Name: Ronald Aylmer Fisher

1. Please write your name in the above space.

2. You need to do all 4 questions. All questions are of equal value (but not necessarily of equal difficulty).

3. Do not turn the page until so instructed. (You will have 90 minutes to work after the examination has been discussed with you.)

4. You may use your crib sheet and your calculator. Otherwise this is a closed book examination.

5. If you do not have enough room for your work in the place provided, use the back of a nearby page. (However, be sure to mark clearly which problem the material on the back of any page refers to.) If you pull the pages apart, sign all pages.

6. Answers should unambiguously state, in words, the approach taken. You should show your work so that partial credit can be given. Poorly described solutions will be penalized. Unsupported answers

7. Good luck!

<table>
<thead>
<tr>
<th>Question</th>
<th>Subject</th>
<th>Points Available</th>
<th>Points Earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Debate on Blood Pressure Medication</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Effect of Education on Economic Development</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Race and Legalized Abortion</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Interest Rates and Building Permits</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>
Question 1) Debate on Blood Pressure Medication (25 points)

Read the article that appeared in the Seattle Times on a “New study stirs debate over blood-pressure medication” and answer the questions posed below.

The article is at:

Note that your answers may be brief and address the article’s reporting of the attached study, not the details of the study itself. If there is not information on a question, leave it blank.

Here are the questions:

a) (12 points)

Briefly address the following basic questions about the article and the underlying study. You only need to address the questions you think are most important to the study.

1. What are the claims made in the article?

The major claim is made in the first sentence: “that inexpensive ‘water pills’ may not be the best treatment for high blood pressure, counter to the findings of a major study two months ago.” Note that the article actually makes the slightly weaker claim not claiming that this is true, only that “New research on more than 6,000 patients shows ...”

2. Are the claims supported in the article? If so, how?

The claims are supported based on the results of the attached study published in The New England Journal of Medicine, and interviews with two researchers.

3. Is there data in the article used to support the claims?

Yes, there is data in the article used to support the claims. It reports the basic proportions of strokes and heart disease from the NEJM article.

4. Is there data in the references for the claims?

Yes. The article is closely bound to the paper describing the original research. It reports on a randomized experiment on 6083 men and women, half of whom were given the (expensive) ACE inhibitor treatment and half of whom were given the less expensive diuretic (i.e., “water pill”) treatment.

The results show the ACE inhibitor had lower rates of cardiovascular events than the diuretic agents.

5. Can you think of alternative explanations for the data in the article? Is there data in the article to support these alternatives?

The study is randomized and used blind assessment. Hence it should be pretty good. The article reports the counts of events and the difference in percentages.

It is important to note that the article and study both show but do not emphasize that the risk was the same for women and lower for men. They also do not emphasize
that the effects on blood pressure were similar.

The risk of suffering a stroke might have more to do with a person’s state and genetic composition than the type of drugs used in controlling high blood pressure and this might explain why diuretics did not show pronounced performance for some cases.

As noted in the article, the previous study may have considered a different population of people - those with less complicated hypertension that the reported study.

6. Does the data presented support the claims?

Note that the rates for men and women combined given ACE inhibitor was 56.1 per 1000 patients-years and was 59.8 for those given the diuretic agent. These rates are pretty close even if statistically they are just different.

It would have been better to indicate that the rates are the same from women and close for men. Pooling both gives a biased view of the actual effects.

Both the study and especially the article seem to over emphasize the actual difference. The article also plays up the conflict with previous research to add drama.

b) (13 points)

Briefly describe each of the following components of the study. You only need to address the questions you think are most important to comment on.

Component 1: The individuals or objects studied and how they were selected

6083 older people with hypertension were selected and half randomly given the (expensive) ACE inhibitor treatment and half of whom were given the less expensive diuretic (i.e., “water pill”) treatment. This is a good approach.

Component 2: The exact nature of the measurements made and the questions asked

The risk of stroke or other “cardiovascular event or death” was recorded over a median of 4.1 years per patient. Other variables of interest were also recorded (e.g., blood pressure). This was done for all subjects with the researchers “blinded” to which treatment the subject was receiving.

Component 3: The setting or context in which the measurement were made

This is in a clinical setting were patients were followed over a number of years. The blinding helps reduce the likelihood of systematic differences in the recording of the blood pressure, etc.

Component 4: The extraneous differences between groups being compared

The randomization should have ensured the two groups were similar in terms of other differences. The study reports that there were no extraneous differences between groups in age, sex or blood pressure. It is possible, that genetic composition may explain why some people are more likely to have a stroke than others, but the randomization should balance out factors that were not adjusted for.

Component 5: The magnitude of any claimed effects of differences

The effects are actually small (56.1 vs. 59.8) and just statistically significant. There was little difference for women in terms of cardiovascular risk and none for either men
or women in terms of blood pressure.

Component 6: The source of the research and the funding

The source of research was not made clear. However it was in Australia and might have been from the drug company that made the ACE inhibitor drug.

Component 7: The researchers who had contact with the participants

This might have been a factor, but blinding was used.
Expanded educational emphasis is thought to benefit national economic development through improved labor force capacities and the creation of new knowledge and capacities. To test this hypothesis researchers considered the percentage growth of real gross domestic product (GDP) from 1995-2000 for counties around the world. The countries were split into those who placed a high emphasis on education and those who placed a low emphasis. This was measured by the number of years of school per person in the country adjusted for the different ages of the populations. The growth for thirteen randomly selected high education counties and thirteen randomly selected low education counties are recorded below:

<table>
<thead>
<tr>
<th>High education</th>
<th>Low education</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.7</td>
<td>18.9</td>
</tr>
<tr>
<td>22.0</td>
<td>13.0</td>
</tr>
<tr>
<td>20.1</td>
<td>18.2</td>
</tr>
<tr>
<td>23.7</td>
<td>15.0</td>
</tr>
<tr>
<td>21.9</td>
<td>13.9</td>
</tr>
<tr>
<td>21.7</td>
<td>12.2</td>
</tr>
<tr>
<td>20.4</td>
<td>15.7</td>
</tr>
<tr>
<td>17.3</td>
<td>13.2</td>
</tr>
<tr>
<td>15.3</td>
<td>19.6</td>
</tr>
<tr>
<td>16.8</td>
<td>13.1</td>
</tr>
<tr>
<td>21.8</td>
<td>18.8</td>
</tr>
<tr>
<td>16.0</td>
<td>19.0</td>
</tr>
<tr>
<td>16.4</td>
<td>14.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>High education</th>
<th>Low education</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean, ($\bar{x}$)</td>
<td>19.16</td>
<td>15.78</td>
</tr>
<tr>
<td>standard deviation, ($s$)</td>
<td>2.97</td>
<td>2.73</td>
</tr>
</tbody>
</table>

a) (10 points)

Based on the axis supplied, construct back-to-back stem-and-leaf plots for the two samples. Be sure to include a legend to indicate the units.

Briefly describe the distributions for the two samples.
The key feature is the bimodal nature of both distributions.

b) (7 points)

Calculate the median, lower quartile, upper quartile, and interquartile range for each sample.

Solution: As $n = 13$, the median is the $(13 + 1)/2 = 7^{th}$ value. The quartiles are the $(7 + 1)/2 = 4^{th}$ values. Using the median of the 6 lowest values (average of 3$^{rd}$ and 4$^{th}$ values) is also acceptable.

<table>
<thead>
<tr>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.1</td>
<td>15.0</td>
</tr>
<tr>
<td>16.4</td>
<td>13.2</td>
</tr>
<tr>
<td>21.8</td>
<td>18.8</td>
</tr>
<tr>
<td>5.4</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Using the alternative method, the quartiles are

<table>
<thead>
<tr>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.2</td>
<td>13.15</td>
</tr>
<tr>
<td>21.85</td>
<td>18.85</td>
</tr>
<tr>
<td>5.65</td>
<td>5.7</td>
</tr>
</tbody>
</table>
c) (5 points)

Below are two parallel box-and-whisker plots, one for each type of country:

![Box-and-whisker plots](image)

Indicate on the plots the location of the median, lower quartile and upper quartile. Briefly describe the features of the High and Low education countries.

**Solution:** The median is at the middle bar of each box. The lower quartile is the left side bar of the box and the upper quartile is the right side bar of the box. The percentage GDP growth for High education countries is skewed toward smaller values while the Low education countries are skewed toward higher values. The median for the High education countries is substantially higher than the median of the Low education countries. The majority of High education countries have percentage GDP growth greater than the Low education countries.

The outlier boundaries are the quartiles plus or minus 3/2 times the IQR. That is:

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>lower boundary</td>
<td>11.7</td>
<td>5.2</td>
</tr>
<tr>
<td>upper boundary</td>
<td>29.9</td>
<td>27.2</td>
</tr>
</tbody>
</table>

Hence in each case the whiskers go out to the minimum and maximum values in the sample.

**d) (3 points)**

A fellow student compares high to low education countries by comparing the means of the samples alone. Is this an adequate comparison? Justify your answer with reference to the above analysis.

**Solution:** No. The mean, a measure of location, is adequate for comparison when the shapes of the distributions are the same and form a single clump. In this case there is a clear gap between small and large values in each sample. This raises the issue that there may be two kinds of high education countries and two kinds of low education countries. This is similar to the situation where the male and female heights are put together in a single batch of numbers.
Question 3) Race and Legalized Abortion (25 points)

A researcher in Alabama is interested in studying the relationship, if any, between a person’s race and their opinion about legalized abortion. They consider two racial groups (“black” and “white”) and two opinions (“favor” and “oppose”). Based on survey data they find that 30% of whites favor legalized abortion, 60% of the respondents are white, and 32% of the respondents are black and favor legalized abortion.

a) (5 points)

Construct a probability tree for this situation, with the first branch being the event “respondent is white.”

**Solution**: The tree diagram is:

<table>
<thead>
<tr>
<th>respondents</th>
<th>respondents</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>race</td>
<td>opinion on abortion</td>
<td>probability</td>
</tr>
<tr>
<td>white</td>
<td>favors</td>
<td>0.6 x 0.3 = 0.18</td>
</tr>
<tr>
<td>0.6</td>
<td>oppose</td>
<td>0.6 x 0.7 = 0.42</td>
</tr>
<tr>
<td>black</td>
<td>favors</td>
<td>0.4 x 0.8 = 0.32</td>
</tr>
<tr>
<td>0.4</td>
<td>oppose</td>
<td>0.4 x 0.2 = 0.08</td>
</tr>
</tbody>
</table>

The nodes in the table represent the respondents characteristic and the links the conditional event. The probabilities of each link are under the nodes. The key conditional probability is

\[ P(F \mid \overline{W}) = \frac{P(F \cap \overline{W})}{P(\overline{W})} = \frac{0.32}{0.40} = 0.80 \]

where \( F \) is the event that the respondent favors legalized abortion, \( W \) is the event that the respondent is white and \( \overline{F} \), and \( \overline{W} \) are the usual complementary events of opposing and being black.

b) (5 points)

Draw a simple Venn diagram for this situation.

**Solution**: The Venn diagram encodes the four probabilities that are the endpoints of the probability tree. Note that the mass in neither event is 0.08.

c) (5 points)

Based on the probability tree and Venn diagram, find the probability that the respondent is white or favors abortion (or both)

**Solution**: \( P(W \cup F) = P(W) + P(F) - P(W \cap F) \)

\[ = P(W \cap F) + P(W \cap \overline{F}) + P(\overline{W} \cap F) = 0.18 + 0.42 + 0.32 = 0.92 \]

Alternatively \( P(W \cup F) = 1 - P(\overline{W} \cap \overline{F}) = 1 - 0.08 = 0.92 \)
d) (4 points)

Find the probability that the respondent is white and opposes abortion.

Solution: From the probability tree:
\[ P(W \cap \overline{F}) = 0.42 \]

e) (4 points)

What is the probability that a black person favors legalized abortion? In symbols, what does this conditional probability represent?

Solution: In symbols, this is the conditional probability
\[ P(F | \overline{W}) = \frac{P(F \cap \overline{W})}{P(\overline{W})} = \frac{0.32}{0.40} = 0.80 \]

f) (2 points)

Are the events “the respondent is white” and “the respondent favors legalized abortion” independent? Briefly explain your answer.

Solution: The events are independent if the event
\[ \{\text{the respondent is white}\} \cap \{\text{the respondent favors legalized abortion}\} \]
is empty. But 18% of the respondents fall into both categories, so the events are not mutually exclusive.

Solution: The events are independent if:
\[ P(\text{the respondent favors legalized abortion given the is white}) = P(\text{the respondent favors legalized abortion}) \]
\[ P(F|W) = P(F) \]

We know that
\[ P(F) = P(F \cap W) + P(F \cap \overline{W}) = 0.32 + 0.18 = 0.5 \]

But
\[ P(F|W) = P(F \cap W)/P(W) = 0.10/0.6 = 0.30 \]

So they are not equal and the events are not independent. This also means that the opinion on legalized abortion is not independent of the person’s race. That is proportionately more white people oppose legalized abortion than black people.
Question 4) Interest Rates and Building Permits (25 points)

The interest rate determines the cost of funds necessary to build a home. The number of building permits issued by local governments in King County is a strong indicator of the number of new homes that will be available in the next six months.

The King County Homebuilders’ Association argued that during periods of high interest rates the number of building permits issued decreases drastically. They then argue that the number of new homes will be correspondingly reduced, leading to a potential housing shortage.

As a result, the homebuilders’ association is lobbying for various home subsidy programs to be applied during periods of high interest rates.

Is there a relationship between interest rates and the number of housing permits? To address this question, we collected data on the interest rates and the number of housing permits for the spring season for twelve successive years.

The scatter plot below relates the interest rates to the number of housing permits issued for the 12 years.

![Scatter Plot](image)

a) (3 points)

Estimate the regression line by eye and draw it on the above scatter plot.

**Solution**: The line is sketched on the above graph. The two points used were the value at $x = 5$ of about 1650 and the value at $x = 2$ of about 1280.

b) (6 points)

Based on the line you have drawn on the plot, estimate the coefficients of the regression line?
**Solution**: Based on the two points, the slope is then about \( \frac{(1280 - 1650)}{(9 - 5)} = -92.5 \) permits per change in percent interest rate. The intercept is the number of permits at zero interest rate. This is way off the left-hand-side of the graph, but we can estimate it from the equation:

\[
y = a + b \times x
\]

or \( a = y - b \times x \). Using the first value from above: \( a = 1650 - (-92.5) \times 5 = 2112.5 \) permits.

The least-squares regression slope is actually \(-89.60\) and the intercept is \(2087\). Your values should be close to these (i.e., about as far as mine are).

c) (4 points)

Estimate the correlation coefficient between the number of building permits and the interest rate. Does it appear that the relationship between the two variables is approximately linear?

**Solution**: This is hard to estimate, but is seems to have a magnitude of about 0.9 and a negative sign as the relationship is decreasing. So -0.9 is a realistic estimate. The true value is -0.94.

d) (2 points)

Does the correlation coefficient indicate a positive or negative association between the two variables? Does it indicate a strong, or weak linear relationship between the two variables?

**Solution**: It is a negative relationship and indicate a strong linear relationship between the number of permits issued and the interest rate.

e) (3 points)

What is the interpretation of the intercept of the regression line?

**Solution**: This is the average number of permits issued in a month where there is zero interest rate. In this case it indicates that this average is a little over 2000.

f) (3 points)

How large a change in the number of building permits is associated with an increase of 2% in interest rates?

**Solution**: This is just twice the slope - that is, about \( 2 \times -90 = -180 \). That is, we expect it to decrease by about 180 permits.

g) (4 points)

Predict the number of building permits for a year when the interest rate is 16%. Based on the model, do you think we can this prediction is valid (approximately)?

**Solution**: As this is off the right-hand-side of the graph, we must use the equation:

\[
y = a + b \times x = 2087 - 90 \times 16 = 647
\]

A number about 650 permits is about right. Of course, we are extrapolating off the end of the graph and so this prediction must be taken with a grain of salt. Without additional information we can not be sure it is valid.