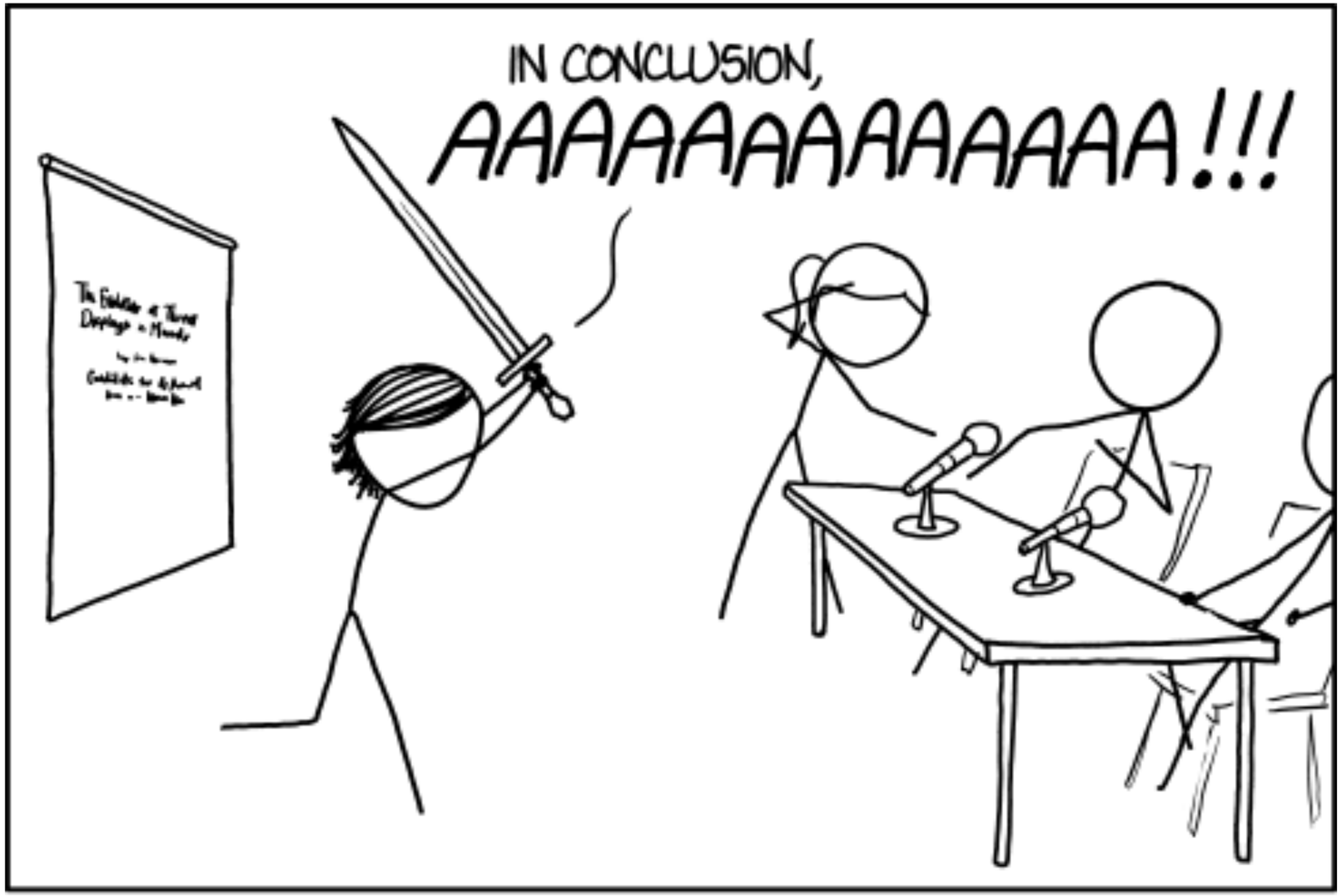


Final Review

12.02.21

https://github.com/katiesevans/IGP_biostatistics/blob/main/final_study_guide.md

Multiple regression

```
Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

MPG = 30.09 - 0.06(HP)

Multiple regression

```
Call:
lm(formula = mpg ~ trans + hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3843 -2.2642  0.1366  1.6968  5.8657

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.584914   1.425094  18.655  < 2e-16 ***
transstick   5.277085   1.079541   4.888 3.46e-05 ***
hp          -0.058888   0.007857  -7.495 2.92e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.909 on 29 degrees of freedom
Multiple R-squared:  0.782,     Adjusted R-squared:  0.767
F-statistic: 52.02 on 2 and 29 DF,  p-value: 2.55e-10
```

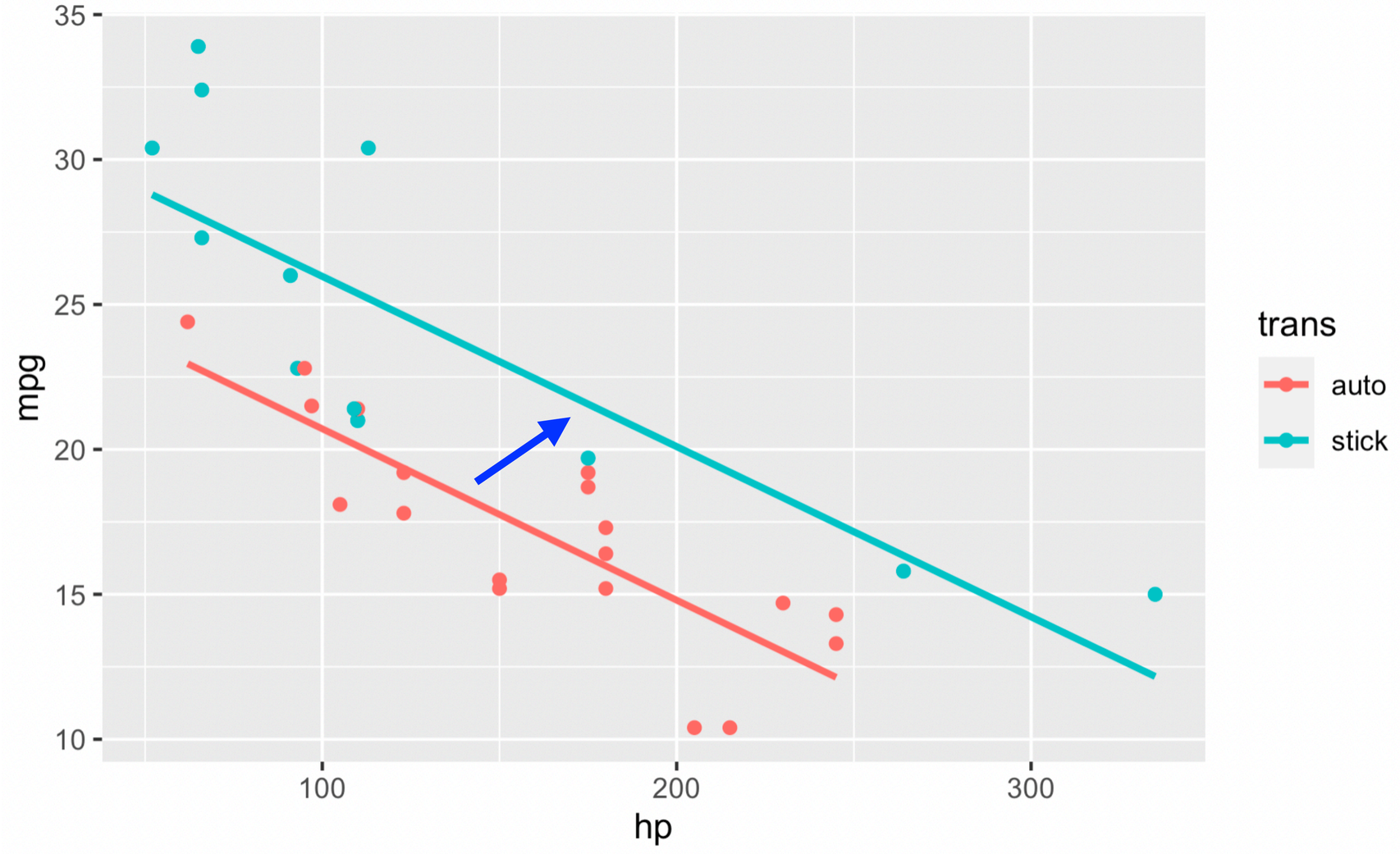
MPG = 26.58 - 0.058(HP) + 5.27(trans)

Default is “auto”

Multiple regression

MPG = 26.58 - 0.058(HP) + 5.27(trans)

Default is “auto”



Multiple regression

```
Call:
lm(formula = mpg ~ hp + factor(gear), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4937 -2.3586 -0.8277  2.2753  7.7287

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   27.88193    2.10908  13.220 1.47e-13 ***
hp            -0.06685    0.01105  -6.052 1.59e-06 ***
factor(gear)4  2.63486    1.55164   1.698 0.100575
factor(gear)5  6.57476    1.64268   4.002 0.000417 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7262
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

```
> unique(mtcars$gear)
[1] 3 4 5
```

Multiple regression

```
Call:
lm(formula = mpg ~ hp + factor(gear), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4937 -2.3586 -0.8277  2.2753  7.7287

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.88193    2.10908  13.220 1.47e-13 ***
hp           -0.06685    0.01105  -6.052 1.59e-06 ***
factor(gear)4  2.63486    1.55164   1.698 0.100575
factor(gear)5  6.57476    1.64268   4.002 0.000417 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7262
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

```
> unique(mtcars$gear)

[1] 3 4 5
```

MPG = 27.88 - 0.066(HP) + 2.63(gear4) + 6.57(gear5)

(Default is gear3)

Gear4 and gear5 = binary (0 or 1)

Multiple regression

```
Call:
lm(formula = mpg ~ hp + factor(gear), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4937 -2.3586 -0.8277  2.2753  7.7287

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.88193    2.10908  13.220 1.47e-13 ***
hp           -0.06685    0.01105  -6.052 1.59e-06 ***
factor(gear)4  2.63486    1.55164   1.698 0.100575
factor(gear)5  6.57476    1.64268   4.002 0.000417 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

MPG = 27.88 - 0.066(HP) + 2.63(gear4) + 6.57(gear5)

HP = 10, gear = 3

MPG = 27.88 - 0.066(10) + 2.63(0) + 6.57(0)

predict(model, data.frame(hp = 10, gear = 3)

HP = 10, gear = 4

MPG = 27.88 - 0.066(10) + 2.63(1) + 6.57(0)

predict(model, data.frame(hp = 10, gear = 4)

Multiple regression

```
> mtcars$gear <- factor(mtcars$gear, levels = c(4,3,5)))
lm(formula = mpg ~ hp + gear, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4937 -2.3586 -0.8277  2.2753  7.7287

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.51679    1.34393  22.707  < 2e-16 ***
hp          -0.06685    0.01105  -6.052 1.59e-06 ***
gear3       -2.63486    1.55164  -1.698   0.1006
gear5        3.93990    2.04730   1.924   0.0645 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7262
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

Multiple regression

These are different because they are compared to a different starting point…

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.51679    1.34393  22.707  < 2e-16 ***
hp          -0.06685    0.01105  -6.052 1.59e-06 ***
gear3       -2.63486    1.55164  -1.698   0.1006
gear5        3.93990    2.04730   1.924   0.0645 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7262
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.88193    2.10908  13.220 1.47e-13 ***
hp           -0.06685    0.01105  -6.052 1.59e-06 ***
factor(gear)4  2.63486    1.55164   1.698 0.100575
factor(gear)5  6.57476    1.64268   4.002 0.000417 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.154 on 28 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7262
F-statistic: 28.41 on 3 and 28 DF,  p-value: 1.217e-08
```

These are the same because it is still the same model…

https://www1.udel.edu/htr/Statistics/Notes816/class14.PDF

Hypothesis testing + *p*-values

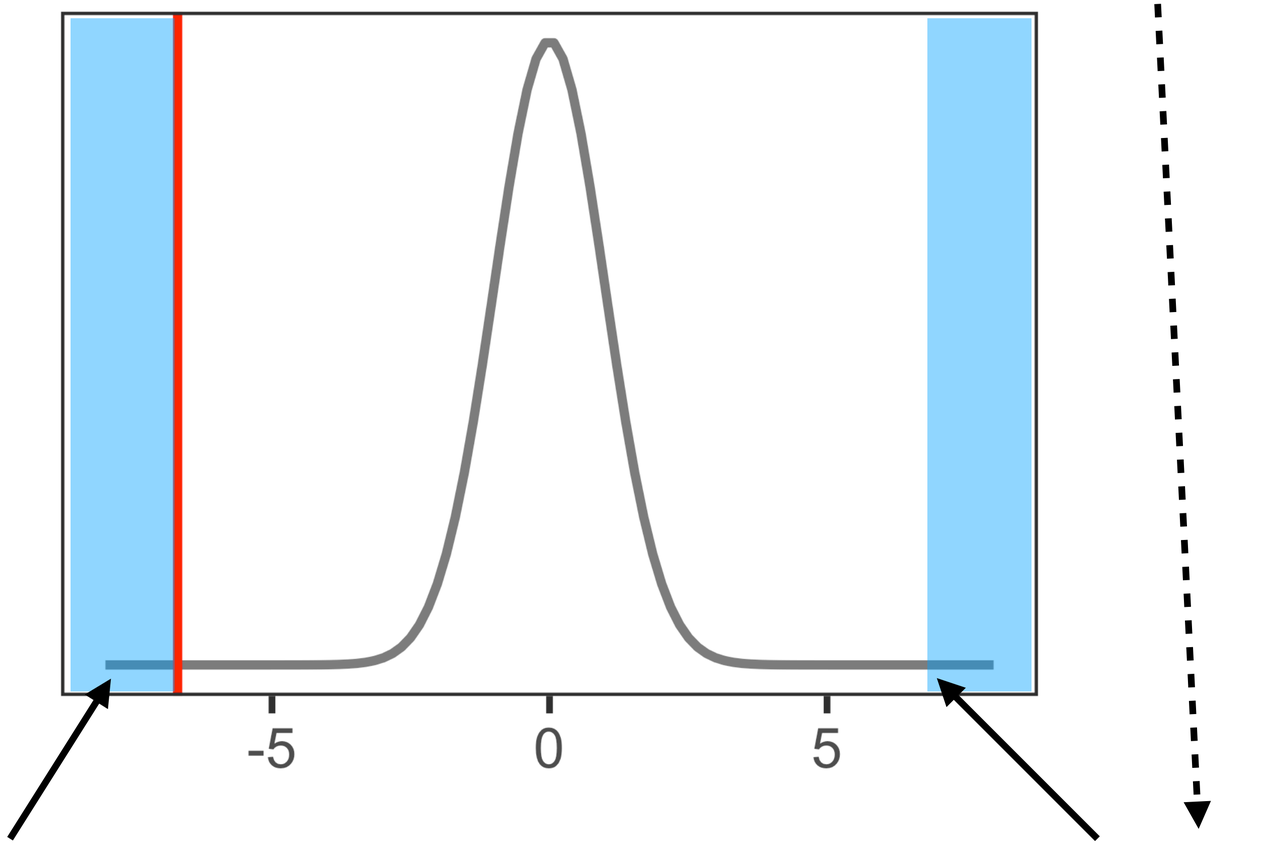
```
Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

pt(6.742, 30, lower.tail = F)



pt(-6.742, 30, lower.tail = T) ⇢ **8.95e-08** **+** **8.95e-08**

Directional hypothesis

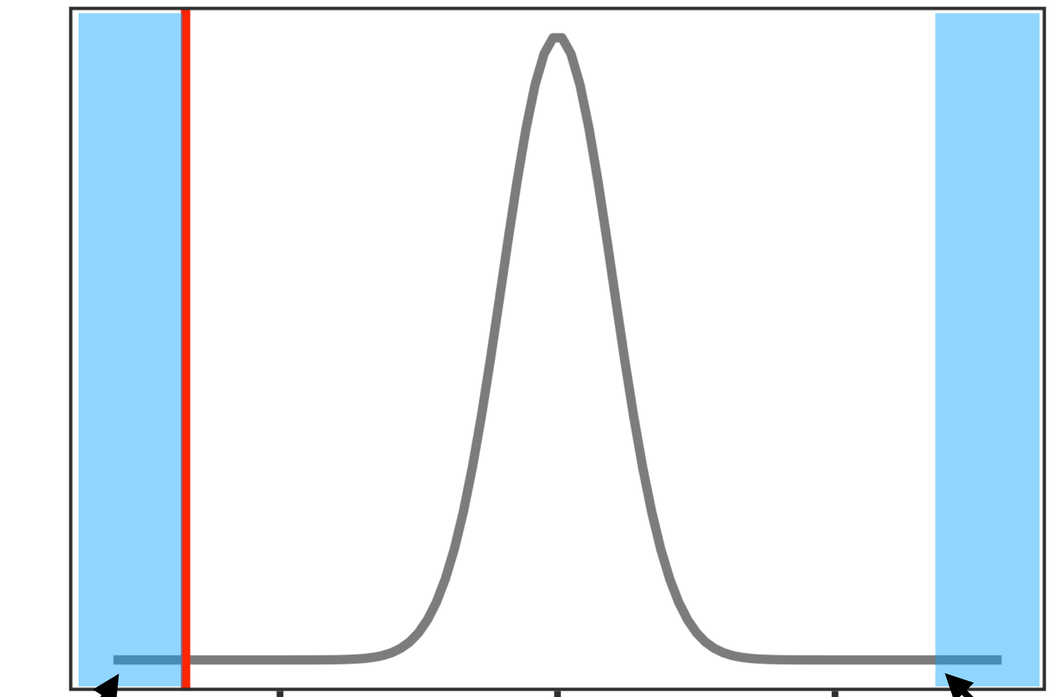
If I only care if the HP has a NEGATIVE regression coefficient…

```
Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```



8.95e-08 **+** **8.95e-08**

Directional hypothesis

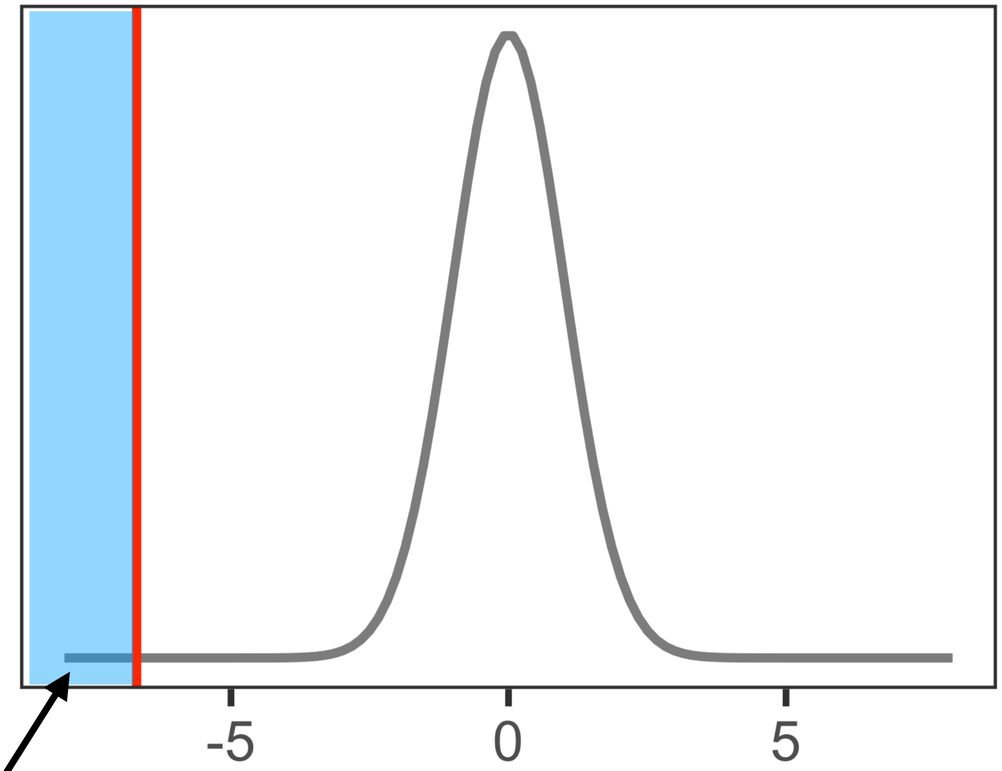
If I only care if the HP has a NEGATIVE regression coefficient…

```
Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```



pt(-6.742, 30, lower.tail = T) ⇢ 8.95e-08 = 1.79e-07 / 2

Directional hypothesis

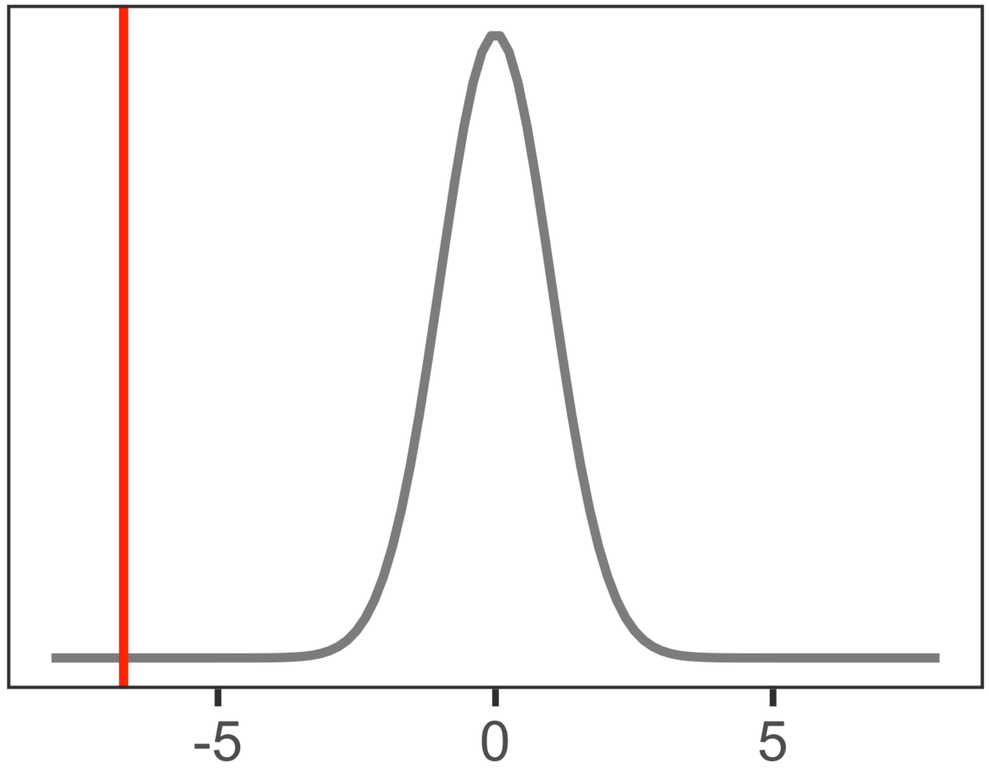
If I only care if the HP has a POSITIVE regression coefficient…

```
Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```



It goes in the wrong direction, so p-value = 0.5 (i.e. not significant)