Name: Ronald Aylmer Fisher

1. Please write your name in the above space.

2. There are 6 practice questions. On the real exam there will be only 4. **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 the statistical tables provided. 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.

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>Power Shortages in Western States</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Incoming Notebooks</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Election and Gambling</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Government Employee Drug Use</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>New Parks and Reinvigorating</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>IRS random audits</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Question 1) Power Shortages in Western States (25 points)

Washington State regulators of hydroelectric power employ a weather prediction strategy to guide their decisions about selling power for the next year to California. They forecast the overall rainfall, which determines the amount of hydroelectric power the state can produce, and adapt their decisions to produce power accordingly.

If the rainfall in the upcoming year is good then the regulator will have a lot of water to produce power, and can sell much of the power to California. If the rainfall is poor, and the regulator has sold a lot of power to California then the reservoirs will be reduced leading to high in-state power costs and environmental problems. In any year the rainfall is either good or bad, and the regulator can either sell a fixed amount of power to California or he can not.

Suppose that a regulator accurately predicts a good year 80% of the time (i.e., predicts that the year will be good in 80% of the years that turn out to be good years). The same regulator accurately predicts a bad year 60% of the time.

In reality, eight of the last ten years have been good years.

You should construct a probability tree for the process on the next page to help in answering questions b) - d).

a) (5 points)

What is the probability that the next year will be good? What concept of probability have you applied to decide this?

Solution: The only information we have on the probability that the next year will be good comes from the knowledge that eight of the last ten years have been good years. Based on the relative frequency concept of probability the probability that the next year will be good is \( \frac{8}{10} = 0.8 \).

b) (5 points)

What is the probability that the regulator predicts a good year and the year is also good?

Solution: To answer this we should construct a probability tree (see the next page). From the tree, we can see that the probability that the outcome that the regulator predicts a good year and the year is also good is 0.64.

c) (5 points)

What is the probability that the regulator will say that the next year will be a good year?

Solution: This is just the probability that the regulator predicts a good year and the year is also good or the regulator predicts a good year and the year is bad. From the tree the probabilities are 0.64 and 0.08, respectively. Thus the probability that the regulator will say that the next year will be a good year is \( 0.64 + 0.08 = 0.72 \).

d) (5 points)

What is the probability that the regulator predicts the next year accurately?

Solution: This is just the probability that the regulator predicts a good year and the year is also good or the regulator predicts a bad year and the year is bad. From the
tree the probabilities are 0.64 and 0.12, respectively. Thus the probability that the regulator will say that the next year will be a good year is $0.64 + 0.12 = 0.76$

(5 points) Construct your probability tree below:
Question 2) Incoming Notebooks  (25 points)

The Dean of the College of Arts and Sciences is thinking about requiring next year’s incoming freshman class to buy a notebook computer for school work.

To see if the current students are generally in favor of this idea, he randomly selects 9 of the current freshman class and asks them personally.

Suppose that, in fact, 55% of the current freshman class is in favor of notebooks. The Dean does not know this, although you may use this figure in your computations here.

a) (3 points)

What is the probability distribution of the number of students that are in favor of the notebook requirement?

Solution: Let X be the random variable describing the number of the nine students selected by the Dean who favor notebooks.

From the description, X is binomial with probability \( p = 0.55 \) and \( n = 9 \) trials.

For X to be binomial we need to assume that the \( n = 9 \) students are chosen independently of each other.

b) (3 points)

What is the expected number of students, out of the 9 interviewed by the Dean, who will be in favor of the notebook requirement?

Solution: The expected value of a binomial random variable is

\[
E(X) = n \times p = 9 \times 0.55 = 4.95
\]

students in favor of notebooks.

c) (3 points)

What is the standard deviation of the number of students, out of the 9 interviewed by the Dean, who will be in favor of the notebook requirement?

Solution: The standard deviation of a binomial random variable is

\[
\sigma = \sqrt{\text{V}(X)} = \sqrt{n \times p \times (1-p)} = \sqrt{9 \times 0.55 \times 0.45} = 1.49
\]

students.

d) (3 points)

What is the standard deviation of the percentage of students out of the 9 interviewed by the Dean, who will be in favor of the notebook requirement?

Solution: The binomial percentage is \( f = X/n \). The standard deviation of a binomial percentage is

\[
\sigma_f = \sqrt{\text{V}(f)} = \frac{p \times (1-p)}{\sqrt{n}} = \frac{0.55 \times 0.45}{\sqrt{9}} = 0.1658 = 16.58\%
\]

of the students.
e) (5 points)

What is the probability that 7 or more of the 9 students will be in favor of the notebook requirement?

**Solution:** The required probability is:

\[
P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9)
\]

\[
= \binom{9}{7}0.55^70.45^2 + \binom{9}{8}0.55^80.45^1 + \binom{9}{9}0.55^90.45^0
\]

\[
= 0.1110 + 0.0339 + 0.0046
\]

\[
= 0.1495
\]

f) (8 points)

Suppose the Dean becomes more ambitious and samples a total of 40 students.

What is the probability that \(7/9 = 77.78\%\) or more of the 40 students interviewed by the Dean will be in favor of the notebook requirement?

**Hint:** Use the normal approximation to the binomial.

**Solution:** Let \(W\) be the number of winners; \(W\) is binomial with \(n = 40\) and \(p = 0.55\). Thus \(W\) has approximately a normal distribution with mean

\[
\mathbb{E}(W) = np = 40 \times 0.55 = 22
\]

and standard deviation

\[
\sigma_W = \sqrt{np(1-p)} = \sqrt{40 \times 0.55(1-0.55)} = 3.15
\]

If the proportion is \(77.78\%\) or more of the 40 students, then this is 32 or more students. We want to know \(P(W \geq 32)\). From the table this is 0.0008. and we’ll add a continuity correction (not needed for the exam!) to convert this to \(P(W \geq 31.5)\). In this case it is 0.0013. That is, the probability that \(7/9 = 77.78\%\) or more of the 40 students interviewed by the Dean will be in favor of the notebook requirement is about one in a thousand.
Question 3) Election and Gambling (25 points)

This question involves two separate scenarios:

a) (15 points)

In the presidential election in November 2004, the vote was very close. In fact George Bush received 51% of the vote compared to John Kerry’s 49%.

Suppose you conduct a random poll of 10 randomly selected voters and ask them if they voted for George Bush.

i) (10 points)

What is the probability that between 4 and 6 (inclusive) of them voted for George Bush?

**Solution:** Let X be the random variable describing the number of votes for George Bush. For X to be binomial we are assuming that all of the $n = 10$ the outcomes are independent of each other. This in insured by the random sampling mechanism.

From the description X is binomial with probability $p = 0.51$ and $n = 10$ trials.

The required probability is then:

$$P(4 \leq X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \binom{10}{4}0.51^40.49^6 + \binom{10}{5}0.51^50.49^5 + \binom{10}{6}0.51^60.49^4$$

$$= 0.1966 + 0.2456 + 0.2130$$

$$= 0.6552$$

ii) (5 points)

Approximately, how different will the percentage who voted for George Bush (from the poll) be from the 51% in the population that voted for George Bush?

**Solution:** The quantity asked for is clearly a measure of spread and the standard deviation in particular. The standard deviation of a binomial proportion is

$$\sigma_f = \sqrt{V(f)} = \frac{\sqrt{p \times (1 - p)}}{\sqrt{n}} = \frac{\sqrt{0.51 \times 0.49}}{\sqrt{10}} = 0.158 = 15.8\%,$$

on average away from the 51% in the population that voted for George Bush.
b) (10 points)

The probability of winning a certain game of chance is 0.25. Suppose that you decide to play the game 625 times. Assume that successive plays of the game are independent.

What is the probability that you will have 125 or fewer winners?

**Solution:** Let $X$ be the number of winners; $X$ is binomial with $n = 625$ and $p = 0.25$. We want to know $P(X \leq 125)$. From the table this is 0.0019.
Question 4) Government Employee Drug Use (25 points)

Illegal drug usage by state government employees can disrupt their lives and lead to poor performance on the job. In Seattle, the usage of (illegal) drugs has been reliably found to be 10% of the state employees. They are very easy to detect in a blood sample, and the test is infallible.

You wish to aid the state government in the screening of a large number of employees for illegal drug use. As an efficiency measure, they combine blood samples from three employees. The combined blood sample is then tested for illegal drugs. If this sample turns out to be negative, then three people are cleared at once. The blood sample will show illegal drug use if at least one of the people is using illegal drugs.

a) (7 points)

Consider any randomly chosen group of three employees. Give the individual probabilities that 0, 1, 2 or 3 of them is using illegal drugs.

Solution: Let X be the random variable describing the number of people using illegal drugs. Then X has a binomial distribution for $n = 3$ trials and probability $p = 0.1$. Using the formula from class $P(X = x) = \binom{3}{x}0.1^x0.9^{3-x}$ we can calculate the probabilities to be:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.729</td>
</tr>
<tr>
<td>1</td>
<td>0.243</td>
</tr>
<tr>
<td>2</td>
<td>0.027</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
</tr>
</tbody>
</table>

b) (3 points)

What is the probability that a combined blood sample from three people will show evidence of illegal drug use?

Solution: The combined sample will show evidence of illegal drug use if any of the people have it. That is, $X > 0$.

$$P(X > 0) = P(X = 1 \ or \ X = 2 \ or \ X = 3) = 0.243 + 0.027 + 0.001 = 0.271$$

c) (3 points)

If a combined blood sample turns up positive all three people are tested. What is the expected number of people tested from each combined blood sample?

Solution: From b) the probability that a combined sample will show evidence of illegal drug use is 0.271. Let $Y$ be the number of people tested from a combined blood sample. Then, from b), the probability $Y = 3$, that is, a combined sample will show evidence of illegal drug use is 0.271. Otherwise nobody is tested so $Y=0$. The expected value of $Y$ is then:

$$\mathbb{E}(Y) = 0.271 \times 3 + (1 - 0.271) \times 0 = 0.813 \text{ people}$$
d) (4 points)
The State is thinking of combining blood samples from four, rather than three people. What is the probability that a combined blood sample from four people will show evidence of illegal drug use?

Solution: This is similar to a). Let Z be the random variable describing the number of people from a sample of 4 who use illegal drugs. Then Z has a binomial distribution for \( n = 4 \) trials and probability \( p = 0.1 \). Using the formula from class \( \Pr(X = x) = \binom{4}{x}0.1^{x}0.9^{4-x} \) we can calculate the probabilities to be:

<table>
<thead>
<tr>
<th>z</th>
<th>( \Pr(Z=z) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.656</td>
</tr>
<tr>
<td>1</td>
<td>0.292</td>
</tr>
<tr>
<td>2</td>
<td>0.049</td>
</tr>
<tr>
<td>3</td>
<td>0.004</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The combined sample will show evidence of illegal drug use if any of the people have it. That is, \( Z > 0 \).

\[ \Pr(Z > 0) = 1 - \Pr(Z = 0) = 1 - 0.656 = 0.344 \]

e) (4 points)
What is the expected number of people re-tested from a combined blood sample?

Solution: Using the same approach as in c), the expected number of people tested is

\[ 0.344 \times 4 + (1 - 0.344) \times 0 = 1.376 \text{ people} \]

f) (4 points)
Would you suggest your company increase its combining to four people from three? Justify your answer with reference to the above analysis.

Solution: Increasing the pooling helps a little.

Suppose the center tests 12n people. If pooling in threes, there will be 4n combined samples each expecting 0.813 people to be tested. So we expect

\[ 4n \times 0.813 = 3.252n \]

people to be tested, plus one test for each combined sample (4n) leading to a total of 7.252n tests.

If pooling in fours, there will be 3n combined samples each expecting 1.376 people to be tested. So we expect

\[ 3n \times 1.376 = 4.127n \]

people to be tested, plus one test for each combined sample (3n) leading to a total of 7.127n tests.

As \( 7.127n < 7.252n \), pooling helps a little.
Question 5) New Parks and Reinvigorating neighborhoods (25 points)

This question involves two separate calculation and estimation scenarios:

a) (12 points)

The King County government is thinking of buying a group of buildings in Seattle and converting the space they cover to a park. They are concerned that the value of the buildings is less than the asking price. To aid in the decision-making they are able to obtain independent valuations on 15 of the buildings in the group. The rule that the government has chosen to decide whether they will purchase the entire group of buildings is the following: “Randomly select 15 buildings. If the sum of the 15 individual valuations exceeds 4.8 million dollars, then buy the group. Otherwise do not buy the group.”

If the true distribution of valuations of the buildings in the group is approximately normal with mean $\mu = 0.36$ million dollars and standard deviation $\sigma = 0.08$ million dollars, what is the probability that the government will buy the group, and hence create the park?

**Solution:** Let $X$ be the random variable that takes the value of a randomly chosen building from the group. Then the independent valuations form a random sample of $n = 15$ valuations, each with this distribution. The distribution of $X$ is approximately normal with mean $\mu = 0.36$ million dollars and standard deviation $\sigma = 0.08$. Hence the distribution of $\overline{X}$ is approximately normal with mean $\mu = 0.36$ million dollars and standard deviation $\sigma = 0.08/\sqrt{15} = 0.02$. It follows that

$$P( \text{“King County buys”} ) = P( \text{“total valuation exceeds 4.8 million dollars”} )$$

$$= P( \overline{X} > \frac{4.8}{15} )$$

$$= P( \overline{X} > 0.32 )$$

$$= 1 - P( \overline{X} < 0.32 )$$

$$= 0.9772$$

Thus it is virtually certain that King County will buy the group.
b) (13 points)

You represent a foundation that funds small community groups who are trying
to reinvigorate city neighborhoods. Only about one in sixteen (0.0625) of the
funded projects produce results that reinvigorate a neighborhood. Suppose you
independently fund 80 groups.

Using the normal approximation to the binomial, what is the probability that
this will result in more than eight neighborhoods being reinvigorated?

**Solution**: Let \( Y \) be the random variable describing the number of reinvigorated neighbor-
hoods funded. Then \( Y \) is binomial with \( n = 80 \) trials each with probability of
reinvigorating a neighborhood \( p = 0.0625 \). Then

\[
P( Y > 8 ) = \sum_{i=9}^{80} \binom{80}{i} 0.0625^i 0.9375^{80-i}
\]

This sum is difficult to calculate directly, but we will use the normal approxi-
mation the binomial to calculate the probability. The mean of \( Y \) is \( n \times p = 80 \times 0.0625 = 5 \) reinvigora-
tions. Note that \( n \times p = 80 \times 0.0625 = 5 \) so that using the normal approximation seems reasonable. The variance of \( Y \)
is \( n \times p \times (1 - p) = 80 \times 0.0625 \times 0.9375 = 4.6875 \). The standard devia-
tion of \( Y \) is then \( \sqrt{4.6875} = 2.165 \). Without the continuity correction the result is \( P(Y > 8) = 0.0829 \). With the continuity correction the result is
\( P(Y > 8.5) = 0.053 \). Hence the probability that you will find more than eight
reinvigorations is 0.053. That is, it is quite unlikely that you will find more than
eight reinvigorations.

The exact probability can be calculated to be 0.0618.
The Internal Revenue Service wants to investigate the amount of tax paid by former UW social science students via its offices in Seattle and Issaquah. The IRS audits 10 randomly selected students from the thousands of students reporting to the Seattle office. The first row of the data below are the amounts of tax paid (in thousands of dollars). The second row reports the amounts paid by 9 randomly selected students from the Issaquah office.

<table>
<thead>
<tr>
<th>Seattle (X)</th>
<th>Issaquah (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>6</td>
</tr>
<tr>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>39</td>
<td>6</td>
</tr>
</tbody>
</table>

The sample means and standard deviations are:

- Seattle: $\bar{X} = 24.00$, $\sigma_X = 12.78$
- Issaquah: $\bar{Y} = 20.00$, $\sigma_Y = 5.15$

Note that these are the standard deviations of the populations of students not the standard errors of the means.

**a) (6 points)**

What is the standard error for the sample mean tax paid to the Seattle office.

**Solution:** A standard error of the sample mean tax paid is

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{12.78}{\sqrt{10}} \approx 4.041$$

as $n = 10$, and $\sigma_X = 12.78$.

**b) (3 points)**

Which of the three distributions: normal, uniform or binomial do you think will provide the best approximation to the shape of the taxes paid? Justify your answer?

**Solution:** As the tax values are close to continuous and bounded below by zero, but not bounded above by an amount, the most likely shape is normal.

**c) (9 points)**

Using your choice of distribution, compute the proportion of students from the Seattle office that will pay less than $20,000 in tax?

**Solution:** We can characterize the normal distribution by its mean and standard deviation estimated by 24 and 12.78, respectively. From the values given in the table below: $P(X < 2000)$ when $\mu = 24$ and $\sigma = 12.78$ is 0.3771.

**d) (7 points)**

The office wants to construct an upper bound on the tax paid by students to the Seattle office. Find the level of tax paid that only 1% of the students will exceed.

**Solution:** This is the value with the area to the right of 1%. This is 53.73 thousand dollars, or $53,7300.$
Some numbers

If $X$ is binomial with $n = 100$ and $p = 0.25$ then the Prob. that $X$ is exactly 50 is 0.0000.

If $X$ is binomial with $n = 100$ and $p = 0.5$ then the Prob. that $X$ is exactly 50 is 0.0796.

If $X$ is binomial with $n = 625$ and $p = 0.25$ then the Prob. that $X$ is exactly 125 is 0.0005.

If $X$ is binomial with $n = 625$ and $p = 0.25$ then the Prob. that $X$ is at most 124 is 0.0014.

If $X$ is binomial with $n = 625$ and $p = 0.25$ then the Prob. that $X$ is at most 125 is 0.0019.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the left of 12.78 is 0.19.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the right of 20 is 0.6229.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the left of 20 is 0.3771.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the right of 10 is 0.8633.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the left of 10 is 0.1367.

If $X$ is normal with mean = 0.36 and Std. Dev. = 0.08 then the Area to the left of 0 is 0.0.

If $X$ is normal with mean = 0.36 and Std. Dev. = 0.08 then the Area to the left of 0.3 is 0.2266.

If $X$ is normal with mean = 0.36 and Std. Dev. = 0.08 then the Area to the left of 0.4 is 0.6915.

If $X$ is normal with mean = 0.36 and Std. Dev. = 0.08 then the Area to the left of 0.32 is 0.0228.

If $X$ is normal with mean = 0.36 and Std. Dev. = 0.08 then the Area to the left of 0.42 is 0.9987.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the left of 53.73 is 0.99.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the right of 24 is 0.5.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the left of 2.979 is 0.05.

If $X$ is normal with mean = 24 and Std. Dev. = 12.78 then the Area to the right of
45.92 is 0.05.

If $X$ is normal with mean = 5 and Std. Dev. = 2.165 then the Area to the right of 8 is 0.0828.

If $X$ is normal with mean = 22 and Std. Dev. = 3.15 then the Area to the right of 32 is 0.0008.

If $X$ is normal with mean = 22 and Std. Dev. = 3.15 then the Area to the right of 31.5 is 0.0013.

If $X$ is normal with mean = 5 and Std. Dev. = 2.165 then the Area to the right of 8.5 is 0.053.