Question 1: Representative Sample (25 Points)

The graph below summarizes the results from a study based on a representative sample of men age 25–64 in 1993, who were working full time that year; the graph shows average income for each age group.

True or false, and explain: the data show that on average, if a man keeps working, his income will increase until age 50 or so, then start decreasing. If false, how do you account for the pattern in the data?

True. The overall education level has increased over the last 50 years, resulting in better paid jobs for younger people. People with higher income might also retire a bit earlier than people with less income, thereby resulting in a lower average towards age 65 that people getting close to 65 will not earn less.


10) False - this is a cross-sectional study. To find out whether men increase in income after an age of 50 years, we need a longitudinal study, e.g., follow a group of men for several years and obtain their income and see if there is a decline in income. Alternatively, we might ask men of age 50 and above about their income during the last 15 years and see whether it did decline (although this might be very unreliable - who keeps IRS records for 15 years or still remembers his income from 15 years ago).

5 The pattern in the plot might be explained by increasing starting salaries. Young men today might earn as much as men that started their worklife 40 years ago had earned 10 or 20 years into their career.
Question 2: Normal Distribution (45 Points)

According to the U.S. Department of Agriculture, one 3-ounce serving of trimmed sirloin beef contains, on average, 7.4 grams of fat. Assume that the amount of fat for such servings closely follows the normal curve, with a standard deviation of 0.4 gram.

Fill the blanks in the statements below and show all the work needed to obtain the answer:

1. (15 Points) The proportion of servings that contain between 6.9 and 7.1 grams of fat is \( \boxed{12.1\%} \).

   \[
   \frac{7.1 - 7.4}{0.4} = \frac{-0.3}{0.4} = -0.75 \text{ s.u.} \quad (3)
   \]

   \[
   \frac{6.9 - 7.4}{0.4} = \frac{-0.5}{0.4} = -1.25 \text{ s.u.} \quad (3)
   \]

   From Table: area between -0.75 and 0.75: \( 54.67\% \)

   \( \text{area between -1.25 and 1.25: } \frac{2 \times 54.67\%}{2} = 27.335\% \) \( \quad (3) \)

   \text{area between -1.25 and 0: } \frac{54.67\%}{2} = 27.335\% \)

   \text{area between -1.25 and -0.75: } 39.435\% - 27.335\% = 12.1\% \quad (3)

2. (15 Points) The proportion of servings that contain more than 8.3 grams of fat is \( \boxed{2.2\%} \).

   \[
   \frac{8.3 - 7.4}{0.4} = \frac{0.9}{0.4} = 2.25 \text{ s.u.} \quad (5)
   \]

   From Table: area between 2.25 and 2.25: 97.56\%

   \[
   \text{area above 2.25: } \frac{100\% - 97.56\%}{2} = 1.22\% \quad (5)
   \]

   \(-3\% \text{ not divided by 2}\)

3. (15 Points) Only 1\% of the servings will contain less than \( \boxed{6.46} \) grams of fat.

   From Table: area between -2.30 and 2.30: 97.86\%

   \text{area between -2.35 and 2.35: } 98.12\%

   \text{We need } x = -2.35 \text{ (or -2.30) in s.u.} \quad (5)

   \text{in original units:}

   \[
   x = -2.35 \times 0.4 + 7.4 = 6.46 \quad (5)
   \]

   \[
   (x = -2.30 \times 0.4 + 7.4 = 6.48) \quad (5)
   \]
Question 3: Controlled Experiment/Observational Study (30 Points)

As part of a study on exercise and health, a group of 1,000 people was followed for 5 years. At the beginning of the study, the researchers asked each person whether they exercised regularly or not. At the end of the study, the researchers measured several health-related variables, and in doing so, they noticed that the death rate for the exercise group was much lower than for the no-exercise group.

1. (10 Points) Is this an example of a controlled experiment or an observational study? Explain.

Worked Out:

They were asked what they did—they weren't told what to do.

2. (20 Points) Does this result necessarily imply that if people who do not exercise start to exercise regularly, they will live longer, on average, than if they do not? Explain clearly.

Worked Out:

No—this result merely shows that the type of people who choose to exercise differ from the type who do not. This is an observational study and there could be many confounding factors. For example, people who exercise might smoke less and it might be smoking that's harmful, not exercise that's beneficial. Taking a smoker or making them exercise might have no impact.
Question 4: Guessing the Correlation Coefficient (40 Points)

Match the four scatterplots with their correlations from the list:

\[ -1.03, -0.99, -0.89, -0.50, -0.05, 0.50, 0.93, 1.03 \]

Plot A: positive  Plot B: positive  Plot C: negative  Plot D: negative

10. Correlation for Plot A: \( r = 0.93 \)
10. Correlation for Plot B: \( r = 0.50 \)
10. Correlation for Plot C: \( r = -0.05 \)
10. Correlation for Plot D: \( r = -0.89 \)

Exploration (not required for your answer):

-1.03 and 1.03 are absolutely impossible as values for the correlation coefficient; only values between -1 and 1 are possible.

Plots A and B show a positive association; plots C and D a negative association.

Plot A shows the strongest association overall: ~ 0.93
Plot D shows an association that is slightly weaker than in A: ~ -0.89
Plot B shows a weak positive association: ~ 0.50
Plot C shows almost no association: ~ -0.05

5 for 1 level off, i.e.,
A: 0.50
B: -0.05, 0.93
C: -0.50, 0.50
D: -0.50, -0.93

2 for correlation but 0 or more levels off

0 for -1.03 or 1.03
Question 5: Regression (60 Points)

Pearson and Lee obtained the following results in a study of about 1,000 families:

average height of husband \( \approx 68 \) inches, SD \( \approx 2.7 \) inches,
average height of wife \( \approx 63 \) inches, SD \( \approx 2.5 \) inches, \( r \approx 0.25 \).

Show all the work needed to obtain the answer.

1. (15 Points) Predict the height of a wife when the height of her husband is 72 inches.

   \[
   x: \text{husband} \quad s.u.x = \frac{x - \text{avg} x}{\text{SD} x} = \frac{72 - 68}{2.7} = 1.48
   \]

   \[
   y: \text{wife} \quad s.u.y = r \cdot s.u.x = 0.25 \cdot 1.48 = 0.37
   \]

   \[
   y = \text{avg} y + s.u.y \cdot \text{SD} y = 63 + 0.37 \cdot 2.5 = 63.925 \approx 64 \text{ inches}
   \]

2. (15 Points) Predict the height of a wife when the height of her husband is 68 inches.

   Nothing to calculate - the regression line goes through the point of averages, i.e., for an average husband, it predicts an average wife, i.e., 63 inches as above in 1. or with explanation.

3. (15 Points) Predict the height of a wife when the height of her husband is unknown.

   63 inches!

   (The best guess if nothing else is known) 10 correct result

4. (15 Points) Predict the height of a husband when the height of his wife is 68 inches.

   \[
   x: \text{wife} \quad s.u.x = \frac{68 - 63}{2.5} = 2
   \]

   \[
   y: \text{husband} \quad s.u.y = 0.25 \cdot 2 = 0.5
   \]

   \[
   y = 68 + 0.5 \cdot 2.7 = 69.35 \text{ inches}
   \]