Part I. (20 points) The following table shows the results of a randomized controlled study on nicotinic acid, which is a drug for coronary (heart) disease.

	Nicotinic Acid		Placebo	
	Number	Deaths	Number	Deaths
Adherers	558	11%	913	15%
Non-adherers	187	26%	382	28%
Total group	745	19%	1295	19%

1. (5 points) What are the treatment and control groups above? Write down a variable included in this study and its values.

Treatment group: those subjects who take nicotinic acid.
Control group: those who take placebo.
Number of deaths in the treatment group (one of the variabl
Values: 0, 1, 2, ..., 745.

2. (5 points) "This experiment started as a randomized controlled experiment, but then changed to an observational study". Do you agree? Explain in four sentences at most.

Since some of the subjects did not adhere to taking their pills regularly, the original design of the experiment was destroyed. The subjects assigned themselves to a group according to their own will. Therefore, this became an observational study.

3. (5 points) What is a confounding variable in this study? Explain in two sentences at most.

"Adherence" is a confounding variable: if a person is an adherer, s/he is also more careful about health issues. Clearly, adherence also affects taking Nicotinic Acid or not; hence affects both variables.

4. (5 points) a) Is this a longitudinal study or not? Explain in one sentence.

The patients are observed/followed for some time so that deaths are recorded, if any, therefore we can suspect that this is a longitudanal study.
(on the other hand, one can think otherwise if only "one time" measurement is taken)

b) Are the controls historical or contemporary, do you think? Explain why in at most one sentence.

The controls are contemporary; because the experiment involves a placebo, not an old drug as an alternative of nicotinic acid.

Part II. (20 points) Consider the following data set:

2.8, 3.0, 2.4, 3.1, 2.8, 2.0, 1.5, 2.9, 3.6, 4.4, 2.8, 2.5, 1.6, 1.9, 3.3, 3.0, 1.0, 2.4, 1.6, 4.1

1. (6 points) Find the mean and the standard deviation of the data set.

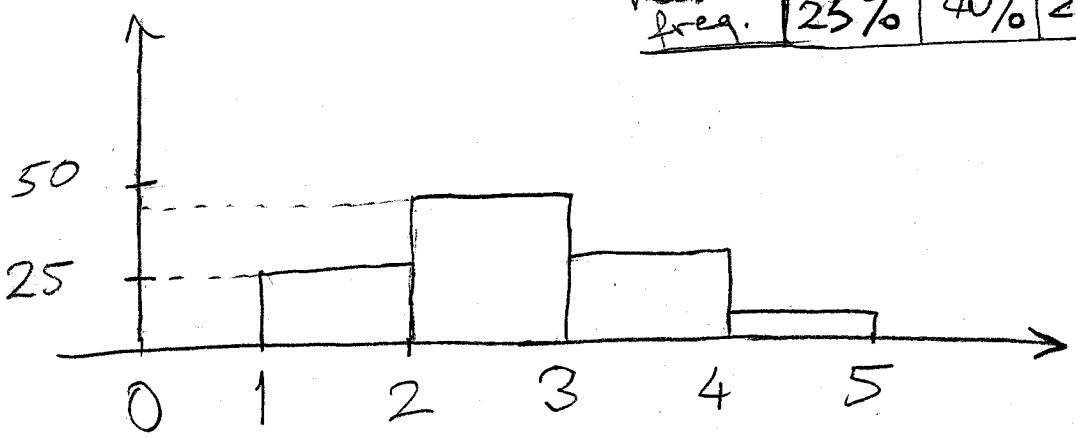
$$\text{mean} = \frac{2.8+3.0+\cdots+4.1}{20} = \frac{52.7}{20} \cong 2.6$$

$$\text{Standard deviation} = \sqrt{\frac{(2.8-2.6)^2+\cdots+(4.1-2.6)^2}{20}}$$

$$= \sqrt{\frac{14.43}{20}} \cong 0.85$$

2. (7 points). Draw a density scale histogram, label the axes.

interval	1-2	2-3	3-4	4-5
frequency	5	8	5	2
rel. freq.	25%	40%	25%	10%



3. (3 points) What percent of the observations do you expect to be within two standard deviations of the mean (average) according to the shape of your histogram? Explain in at most one sentence.

Since the histogram is approximately bell-shaped, we expect about 95% of the observations within 2 SD's of the average.

4. (4 points) Exactly, what percent of the observations is within two standard deviations of the mean in the given data set?

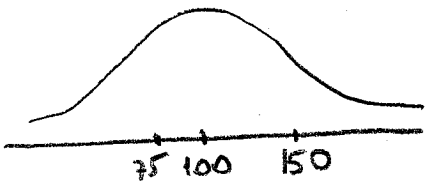
mean = 2.6 $\Rightarrow$ $2.6 \mp 2(0.85)$
SD = 0.85 $\Rightarrow$ between 0.9 and 4.3

$\Rightarrow$ only 4.4 is excluded $\Rightarrow$ 19 observations in this interval.

$\Rightarrow \frac{19}{20} = 95\%$ (good coincidence!)

Part III. (20 points) According to the association Romance Writers of America, romance fiction books generate 100 million dollars in sales on the average every month in the United States (US), with a standard deviation of 35 million dollars. Assume that the sales figures of romance fiction books per month follow a normal distribution approximately.

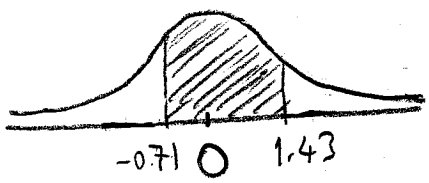
1. (8 points) Approximately, what is the chance that the romance fiction book sales will be between 75 and 150 million dollars next month?



$\frac{75-100}{35} \cong -0.71$ $\frac{150-100}{35} \cong 1.43$

Table → 1.43 → 85.29 %

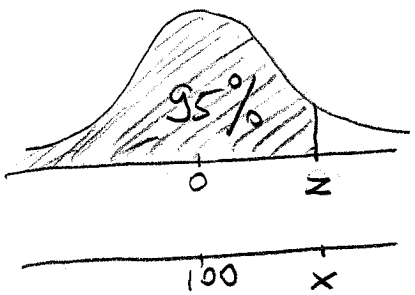
−0.71 → 51.61 %



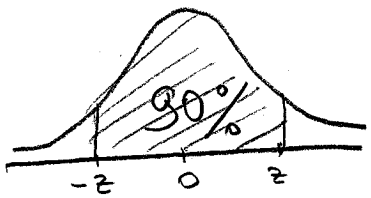
shaded area: $\frac{85.29 + 51.61}{2}$ %

$= 68.45\%$

2. (6 points) The 95th percentile of the sales distribution for romance books is considered extraordinary. What is this sales figure?



$\Rightarrow$



$\Rightarrow z \cong 1.65$ from table.

$\Rightarrow \frac{x-100}{35} = 1.65 \Rightarrow x = 157.75$ million dollars

3. (4 points) The sales figures of the next year will be estimated as follows. The booksellers which are members of Romance Writers of America, will be sent an interview to respond their sales figures of that year.

   (a) Why isn't this survey reliable to estimate the romance fiction sales of US? (at most two sentences)

Because the survey was sent to only to those booksellers which are members of Romance Writers of America, not to a random sample from all booksellers of U.S.

   (b) Is the survey representative of all booksellers, which are members of Romance Writers of America? What may go wrong? (at most one sentence)

Not either, because of nonresponse bias as this is a mailed survey and not an interview.

Part IV. (20 points) A recent research report states that 20% of all employed philosophers are academicians (=working in a university).

1. (3 points) In a random sample of 100 employed philosophers, how many of them are expected to be academicians?

$100 \times 20\% = 20$ academicians are expected.

2. (6 points) True or false, and explain in at most two sentences each:

   a) The percentage of academicians in a sample of size 100 will be exactly 20%.

False. Due to randomness in sampling process, it may be a different percentage.

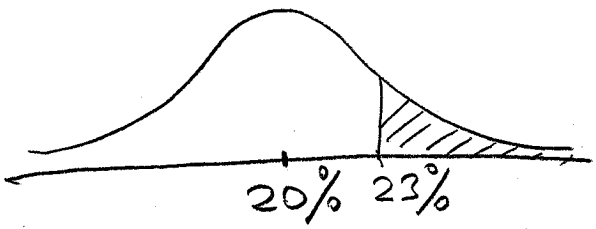
   b) The expected value of the percentage of academicians in a sample of size 100 is 20%, give or take (plus-minus) 4%.

False. Expected value is the population percentage. 20% is correct, but there is no error on it!

3. (8 points) What are the chances that the percentage of academicians in a random sample of 100 employed philosophers is greater than 23?

$\frac{23}{100} = 0.23 = 23\%$

$SE = \sqrt{\frac{(0.20)(0.80)}{100}} = 0.04 = 4\%$

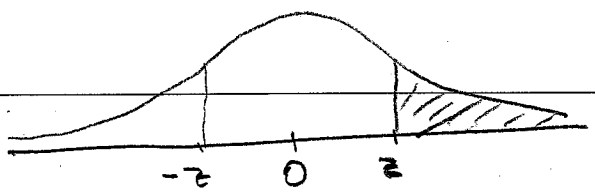


$\Rightarrow z = \frac{23\% - 20\%}{4\%} = 0.75$

Table $\Rightarrow 54.67\%$

Shaded area $= \frac{100 - 54.67}{2}\%$

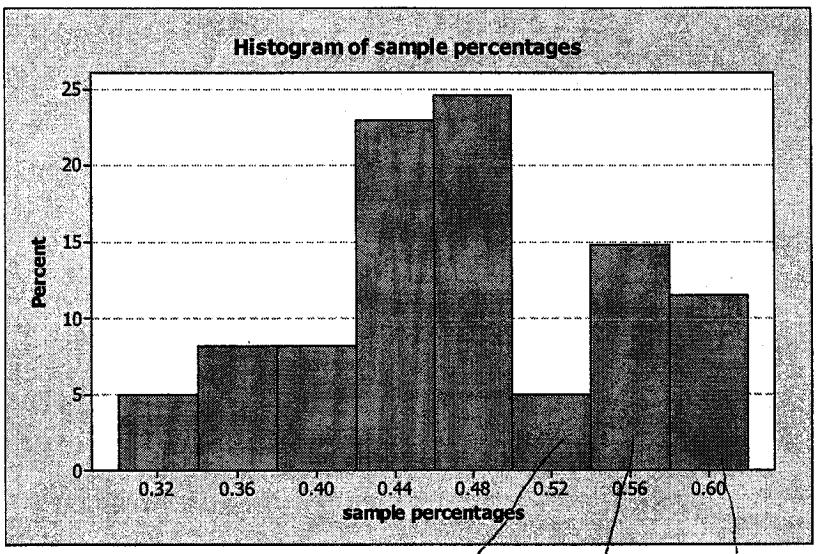
$= 22.65\%$



4. (3 points) Suppose that a random sample of size 100 is taken and it is found that 32 of them work in academia. Fill in the blanks:

For the number of academicians in the sample, the observed value is 32, but the expected value is 20.

Part V. (20 points) The following is a relative frequency histogram for the percentage of men in (many!) samples of size 80. Suppose that the percentage of men in the population is 46%.



1. (5 points) What percent of the sample percentages is greater than 50%, approximately from the histogram?

Approximately, $5\% + 15\% + 12\%$

$= 32\%$

2. (5 points) In which interval is the 10th percentile of the sample percentage distribution?

Interval $(0.30, 0.34) \to 5\%$ $\Rightarrow 5+8 = 13\%$

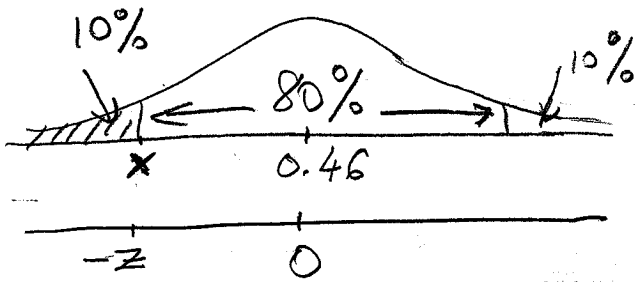
$(0.34, 0.38) \to 8\%$

So 10th percentile is somewhere between 0.34 and 0.38 (34% and 38%)

3. (5 points) Is normal curve a good approximation for the above histogram? Explain in 1 sentence.

Yes. By theory, the histogram of percentages follows a normal curve for large n (here 80) and indeed the histogram looks normal approximately.

4. (5 points) Find the 10th percentile of the sample percentage distribution, approximately.



$z = 1.30$

$SE = \sqrt{\frac{(0.46)(0.54)}{80}} \cong 5.6\%$

$\frac{x - 0.46}{0.056} = -1.30$

$\Rightarrow \boxed{x \cong 0.39}$