

# Coefficient-Level Inference

2019-06-28

## Introduction and Research Question

In this set of notes, you will continue your foray into regression analysis. To do so, we will use the *keith-gpa.csv* data to examine whether time spent on homework is related to GPA. The data contain three attributes collected from a random sample of  $n = 100$  8th-grade students (see the [data codebook](#)). To begin, we will load several libraries and import the data into an object called *keith*.

## Preparation

```
# Load libraries
library(broom)
library(corr)
library(dplyr)
library(ggplot2)
library(readr)

# Read in data
keith = read_csv(file = "~/Documents/github/epsy-8251/data/keith-gpa.csv")
head(keith)
```

```
# A tibble: 6 x 3
  gpa homework parent_ed
<dbl>   <dbl>   <dbl>
1    78         2       13
2    79         6       14
3    79         1       13
4    89         5       13
5    82         3       16
6    77         4       13
```

As in previous analyses, we begin the regression analysis by looking at the marginal distributions of any predictors and the outcome, as well as summary statistics of center and variation. We also examine a scatterplot of the outcome versus each of the predictors and the bivariate correlations (syntax not shown).

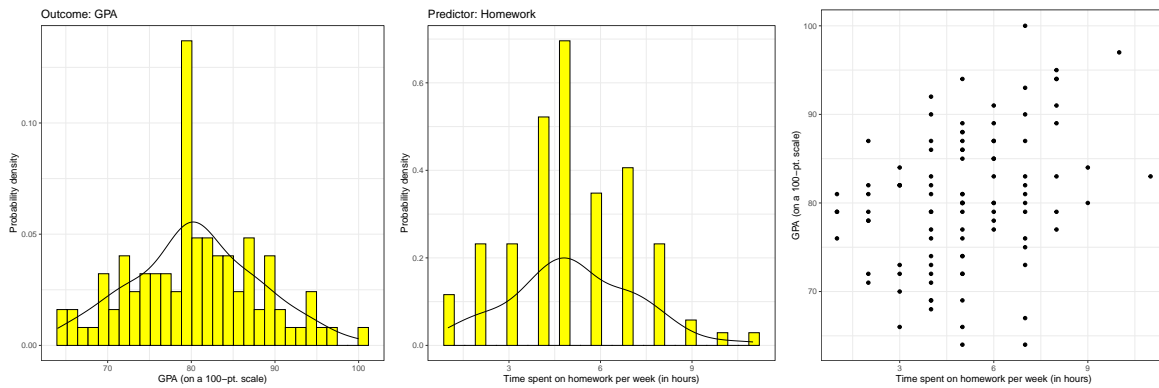


Figure 1. Density plots of the marginal distributions of GPA and time spent on homework. The scatterplot showing the relationship between GPA and time spent on homework is also shown.

Table 1  
Summary Measures for 8th-Grade Students' GPA and Time Spent on Homework

Measure	M	SD
GPA	80.47	7.62
Time spent on homework	5.09	2.06

Table 2  
Correlations between 8th-Grade Students' GPA and Time Spent on Homework

Measure	1	2
1. GPA	—	
2. Time spent on homework	.33	—

We might describe the results of this analysis as:

The marginal distributions of GPA and time spent on homework are both unimodal. The average amount of time these 8th-grade students spend on homework each week is 5.09 hours (SD = 2.06). These 8th-grade students have a mean GPA of 80.47 (SD = 7.62) on a 100-pt scale. There is a moderate, positive, linear relationship between time spent on homework and GPA for these students ( $r = .33$ ). This suggests that 8th-grade students who spend less time on homework tend to have lower GPAs, on average, than students who spend more time on homework.

We will also fit a model by regressing income on education level and storing those results in an object called `lm.1`.

```
# Fit regression model
lm.1 = lm(gpa ~ 1 + homework, data = keith)
lm.1
```

Call:

```
lm(formula = gpa ~ 1 + homework, data = keith)
```

Coefficients:

(Intercept)	homework
74.290	1.214

The fitted equation is:

$$\hat{GPA} = 74.29 + 1.21(\text{Time spent on homework}),$$

Summarizing this,

This model estimated mean GPA for all 8th-grade students who spend 0 hours a week on homework is 74.29. Each additional hour 8th-grade students spend per week on homework is associated with a difference in GPA of 1.21, on average. Differences in time spent on homework explains 10.7% of the variation in students' GPAs. All this suggests that time spent on homework is related to GPA for the  $n = 100$  8th-graders in the sample.

## Statistical Inference

What if we want to understand the relationship between education level and income for a larger population of 8th-grade students, say all of them in the district? The problem is that if we had drawn a different sample of  $n = 100$  8th-grade students, all the regression estimates ( $\hat{\beta}_0$ ,  $\hat{\beta}_1$ , and  $R^2$ ) would be different than the ones we obtained from our sample. This makes it difficult to say, for example, how does the conditional mean GPA differs for students with differing amounts of time spent on homework. In our observed sample,  $\hat{\beta}_1$  was 1.21. But, had we sampled different students, we might have found that  $\hat{\beta}_1$  was 2.03. And a different random sample of employees we might have produced a  $\hat{\beta}_1$  of 0.96.

This variation in the estimates arises because of the random nature of the sampling. One of the key findings in statistical theory is that the amount of variation in estimates under random sampling is completely predictable (this variation is called *sampling error*). Being able to quantify the sampling error allows us to provide a more informative answer to the research question. For example, it turns out that based on the quantification of sampling error in our example, we believe that the actual  $\beta_1$  is between 0.51 and 1.92.

Statistical inference allows us to learn from incomplete or imperfect data Gelman & Hill (2007). In many studies, the primary interest is to learn about one or more characteristics about a population. These characteristics must be estimated from sample data. This is the situation in our example, where we have only a sample of 8th-grade students and we want to understand the relationship between time spent on homework and GPA for ALL 8th-grade students in the district.

In the example, the variation in estimates arises because of sampling variation. It is also possible to have variation because of imperfect measurement. This is called *measurement error*. Despite these being very different sources of variation, in practice they are often combined (e.g., we measure imperfectly and we want to make generalizations).

Regardless of the sources of variation, the goals in most regression analyses are two-fold:

1. Estimate the parameters from the observed data; and
2. Summarize the amount of uncertainty (e.g., quantify the sampling error) in those estimates.

The first goal we addressed in the notes on description. It is the second goal that we will explore in these notes.

## Quantification of Uncertainty

Before we talk about estimating uncertainty in regression, let me bring you back in time to your *Stat I* course. In that course, you probably spent a lot of time talking about sampling variation for the mean. The idea went something like this: Imagine you have a population that is infinitely large. The observations in this population follow some probability distribution. (This distribution is typically unknown in practice, but for now, let's pretend we know what that distribution is.) For our purposes, let's assume the population is normally distributed with a mean of  $\mu$  and a standard deviation of  $\sigma$ .

Sample  $n$  observations from that population. Based on the  $n$  sampled observations, find the mean. We will call this  $\hat{\mu}_1$  since it is an estimate for the population mean (the subscript just says it is the first sample). In all likelihood,  $\hat{\mu}_1$  is not the exact same value as  $\mu$ . It varies from the population mean because of sampling error.

Now, sample another  $n$  observations from the population. Again, find the mean. We will call this estimate  $\hat{\mu}_2$ . Again, it probably varies from  $\mu$ , and may be different than  $\hat{\mu}_1$  as well. Continue to repeat this process: randomly sample  $n$  observations from the population; and find the mean.

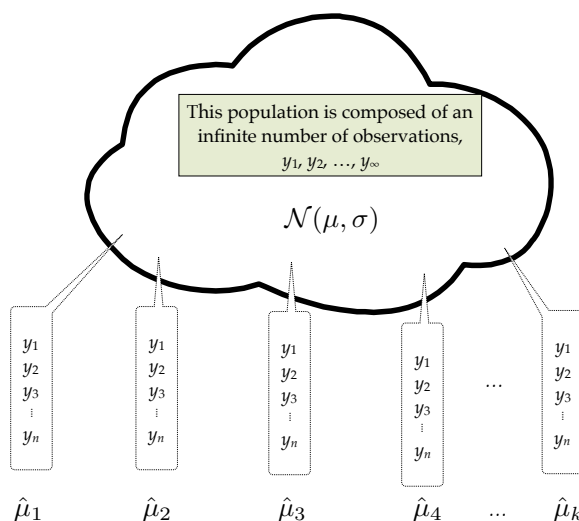


Figure 2. Thought experiment for sampling samples of size  $n$  from the population to obtain the sampling distribution of the mean.

The distribution of the sample means, it turns out, is quite predictable using statistical theory. Theory predicts that the distribution of the sample means will be normally distributed. It also predicts that the mean, or *expected value*, of all the sample means will be equal to the population mean,  $\mu$ .<sup>1</sup> Finally, theory predicts that the standard deviation of this distribution, called the *standard error*, will be equal to the population standard deviation divided by the square root of the sample size. Mathematically, we would write all this as,

$$\hat{\mu}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

The important thing is not that you memorize this result, but that you understand that the process of randomly sampling from a known population can lead to predictable results in the distribution of statistical summaries (e.g., the distribution of sample means). The other crucial thing is that there the sampling variation can be quantified. The standard error is the quantification of that sampling error. In this case, it gives a numerical answer to the question of how variable the sample mean will be because of random sampling.

## Quantification of Uncertainty in Regression

We can extend these ideas to regression. Now the thought experiment goes something like this: Imagine you have a population that is infinitely large. The observations in this population have two attributes, call them  $X$  and  $Y$ . The relationship between these two attributes can be expressed via a regression equation as:  $\hat{Y} = \beta_0 + \beta_1(X)$ . Randomly sample  $n$  observations from the population. This time, rather than computing a mean, regress the sample  $Y$  values on the sample  $X$  values. Since the sample regression coefficients are estimates of the population parameters, we will write this as:  $\hat{Y} = \hat{\beta}_{0,1} + \hat{\beta}_{1,1}(X)$ . Repeat the process. This time the regression equation is:  $\hat{Y} = \hat{\beta}_{0,2} + \hat{\beta}_{1,2}(X)$ . Continue this process an infinite number of times.

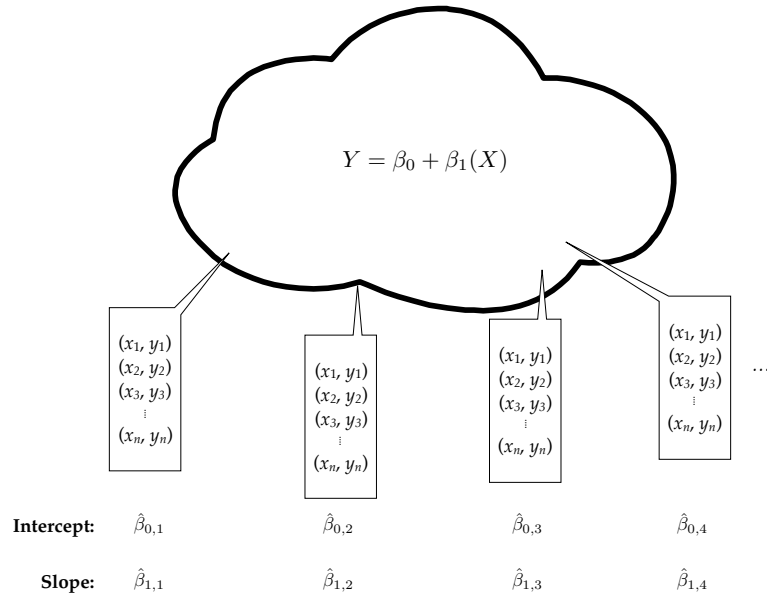


Figure 3. Thought experiment for sampling samples of size  $n$  from the population to obtain the sampling distribution of the regression coefficients.

<sup>1</sup>Mathematically, we would write  $E(\hat{\mu}) = \mu$ .

Statistical theory again predicts the characteristics of the two distributions, that of  $\hat{\beta}_0$  and that of  $\hat{\beta}_1$ . The distribution of  $\hat{\beta}_0$  can be expressed as,

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \sigma_\epsilon \sqrt{\frac{1}{n} + \frac{\mu_X^2}{\sum (X_i - \mu_X)^2}}\right).$$

Similarly, the distribution of  $\hat{\beta}_1$  can be expressed as,

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma_\epsilon}{\sigma_x \sqrt{n-1}}\right).$$

Again, don't panic over the formulae. What is important is that theory allows us to quantify the variation in both  $\hat{\beta}_0$  and  $\hat{\beta}_1$  that is due to sampling error. In practice, our statistical software will give us the numerical estimates of the two standard errors.

## Obtaining SEs for the Regression Coefficients

To obtain the standard errors for the regression coefficients, we will fit the regression using the `lm()` function and save the output into an object, as we did previously. Now, however, we will use the `tidy()` function from the **broom** package to display the fitted regression output.

```
# Display the coefficient-level output
tidy(lm.1)
```

```
# A tibble: 2 x 5
  term      estimate std.error statistic  p.value
  <chr>      <dbl>    <dbl>    <dbl>   <dbl>
1 (Intercept)  74.3      1.94     38.3 1.01e-60
2 homework    1.21     0.354     3.43 8.85e- 4
```

In the displayed output, we now obtain the estimates for the standard errors in addition to the coefficient estimates. We can use these values to quantify the amount of uncertainty due to sampling error. For example, the estimate for the slope, 1.21, has a standard error of 0.35. One way to envision this is as a distribution. Our best guess (mean) for the slope parameter is 1.21. The standard deviation of this distribution is 0.35, which indicates the precision (uncertainty) of our estimate.

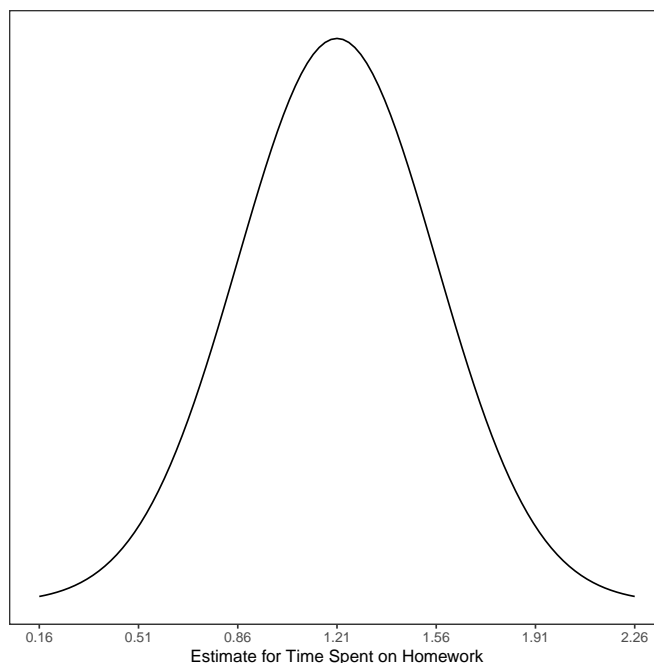


Figure 4. Sampling distribution of the slope coefficient. The distribution is normal with a mean of 1.21 and a standard error of 0.35.

In the social sciences, it is typical to express uncertainty as  $\pm 2(SE)$ . Here we would say that because of sampling variation, the slope is likely between 0.51 and 1.91. Interpreting this, we might say,

For all 8th-graders in the district, each one-hour difference in time spent on homework per week is associated with a difference in overall GPA between 0.51 and 1.91, on average.

Similarly, we could express the uncertainty in the intercept as,

$$74.29 \pm 2(1.94) = [70.41, 78.17]$$

Interpreting this, we might say,

The average GPA for all 8th-grade students in the district who spend zero hours per week on homework is between 70.41 and 78.17.

We can use the `confint()` function to obtain these limits directly. We just provide the fitted regression object as the input to this function.<sup>2</sup>

```
confint(lm.1, level = 0.95)
```

```

                2.5 %    97.5 %
(Intercept) 70.4359280 78.143426
homework    0.5116668  1.916751

```

<sup>2</sup>The actual limits from the `confint()` function are computed using a multiplier that is slightly different than two; thus the discrepancy between our off-the-cuff computation earlier and the result from R. Using a multiplier of two is often close enough for practical purposes, especially when the sample size is large.

## Hypothesis Testing

Some research questions point to examining whether the value of some regression parameter differs from a specific value. For example, it may be of interest whether a particular population model (e.g., one where  $\beta_1 = 0$ ) could produce the sample result of a particular  $\hat{\beta}_1$ . To test something like this, we state the value we want to test in a statement called a *null hypothesis*. For example,

$$H_0 : \beta_1 = 0$$

The hypothesis is a statement about the population. Here we hypothesize  $\beta_1 = 0$ . It would seem logical that one could just examine the estimate of the parameter from the observed sample to answer this question, but we also have to account for sampling uncertainty. The key is to quantify the sampling variation, and then see if the sample result is unlikely given the stated hypothesis.

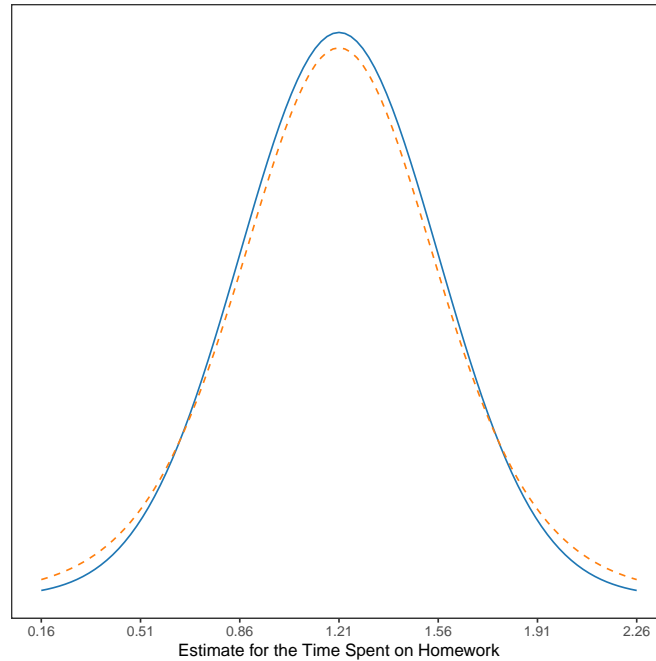
One question of interest may be: Is there evidence that the average GPA differs for different amounts of time spent on homework? In our example, we have a  $\hat{\beta}_1 = 1.21$ . This is sample evidence, but does 1.21 differ from 0 more than we would expect because of random sampling? If it doesn't, we cannot really say that the average GPA differs for different amounts of time spent on homework. To test this, we make an assumption that there is no relationship between time spent on homework and GPA, in other words, the slope of the line under this assumption would be 0. Before we talk about how to test this, we need to introduce one wrinkle into the procedure.

## Estimating Variation from Sample Data: No Longer Normal

In theory, the sampling distribution for two regression coefficients were both normally distributed. This is the case when we know the variation parameters in the population. For example, for the sampling distribution of the slope to be normally distributed, we would need to know  $\sigma_\epsilon$  and  $\sigma_x$ .

In practice these values are typically unknown and are estimated from the sample data. Anytime we are estimating things we introduce additional uncertainty. In this case, the uncertainty affects the shape of the sampling distribution.





*Figure 5.* Comparison of two distributions. The normal distribution (solid, blue) and one with additional uncertainty (dashed, orange).

Compare the normal distribution (solid, blue) to the distribution with additional uncertainty (dashed, orange). From the figure you can see that the additional uncertainty slightly changed the shape of the distribution from normal.

- It is still symmetric and unimodal (like the normal distribution).
- The additional uncertainty makes more extreme values more likely than they are in the normal distribution.
- The additional uncertainty makes values in the middle less likely than they are in the normal distribution.

It is important to note that the amount of uncertainty affects how closely the shape of the distribution matches the normal distribution. And, that the sample size directly affects the amount of uncertainty we have. All things being equal, we have less uncertainty when we have larger samples. The following figure illustrates this idea.

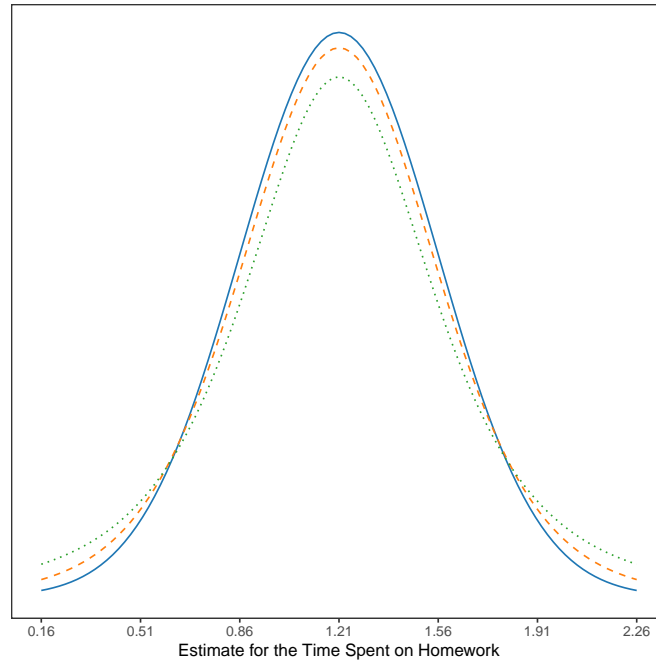


Figure 6. The normal distribution (solid, blue) and two with additional uncertainty; one based on  $n=10$  (dashed, orange), and the other based on  $n=4$  (dotted, green).

## The $t$ -Distribution

As pointed out, the distributions with uncertainty introduced from using a sample of data are not normally distributed. Thus, it doesn't make sense to use a normal distribution as a model for describing the sampling variation. Instead, we will use a  $t$ -distribution; a family of distributions that have several advantageous properties:

- They are unimodal and symmetric.
- They have more variation (uncertainty) than the normal distribution resulting in a distribution that has thicker tails and is shorter in the middle than a normal distribution.
- How thick the tails are and how short the middle of the distribution is, is related to the sample size.

Specifically, the  $t$ -distribution is unimodal and symmetric with a mean of 0. The variance of the distribution (which also specifies the exact shape), is

$$\text{Var} = \frac{df}{df - 2}$$

for  $df > 2$  where  $df$  is referred to as the *degrees of freedom*.

## Back to the Test

Recall that we are interested in testing the following hypothesis,

$$H_0 : \beta_1 = 0$$

To test this we compute the number of standard errors that our observed slope ( $\hat{\beta}_1 = 1.21$ ) is from the hypothesized value of zero (stated in the null hypothesis). Since we already obtained the standard error for the slope ( $SE = 0.354$ ), we just use some straight-forward algebra to compute this:

$$\frac{1.21 - 0}{0.354} = 3.42$$

Interpreting this, we can say,

The observed slope of 1.21 is 3.42 standard errors from the expected value of 0.

This value is referred to as the observed  $t$ -value. (It is similar to a  $z$ -value in the way it is computed; it is standardizing the distance from the observed slope to the hypothesized value of zero. But, since we had to estimate the SE using the data, we introduced additional uncertainty; hence a  $t$ -value.)

We can evaluate this  $t$ -value within the appropriate  $t$ -distribution. For regression coefficients, the  $t$ -distribution we will use for evaluation has degrees of freedom that are a function of the sample size and the number of coefficient parameters being estimated in the regression model, namely,

$$df = n - (\text{number of parameters}).$$

In our example the sample size ( $n$ ) is 100, and the number of coefficient parameters being estimated in the regression model is two ( $\beta_0$  and  $\beta_1$ ). Thus,

$$df = 100 - 2 = 98$$

Based on this, we will evaluate our observed  $t$ -value of 3.42 using a  $t$ -distribution with 98 degrees of freedom. Using this distribution, we can compute the probability of obtaining a  $t$ -value (under random sampling) at least as extreme as the one in the data under the assumed model. This is equivalent to finding the area under the probability curve for the  $t$ -distribution that is greater than or equal to 3.42.<sup>3</sup> This is called the  $p$ -value.

---

<sup>3</sup>We actually compute the area under the probability curve that is greater than or equal to 3.42 AND that is less than or equal to  $-3.42$ .

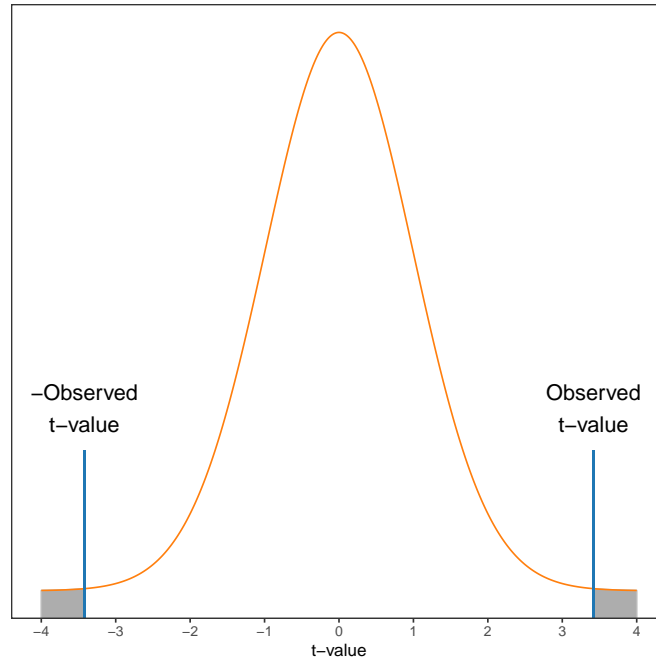


Figure 7. Plot of the probability curve for the  $t(98)$  distribution. The shaded area under the curve represents the  $p$ -value for a two-tailed test evaluating whether the population slope is zero using an observed  $t$ -value of 73.42.

The  $p$ -value is computed for us and displayed in the `tidy()` output, along with the  $t$ -value (provided in the `statistic` column). In our example,  $p = 0.000885$ . (Note that the  $p$ -value might be printed in scientific notation. For example, it may be printed as  $8.85e-04$ , which is equivalent to  $8.85 \times 10^{-4}$ .) To interpret this we would say,

The probability of observing a  $t$ -value of 3.42, or a  $t$ -value that is more extreme, under the assumption that  $\beta_1 = 0$  is 0.000885.

This is equivalent to saying:

The probability of observing a sample slope of 1.21, or a slope that is more extreme, under the assumption that  $\beta_1 = 0$  is 0.000885.

This is quite unlikely, so it serves as evidence against the hypothesized model. In other words, it is likely that  $\beta_1 \neq 0$ .

## Testing the Intercept

The hypothesis being tested for the intercept is  $H_0 : \beta_0 = 0$ . The `tidy()` output also provides information about this test:

```
tidy(lm.1)
```

```
# A tibble: 2 x 5
  term      estimate std.error statistic  p.value
<chr>      <dbl>      <dbl>      <dbl>    <dbl>
1 (Intercept)  74.3        1.94       38.3  1.01e-60
2 homework    1.21        0.354       3.43  8.85e- 4
```

The results indicate that the observed intercept of 74.28 is 38.26 standard errors from the hypothesized value of 0;

$$t = \frac{74.28 - 0}{1.94} = 38.26$$

Assuming the null hypothesis that  $\beta_0 = 0$  is true, the probability of observing a sample intercept of 74.28 or one that more extreme, is  $1.01 \times 10^{-60}$ . (Any  $p$ -value less than .001 is typically reported as  $p < .001$ .) This is evidence against the hypothesized model. Because of this, we would say the empirical data are inconsistent with the hypothesis that  $\beta_0 = 0$ ; it is unlikely that the intercept in the population is zero.

## “Statistical Significance”: An Outdated Idea for Research

You may have read papers or taken statistics courses that emphasized the language “statistically significant”. This adjective was typically used when the empirical evidence was inconsistent with a hypothesized model, and the researcher subsequently “rejected the null hypothesis”. In the social sciences this occurred when the  $p$ -value was less than or equal to 0.05.

In 2019, the American Statistical Association put out a special issue in one of their premier journals, stating,

“...it is time to stop using the term “statistically significant” entirely. Nor should variants such as “significantly different,” “ $p < 0.05$ ,” and “nonsignificant” survive, whether expressed in words, by asterisks in a table, or in some other way. Regardless of whether it was ever useful, a declaration of “statistical significance” has today become meaningless.” (Wasserstein, Schirm, & Lazar, 2019, p. 2)

They went on to say,

“...no  $p$ -value can reveal the plausibility, presence, truth, or importance of an association or effect. Therefore, a label of statistical significance does not mean or imply that an association or effect is highly probable, real, true, or important. Nor does a label of statistical nonsignificance lead to the association or effect being improbable, absent, false, or unimportant.” (Wasserstein et al., 2019, p. 2)

This is not to say that  $p$ -values should not be reported; they should. But rather that we should not arbitrarily dichotomize a continuous measure into two categories whose labels are at best meaningless and at worst misleading. The goal of scientific inference (which is much broader than statistical inference for a single study) is replicability and empirically generalizable results and findings. And, as Hubbard, Haig, & Parsa (2019) point out, declaring findings as “significant” or “not significant” works in direct opposition to the broader culmination of knowledge and evidence in a field.

Instead, we want to begin to see the  $p$ -value as a measure of incompatibility between the empirical data and a very specific model, one in which a certain set of assumptions are true. Both the empirical data (which are unique to the specific study) and the model's set of assumptions often make the  $p$ -value unique to the specific study carried out and less useful in the broader goal of scientific inference. As such we need to come to view the  $p$ -value for what it is, one measure of evidence, for one very particular model, in one very localized study. As Ron Wasserstein reminds us,

“Small  $p$ -values are like a right-swipe in Tinder. It means you have an interest. It doesn't mean you're ready to book the wedding venue.”

## Confidence Intervals as Compatibility Intervals

To build on this, let's return to the reporting and interpretation of confidence intervals. In our example, the 95% CI was:

```
confint(lm.1, level = 0.95)
```

	2.5 %	97.5 %
(Intercept)	70.4359280	78.143426
homework	0.5116668	1.916751

One way of interpreting this interval is that every value in the interval is a parameter value that is reasonably compatible with the empirical data. For example, in considering the CI for the slope parameter, population slope ( $\beta_1$ ) values between 0.51 and 1.92 are all reasonably compatible with the empirical data (with the caveat that, again, all the assumptions used to create the interval are satisfied). As applied researchers, we should describe the practical implications of all values inside the interval, especially the observed effect (or point estimate) and the limits.

For us this means describing the practical implications of the true slope being 1.21, as low as 0.51, and as high as 1.92. Are these meaningful differences in GPA (measure on a 100-pt. scale)? Given that the SD for GPA was 7.62, a one-hour difference in time spent on homework is associated with at most a 0.25 SD difference in GPA or as little as a 0.07 SD difference in GPA. This is not a large difference, however whether it is meaningful depends on previous research about GPA.<sup>4</sup>

Confidence intervals help us keep an open-mind about uncertainty, after all they suggest several values that are compatible with the empirical data. However, they can also be misleading. Amrhein, Greenland, & McShane (2019) point out four key points for us to remember as we use CIs:

- Just because the interval gives the values most compatible with the data, given the assumptions, it doesn't mean values outside it are incompatible; they are just less compatible.
- Not all values inside are equally compatible with the data, given the assumptions. The point estimate is the most compatible, and values near it are more compatible than those near the limits.
- Like the 0.05 threshold from which it came, the default 95% used to compute intervals is itself an arbitrary convention.
- Last, and most important of all, be humble: compatibility assessments hinge on the correctness of the statistical assumptions used to compute the interval. In practice, these assumptions are at best subject to considerable uncertainty.

---

<sup>4</sup>It turns out this is quite a complicated question and the effects of homework depend on a variety of student factors, including age, culture, household income, etc. Many studies have also found a non-linear effect of homework, indicating there may be an optimum amount for some groups of students.

## Coefficient Plot

One plot that gives a great deal of information about the estimates of the regression coefficients and the associated uncertainty is a *coefficient plot*. This plot, recommended by Gelman & Hill (2007), is a graphical representation of the information provided in the `tidy()` output.

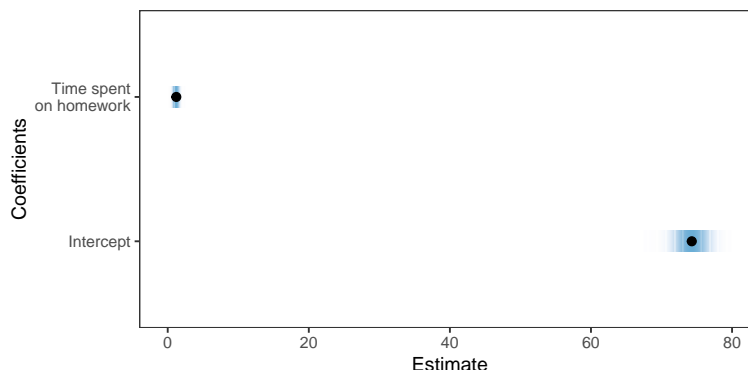


Figure 8. Coefficient plot for the model regressing GPA on time spent on homework. Uncertainty based on the 95% confidence intervals are displayed.

The coefficient plot shows the estimates of the regression coefficients (dots) and the uncertainty in those estimates via the confidence intervals (blue shading). Notice that darker shading is associated with parameter values that are more probable given the empirical data; the sample estimates are the most likely values for the regression parameters.

To create a coefficient plot, we will use the `stat_confidence_density()` function from the **ungeviz** package. You need to install the **ungeviz** package from `syntaxis` since it is not currently available on CRAN. To do this you need to use the `install_github()` function from the **devtools** package. The syntax to install **ungeviz** is:

```
library(devtools)
install_github("wilkelab/ungeviz")
```

Once **ungeviz** is installed and loaded we have access to the `stat_confidence_density()` function. This function, which we will use as a layer in `ggplot()` will be used to create the coefficient plot. The syntax I used to create the coefficient plot above is:

```
# Load ungeviz library
library(ungeviz)

# Create plot
ggplot(data = tidy(lm.1), aes(x = estimate, y = term)) +
  stat_confidence_density(aes(moe = std.error, confidence = 0.68, fill = stat(ndensity)),
    height = 0.15) +
  geom_point(aes(x = estimate), size = 2) +
  scale_fill_gradient(low = "#eff3ff", high = "#6baed6") +
  theme_bw() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  ) +
  scale_x_continuous(name = "Estimate", limits = c(0, 80)) +
  scale_y_discrete(name = "Coefficients", labels = c("Intercept", "Time spent\non homework"))
```

Note that the data we use in the `data=` argument to create the plot is the coefficient-level output from the `tidy()` function. The `stat_confidence_density()` layer requires an aesthetic mapping that relates `moe=` to some margin of error, in our case, the `std.error` column of the `tidy()` dataframe. Since one standard error is the margin of error on a 68% confidence interval, we also include `confidence=0.68` in the `aes()` function. Lastly, the argument `fill = stat(ndensity)` is included in `aes()` to allow for colored shading in the interval.

The remainder of the syntax is making the plot pretty. The `scale_fill_gradient()` layer sets the two colors to interpolate between for the shading. In our case those give a light and dark blue color. You also may need to toy with the limits on the `x`-variable to encompass the entire interval in the plot.

There are several variations of this plot. For example, below I omit the intercept from this plot. To do this, I use the `filter()` function to omit the row in the `tidy()` output that includes the intercept. I also change the limits on the `x`-axis to better fit the homework interval. Here I also dropped the `fill = stat(ndensity)` from the aesthetic mapping in the `stat_confidence_density()` layer, which gives a greyscale plot.

```
# Omit intercept
my_coefs = tidy(lm.1) %>%
  filter(term != "(Intercept)")

# Create plot
ggplot(data = my_coefs, aes(x = estimate, y = term)) +
  stat_confidence_density(aes(moe = std.error, confidence = 0.68), height = 0.15) +
  geom_point(aes(x = estimate), size = 2) +
  theme_bw() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  ) +
  scale_x_continuous(name = "Estimate", limits = c(0, 2)) +
  scale_y_discrete(name = "Coefficients", labels = c("Time spent\non homework"))
```

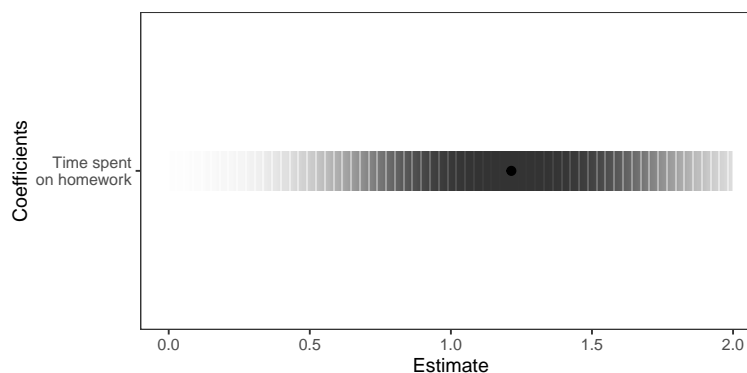


Figure 9. Coefficient plot for the model regressing income on education. Uncertainty based on the 95% confidence interval is displayed. Note that the intercept has been omitted.



## References

- Amrhein, V., Greenland, S., & McShane, B. (2019). Comment: Retire statistical significance. *Nature*, 567, 305–307. doi:[10.1038/d41586-019-00857-9](https://doi.org/10.1038/d41586-019-00857-9)
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Hubbard, R., Haig, B. D., & Parsa, R. A. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician*, 1–9. doi:[10.1080/00031305.2018.1464947](https://doi.org/10.1080/00031305.2018.1464947)
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond " $p < 0.05$ ". *The American Statistician*, 73(sup1), 1–19. doi:[10.1080/00031305.2019.1583913](https://doi.org/10.1080/00031305.2019.1583913)