Show all your work to be eligible for partial credit and round your **final** answers to **two decimal** places in **all of the exam questions**.

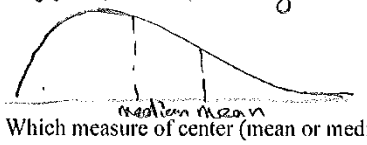
Question 1 (15 points)

The below output is from a computer software and shows descriptive statistics for the data on duration of the phone calls (in seconds), which has a distribution with a single mode.

Statistic	Result
Mean	196.5750
Standard Deviation	342.0215
Median	103.5000
First Quartile	54.5000
Third Quartile	200
Mode	9
N	80

(a) What is the best description of the data distribution? (i.e., is it symmetric, skewed right or skewed left?)

The data distribution is skewed right, since mean is substantially larger than the median.



(b) Which measure of center (mean or median) is more appropriate, and why?

median is more appropriate, since it is robust to outliers (and there are very large values (i.e., outliers) as the data is skewed right)

(c) Find IQR. Also which measure of spread (standard deviation or IQR) is more appropriate, and why?

IQR = 200 − 54.5 = 145.5

IQR is more appropriate, since it is robust to outliers.

(d) There is a claim by a calling company that mean duration of phone calls actually should exceed (aşmak) 185 seconds. Test this claim based on this data using $\alpha = 0.05$.

$H_0: \mu = 185$ $H_a: \mu > 185$

$n = 80 > 30$ is large, so use z-test

$$TS = \frac{196.575 - 185}{342.0215/\sqrt{80}} = \frac{11.575}{38.2392} = 0.30$$

DR: Reject H_0, if $TS > z_\alpha = z_{0.05} = 1.645$

since $0.30 \not> 1.645$, we do not reject H_0.

The mean duration of calls is not significantly greater than 185 sec.

Question 2 (20 points):
Sixteen mice were randomly assigned to a treatment (trt) group or a control (ctrl) group. Shown are their survival (hayatta kalma) times, in days, following a test surgery (ameliyat). Assume the population variances are not equal and use df=10 in your calculations.

Group	data	Sample size	Sample mean	Sample standard deviation
Trt	94, 38, 23, 197, 99, 16, 141	7	86.86	66.77
Ctrl	52, 10, 40, 104, 51, 27, 146, 30, 46	9	56.22	42.48

14 (a) Did the treatment prolong (uzatmak) survival? Use $\alpha = 0.025$.

	group1 trt	group2 ctrl
Mean	86.86	56.22
SD	66.77	42.48
n	7	9

and $\sigma_1 \neq \sigma_2$, but $df = 10$ is given.

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 > \mu_2 \Rightarrow \mu_1 - \mu_1 > 0$

$n_1 = 7$ & $n_2 = 9$ are both small, so we use t-test.

$$\text{std error} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{66.77^2}{7} + \frac{42.48^2}{9}} = \sqrt{839.396} = 28.94$$

$$T.S. = \frac{(\bar{x} - \bar{y}) - 0}{\text{std. error}} = \frac{86.86 - 56.22}{28.94} = \frac{30.64}{28.94} = 1.06$$

DR: Reject H_0 if $TS > t_{10, 0.025} = 2.228$

Since $1.06 \not> 2.228$, we do not reject H_0.

So, the treatment did not significantly prolong survival

6 (b) Also construct a 95% CI for the mean difference of survival times (as treatment minus control group).

95% CI on $\mu_1 - \mu_2$ is $(\bar{x} - \bar{y}) \mp t_{10, 0.025}(\text{std. err})$

$= 30.64 \mp 2.228(28.94)$

$= 30.64 \mp 64.48$

$= (-33.84, 95.12)$

Question 3: (15 points)
The following table indicates the mileage of used Toyota cars and their current mileage (=how many miles it has been driven). These are 3-year old Toyota Camry's and there are 100 cars in total (not all data are shown here).

Car	Mileage (x1000 miles)	Price (in thousand dollars)
1	37.4	14.6
2	44.8	14.1
3	45.8	14.0
.		
.		
.		
100	36.4	14.3

(a) Find the regression line by using the following information: $\sum x_i = 3601.1$, $\sum y_i = 1484.1$, $\sum x_i y_i = 53155.93$, $\sum x_i^2 = 133986.59$.

$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$ $\quad SS_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

$= 53155.93 - \frac{(3601.1)(1484.1)}{100}$

$= \frac{-288}{4,307.38}$ $\quad \cong -288$

$\cong -0.07$ $\quad SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$= 133986.59 - \frac{(3601.1)^2}{100}$

$\cong 4307.38$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 17.36$

$\hat{y} = 17.36 - 0.07x$

(b) Estimate the price of a Toyota Camry with mileage 46000 miles.

$\hat{y} = 17.36 - 0.07(46)$

$= 14.14$ ths. dollars $= 14140$ dollars

(c) Does mileage affect the selling price? Explain how by interpreting the slope.

Since the slope is -0.07, for every thousand mile driven price decreases by 70 dollars.
In short, as miles driven increases, price decreases.

Question 4: (20 points)
The dean of the College of Administrative Sciences and Economics wanted to know whether the graduates of his college used statistical inference techniques during their first year of employment after graduation. He surveyed 304 graduates and found that 204 used statistical inference in the first year of their employment.

(a) Estimate with 90% confidence the proportion of all business school graduates who use their statistical education within a year of graduation.

$\hat{p} = \frac{204}{304} = 0.67$ $\quad \alpha = 0.10 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.645$

$\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.67)(0.33)}{304}} = 0.027$

$p = \hat{p} \pm z_{\alpha/2}\,\sigma_{\hat{p}} = 0.67 \pm 1.645 \cdot 0.027 = (0.63, 0.71)$

(b) What should be the sample size to estimate the proportion within 0.05 error margin (sampling error) if you assume p=1/2.

$z_{\alpha/2}\sqrt{\frac{pq}{n}} = SE \Rightarrow n = \frac{(z_{\alpha/2})^2 (pq)}{(SE)^2} = \frac{(1.645)^2 (\frac{1}{2}\cdot\frac{1}{2})}{(0.05)^2} \approx 271$

(c) Do the data given in the question indicate that the proportion is less than 0.80? Answer by performing a test of hypothesis at $\alpha = 0.05$

$H_0: p = 0.80$ $\quad \alpha = 0.05$

$H_a: p < 0.80$ $\quad \sigma = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.80)(0.20)}{304}} = 0.02$

test statistic: $z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.67 - 0.80}{0.02} \approx -5.65$

$z_{0.05} = 1.645$

$z = -5.65 < -z_{0.05} = -1.645 \Rightarrow$ Reject null hypothesis

The proportion is significantly less than 0.80.

Question 5 (15 points):
The difference in the salary of MBA (=master of business administration) graduates have been recorded as follows. Two groups of Finance and Marketing majors, each of size 25, are chosen so that a Finance major and a Marketing major graduate are matched according to their graduation GPA's. Then, the difference in their salaries is recorded as "finance major's salary" minus "marketing major's salary". The mean difference is 5065 dollars and the standard deviation of all differences is 6647 dollars.

(a) Test at $\alpha = 0.01$ if there is a significant difference between the salaries of the two majors.

$H_0: \mu_d = 0$

$H_a: \mu_d \neq 0$ where $\mu_d = \mu_{Finance} - \mu_{Marketing}$

$$t^* = \frac{\bar{x}_d - 0}{s_d/\sqrt{n}} = \frac{5065}{6647/\sqrt{25}} \cong 3.81$$

$t_{0.005,24} = 2.797$ $(\alpha = 0.01 \Rightarrow \alpha/2 = 0.005)$

Since $|3.81| > 2.797$, reject H_0.

Yes, there is a significant difference between the salaries of the two majors.

(b) Find p-value as well.

Since the last value in the table for d.f. 24 is $t_{0.0005} = 3.745$, we find that $P(t > 3.81) < 0.0005$ and p-value $< 2\,(0.0005) = 0.001$

Question 6 (15 points):

Ice cream companies aim to predict the demand for different flavors of ice cream. In a recent poll, 38% of people said that chocolate was their favorite flavor.

(a) In a random sample of 8 people, what is the probability that at most 6 of them prefer chocolate?

$$P(X \leq 6) = 1 - P(X=7) - P(X=8)$$
$$= 1 - \binom{8}{7}(0.38)^7(0.62)^1 - \binom{8}{8}(0.38)^8(0.62)^0$$
$$\approx 0.9939$$

(b) In a random sample of 180 people, what is the probability that at least 80 of them prefer chocolate? Find an approximate answer.

$$\mu = np = 180(0.38) = 68.4 \qquad \sigma^2 = npq = (68.4)(0.62) = 42.408$$
$$\Rightarrow \sigma = \sqrt{npq} \approx 6.51$$
$$P(X \geq 80) \approx P\left(Z \geq \frac{79.5 - 68.4}{6.51}\right)$$
$$\approx P(Z \geq 1.71)$$
$$= 0.5 - 0.4564$$
$$= 0.0436$$