Problem Set 6

Daniel Deneau

Problems 6.5, 6.7, 6.15, 6.17, 6.21, 6.27, 6.34, 6.38

Problem 6.5:

a) False, the confidence level is for the population proportion, not the sample proportion.
b) True, that is the confidence interval with a 2% margin of error.
c) True, a 95% confidence interval means that 95% of the samples include the true proportion.
d) True, the margin of error with a sample size of 5560 is 1%

```
p <- 0.5
n = 1390
z = 1.96

margin_of_error = z * sqrt((p*(1-p))/n)
n = 1390*4
margin_of_error_2 = z *sqrt((p*(1-p))/n)
margin_of_error
```

```
## [1] 0.02628565
```

```
margin_of_error_2
```

```
## [1] 0.01314282
```

e) True. Since it was a 95% confidence interval.

Probelm 6.7:

The margin of error using a 95% confidence interval is: 0.045

```
p = 0.56
n = 600
z = 1.96
z * sqrt((p*p)/n)
```

```
## [1] 0.04480933
```

Problem 6.15:

An appropriate sample size for a 1% margin of error with 90% confidence would be: n = 9508

```r
p = 0.55
z = 1.645
m = 0.01

(1.96^2) * ((p*(1-p))/(m^2))
```

```
## [1] 9507.96
```

Problem 6.17:

Since the restaurant was the same for testing both the provocative and conservative scenario, the data for both proportions is not independent. Therefore, the indepence criteria is not met.

6.21:

a) We are 95% confident that the difference between the proportion of Democrats and Independents who support a National Health Plan is in the interval (0.1821, 0.2978) between 18% and 30%
b) True, since the proportion of democrats was 79% vs 55% for independents.

```r
pd = .79
pi = .55
nd = 347
ni = 617
z = 1.95
standard_error_d_squared = (pd*(1-pd))/nd
standard_error_i_squared = (pi*(1-pi))/ni
standard_error = sqrt(standard_error_d_squared + standard_error_i_squared)
confidence_interval <-c((pd - pi) - z*standard_error,(pd-pi) + z*standard_error)
confidence_interval
```

```
## [1] 0.1821789 0.2978211
```

6.27:

The null hypothesis is that both occupations have the same proportion of sleep deprivation. The alternate hypothesis is that they do not have the same proportion of sleep deprivation. Assuming the survey was done independent and randomly among the groups, the success-failure criteria is also met:

```r
n_o = 180
n_d = 203
p_o = 29/n_o
p_d = 35/n_d
p_o * n_o
```

```
## [1] 29
```

```r
(1-p_o)*n_o
```

```
## [1] 151
```

```
p_d * n_d
```

```
## [1] 35
```

```
(1-p_d)*n_d
```

```
## [1] 168
```

Meaning the difference between the proportions can be modeled using a normal distribution. The standard error is:

```
SE_p_o = (p_o*(1-p_o))/n_o
SE_p_d = (p_d*(1-p_d))/n_d
standard_error = sqrt(SE_p_o + SE_p_d)
```

The null value is:

```
null_value <- ((p_o*n_o)+(p_d*n_d))/(n_o + n_d)
```

Since the null hypothesis is that the difference between the proportions is 0. And the Z-Value is:

```
point_est = p_d - p_o
null_value <- ((p_o*n_o)+(p_d*n_d))/(n_o + n_d)
z_value <- (point_est - null_value) / standard_error
z_value
```

```
## [1] -4.086205
```

The tail area is:

```
2*pnorm(z_value)
```

```
## [1] 4.384862e-05
```

At a significance level of 0.05, we reject the null hypothesis and conclude that there is evidence of a difference in the proportions of sleep deprivation between truck drivers and train operators.

6.34:

a)The null hypothesis is that the proportions of forage sites is equal, and the barking deer does not prefer one habitat over others. The alternate hypothesis is that the proportions of forage sites are not equal, and the barking dear does prefer certain habitats over others. b) The Chi-squared test for a one-way table. c) Each visit to a site is independent of each other visit, and it is expected that each site could be visited at least five times. So the assumptions and conditions are satisfied.

```
e_w = 426 * 0.048
e_g = 426 * 0.147
e_f = 426 * 0.396
e_o = 426 * (1 - (0.048 + 0.147 + 0.396))
e_w
```

```
## [1] 20.448
```

```
e_g
```

```
## [1] 62.622
```

```
e_f
```

```
## [1] 168.696
```

```
e_o
```

```
## [1] 174.234
```

d) With a p__value of 2.2e-16 below a significance level of 0.05, we can reject the null hypothesis and say there is evidence that the barking dear does prefer certain habitats over others.

```
expected <- c(0.048, 0.147, 0.396, 0.409)
observed <- c(4, 16, 61, 345)
chisq.test(x = observed, p = expected)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed
## X-squared = 284.06, df = 3, p-value < 2.2e-16
```

6.38:

a)The null hypothesis is that the types of treatments do not effect the performance of the treatment. The alternate hypothesis is that different treatments do affect the performance. Since the experiment was randomized, and each group was given treatment independently,the ‘ independence criteria is met. There are also at least 5 expected counts for each cell in the table:

```
table_total = 52 + 31 + 42 + 2 + 24 + 14
row_0 = 52 + 2
row_1 = 31 + 24
row_2 = 42 + 13
col_0 = 51 + 31 + 42
col_1 = 2 + 24 + 14
(row_0 * col_0) / table_total
```

```
## [1] 40.58182
```

```
(row_1 * col_0) / table_total
```

```
## [1] 41.33333
```

```
(row_2 * col_0) / table_total
```

```
## [1] 41.33333
```

```
(row_0 * col_1) / table_total
```

```
## [1] 13.09091
```

```
(row_1 * col_1) / table_total
```

```
## [1] 13.33333
```

```
(row_2 * col_1) / table_total
```

```
## [1] 13.33333
```

b) Since the p-value is much less than a significance level of 0.05, we can reject the null hypothesis and say that the type of treatment does have an effect on the performance of the treatment.