**Chapter 5 Exercises**

**Concepts**

1. The choice of the value of 𝛂 influences the likelihood of Type I and Type II errors. If 𝛂 is increased, how does the likelihood of either type change? In what situation might an experimenter opt for a very small value of 𝛂?
2. A researcher wants to explore the distribution of an exponentially distributed variable like wealth. Will her test statistic still be normally distributed? Why or why not?
3. Why can it be misleading to assume that spatial data observations are independent? How can this assumption influence the conclusions a researcher draws?
4. A debate has emerged in the scientific community about the use and the mis-use of   
   *p*-values. What is a *p*-value and how might it be mis-used or mis-represented when make conclusions?

**Exercises**

1. What portion of the normal distribution is associated with the following ranges:
   1. Obtaining a *z*-value greater than *z* = 1.32
   2. Obtaining a *z*-value of less than z *=* -0.63
   3. Obtaining a *z*-value between *z* = 1.57 and *z =* 2.02
   4. Obtaining a *z*-value between *z =* -0.25 and *z =* 0.25
2. What are the critical values associated with the following situations using the *t-*distribution
   1. *A one-sided hypothesis test using p =* 0.10, *df* = 15
   2. *A one-sided hypothesis test using p = 0.05, df = 21*
   3. *A two-sided hypothesis test p = 0,01, df = 49*
3. Formulate the null and alternative hypotheses for each of the following scenarios.
   1. An advocacy group has approach a legislator claiming that the average amount of government assistance received by people who have not finished college are higher than the average for the population as a whole.
4. A sample of 120 students in a high school reveals a mean height of 1.62 meters and a standard deviation of 7 centimeters. Find 90% and 99% confidence intervals for the true mean height of the student population.
5. School district administrators are curious whether there is a difference between the rate of participation of different schools in a regional science fair. At one school, 8 of 40 students surveyed report that they plan to participate, while 25 of 117 students surveyed at another school plan to participate. Does the evidence support the conclusion that a difference in participation rates exists? State the null and alternative hypotheses, and use a Type I error level of 0.05. Find the test statistic, critical value, and *p*-value, and state your conclusion.
6. Within a large company, approximately 37% of employees are women. The engineering department is surveys, and of a sample of n = 42, 19% of the team members are women. Using a significance level of 𝛂=0.05, test the hypothesis that the gender ratio of the company’s engineering team does not differ from that of the company overall. State the null and alternative hypotheses, and give the critical value of the test statistic. State your conclusion, complete with *p*-value.
7. In a city where the mean age is 33.1, a sample of 40 residents in a particular neighborhood has a mean age is 30.5 with standard deviation *s* = 8.6. Does the evidence support the hypothesis that neighborhood residents tend to be younger than city residents overall? State the null and alternative hypotheses, give the critical value, and determine the test statistic. What is your conclusion? What is the *p*-value?
8. A gym wants to know how many times per year its pass-holding patrons tend to come in and exercise, within plus or minus 2 visits and at a 95% confidence level. A pilot study suggests that the variance in visits is approximately 30 visits per year. How many patrons should they sample?
9. State whether the probability of rejecting the null hypothesis is unnaturally low or unnaturally high in each of the following situations:
   1. A researcher believes that the unemployment rate in a country is a certain value and attempts to test whether the evidence supports this, ignoring the variation in employment rates among different regions.
   2. A school wants to determine whether its standardized test preparation program has in fact influenced the average student score on the test. They decide to compare the set of student test scores from before the program to the test scores of the same students after undertaking the program.
   3. A medical doctor is trying to test whether the fatality rate of a new strain of a particular disease has increased, and does not consider the fact that children react differently to the new strain than do adults.
10. As part of a class action lawsuit, the residents within a radius of 9 km of an industrial plant are surveyed regarding whether they have suffered from a particular illness. The average distance between the center of the study area and the 40 reported instances of that illness is 5.1 km. Test the null hypothesis that the individuals are randomly distributed around the center. What is the *p*-value?
11. Researchers believe that a tagged wild bird has made its nest in a particular tree. After careful observation over a square area of 49 km2 which is centered around the tree, they record 74 sightings of the bird, with a mean distance between the sightings and the tree being 2.5 km. Test the researchers’ null hypothesis that the sightings are randomly distributed around the center of the area using a two-tailed test and 𝛂=.05. What is the

*p*-value?

1. A city council wants to know about the usage of a public library. Specifically, they are interested in how many times children from different neighborhoods utilize the library every year. For neighborhood A, the sample of 11 individuals has a mean of 47.6 and a standard deviation of 7.2. For neighborhood B, the sample of 16 individuals has a mean of 52.3 and 8.7. Should a researcher assume these variances are equal? Use an *F*-test with 𝛂=.05. Then, choose the appropriate test for the hypotheses and test it.
2. A tourism geographer is interested in understanding differences in the number of visits white and non-white residents from a target make to national parks. From a survey of 23 white residents revealed that group make an average of 2.7 park trips per year with a variance of 0.9 trips. The same survey of 17 non-white residents shows that group makes an average of 2.3 visits per year with a variance of 1.1 trips. Because this geographer is exploring the data he has decided to use the 10% significance level for his testing. Does there appear to be a difference in park visitation between the groups?

**Chapter 5 Solutions**

* 1. 1-9066=0.0944
  2. 0.7357
  3. 0.9783-0.9418=0.0365
  4. 0.5987-0.4013=0.1974

a. *1.34*

1. *1.72*
2. *2.76*

*a. Ho: xno college ≤ µ Ha xno college > µ*

*b. Ho: xpollutiion ≤ 4.2 Ha xpollution > 4.2*

1. For a 90% confidence interval, we use z𝛂/2 = 1.645; for a 99% confidence interval, z𝛂/2 = 2.575. Thus the 90% CI is xmean ± z𝛂/2 \* (s/sqrt(n)) = 1.62 ± 1.645 ( .07 / sqrt(120)) = 1.62 ± .011, or (1.609, 1.631). The 99% CI is 1.62 ± 2.575 ( .07 / sqrt(120)) = 1.62 ± .016, or (1.604, 1.636)
2. HO: p1 = p2 HA: p1 < p2 The critical value is *z* = 1.645. p1 = 8/40 = 0.200, while p2 = 25 / 117 = 0.214. p = (8 + 25) / (40 + 117) = 0.210. Thus 𝛔p = sqrt( p(1-p)/n1 + p(1-p)/n2) = sqrt(.166/40 + .166/117) = 0.075. Thus our value of *z* = (p2 - p1) / 𝛔p = (0.214 - 0.200) / 0.075 = 0.187. This is less than the critical value of 1.645, and we fail to reject the null hypothesis. In this case, p = .426.
3. HO: 𝞀 = .37 HA: 𝞀 ≠ .37. The critical value is z = 1.645, and z = (p - 𝞀0) / sqrt(𝞀0(1 - 𝞀0) / n) = (.19 - .37) / sqrt(.37(1-.37) / 42) = -2.416. This is outside of the limits of our critical values, and we reject the null hypothesis. The probability of getting a statistic more extreme than this is p = .008.
4. HO: 𝝁 = 33.1 HA: 𝝁 ≠ 33.1. Choosing 𝛂=.05, the critical value is z = 1.645. Thus, z = (xmean - 𝝁) / (s / sqrt(n)) = (30.5 - 33.1) / (8.6 / sqrt(40)) = -1.912. This is beyond the critical value of z, and the evidence supports the hypothesis that 𝝁 ≠ 33.1. Specifically, p = 0.279.
5. n = z𝛂2 s2 / W2 = 1.962 \* 30 / [22] = 29 patrons.
6. a) high b) high c) low
7. z = (dmean - E[d]) / sqrt(V[d] / n) = (dmean - (2 \* r / 3)) / sqrt((r2 / 18) / n) = (5.1 - (2 \* 9 / 3)) / sqrt(92 / 18) / 40) = -2.68. This is significantly less than the critical value of -1.96, with *p* = .004. We reject the null hypothesis that points are randomly distributed around the center.
8. The area has a side *s* = sqrt(49) = 7. z = (dmean - E[d]) / sqrt(V[d] / n) = (dmean - 0.383*s*) / sqrt(.02*s*2 / n) = (2.5 - 0.383 *\* 7*) / sqrt(.02 \* 49 / 74) = -1.57. This is within the range of values between -1.96 and 1.96, with *p* = .058. We fail to reject the null hypothesis that points are randomly distributed around the center.
9. *F* = 8.72 / 7.22 = 1.46 < *Fcrit* = *F0.05,15,10* = 2.85, so we fail to reject the assumption of equal variances. Thus, *sp* = 8.1 and *t* = 1.529. Using the t-table with 25 degrees of freedom, the critical t-value for a two-tailed test implies that *t* should be greater than 2.06 or less than -2.06. Thus, we fail to reject the null hypothesis that the two neighborhoods have equivalent rates of library usage.
10. *F=1.43/1.2=1.19 <Fcirt=F0.05,19,15=2.40, which implies that we fail to reject the hypothesis of equal variances. The critical value for the t-statistic with df=20+16-2=34 is Tcrit0.05,34=-1.697. Notice that this critical value is relatively similar to that of the normal distribution as the pooled sample size is relatively large. Calculating the pooled variance we find*

*and using the pooled variance we calculate the t-stat as,*

*The t-statistic is between the 10% critical value of a two-sided test of -1.69 and 1.69. Therefore we fail to reject the null hypothesis.*