R12 - Pvalues

HCI/PSYCH 522
Iowa State University

April 21, 2022

p-values

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$.

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}}$$

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation.

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation. Mathematically

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation. Mathematically

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

One-sample t-test

Let $Y_i \overset{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation. Mathematically

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

If the null model is true,

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation. Mathematically

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

If the null model is true, then t_{n-1} has a Student's t-distribution with $n-1$ degrees of freedom

One-sample t-test

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with null model that assumes $\mu = 0$. Use test statistic

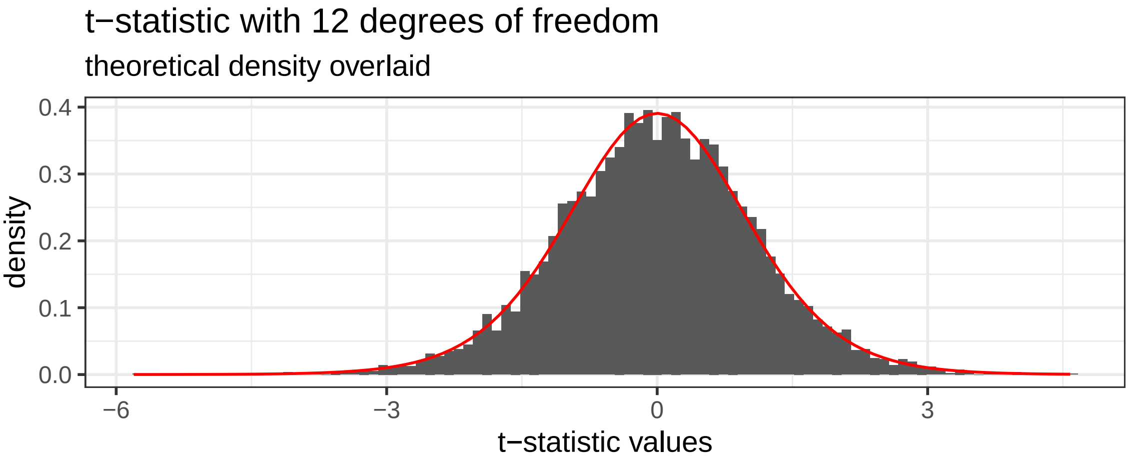
$$t_{n-1} = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{\overline{y}}{s/\sqrt{n}}$$

where $\overline{y}$ is the sample average and s is the sample standard deviation. Mathematically

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

If the null model is true, then t_{n-1} has a Student's t-distribution with $n-1$ degrees of freedom when we think about repeatedly getting new data.

Student's t-distributions



p-values

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

p-values

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

p-values

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

```
tstat <- 2.5
```

Suppose we collected data with a sample size of $n = 13$

p-values

Definition

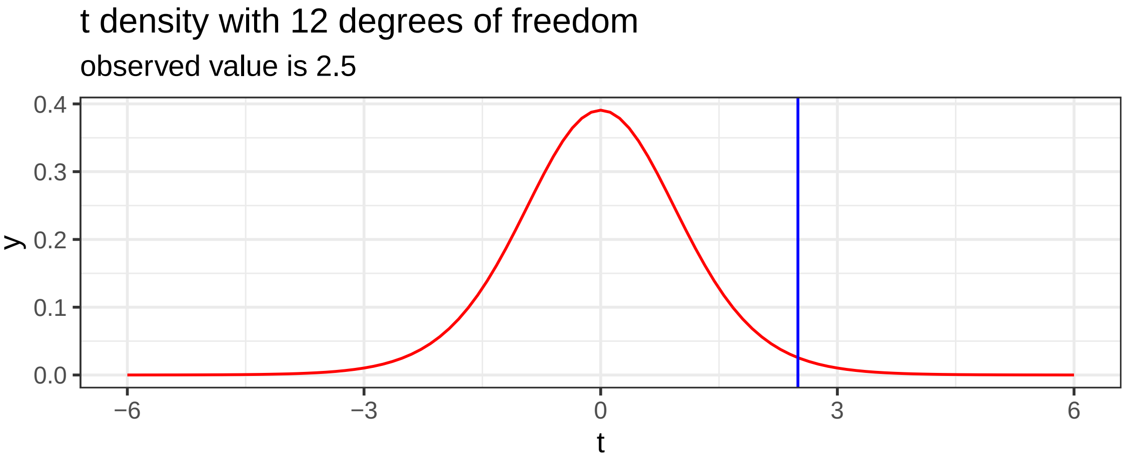
A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

```
tstat <- 2.5
```

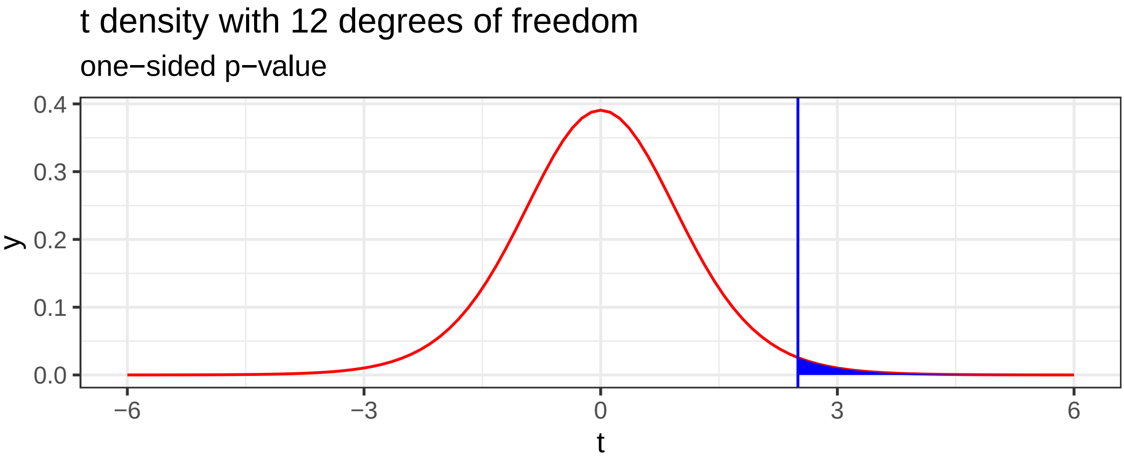
Suppose we collected data with a sample size of $n = 13$ and calculated the t-statistic and found it to be 2.5.

As or more extreme regions

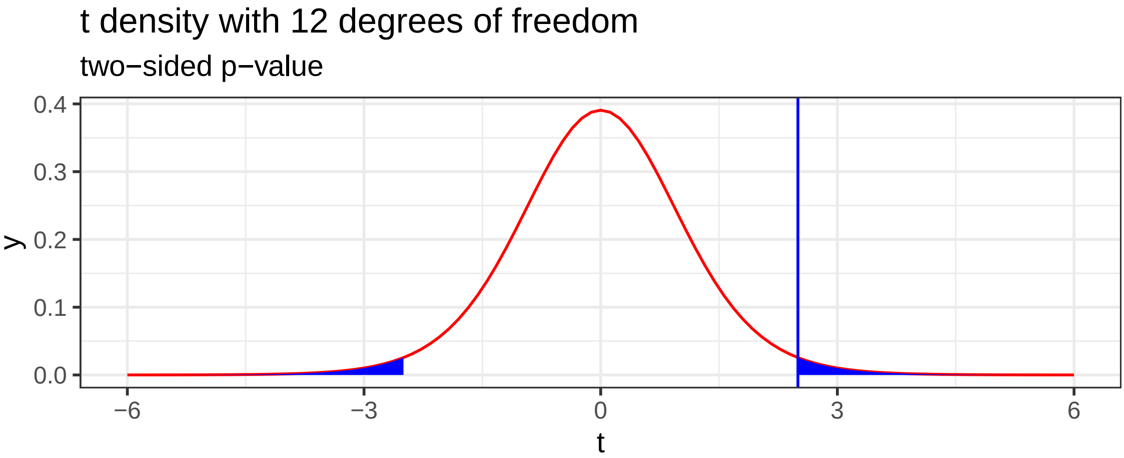
As or more extreme regions



As or more extreme regions



As or more extreme regions



p-value summary

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

p-value summary

Definition

A p-value is the probability of observing a test statistic as or more extreme than the value observed if the null model is true.

Small p-values provide evidence against the null model.

Regression model

Regression model with categorical variable:

$$Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero

Regression model

Regression model with categorical variable:

$$Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero and
- β_p, $p > 0$ is the mean increase in the dependent variable for a one-unit increase in the p^{th} independent variable when all other independent variables are held constant.

Regression model

Regression model with categorical variable:

$$Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero and
- β_p, $p > 0$ is the mean increase in the dependent variable for a one-unit increase in the p^{th} independent variable when all other independent variables are held constant.

Default p-values provided have null models where

- $\beta_p = 0$ for some p

Regression model

Regression model with categorical variable:

$$Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero and
- β_p, $p > 0$ is the mean increase in the dependent variable for a one-unit increase in the p^{th} independent variable when all other independent variables are held constant.

Default p-values provided have null models where

- $\beta_p = 0$ for some p (t-tests with $n - (p+1)$ degrees of freedom)

Regression model

Regression model with categorical variable:

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero and
- β_p, $p > 0$ is the mean increase in the dependent variable for a one-unit increase in the p^{th} independent variable when all other independent variables are held constant.

Default p-values provided have null models where

- $\beta_p = 0$ for some p (t-tests with $n - (p+1)$ degrees of freedom)
- some set of βs are 0

Regression model

Regression model with categorical variable:

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}, \sigma^2)$$

where

- β_0 (intercept) is the mean value of the dependent variable Y_i when all independent variables are zero and
- β_p, $p > 0$ is the mean increase in the dependent variable for a one-unit increase in the p^{th} independent variable when all other independent variables are held constant.

Default p-values provided have null models where

- $\beta_p = 0$ for some p (t-tests with $n - (p + 1)$ degrees of freedom)
- some set of βs are 0 (F-tests).

Regression output in R - continuous independent variables

Regression output in R - continuous independent variables

```
m <- lm(log(count) ~ acreage + do2 + maxdepth + no3 + so4 + temp, data = longnosedace)
summary(m)
```

```
##
## Call:
## lm(formula = log(count) ~ acreage + do2 + maxdepth + no3 + so4 +
##     temp, data = longnosedace)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62524 -0.59964  0.04707  0.72422  1.80151
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.745e+00  1.606e+00  -2.331  0.02312 *
## acreage      5.051e-05  1.643e-05   3.074  0.00318 **
## do2          3.837e-01  1.337e-01   2.871  0.00565 **
## maxdepth     9.059e-03  4.319e-03   2.097  0.04019 *
## no3          2.155e-01  7.223e-02   2.984  0.00411 **
## so4          9.092e-03  1.872e-02   0.486  0.62887
## temp         9.026e-02  4.203e-02   2.147  0.03581 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9922 on 60 degrees of freedom
```

Regression output in R - categorical independent variables

```
mouse <- read_csv("mouse.csv", show_col_types = FALSE)
m <- lm(Skill ~ Mouse, data = mouse)
summary(m)

##
## Call:
## lm(formula = Skill ~ Mouse, data = mouse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.5167  -3.3857   0.8143   5.1833  10.0143
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   32.6912     0.8846  36.958  < 2e-16 ***
## MouseDell                     -5.2892     1.3010  -4.065 5.95e-05 ***
## MouseMamba (Wired)             9.6060     1.1877   8.088 1.06e-14 ***
## MouseMamba (Wireless)          6.9945     1.2565   5.567 5.25e-08 ***
## MouseViper (Wired, light)     12.4254     1.2352  10.059  < 2e-16 ***
## MouseViper (Wired)            10.1945     1.2565   8.113 8.88e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.678 on 343 degrees of freedom
## Multiple R-squared:  0.4543,Adjusted R-squared:  0.4463
## F-statistic:  57.1 on 5 and 343 DF,  p-value: < 2.2e-16
```

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$X_{i,1} = \mathrm{I}(\text{Mouse for observation } i \text{ is Dell})$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$X_{i,1} = \mathrm{I}(\text{Mouse for observation } i \text{ is Dell})$$
$$X_{i,2} = \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)})$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$X_{i,1} = \mathrm{I}(\text{Mouse for observation } i \text{ is Dell})$$
$$X_{i,2} = \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)})$$
$$X_{i,3} = \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wireless)})$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$\begin{aligned}
X_{i,1} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Dell}) \\
X_{i,2} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)}) \\
X_{i,3} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wireless)}) \\
X_{i,4} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired, light)})
\end{aligned}$$

Regression model with a categorical variable

Regression model

$$Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$\begin{aligned}
X_{i,1} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Dell}) \\
X_{i,2} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)}) \\
X_{i,3} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wireless)}) \\
X_{i,4} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired, light)}) \\
X_{i,5} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired)})
\end{aligned}$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$\begin{aligned}
X_{i,1} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Dell}) \\
X_{i,2} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)}) \\
X_{i,3} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wireless)}) \\
X_{i,4} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired, light)}) \\
X_{i,5} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired)})
\end{aligned}$$

Regression model with a categorical variable

Regression model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,4} + \beta_5 X_{i,5}, \sigma^2)$$

$$\begin{aligned}
X_{i,1} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Dell}) \\
X_{i,2} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wired)}) \\
X_{i,3} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Mamba (Wireless)}) \\
X_{i,4} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired, light)}) \\
X_{i,5} &= \mathrm{I}(\text{Mouse for observation } i \text{ is Viper (Wired)})
\end{aligned}$$

An F-test has null model with $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

ANOVA - categorical independent variables

```
mouse <- read_csv("mouse.csv", show_col_types = FALSE)
m <- lm(Skill ~ Mouse, data = mouse)
drop1(m, test="F")

## Single term deletions
##
## Model:
## Skill ~ Mouse
##        Df Sum of Sq   RSS    AIC F value    Pr(>F)
## <none>                15297 1331.3
## Mouse   5     12734 28031 1532.7  57.104 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression output in R - categorical independent variables

```
m <- lm(breaks ~ wool + tension, data = warpbreaks)
summary(m)

##
## Call:
## lm(formula = breaks ~ wool + tension, data = warpbreaks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.500  -8.083  -2.139   6.472  30.722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.278      3.162  12.423  < 2e-16 ***
## woolB         -5.778      3.162  -1.827 0.073614 .
## tensionM     -10.000      3.872  -2.582 0.012787 *
## tensionH     -14.722      3.872  -3.802 0.000391 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.62 on 50 degrees of freedom
## Multiple R-squared:  0.2691,Adjusted R-squared:  0.2253
## F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
```

ANOVA

```
mouse <- read_csv("mouse.csv", show_col_types = FALSE)
m <- lm(Skill ~ Mouse, data = mouse)
drop1(m, test="F")

## Single term deletions
##
## Model:
## Skill ~ Mouse
##        Df Sum of Sq   RSS    AIC F value    Pr(>F)
## <none>               15297 1331.3
## Mouse   5     12734 28031 1532.7  57.104 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA

```
m <- lm(breaks ~ wool + tension, data = warpbreaks)
drop1(m, test = "F")

## Single term deletions
##
## Model:
## breaks ~ wool + tension
##         Df Sum of Sq    RSS    AIC F value   Pr(>F)
## <none>                6747.9 268.71
## wool     1    450.67 7198.6 270.20  3.3393 0.073614 .
## tension  2   2034.26 8782.1 278.94  7.5367 0.001378 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA for interactions

```
m <- lm(breaks ~ wool + tension + wool:tension, data = warpbreaks)
drop1(m, test = "F")

## Single term deletions
##
## Model:
## breaks ~ wool + tension + wool:tension
##              Df Sum of Sq    RSS    AIC F value  Pr(>F)
## <none>                    5745.1 264.02
## wool:tension  2    1002.8 6747.9 268.71  4.1891 0.02104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression with interactions

```
m <- lm(breaks ~ wool + tension + wool:tension, data = warpbreaks)
summary(m)

##
## Call:
## lm(formula = breaks ~ wool + tension + wool:tension, data = warpbreaks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.5556  -6.8889  -0.6667   7.1944  25.4444
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      44.556      3.647  12.218 2.43e-16 ***
## woolB           -16.333      5.157  -3.167 0.002677 **
## tensionM        -20.556      5.157  -3.986 0.000228 ***
## tensionH        -20.000      5.157  -3.878 0.000320 ***
## woolB:tensionM   21.111      7.294   2.895 0.005698 **
## woolB:tensionH   10.556      7.294   1.447 0.154327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.94 on 48 degrees of freedom
## Multiple R-squared:  0.3778,Adjusted R-squared:  0.3129
## F-statistic: 5.828 on 5 and 48 DF,  p-value: 0.0002772
```

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true,

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
- Proper inference requires full reporting and transparency.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
- Proper inference requires full reporting and transparency.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
- Proper inference requires full reporting and transparency.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

American Statistical Association Statement on p-values

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
- Proper inference requires full reporting and transparency.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.