{372.8}{\sqrt{16}} = 60,137.14 \pm 93.2
\]

\[
90\% \text{ CI for } \mu: (59,973.76, 60,300.52)
\]

Since 60,000 falls in the interval, we fail to reject \( H_0: \mu \leq 60,000 \) at \( \alpha = .05 \)

\[
(\Rightarrow \text{ p-value is not less than } .05)
\]