KEY

Part I. (20 points) An experiment focuses on the following research question about frustration:

"Are type A personality subjects more frustrated than type B personality subjects on the average?"

where type A personality refers to competitive and ambitious people, and type B personality subjects are more relaxed and easy-going. Both groups are put in a frustrating situation and each subject has received a score of 0 if the frustration level is low and 1 if it is high. The following table summarizes the data.

	sample size	the number of frustrated subjects (those with score 1)
Type A	32	16
Type B	33	15

1. (15 points) Answer the question above statistically by showing all steps of a hypothesis test and stating your conclusion in plain English.

(1) $H_0 : p_A - p_B = 0$

$H_a : p_A - p_B > 0$

(2) $\bar{p}_A = \frac{16}{32} = 50\%$, $\bar{p}_B = \frac{15}{33} = 45\%$

$\bar{p}_{diff} = 50\% - 45\% = 5\% = 0.05$

(3) $SE_A = \sqrt{\frac{0.5 \times 0.5}{32}} = 0.0884$

$SE_B = \sqrt{\frac{0.45 \times 0.55}{33}} = 0.0866$

$SE_{diff} = \sqrt{SE_A^2 + SE_B^2}$

$SE_{diff} \cong 0.12$

(4) $z = \frac{0.05}{0.12} = 0.42$

↳ Normal Table $\cong 31\%$

(5) P-value $= \frac{100\% - 31\%}{2} = 34.5\%$

(6) P-value $\gg 5\%$

i) Do not reject the null.

ii) We don't have enough evidence to say that type A is more frustrated than type B.

2. (5 points) Suppose the experimenter wishes to see the effect of the frustrating situation in a different experiment. He takes a random sample of a mixed group of personalities and measures their frustration level before and after the subjects are exposed to the frustrating situation. Is there a two-sample test covered in Math 202 which is applicable in this case? If yes, which one, if no, why not?

No, the samples are not independent. Because both measurements are taken from the same subjects (There exists a two-sample test for dependent samples, but it is not covered in Math 202)

Part II. (20 points) The following MINITAB output summarizes the data on a financial indicator (the ratio of sales to loans) of 10 companies randomly chosen from the textile sector. Previous studies show that this financial indicator has a normal distribution. If the average value of the indicator is less than 4, then the sector is said to be in financial difficulty.

Variable	N	Mean	SE Mean	StDev
Fin. Indicator	10	3.31	0.34	1.075

1. (10 points) Decide if the textile sector is in financial difficulty or not by performing a test of significance. Show all steps and state your conclusion in plain English.

(1) $H_0: \mu = 4$

$H_a: \mu < 4$

(2) $SE = 0.34$

$t = \frac{3.31 - 4}{0.34} \cong -2$

From the t-table: degrees of freedom $= N - 1 = 9$

(3) $2.5\% < P\text{-value} < 5\%$

(4) Since P-value $< 5\%$ i) reject the null

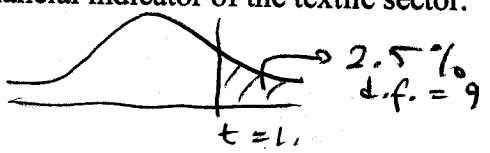
ii) The sector is in significant financial difficulty.

2. (7 points) Find a 95% CI for the average financial indicator of the textile sector.

mean : 3.31

t : 2.26

SE : 0.34



$\Rightarrow$ CI: $[3.31 \mp 2.26 \times 0.34]$

$[2.54, 4.08]$

3. (3 points) If 40 different independent random samples are taken and a 95% CI is constructed from each, how many of these confidence intervals are expected to cover the true average indicator of the textile sector.

$40 \times 95\% = 38$

Part III. (20 points) In a study, several twins were followed for their exercising habits and developing high blood pressure. Among each pair of twins, one of the twins exercised regularly and the other did not as part of their daily routines. There was no other major difference in their lifestyles (nutrition, smoking, etc.).

Suppose that each twin in a pair is equally likely to develop high blood pressure first. However, the result is as follows among 8 pairs of twins who develop high blood pressure.

	Regularly exercising twin (in the pair) develops high blood pressure first	The twin (in the pair) who does not exercise regularly develops high blood pressure first
High blood pressure	1 pair	7 pairs

1. (8 points) What is the chance of having 7 or more pairs out of 8, in which the twin who does not exercise develops high blood pressure first?

$$\binom{8}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right) + \binom{8}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0 = 0.0352$$

$\cong 3.5\%$ chance

2. (6 points) Can the difference between the blood pressure levels of the exercising and non-exercising twins be explained by i) chance, ii) genetics or iii) health effects of exercising? In your discussion, mention i), ii) and iii) explicitly.
Hint: Use your answer to Question 2; no need to do calculations here.

↳ iii) Health effects, not chance (P-value = 3.5

↳ genetics is a controlled variable as they're tw

3. (3 points) In the statement "Only in 1 pair out of 8, the regularly exercising twin developed high blood pressure first", a statistic or a parameter mentioned? What is its value?

statistics, because the statement is about a sample. Its value is "1 out of 8".

4. (3 points) In about how many pairs out of 8 do you expect the twin who exercises regularly to develop high blood pressure first, if you assume that each twin in a pair is equally likely to develop high blood pressure first?

$$8 \times \frac{1}{2} = 4$$

Part IV. (20 points) According to a recent article in a newspaper, 38% of the school teachers work in a second job for extra income. The news is based on a random sample of 49 teachers.

1. (8 points) Construct a 90% confidence interval for the proportion of school teachers who work in a second job.

$p = 38\%$, $n = 49$

$SE = \sqrt{\frac{0.38 \times 0.62}{49}} = 0.07 = 7\%$

90 % $\xrightarrow[z\text{-value}]{\text{Normal Table}}$ 1.65

CI: $[\ 38\% \mp 1.65 \times 7\%\ ]$

$[\ 26.45\%\ ,\ 49.55\%\ ]$

2. (12 points) It is known that 28% work in a second job within the population of all working people in Turkey. Can the difference between this and the percentage of school teachers who work in a second job be explained by chance? Show all steps of a hypothesis test to answer this question and state your conclusion in plain English.

(1) H_0: $p = 28\%$

H_1: $p > 28\%$

(2) $SE = \sqrt{\frac{(0.28) \times (0.72)}{49}} = 0.0$

$= 6\%$

(3) $z = \frac{\bar{p} - p}{SE} = \frac{38\% - 28\%}{6\%} = 1.67 \xrightarrow{\text{Normal Table}} \sim 90\%$

(4) P-value $= \frac{100 - 90}{2} \cong 5\%$

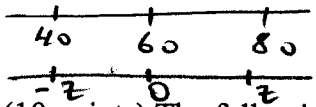
(5) i) Reject H_0 since P-value is barely 5% (a little less in fact)

ii) The difference is barely significant. It cannot be explained by chance.

Part V. (20 points) On the basis of historical data, Math 201 (Statistics) course grades can be approximated by a normal curve with mean 60 and standard deviation 20.

1. (6 points) Find the expected percentage of grades of Math 201 that fall into the following intervals.

Grades	less than 40	between 40 and 80	greater than 80
Percentage of grades	16 %	68 %	16 %



$z = \frac{80-60}{20} = 1$ Normal Table → 68 %

2. (10 points) The following shows the frequency of Math 202 (Statistics for Social Sciences) course grades from a sample of size 50.

Grades	less than 40	between 40 and 80	greater than 80
Frequency	15	18	17
	30 %	36 %	34 %

Is the distribution of Math 202 grades significantly different from Math 201 grades? Show all steps of a hypothesis test and state your conclusion in plain English.

	Math 201 (exp.)		Math 202 (obs.)		diff.	diff²
40 <	16 %	8	30 %	15	-7	49
40-80	68 %	34	36 %	18	1 6	256
80 >	16 %	8	34 %	17	-9	81

$\chi^2 = \frac{81}{8} + \frac{256}{34} + \frac{49}{8} \cong 23.8$ deg. of freedom = 2

χ^2-table : P-value < 1 % ⇒ Math 202 grades differ highly significantly from Math 201

3. (4 points) The data for Math 202 students also included different majors as given in the following table. Find the conditional distribution of grades <u>for history majors</u>.

	less than 40	between 40 and 80	greater than 80	
History	1	4	3	8
Psychology	8	7	6	21
Sociology	6	7	8	21
				50
Marg. Prob: History	1/50 = 2 %	4/50 = 8 %	3/50 = 6 %	16 %
Cond. Prob: History →	2 % / 16 % = 12.5 %	8 % / 16 % = 50 %	6 % / 16 % = 37.5 %	100 %
(History OR →)	1/8 = 12.5 %	4/8 = 50 %	3/8 = 37.5 %	)