Practice Problems for Test II

**Penicillin study.** A biomedical research firm developed a new penicillin manufacturing process, and offered our company an exclusive right to use their method. The research firm provided us with their report. The penicillin yield (unit/mg) were measured for seven types of base blend (B1 to B7) to produce penicillin (see the box plots below). Method I and II refer respectively to the current method and the new method developed by the research firm.

![Box plots for penicillin study](image)

**Trout hemoglobin measurement.** The data in the following table show the measurements of hemoglobin (grams per 100 ml) in the blood of brown trout. The trout were placed at random in four different troughs. The fish food added to the troughs contained, respectively, 0, 5, 10, and 15 grams of sulfamerazine per 100 pounds of fish (coded T1, T2, T3, and T4). The measurements were made on ten randomly selected fish from each trough after 35 days (see the box plots below).

![Box plots for trout hemoglobin measurement](image)

Given here is the ANOVA table obtained from the measurement data.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>test.statistic</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>3</td>
<td>26.80275</td>
<td>8.93425</td>
<td>5.695543</td>
<td>0.002684745</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>56.471</td>
<td>1.568639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>83.27375</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Hubble constant.** In 1929 Edwin Hubble investigated the relationship between distance and velocity of extra-galactic nebulae (celestial objects). He published the data about how galaxies are moving away from us no matter which direction we look, and hypothesized the so-called “Hubble's law” as follows:

Velocity = (Hubble's constant) * Distance

Given here are the scatter plot with fitted model (below left) and the residual plot (below right) which were produced from 24 data points Hubble published in 1929.

Results of linear regression are summarized in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.error</th>
<th>test.statistic</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-40.78365</td>
<td>83.43887</td>
<td>-0.4887848</td>
<td>0.6298312</td>
</tr>
<tr>
<td>Slope</td>
<td>454.1584</td>
<td>75.23711</td>
<td>6.036362</td>
<td>4.48E-06</td>
</tr>
</tbody>
</table>
**Problem 1.** Currently in making aluminum castings, an average of 3.5 ounces per casting must be trimmed off and recycled as a raw material. A new manufacturing procedure has been proposed to reduce the amount of aluminum that must be recycled in this way. For a sample of 12 castings made with the new process, the following table shows the summary statistics for the weights of aluminum trimmed and recycled.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D</th>
<th>L.Quartile</th>
<th>Median</th>
<th>U.Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>3.15</td>
<td>0.9070431</td>
<td>2.575</td>
<td>3.25</td>
<td>3.65</td>
</tr>
</tbody>
</table>

1) To test whether the new process reduces the amount of trimmed aluminum, state the *null hypothesis* using the population mean $\mu$ of trimmed aluminum from the new procedure. The test is to find whether $\mu < 3.5$ or not. Then the null hypothesis must be that $\mu = 3.5$. (In the test you must also choose the alternative hypothesis “$\mu \neq 3.5$,” or “$\mu < 3.5$”)

2) We conclude that there is no evidence that the new process reduces the amount of trimmed aluminum. What was your decision regarding the null hypothesis? The null hypothesis could not be rejected.

3) We have chosen the significance level of 0.05. In justifying your answer above, which of the following is possibly the *correct* statement about the p-value? (iv) [The finding not to reject the null hypothesis with significance level 0.05 should be implied by (iv).]

   i. p-value is less than 0.01.
   ii. p-value is between 0.01 and 0.05
   iii. p-value is equal to 0.05.
   iv. p-value is greater than 0.05.

**Problem 2.** Answer the following questions regarding penicillin study (see Penicillin study).

4) Here $\mu_1$ and $\mu_2$ denote the respective mean penicillin yield from Method I and II. State the *null and alternative hypothesis* which is suitable for the purpose of study. The null hypothesis is $\mu_1 = \mu_2$, and the alternative hypothesis is $\mu_1 < \mu_2$. (Or, you may choose the alternative hypothesis that $\mu_1 \neq \mu_2$.)

5) The penicillin study report says that the p-value is 0.018. State your choice of the significance level and your finding regarding the null hypothesis. If you choose the significance level 0.01 then we fail to reject the null hypothesis. If you choose the significance level 0.05 or 0.1 then we reject the null hypothesis.

6) Write your own conclusion for the study in a way consistent with the previous answer. Do not use statistical terms in your conclusion. The result with significance level 0.01 indicates no evidence regarding the benefit of new method. Observe in the box plots that Method II yield is higher. Thus, if you choose the significance level 0.05 or 0.1 then you conclude that there is some benefit of adapting the new method.
Problem 3. A study compares hospital stays (in days) between HMO (health-maintenance organization) patients and non-HMO patients, and the summary statistics are obtained. A researcher is interested in whether the average length of hospital stays (in days) are different between the two groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample size</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMO</td>
<td>50</td>
<td>3.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Non-HMO</td>
<td>50</td>
<td>4.2</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Here \( \mu_1 \) and \( \mu_2 \) denote the respective mean hospital stay for HMO and non-HMO.

7) State the null and alternative hypothesis to investigate the researcher’s inquiry. The null hypothesis is \( \mu_1 = \mu_2 \), and the alternative hypothesis is \( \mu_1 \neq \mu_2 \).

8) The result of test shows that the p-value is 0.005. State your choice of the significance level, and explain your finding regarding the null hypothesis. The null hypothesis is rejected.

9) Write your own conclusion for the researcher’s inquiry. Do not use statistical terms in your conclusion. There is a difference in hospital stay between the two groups. Since (i) the difference is statistically significant and (ii) HMO has the shorter average stay, HMO does a better job by managing a shorter hospital stay.

Problem 4. Answer the following questions regarding trout hemoglobin (see Trout hemoglobin measurement).

10) The researchers want to determine whether sulfamerazine has any effect on the hemoglobin content of trout blood. Write your observation based on the box plots in Data sheet. The box plots indicate that there is no particular trend. Hemoglobin measurements are the lowest when no sulfamerazine was added. But the initially high measurement of hemoglobin decreases as the content of sulfamerazine increases. Overall, it suggests some effect on the hemoglobin level caused by sulfamerazine.

11) Write their null and alternative hypothesis. The null hypothesis is that all the groups have the same mean value, and the alternative hypothesis is that their mean values are different.

12) Explain how ANOVA table would help you find a result regarding the hypothesis test. According to the ANOVA table the ratio of mean squares (MS) is 5.7. The ratio 5.7 is high enough to suggest that the hemoglobin measurements of individual groups are different. Furthermore, the p-value 0.0027 indicates the rejection of null hypothesis, and provides statistical evidence for the observation made initially.

13) Choose the significance level of your own, and state the result regarding the null hypothesis.
Since the p-value is very small, the null hypothesis can be rejected.

14) Write your finding for general audience. Do not use statistical terms in your conclusion. **We can conclude that sulfamerazine additive makes a difference in hemoglobin level of brown trout, and therefore, that sulfamerazine has some effect on the hemoglobin content.**

**Problem 5.** Answer the following questions regarding the distance and the velocity of extra-galactic nebulae (celestial objects). (see **Hubble constant**).

15) Which variable, Velocity or Distance, should be the explanatory variable? **Distance.**

16) We test the null hypothesis that the slope is zero. What should you find about the null hypothesis? Choose the significance level and justify your answer. **The p-value of \(4.5 \times 10^{-6}\) is very small and the null hypothesis must be rejected.**

17) The test was two-sided. Can you support that there is a positive relationship between the distance and the velocity of extra-galactic nebulae? Justify your answer. **Since (i) the slope should not be zero by the hypothesis test and (ii) the estimated slope of 454.2 is positive, there is a positive correlation.**

18) We tested the null hypothesis that the intercept is zero, and obtained the p-value of 0.63. What should you find about the null hypothesis? Choose the significance level and justify your answer. **Since the p-value is quite large, we cannot reject the null hypothesis, suggesting that the intercept could be zero.**

19) Does the Hubble's law seem appropriate from the data? Justify your answer. **Since we have failed to reject the null hypothesis that the intercept could be zero, we find no evidence to reject the (null) hypothesis that the intercept is zero. That is, there is no evidence against the Hubble’s law asserting the intercept being zero.**

20) We have obtained the 95% confidence interval (302, 562) for the slope. The Hubble's constant is now thought to be about 75. Does the data published in 1929 support this Hubble's constant up to date? Justify your answer. **No, since the confidence interval (302, 562) for the slope does not contain the value 75.**