

QSCI 381 AB - AUT 22
Homework #6 - Vinsensius

The PDF file of the code is attached at the end of the HW

Question 1

Hawai'i is the most isolated archipelago geographically on the planet. With the arrival of people Hawai'i has become a hotspot of biological invasion with dramatic changes to the native ecosystems. The introduction of predators (mongoose and cats) has been particularly devastating to native birds.

Goltz et al. (2008) conducted a feral cat home range study on Mauna Kea, Hawai'i. They were able to reliably estimate the home range of six males.

The results are as follows (in square kilometers):

home ranges: 11.67, 20.5, 12.79, 11.76, 8.75, 8.31

Throughout this question you may assume that cat home ranges are normally distributed.

Reference: Goltz, Daniel M., et al. "Home range and movements of feral cats on Mauna Kea, Hawai'i." (2008).

We want to determine confidence intervals for male feral cat home ranges on Mauna Kea, Hawai'i.

1a

For the data of male cat home ranges, enter the data into R and calculate and report the following:

1. sample mean (1 pt)
2. standard deviation (1 pt)
3. sample size (1 pt)

Sample Mean	12.29 km ²
Sample Standard Dev	4.40 km ²
Sample Size	6

1b

Given the information provided, can the central limit theorem be applied in this scenario? (2 pts)

Yes, we can since the data is assumed to be normally distributed despite having $n < 30$.

1c

For a confidence level of 90% calculate the:

- i. Margin of error (4 pts)
- ii. The confidence interval bounds (3 pts)

E	3.62
CI bounds	$8.70 < \mu < 15.91$

1d

Provide a written statement regarding the range of values that are plausible with 90% confidence for the mean home range of males cats given this information (2 pts)

We estimate with 90% confidence that the true population mean is within $[8.70, 15.91]$.

1e

Repeat these calculations for a confidence level of 99%, providing the 99% confidence interval, and identify whether a mean home range of 20 square km is plausible to a 99% confidence (5 pts)

E	7.24 km ²
CI bounds	$5.06 < \mu < 19.54$

The mean home range of 20 km² is not plausible to 99% confidence interval.

Question 2: z-test

It is claimed that only 37% of American citizens have a valid US passport. To test this claim, you survey 348 and get the following data

$$n = 348$$

$$x = 141 \text{ (responded Yes: has a valid passport)}$$

2a

Write out the null and alternate hypotheses of this claim (2 pts)

H_0	$p = 0.37$
H_A	$p \neq 0.37$

2b

Calculate the test statistic, and the standardized test statistic for performing a one-sample z-test on this data (4 pts)

\hat{p}	0.405
z	1.36

2c

Calculate the critical value, z_{crit} , for rejecting the null hypothesis at a significance level $\alpha = 0.01$ (2 pts)

The test that will be performed is two-tailed test. So, the $z_{\text{crit}} = -2.58$.

2d

Given your answers in (a) – (c) what would you conclude regarding the initial claim. In your answer make sure to refer to the components of this test (statistic, claim, test thresholds) to support your answer (3 pts)

The rejection range is when $z < -2.58$ or $z > 2.58$. The z-score that we get in part (b) does not fall within the rejection region. So, we have no sufficient evidence to reject the null hypothesis.

Question 3: t-test

In this question we will look at a dataset of morphological measurements taken from two species of seaduck: Surf scoter and White-winged scoter. We will be using this dataset to examine mean bill lengths (mm).

Download the data from canvas and read it into R

```
scot <- read.csv(file="qsci_scooters.csv")
```

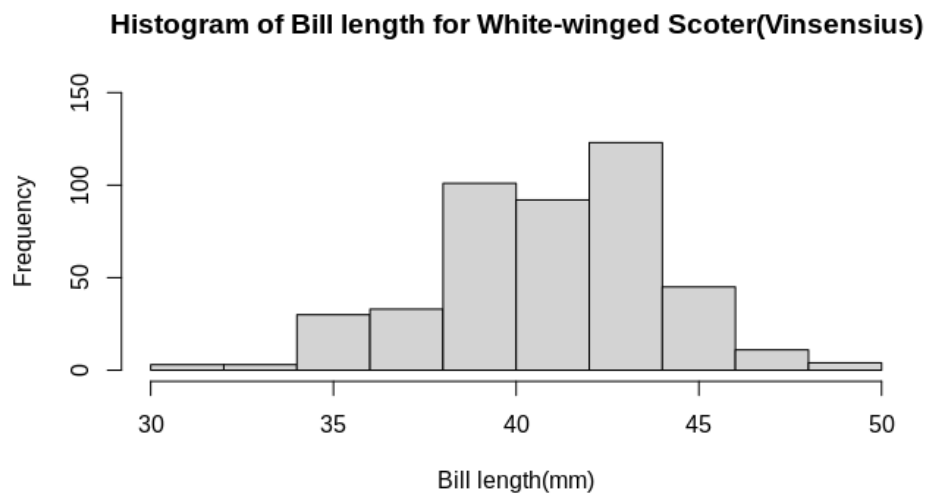
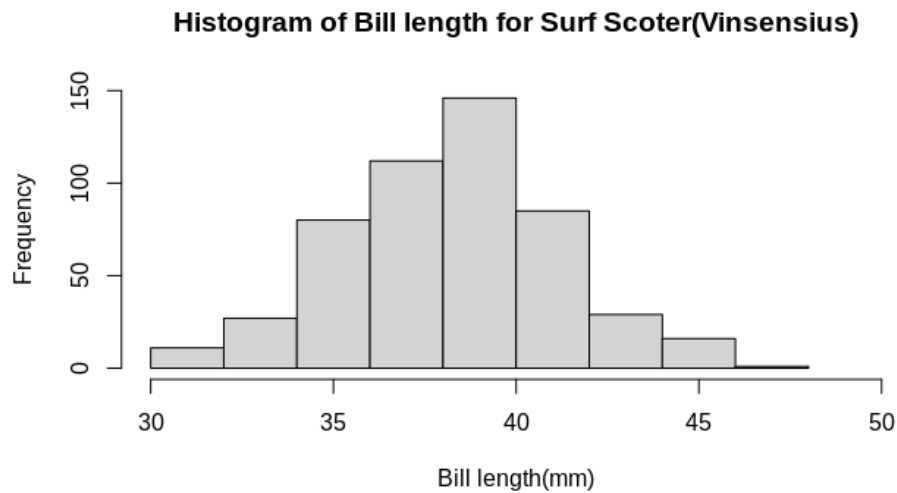
3a

Calculate summary statistics (mean, sd, and sample size) of bill length for each species given the sample data (6 pts)

	Surf Scoter	White-winged Scoter
Mean (mm)	38.74	41.44
Standard Dev. (mm)	2.98	3.08
Sample Size	507	445

3b

For each species, plot a histogram of bill lengths, making sure to label axes accordingly. Based on these do you think bill length measurements are normally distributed for each species? (5 pts)



Based on the figures above, the histogram of Surf scoter is more normally distributed. For the white-winged scoter, the histogram looks like left-skewed distribution.

3c

I claim that the mean length of Surf Scoter bills is at least 39 mm. What are the null and alternate hypotheses related to this claim, and which hypothesis is represented by my claim (3 pts)

H_0	$\mu \geq 39$
H_A	$\mu < 39$

The claim is supported by the null hypothesis.

3d

For the claim in (c) related to Surf Scoters, use the summary statistics calculated in (a) to calculate the

- i. Test statistic (2 pts)
- ii. Standard error of the test statistic (2 pts)
- iii. Standardised test statistic (1 pt)
- iv. The critical value for rejecting the null hypothesis at a significance level $\alpha = 0.05$ (3 pts)
- v. The p-value of the observed test statistic (2 pts)

\bar{x}	38.74
SE	0.132
t	-1.96
t_{crit} (left-tailed test)	-1.65
p-val	0.0255

3e

Given the answers to part (d) draw a conclusion regarding the original claim at a significance level $\alpha = 0.05$ (2 pts)

The rejection region will be $t < -1.65$. Thus, we will reject the null hypothesis in favor of alternate hypothesis. So, the claim about 39 mm being the mean of Surf scoter bill length is not true.

```
## Q1
```

```
home_range_dat = c(11.67, 20.5, 12.79, 11.76, 8.75, 8.31 )
```

```
home_range_mean = mean(home_range_dat)
```

```
home_range_std = sd(home_range_dat)
```

```
home_range_size = length(home_range_dat)
```

```
t_st = abs(qt(0.05,home_range_size-1))
```

```
E = t_st*home_range_std/sqrt(home_range_size)
```

```
up_lim_1 = home_range_mean+E
```

```
low_lim_1 = home_range_mean-E
```

```
t_st_2 = abs(qt(0.005,home_range_size-1))
```

```
E_2 = t_st_2*home_range_std/sqrt(home_range_size)
```

```
up_lim_2 = home_range_mean+E_2
```

```
low_lim_2 = home_range_mean-E_2
```

```
##### Q2
```

```
n_q2 = 348
```

```
p_hat = 141/348
```

```
p = 0.37
```

```
expected_mean = n_q2*p
```

```
expected_std = n_q2*(1-p)
```

```
SE_q2 = sqrt(p*(1-p)/n_q2)
```

```
z_score_q2 = (p_hat - p )/SE_q2
```

```
z_crit_q2 = qnorm(0.005)
```

```
##### Q3
```

```
scot_dat <- read.csv(file="qsci_scooters.csv")
```

```
surf_scooter_dat = scot_dat[scot_dat$Species == 'Surf Scoter',]
```

```
white_scooter_dat = scot_dat[scot_dat$Species == 'White-winged Scoter',]
```

```
surf_mean = mean(surf_scooter_dat$bill_length)
```

```
surf_std = sd(surf_scooter_dat$bill_length)
```

```
white_mean = mean(white_scooter_dat$bill_length)
```

```
white_std = sd(white_scooter_dat$bill_length)
```

```
#x_surf = seq(min(surf_scooter_dat$bill_length),  
#             max(surf_scooter_dat$bill_length), length=100)
```

```
#y_surf = dnorm(x_surf, surf_mean,surf_std)
```

```
#x_white = seq(min(white_scooter_dat$bill_length),  
#             max(white_scooter_dat$bill_length), length=100)
```

```
#y_white = dnorm(x_white, white_mean,white_std)
```

```
hist(surf_scooter_dat$bill_length,  
     main = "Histogram of Bill length for Surf Scoter(Vinsensius)"  
     ,ylim = c(0,150),xlim = c(30,50),xlab = "Bill length(mm)")
```

```
#points(x_surf, y_surf*1100, col='green', pch=12)  
hist(white_scooter_dat$bill_length,  
     main = "Histogram of Bill length for White-winged Scoter(Vinsensius)"  
     ,ylim = c(0,150),xlim = c(30,50),xlab = "Bill length(mm)")
```

```
#points(x_surf, y_surf*1100, col='green', pch=12)
```

```
n_surf = NROW(surf_scooter_dat)
```

```
SE_surf = surf_std/sqrt(n_surf)

mu_surf = 39

t_score = (surf_mean-mu_surf)/SE_surf

t_crit = qt(0.05,n_surf-1)

p_val_q3 = pt(t_score,n_surf-1)
```