Hickman_Midterm2Practice

Keith Hickman

November 8, 2017

Problem 1

(a) What is the experimental unit? The wolf spider is the unit, the measurement is walking speed.

(b) Give null and alternative hypotheses for an appropriate two-tailed t-test. If the null hypothesis is true, what is the distribution of the t-statistic? The null $H_0 : \mu = 0$

The researchers are interested in showing that praying mantis excrement has an effect on the walking speed of spider, so that's our alternative hypothesis, $H_1 : \mu \neq 0$, or that there will be some effect on the walking speed. This leaves $H_0 : \mu = 0$ at $\alpha = .05$

The experiment does not specify a direction, e.g. faster or slower walking speeds, so we will use a two-tailed test here (also, the question specifies it).

(c) The P-value (significance probability) was calculated to be 0.053, so the null hypothesis was not rejected. From this and the other information given, is it correct to conclude that we are sure that wolf spiders' walking speed is not affected by praying mantis excrement? Explain.

We can't reject the null for this experiment outright, but with a p-value so close to the significance level, rejecting the null and calling it a day isn't advisable either. Best solution is to gather more data.

Problem 2

We should use the binom probabilities.

```
p.hat <- 1844/3600
upper_bound <- p.hat + 1.96 * (sqrt(p.hat*(1-p.hat)/3600))
lower_bound <- p.hat - 1.96 * (sqrt(p.hat*(1-p.hat)/3600))
upper_bound
```

```
## [1] 0.5285507
```

```
lower_bound
```

```
## [1] 0.4958938
```

```
p.hat2 <- 828/1560
upper_bound2 <- p.hat2 + 1.96 * (sqrt(p.hat*(1-p.hat)/1560))
lower_bound2 <- p.hat2 - 1.96 * (sqrt(p.hat*(1-p.hat)/1560))
upper_bound2
```

```
## [1] 0.5555739
```

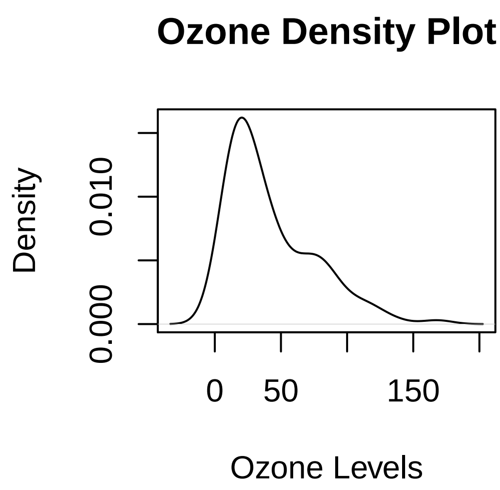
```
lower_bound2
```

```
## [1] 0.5059645
```

```
ozone <- read.csv("C:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module12/ozone.txt")
ozone <- ozone[,1]
summary(ozone)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   18.00   31.00   42.14   63.50  168.00
```

```
plot(density(ozone), xlab="Ozone Levels",main="Ozone Density Plot")
```



The variable is not normally distributed: it is right-skewed.

95% confidence interval:

```
mu <- mean(ozone)
se <- sd(ozone)/sqrt(116)
mu
```

```
## [1] 42.13913
```

```
se
```

```
## [1] 3.076237
```

```
upper <- mu + 1.96 * se
upper
```

```
## [1] 48.16855
```

```
lower <- mu - 1.96 * se
lower
```

```
## [1] 36.10971
```

Notes: be careful about calculating standard error: standard error is the standard deviation of the sample/the square root of n. To find confidence interval in an iid variable:

mean(variable) +/- conf.interval * sd(variable)/sqrt(n)

Problem 4: What parameters are we going for?

1) Mean/median or proportion or delta (mean_1 - mean_2)
2) Can we assume normality? Verify normality via use of plots.

- Calculate Z or T stats.
- If n is large, use Z.
- If proportion is parameter, use Z.

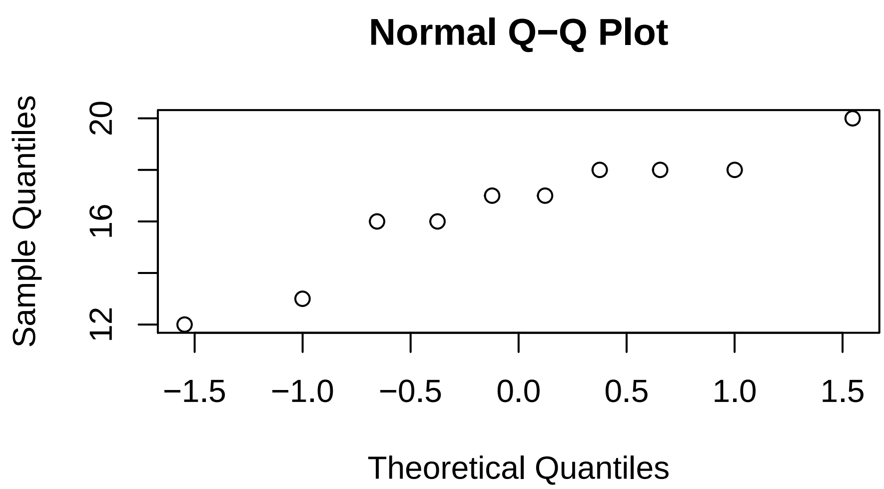
- If mean/delta is param:
  - if you know SD, use Z
  - if you don't know SD, use T stat.

3) Make your inference

- Point estimation
- Set Estimation
- Hypothesis testing?

Test null hypothesis for $\mu \leq 16$. Find p-value, give conclusion. $H_0 : \mu \leq 16$, $H_1 : \mu > 16$. We'll want to show that the average age of the breakfast club older than 16.

```
ages = c(18, 13, 18, 16, 12, 16, 17, 17, 20, 18)
qqnorm(ages)
```




```
sd(ages)
```

```
## [1] 2.415229
```