Week 9: t distribution and t-tests

Stat 201: Statistics I

March 10, 2019

R Distribution Functions

R provides four functions for many commonly used distributions. These functions have names that begin with a letter (`d`, `p`, `q` or `r`) followed by the distribution name or an abbreviation of the name (i.e. `norm` for normal).

Function	Binomial	Normal	t
Density	`dbinom`	`dnorm`	`dt`
Probability	`pbinom`	`pnorm`	`pt`
Quantile	`qbinom`	`qnorm`	`qt`
Random	`rbinom`	`rnorm`	`rt`

t distributions in R

- The t distribution is similar to the standard normal distribution.
- Like the standard normal, it is centered at 0 and standardized by the standard error.
- The degrees of freedom $(n - 1)$ of the distribution must be specified.
- Note: the t distribution functions have a parameter `ncp` to specify distributions that are not centered at 0. This is rarely used and can be ignored for this class.

Critical t values

The critical value $t_{\alpha,df}$ is the value of a t distribution with df degrees of freedom that delineates an upper probability of α from the rest of the distribution. This is a quantile of an upper proportion.

- What are the critical values $t_{\alpha,df}$ and $t_{\alpha/2,df}$ for $\alpha = 0.05$ and 15 degrees of freedom?

```
alpha <- 0.05
df <- 15
qt(alpha, df, lower.tail = FALSE)
```

```
## [1] 1.75305
```

```
qt(alpha/2, df, lower.tail = FALSE)
```

```
## [1] 2.13145
```

One sample t-tests with summary statistics

In order to conduct a t-test with summary statistics (sample mean and standard deviation), the t-score is calculated and the p-value is found from the `pt` function.

- Test whether the population mean is not equal to 80 with a sample with mean $\bar{x} = 76$, standard deviation of 6.4 and sample size of $n = 31$, at a significance level of $\alpha = 0.05$.

```
alpha <- 0.05
n <- 31
x.bar <- 76
sd <- 6.4

t <- (x.bar - 80) / (sd / sqrt(n))
t
```

```
## [1] -3.479853
```

One sample t-tests with summary statistics, cont

- Test whether the population mean is not equal to 80 with a sample with mean $\bar{x} = 76$, standard deviation of 6.4 and sample size of $n = 31$, at a significance level of $\alpha = 0.05$.
  - Remember, for two-sided tests the p-value is $2 \times P(T > t)$, or, as in this case, $2 \times P(T < t)$ if the test statistic t is negative.

```
p.value <- 2 * pt(t, df=n-1)
p.value
```

```
## [1] 0.001557881
```

Because $p = 0.002 < \alpha = 0.05$, reject the null hypothesis. There is evidence that the population mean is not 80.

t.test function

If we have sample data, we could conduct tests by calculating summary statistics, the t-score and p-value. However, R provides a simple function for t-tests with data, `t.test`.

```
t.test(x, mu = 0, alternative = c("two.sided", "less", "greater"))
```

- `x` is a vector of sample values
- `mu` is the value the sample is being compared against.
- `alternative` specifies the form of the alternative hypothesis.

t.test function output

```r
t.test(x, mu=5, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  x
## t = -2.0997, df = 19, p-value = 0.02467
## alternative hypothesis: true mean is less than 5
## 95 percent confidence interval:
##      -Inf 4.816103
## sample estimates:
## mean of x
##  3.958016
```

t.test function output, cont.

- After the title and specification of the data, the next line of output displays the test statistic, degrees of freedom and p-value of the test.

```
## t = -0.19788, df = 19, p-value = 0.4226
```

- A description of the alternative hypothesis of the test is next.

```
## alternative hypothesis: true mean is less than 5
```

- The confidence interval of the mean is provided.

```
## 95 percent confidence interval:
##      -Inf 4.816103
```

- Finally, the sample mean is given.

```
## mean of x
##  3.958016
```

One sample t-tests with data

- Test whether the sample of trees in the built-in data set `trees` come from a population with a mean height of 80, at a significance level of $\alpha = 0.05$.

```
t.test(trees$Height, mu=80)
```

```
##
##  One Sample t-test
##
## data:  trees$Height
## t = -3.4952, df = 30, p-value = 0.001496
## alternative hypothesis: true mean is not equal to 80
## 95 percent confidence interval:
##  73.6628 78.3372
## sample estimates:
## mean of x
##        76
```

Two independent samples t-tests

t-tests on two independent samples can be conducted by providing a second vector of sample values to the `t.test` function.

- Test whether men's and women's heights from the `heights` data set on D2L are different.

```
t.test(heights$men, heights$women)
```

```
##
##  Welch Two Sample t-test
##
## data:  heights$men and heights$women
## t = 4.7583, df = 47.894, p-value = 1.834e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.009595 9.878405
## sample estimates:
## mean of x mean of y
##    70.236    63.292
```

Paired samples t-tests

t-tests on paired samples need to add the parameter `paired = TRUE`.

- Test whether exam scores from the `scores` data set on D2L improved from the midterm to the final exam.

```
t.test(scores$final, scores$mid, paired=TRUE, alternative="greater")
```

```
##
##  Paired t-test
##
## data:  scores$final and scores$mid
## t = 3.8177, df = 41, p-value = 0.0002237
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.250066      Inf
## sample estimates:
## mean of the differences
##                 4.02381
```