Statistics 401C
Examination 1 (100 points)

Name ____________________

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

1. (12) PCB’s are a class of chemicals often found near disposal sites of electrical devices. They tend to concentrate in human fat and has been associated with a number of health problems. The following are concentrations of PCB (ng/g) in fat of a sample of 14 adults (shown in increasing order of value):

520, 650, 820, 940, 1100, 1250, 1300, 1450, 1600, 1720, 1810, 1940, 2310, 2540

(a) Give the 0.5 quantile, Q(0.5), of these data.

(b) Give the first quartile, Q(0.25), of these data.

(c) Give the interquartile range, IQR, of these data.

(d) Calculate Q(0.25)-1.5× IQR:

(e) Calculate LAV: ___________ and UAV: ___________

(f) Draw the Box Plot of these data.

2. (12) Suppose that 8% of the diodes used in motion detectors in a large case to be shipped from a factory are defective. (First, define the Binomial random variable X).

(a) Compute the probability that at most 1 out of 10 diodes randomly selected from this case are defective.

\[ P(X \leq 1) = P(X = 0) + P(X = 1) \]

\[ = \binom{10}{0}(0.08)^0(0.92)^{10} + \binom{10}{1}(0.08)^1(0.92)^9 \]

\[ = (0.92)^{10} + 10 \times 0.08 \times (0.92)^9 = 0.4344 + 0.3778 = 0.8121 \]

(b) An entire case will be rejected if more than 10 out of 100 of a random sample of 100 diodes tested from the case are defective. Use the Normal approximation to the Binomial to find the probability of rejecting a case if a random sample of 100 diodes from a case are tested.

\[ n = 100 \quad \bar{x} = 0.08 \quad n\bar{x} = 8 \quad n(1-x) = 92 \]

Use normal approx. will cont. correction

\[ \text{Need } P(Y > 10) = P(Y > 11) \]

\[ Y \sim N(0.8, 7.36) \quad \text{so } P(Y > 11) \approx P \left( z > \frac{10.5 - 0.8}{\sqrt{7.36}} \right) \]

\[ = P(z > \frac{9.7}{0.706}) = 1 - 0.8212 = 0.1788 \]
3. (24) The selling price of a particular brand of a high performance workstation among dealers nationwide is assumed to be normally distributed with mean $\mu = 2,400$ and standard deviation $\sigma = 500$.

$$X = \text{price of workstation} \sim N(2400, 500^2)$$

(a) What is the probability that a high performance workstation of the same brand, chosen randomly from a dealer, will cost less than $2,300$?

$$P(X < 2300) = P \left( Z < \frac{2300 - 2400}{500} \right) = P \left( Z < -\frac{1}{2} \right) = .4207$$

(b) Use the answer in part a) to compute the probability that at least four of five randomly selected high performance workstations of the same brand from dealers nationwide will cost less than $2,300$.

$$Y = \# \text{ in 5 that cost less than } 2300 \sim Bin(5, \pi) \text{ where } \pi = .4207$$

$$P(Y \geq 4) = P(Y = 4) + P(Y = 5) = \frac{5!}{4!1!} (.4207)^4 (.5793)^1 + \frac{5!}{5!0!} (.4207)^5 (.5793)^0$$

$$= .098733 + .013178 = .1119$$

(c) The manufacturer of this workstation suspects that sales of their product is not competitive with similar machines in some areas because dealers have marked the price above the 90th percentile cut-off of the price distribution. What is this cut-off price?

Need $c$ where $P(X > c) = .1 \implies P \left( Z > \frac{c - 2400}{500} \right) = .1$

$$\implies \frac{c - 2400}{500} = .8 \implies c = 3040 \text{ i.e. }$3040 is high

(d) Compute the probability that the average price of 100 high performance workstations randomly selected from dealers nationwide will cost between $2,300 and $2,450.

$$\bar{X} \sim N(2400, \frac{500^2}{100})$$

$$P(2300 < \bar{X} < 2450) = P \left( \frac{2300 - 2400}{500/\sqrt{10}} < Z < \frac{2450 - 2400}{500/\sqrt{10}} \right)$$

$$= P(-2 < Z < 1) = .4772 + .3413 = .8185$$
4. (24) Suppose that the observed readings of precision thermometers of a particular brand when the actual temperature measured is 65°F, has mean $\mu$ and variance $\sigma^2$. For the next four parts assume that a value of $\sigma^2 = 5.29$ is available from past data.

(a) Describe the sampling distribution of the sample mean random variable $\bar{Y}$ for a sample of size $n = 36$, assuming that the distribution of readings from the population of thermometers is unknown. Justify your answer.

$$\bar{Y} \sim N(\mu, \sigma^2/n) \quad \frac{\sigma^2}{n} = \frac{5.29}{36}$$

By CLT as $n$ is large.

(b) If you desire to estimate the population mean $\mu$ to be within $\pm 1^\circ F$ of the true value with 95% confidence, how many thermometers should you sample from the above population?

$$E = 1 \quad \delta_{.025} = 1.96 \quad \sigma^2 = 5.29$$

$$n = \left(\frac{\delta_{.025}}{E}\right)^2 \frac{\sigma^2}{\delta^2} = \left(\frac{1.96}{1}\right)^2 \frac{5.29}{1} = 20.32 \approx 21$$

The company statistician obtains a sample of 36 thermometers, and tests each to check for accuracy at 65°F. The sample mean of the readings of the 36 thermometers is $\bar{y} = 63.8$.

(c) Construct a 95% confidence interval estimate of $\mu$. The management of the company wants you to construct a 99% confidence interval in order to obtain a more accurate estimate. Is this a reasonable proposal? Explain to the management why you agree or disagree.

$$\bar{y} \pm \delta_{.01} \cdot \frac{\sigma}{\sqrt{n}}$$

$$63.8 \pm 1.96 \cdot \frac{\sqrt{5.29}}{\sqrt{36}} = 63.8 \pm (1.96) \left(\frac{2.3}{6}\right)$$

The 99% CI will be wider than the 95% CI thus making the interval estimate less accurate.

(d) Test $H_0: \mu = 65$ vs. $H_a: \mu \neq 65$ using $\alpha = .01$. State the rejection region and your decision clearly.

$$z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} = \frac{63.8 - 65}{2.3 / 6} = -3.13$$

$$z_{.005} = 2.575$$

R. R. $| z | > 2.575$

Since $| z_0 | = 3.13$ is in the R. R., Rej $H_0$ at $\alpha = .01$
5. (28) If a new process for mining copper is to be put into full operation, it must produce an average of more than 50 tons of ore per day. A 15-day trial period was carried out and daily yields were recorded in tons per day. A sample mean of \( \bar{y} = 74.2 \) and a sample standard deviation \( s = 44.2 \) were calculated from these data. Assume that the distribution of the yields per day is nearly Normal with mean \( \mu \).

(a) Write down appropriate null and alternative hypotheses needed in terms of \( \mu \) to test whether on the average, the mine produces more than 50 tons of ore per day under the new process. Calculate a test statistic to test this hypothesis, state the rejection region, and your conclusion using \( \alpha = .05 \).

\[
H_0: \mu \leq 50 \quad \text{vs} \quad H_a: \mu > 50
\]

\[
t_c = \frac{74.2 - 50}{44.2/\sqrt{15}} = 2.12
\]

\[
t_{.05, 14} = 1.761 \Rightarrow \text{R.R.: } t > 1.761
\]

Since \( t_c \) is in the R.R., 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.

\[
p = P_{H_0}(T_{14} > 2.12)
\]

probability of obtaining a test value as large as 2.12 if \( H_0 \) is true.

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

From table under d.f. = 14,

\[ .025 < p < .05 \]

iii. Show how to use the p-value you computed to test the hypothesis you stated in part (a).

Since \( p < .05 \), reject \( H_0 \) at \( \alpha = .05 \).

(c) Calculate a 90\% confidence interval for \( \mu \). Use this interval to test the hypothesis you gave in part (a) (i.e., say if you would reject or fail to reject \( H_0 \) using the confidence interval). Explain how you arrived at your decision. What is the \( \alpha \)-level of your test?

\[
90\% \text{ CI: } 14.2 \pm t_{.05, 14} \times \frac{44.2}{\sqrt{15}} \]

\[
1.761
\]

\[
(54.1, 94.3)
\]

Since 50 is not in this interval, we

\[
\text{reject } H_0 \text{ at } \alpha = .05.
\]