Solutions to Exam 1

1. A study compared three treatments for dandruff and a placebo. The treatments were 1% pyrithone zinc shampoo (PyrI), the same shampoo but with instructions to shampoo twice (PryII), 2% ketoconazole shampoo (Keto), and a placebo (Placebo). After six weeks of treatment, eight sections of the scalp were examined and given a score for flaking on a scale from 1 to 10. The scores of the eight sections were added to obtain an overall "flaking score." There were 112 subjects in the PyrI group, 109 in the the PyrII group, 106 in the Keto group and 28 in the Placebo group. Numerical summaries are given below.

PyrI group:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.00	17.00	17.00	17.39	18.00	20.00

PyrII group:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.0	16.0	17.0	17.2	18.0	21.0

Keto group:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.00	15.00	16.00	16.03	17.00	18.00

Placebo group:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
26.00	28.00	29.00	29.39	31.00	32.00

(a) Use the number summaries to sketch box plots of the scores for each of the groups in the same graph.

(b) Are there any trends that the graph in the previous part reveals? Explain your answer.

***Solution**:* The box plots are shown in Figure 1.

The flaking scores for the three dandruff treatment groups are considerably lower than that for the placebo group. There doesn't seem to much difference between the two pyrithone zinc treatments. The ketoconazole shampoo appears to work a little better than the pyrithone zinc shampoo. □

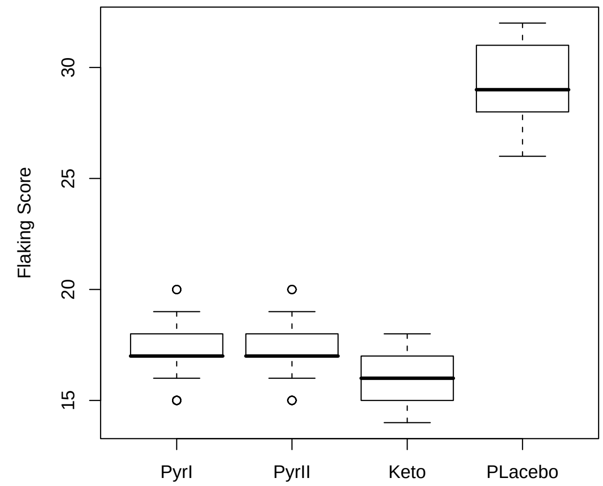

Figure 1: Box Plots of Flaking Scores for Various Dandruff Treatments

2. The following questions relate to significance tests.

(a) Define the p–value of a significance test.

Answer: The p–value for a significance test is the probability that the test statistic will take on the observed value, or more extreme values, under the assumption that the null hypothesis, H_o, is true. □

(b) Explain why a significance test that is significant at the 1% significance level must also be significant at the 5% level.

Answer: If the results are significant at the 1% significance level, then the p–value is less than 0.01. Consequently, the p–value is also less than 0.05, and so the results are significant at the 5% level as well. □

(c) A significance test is run on a given data set and you are told that the data provide evidence to reject the null hypothesis at the significance level of 5%. Is the test significant at the 1% level? Explain your answer.

Answer: If the p–value is less than 0.05, it does not follow that it is less than 0.01 necessarily. Therefore, the data could be significant at the 5% level, but not at the 1% level. □

(d) When asked to explain the meaning of "statistically significant at the 5% level," a student says that "there is a 1 in 20 chance that the null hypothesis is true." Is this a correct explanation of statistical significance? Explain your answer.

Answer: No, the student's explanation is not the correct one according to what we have learned in class. When determining statistical significance, the null hypothesis is assumed to be true. Based on that assumption, a p–value is computed or estimated. Statistical significance at the 5% level then means that the p–value, computed under the assumption that H_o is true, is less than 0.05; that is, the chances that the test statistic will take on the observed value, or more extreme values, under the assumption that H_o is true are less than 1 in 20. □

3. On January 28, 1986, the space shuttle Challenger exploded shortly after take-off and disintegrated over the Atlantic Ocean. It was later discovered that the explosion was due to fuel leakage caused by an O–ring seal failure. The night before, engineers at the company which built the shuttle warned NASA that the launch should be postponed due to predicted low temperatures. Low temperatures had been linked to O–ring seals failures. The data in Table 1 are taken from a graph in Richard P. Feynman's autobiographical book: *What do you care what other people think?* (New York: W. W. Norton, 1988). Feynman was a member of the commission that investigated the accident.

Launch Temperature																				
Below 65°F	1	1	1	3																
Above 65°F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2

Table 1: Number of O–ring failures in 24 space shuttle flights prior to 1/28/1986

(a) What is the question that needs to be answered based on analysis of the data in Table 1?

Answer: Is the number of O–ring seal failures affected by temperature? or, Are O–ring seal failures more frequent in days with temperature below 65°F. □

(b) State an appropriate null hypothesis for a significance test based on the data.

Answer: H_o : The number of O–ring failure incidents does not depend on whether the temperature is above or below 65°F □

(c) What test statistic would you use in the significance test?

Answer: We can take as a test statistic the total number of O–ring failure incidents in four days of temperature lower than 65°F. □

(d) Describe a randomization procedure that you would use to estimate the p–value for the test.

Answer: If we assume that H_o is true, then any four of the 24 numbers listed in Table 1 could have occurred on days when the temperature was below 65°F. Thus, we can select any four of the 24 numbers in Table 1 to simulated the four days in which the temperature was below 65°F. We can then add the total number of O–ring incidents in each random sample of 4 and compare with the total number that actually occurred in the four days of temperatures below 65°F. □

4. Refer to the data provided in Problem 3.

The histogram in Figure 2 shows the distribution of the number of O–ring failure incidents in 10,000 random samples of size 4 selected without replacement from a vector containing all the numbers in Table 1. The total number of incidents in the four days in each sample were stored in a vector called `Nfail`.

(a) Use the histogram in Figure 2 to estimate the p value for the significance test you formulated in Problem 3.

Solution: The p–value for the test formulated in the previous problem is the probability that in the a random sample of size 4 drawn without replacement from the 24 numbers in Table 1, the sum of the number of incidents in the sample is 6 or larger. This probability can be estimated by computing the proportion of times in the 10,000 simulations that the total sum of the numbers in each sample is 6 or above. This can be done by adding the frequencies corresponding to 6 and 7 and dividing by 10,000. We obtain that

$$p\text{–value} \approx \frac{90}{10000} = 0.009.$$

□

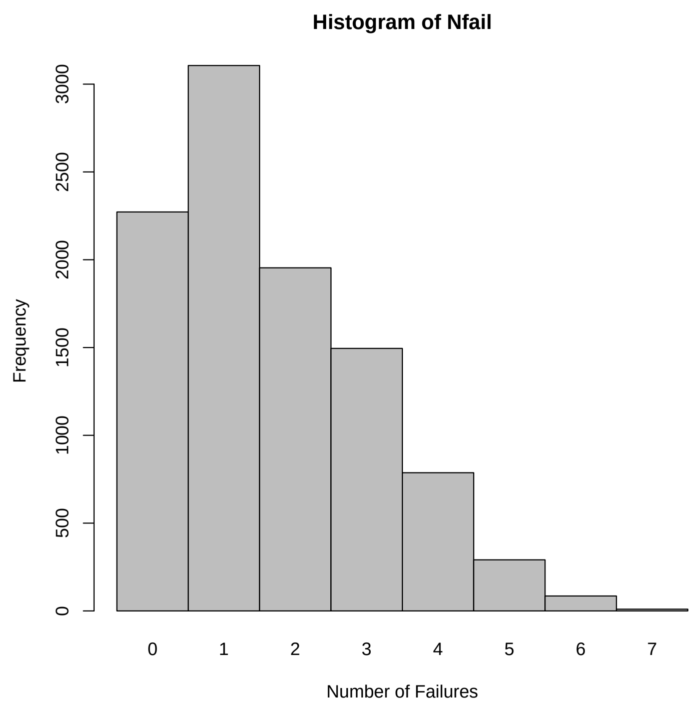

Figure 2: Distribution of simulated O–ring incidents in low temperature days

(b) What do you conclude?

Solution: The estimate for the p–value is below the 1% level of significance. It then follows that the data in Table 1 are statistically significant at the 1% significance level. Consequently, the null hypothesis can be rejected and we can conclude that there is strong evidence that the number of O–ring failures is associated with the temperature in the days of the 24 launches depicted in Table 1. □

5. Refer to the data provided in Problem 3.

Let N denote the number of O–ring failures in a random sample, without replacement, from the 24 observations in Table 1. The probability distribution for N is given in Table 2.

N	0	1	2	3	4	5	6	7
Probability	0.22398	0.31997	0.19198	0.14399	0.08046	0.02974	0.00894	0.00094

Table 2: Probability Distribution for N

(a) Compute the expected value of N. Give an interpretation of this result.

Solution: Compute

$$\begin{aligned} E(N) &= 0(0.22398) + 1(0.31997) + 2(0.19198) + 3(0.14399) \\ &\quad + 4(0.08046) + 5(0.02974) + 6(0.00894) + 7(0.00094), \end{aligned}$$

which is about 1.67. Thus, under the assumption that there is not association between the number of incidences of O–failure and the temperature on the day of the launch, then, on average, we expect to see about 1.67 failures per launch. □

(b) Use the formula $\text{Var}(N) = E(N^2) - [E(N)]^2$ to find the standard deviation of N.

Solution: Compute

$$\begin{aligned} E(N^2) &= 0^2(0.22398) + 1^2(0.31997) + 2^2(0.19198) + 3^2(0.14399) \\ &\quad + 4^2(0.08046) + 5^2(0.02974) + 6^2(0.00894) + 7^2(0.00094), \end{aligned}$$

or about 4.78. It then follows that

$$\text{Var}(N) \approx 4.78 - (1.67)^2 \approx 1.99.$$

Hence, the standard deviation of N is about 1.41.

□