Homework 5

Tyler Stevenson

2020-11-01

Instructions

Use the R Markdown version of this file to complete and submit your homework. Items in **bold** require an answer. Make sure you change the author in the header to your own name.

Conceptual Questions

While these questions are labelled "Conceptual" you may, and probably should, use R to answer them.

1. Ruchdeschel et al. Ruckdeschel, Shoop, and Kenney (2005) claim that the sex ratio of Ridley's sea turtle (a very rare and endangered sea turtle) moved from a male biased ratio to a female biased ratio.

They recorded the sex of stranded seas turtles on Cumberland Island. From 1983 to 1989 there were 16 males and 10 females. From 1990 to 2001 there were 19 males and 56 females.

Is there evidence that the sex ratio of Ridley's sea turtles was male biased in the period 1983-1989, and female biased in the period 1990-2001?

For each of this period, conduct an appropriate test, construct a confidence interval and write a summary with your conclusions in the context of the study.

Test 1: Is there evidence that the sex ratio of Ridley's sea turtles was male biased in the period 1983-1989?

Null: The true proportion of male to female turtles is 0.5.

Alternative: the true proportion of male to female turtles is greater than 0.5.

Test 2: Is there evidence that the sex ratio of Ridley's sea turtles was female biased in the period 1990-2001?

Null: The true proportion of female to male turtles is 0.5.

Alternative: The true proportion of female to male turtles is greater than 0.5.

```
male_1983_89 <- prop.test(x = 16, alternative = c("greater"), n = 26, p = 0.5)
female_1990_01 <- prop.test(x = 56, alternative = c("greater"), n = 75, p = 0.5)

print("Test 1")
```

```
## [1] "Test 1"
```

```
male_1983_89
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  16 out of 26, null probability 0.5
## X-squared = 0.96154, df = 1, p-value = 0.1634
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.4361819 1.0000000
## sample estimates:
##         p
## 0.6153846
```

```r
print("Test 2")
```

```
## [1] "Test 2"
```

```r
female_1990_01
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  56 out of 75, null probability 0.5
## X-squared = 17.28, df = 1, p-value = 1.613e-05
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.649371 1.000000
## sample estimates:
##         p
## 0.7466667
```

With 95% confidence, the true sex ratio of Ridley's sea turtles is between 0.44 and 1.0 male biased for the years 1983-1989. We fail to reject the null hypothesis that the sex ratio of Ridley's sea turtles was male biased in the years 1983-1989.

With 95% confidence, the true sex ratio of Ridley's sea turtles is between 0.65 and 1.0 female biased for the years 1990-2001. With a very low p-value, and a confidence interval that excludes the null value, we have strong evidence to reject the null hypothesis.

In summary, it is unlikely that the sex ratio of Ridley's sea turtles was male biased in the years 1983-1989. However, in the years 1990-2001, it is likely that the sex ratio of Ridley's sea turtles was female biased.

2. *(From Ex 6. Chapter 4 Statistical Methods. Freund, R.; Mohr, D; Wilson,W. (2010))*

Average systolic blood pressure of a normal male is supposed to be about 129. Measurements of systolic blood pressure on a sample of 12 adult males from a community whose dietary habits are suspected of causing high blood pressure are (in R ready format):

```r
bp <- c(115, 134, 131, 143, 130, 154, 119, 137, 155, 130, 110, 138)
```

Do the data justify the suspicions regarding the blood pressure of this community?

Conduct an appropriate test, construct a confidence interval and write a summary with your conclusions in the context of the study.

Null Hypothesis: The average systolic blood pressure of an adult male from a suspected community is 129.

Alternative Hypothesis: The average systolic blood pressure of an adult male from a suspected community is not 129.

```
t.test(bp, mu=129, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  bp
## t = 0.9939, df = 11, p-value = 0.3416
## alternative hypothesis: true mean is not equal to 129
## 95 percent confidence interval:
##  124.142 141.858
## sample estimates:
## mean of x
##       133
```

With 95% confidence, we fail to reject the null hypothesis that the average systolic blood pressure of an adult male from a suspected community is 129.

3. *(Adapted From Ex 22. Chapter 4 Statistical Methods. Freund, R.; Mohr, D; Wilson,W. (2010))*

The following data gives the average pH in rain/sleet/snow for the two-year period 2004-2005 at 20 rural sites on the U.S. West Coast. (Source: National Atmospheric Deposition Program).

```
rain <- c(5.335, 5.345, 5.380, 5.520, 5.360, 6.285, 5.510, 5.340,
          5.395, 5.305, 5.190, 5.455, 5.350, 5.125, 5.340, 5.305,
          5.315, 5.330, 5.115, 5.265)
```

Is there evidence the median pH is not 5.4?

Conduct an appropriate test, construct a confidence interval and write a summary with your conclusions in the context of the study.

```
M <- median(rain)
M
```

```
## [1] 5.34
```

```
rain_less_M = rain<M
#rain_less_M
#rain_greater_M = rain>M
n = length(rain)


prop.test(sum(rain_less_M), n, p=0.5)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(rain_less_M) out of n, null probability 0.5
## X-squared = 0.05, df = 1, p-value = 0.8231
## alternative hypothesis: true p is not equal to 0.5
```

```
## 95 percent confidence interval:
##  0.2382855 0.6795196
## sample estimates:
##    p
## 0.45
```

```
#prop.test(sum(rain_greater_M), n, p=0.5)

lower_index <- round(n/2 - (qnorm(0.975)*sqrt(n))/2)
upper_index <- round(n/2 + (qnorm(0.975)*sqrt(n))/2 + 1)
#c(lower_index, upper_index)
rain_sorted = sort(rain);
rain_sorted[c(lower_index, upper_index)]
```

```
## [1] 5.305 5.380
```

There is significant evidence that the population median pH for rain is not 5.4. It is estimated the population median for this rain pH is 5.34. With 95% confidence, the population median is between 5.305 and 5.380.

R Question

This question explores the difference between the Normal distribution and t-distribution as reference distributions for a two sample comparison.

Begin by setting the seed to 1908:

```
set.seed(1908)
```

Then use `rexp()` to draw a sample of size 10 from an Exponential distribution with rate parameter 1:

```
exp_sample <- rexp(n = 10, rate = 1)
mean_1 <- mean(exp_sample)
std_1 <- sd(exp_sample)
```

a) It is helpful to be able to picture the Exponential distribution, so follow the steps below to **plot the distribution function curve.**

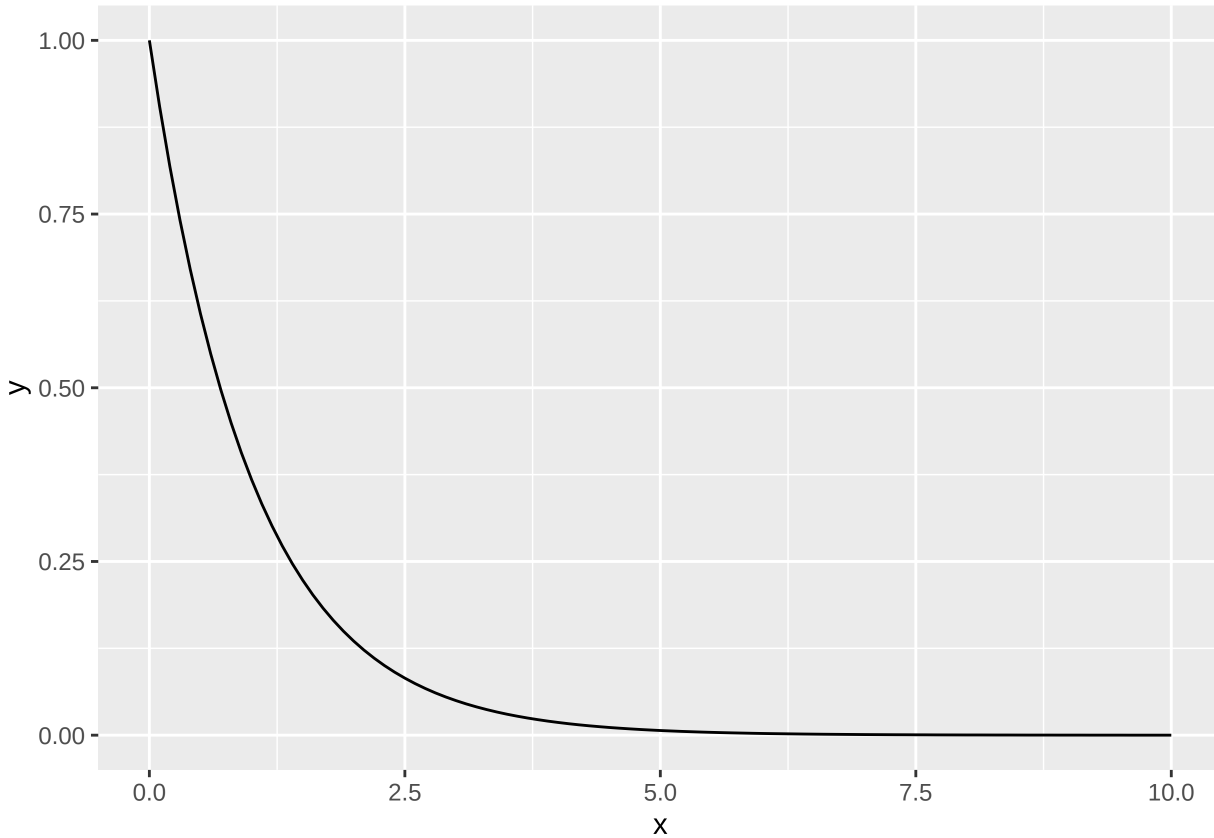
- First you need to create a vector of x-axis values, called `x`. The function `seq()` creates a sequence, and has as arguments `from`, `to`, and `by`.
- Then you need to find the values of the Exponential(1) distribution at those x-axis values. The function `dexp(x, rate = 1)` gives the value of the Exponential(1) distribution for the values stored as the vector `x`. Store these values as `y`.
- Then use `qplot(x, y, geom = "line")` to create a plot. Remember to load the `ggplot2` package.

```
library(ggplot2)

x <- seq(from = 0, to = 10, by = 0.1)

y <- dexp(x, rate = 1)

qplot(x,y,geom="line")
```



b) Run a t-test on your size 10 sample, for a null hypothesis that $\mu = 2$, against a two sided alternative. **Write a non-technical summary that includes an interpretation of the p-value and 95% confidence interval.**

```
exp_sample_t_test <- t.test(exp_sample, mu=2, conf.level = 0.95)
two_sided_t <- t.test(exp_sample, alternative=c("two.sided"), mu=2, conf.level = 0.95)
exp_sample_t_test
```

```
##
##  One Sample t-test
##
## data:  exp_sample
## t = -1.6165, df = 9, p-value = 0.1404
## alternative hypothesis: true mean is not equal to 2
## 95 percent confidence interval:
##  0.7924928 2.2010150
## sample estimates:
## mean of x
##  1.496754
```

```
two_sided_t
```

```
##
##  One Sample t-test
```

```
##
## data:  exp_sample
## t = -1.6165, df = 9, p-value = 0.1404
## alternative hypothesis: true mean is not equal to 2
## 95 percent confidence interval:
##  0.7924928 2.2010150
## sample estimates:
## mean of x
##  1.496754
```

Summary of Results: There is strong evidence that the true mean is not equalt to 2. Since the p-value is so low at 0.001563 and mu is outside of the confidence interval, it is unlikely that the true population mean is 2.

c) **Calculate a z-statistic (continue to use the sample SD, not population SD), for the same hypotheses as part b), and a p-value based on the normal distribution.**

```r
Z <- ((mean(exp_sample) - 2) / (sd(exp_sample)/sqrt(length(exp_sample))))
P <- 2 * pnorm(abs(Z), mean = 0, sd = 1, lower.tail = FALSE)

Z
```

```
## [1] -1.616477
```

```r
P
```

```
## [1] 0.1059913
```

d) **If the test statistic is the same for both tests, why is the p-value different?**

The test statistic is the same for both tests. However, the p-value is different. The reason for this difference is that the p-value is determined from different statistic tables (t-table and Z-table). Further, the p-value is found through different calculations for each distribution.

e) **Which is more appropriate in real life, where the population standard deviation is usually unknown?**

The student's t-distribution will be more appropriate for this sample in real life. By using s for an estimate of sigma as we do in a Z-test, we introduce a lot of extra uncertainty. The t-distribution helps limit this uncertainty. Further, due to the small sample size, the t-test is also a more reasonable test.

References

Ruckdeschel, Carol, C Robert Shoop, and Robert D Kenney. 2005. "On the Sex Ratio of Juvenile Lepidochelys Kempii in Georgia." *Chelonian Conservation and Biology* 4 (4): 858–61.