Lab 4

Summer Heschong

2025-02-05

#setup

#Tidy Functions

```
#rename columns
weather_data <- weather_data %>%
  rename(daily_precip_mm = daily_precip)

#case_when
#create new column to indicate snowy days
weather_data <- weather_data %>%
#when the air temperature is below freezing
mutate(snow_days = case_when(mean_airtemp < 0 &
#and there is precipitation, we classify the day as SNOWY
                               daily_precip_mm > 0 ~ 'SNOW',
#alternatively, for all other cases, there's no snow
                            TRUE ~ 'NO SNOW'))
```

#Confidence Intervals

```
#save output of t.test() to new variable name
precip_test <- t.test(weather_data$daily_precip_mm,
                      conf.level = 0.90)
#examine the result
precip_test
```

```
##
##  One Sample t-test
##
## data:  weather_data$daily_precip_mm
## t = 35.867, df = 10750, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  0.9073396 0.9945672
## sample estimates:
## mean of x
## 0.9509534
```

```
#save output of binom.test() to new variable name.
vote_test <- binom.test(x = 187, n = 254,
                        conf.level = 0.90)
vote_test
```

```
##
##  Exact binomial test
##
## data:  187 and 254
## number of successes = 187, number of trials = 254, p-value = 2.681e-14
## alternative hypothesis: true probability of success is not equal to 0.5
## 90 percent confidence interval:
##  0.6868405 0.7814295
## sample estimates:
## probability of success
##              0.7362205
```

#One-Sample Hypothesis Tests

```
#null hypothesis: the mean daily precipitation is 2mm
#alternative hypothesis: the mean daily precipitation is not 2 mm

#one-sample t-test
precip_ttest1 <- t.test(weather_data$daily_precip_mm,
                        mu = 2,
                        conf.level = 0.99,
                        alternative = 'two.sided')
precip_ttest1
```

```
##
##  One Sample t-test
##
## data:  weather_data$daily_precip_mm
## t = -39.567, df = 10750, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 2
## 99 percent confidence interval:
##  0.8826482 1.0192586
## sample estimates:
## mean of x
## 0.9509534
```

```
#99% CI means alpha = 0.01

#p < 2.2e-16 so p < alpha

#we choose to reject the null hypothesis and retain the alternative
#hypothesis that mean daily precipitation is not 2 mm per day,
#because p < 0.01
```

```
#null hypoth: the mean daily precipitation is less than 1mm
#alternative hypoth: the mean daily precip is greater than 1mm

#one sample t-test
precip_ttest2 <- t.test(weather_data$daily_precip_mm,
                        mu = 1,
                        conf.level = 0.99,
                        alternative = 'greater')
precip_ttest2
```

```
##
##  One Sample t-test
##
## data:  weather_data$daily_precip_mm
## t = -1.8499, df = 10750, p-value = 0.9678
## alternative hypothesis: true mean is greater than 1
## 99 percent confidence interval:
##  0.8892656       Inf
## sample estimates:
## mean of x
## 0.9509534
```

```
#99% CI means alpha = 0.01

#p = 0.9678 so p > alpha

#we choose to retain the null hypothesis that mean daily precip
#is less than 1mm per day, because p > 0.01.
```

#Two-Sample Hypothesis Test

```
#first use case_when function to create a new column that categorizes
#records based on when they were collected
weather_data <- weather_data %>%
  mutate(time_period = case_when(date < '2003-01-01' ~ 'early',
                                 date >= '2003-01-01' ~ "late",
                                 TRUE ~ NA))
#we added the TRUE ~ NA athe the end in order to spit out an NA for
#any row that does not adhere to any of the logical statements we've provided.

#Create an 'early' dataset using the filter function
early_data <- weather_data %>%
  filter(time_period == 'early')

#create a late dataset
late_data <- weather_data %>%
  filter(time_period == 'late')

#first test for equal variances
precip_vartest <- var.test(x = early_data$daily_precip_mm,
                           y = late_data$daily_precip_mm,
                           alternative = 'two.sided',
                           conf.level = 0.95)
precip_vartest
```

```
##
##  F test to compare two variances
##
## data:  early_data$daily_precip_mm and late_data$daily_precip_mm
## F = 0.9251, num df = 4958, denom df = 5791, p-value = 0.004498
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.8768411 0.9761337
```

```
## sample estimates:
## ratio of variances
##           0.925104
```

```r
#p < 0.05 so we reject the null hypoth and retain the alternative hypot that the variances are not equa

#perform t-test calling data from new datasets. are the mean daly precips different from before and aft
precip_ttest3 <- t.test(early_data$daily_precip_mm,
                        late_data$daily_precip_mm)
#r assumes the difference in means is 0 (mu = 0), and conf.level = 0.95 and we are interested in a two-
precip_ttest3
```

```
##
##  Welch Two Sample t-test
##
## data:  early_data$daily_precip_mm and late_data$daily_precip_mm
## t = -1.2634, df = 10605, p-value = 0.2065
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.17093029  0.03694546
## sample estimates:
## mean of x mean of y
## 0.9148619 0.9818543
```