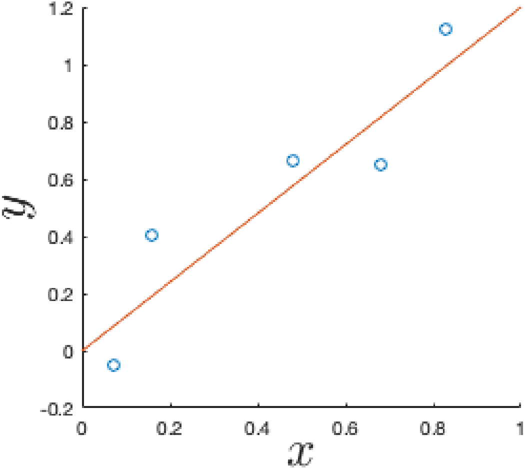
Linear Models II

BIOE 210

Review: A *noisy* linear system

x^{true}	y^{true}
0.07	-0.05
0.16	0.40
0.48	0.66
0.68	0.65
0.83	1.12



A matrix formalism for linear models

Let's write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}$$

A matrix formalism for linear models

Let’s write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}$$

$$\begin{pmatrix} -0.05 \\ 0.40 \\ 0.66 \\ 0.65 \\ 1.12 \end{pmatrix} = \begin{pmatrix} 1 & 0.07 \\ 1 & 0.16 \\ 1 & 0.48 \\ 1 & 0.68 \\ 1 & 0.83 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

A matrix formalism for linear models

Let's write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}$$

$$\begin{pmatrix} -0.05 \\ 0.40 \\ 0.66 \\ 0.65 \\ 1.12 \end{pmatrix} = \begin{pmatrix} 1 & 0.07 \\ 1 & 0.16 \\ 1 & 0.48 \\ 1 & 0.68 \\ 1 & 0.83 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Solving the linear system

A few points about $\mathbf{y} = \mathbf{X}\beta + \epsilon$:

- The unknowns are β, not $\mathbf{X}$.
- The coefficient matrix $\mathbf{X}$ is called the *model matrix*.
- The model matrix $\mathbf{X}$ is rarely square.

Solving the linear system

A few points about $\mathbf{y} = \mathbf{X}\beta + \epsilon$:

- The unknowns are β, not $\mathbf{X}$.
- The coefficient matrix $\mathbf{X}$ is called the *model matrix*.
- The model matrix $\mathbf{X}$ is rarely square.

The solution to this system that minimizes the errors in ϵ is

$$\beta = \mathbf{X}^{+}\mathbf{y}$$

where $\mathbf{X}^{+}$ is the *pseudoinverse* of $\mathbf{X}$.

The intercept

The linear model

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

has intercept (or *grand mean*) β_0.

The intercept

The linear model

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

has intercept (or *grand mean*) β_0.

Really, we should write the model as

$$\mathbf{y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon.$$

The intercept

The linear model

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

has intercept (or *grand mean*) β_0.

Really, we should write the model as

$$\mathbf{y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon.$$

Or, in vector form

$$\mathbf{y} = \begin{pmatrix} \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon.$$

Models without an intercept

If we wanted to fit a model without an intercept, we would write

$$\mathbf{y} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon.$$

Models without an intercept

If we wanted to fit a model without an intercept, we would write

$$\mathbf{y} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon.$$

Why would we not want an intercept?

- When all inputs $\mathbf{x}_i = 0$, the response $y = \beta_0$.

Models without an intercept

If we wanted to fit a model without an intercept, we would write

$$\mathbf{y} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon.$$

Why would we not want an intercept?

- When all inputs $\mathbf{x}_i = 0$, the response $y = \beta_0$.
- If we know our system has zero response without an input, we don't include an intercept.

Models without an intercept

If we wanted to fit a model without an intercept, we would write

$$\mathbf{y} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon.$$

Why would we not want an intercept?

- When all inputs $\mathbf{x}_i = 0$, the response $y = \beta_0$.
- If we know our system has zero response without an input, we don't include an intercept.
- This is rare, so most models include an intercept.

Prediction vs. Inference

Prediction uses a model to find the response of inputs we've never seen before.

Inference uses a model to understand what inputs determine the response.

How accurate are a model's predictions?

How accurate are a model's predictions?

Our training data include a design matrix **X** and a vector of responses **y**. Each has *n* observations (rows).

How accurate are a model's predictions?

Our training data include a design matrix $\mathbf{X}$ and a vector of responses $\mathbf{y}$. Each has n observations (rows).

We fit the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ to find the best parameters β.

How accurate are a model's predictions?

Our training data include a design matrix $\mathbf{X}$ and a vector of responses $\mathbf{y}$. Each has n observations (rows).

We fit the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ to find the best parameters β.

We feed the training data back into the model to find the residuals:

$$\begin{aligned} \epsilon &= \mathbf{y} - \mathbf{X}\beta \\ &= \mathbf{y}^{\text{true}} - \mathbf{y}^{\text{pred}} \end{aligned}$$

How accurate are a model's predictions?

Our training data include a design matrix **X** and a vector of responses **y**. Each has *n* observations (rows).

We fit the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ to find the best parameters β.

We feed the training data back into the model to find the residuals:

$$\begin{aligned} \epsilon &= \mathbf{y} - \mathbf{X}\beta \\ &= \mathbf{y}^{\text{true}} - \mathbf{y}^{\text{pred}} \end{aligned}$$

We quantify the accuracy using the *root mean squared error* of the residuals.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{\text{true}} - y_i^{\text{pred}}\right)^2}$$

Prediction intervals

A model's predictions should be reported ± the RMSE.

Prediction intervals

A model's predictions should be reported ± the RMSE.

The 95% confidence interval of a prediction is $\pm 2 \times$ RMSE.

Prediction intervals

A model's predictions should be reported ± the RMSE.

The 95% confidence interval of a prediction is ±2 × RMSE.

Example: A model to predict pulse rate has RMSE of 12 bpm. If the model predicts 68 bpm for a patient, the 95% confidence interval is

$$[68 - 2 \times 12,\ 68 + 2 \times 12] = [44, 92]\ \text{bpm}$$

Prediction intervals

A model's predictions should be reported ± the RMSE.

The 95% confidence interval of a prediction is $\pm 2 \times$ RMSE.

Example: A model to predict pulse rate has RMSE of 12 bpm. If the model predicts 68 bpm for a patient, the 95% confidence interval is

$$[68 - 2 \times 12,\ 68 + 2 \times 12] = [44, 92] \text{ bpm}$$

Remember: If you transformed your model, the RMSE will be in the transformed units!

Model inference

Model inference involves asking two questions about each of the models inputs.

1. How large of an effect does this input have on the output?
2. How confident are we in our estimate of this effect?

Model inference

Model inference involves asking two questions about each of the models inputs.

1. How large of an effect does this input have on the output?
2. How confident are we in our estimate of this effect?

Consider the two input model

$$\mathbf{y} = 1.2 - 3.6\mathbf{x}_1 + 0.8\mathbf{x}_2 + \epsilon.$$

Model inference

Model inference involves asking two questions about each of the models inputs.

1. How large of an effect does this input have on the output?
2. How confident are we in our estimate of this effect?

Consider the two input model

$$\mathbf{y} = 1.2 - 3.6\mathbf{x}_1 + 0.8\mathbf{x}_2 + \epsilon.$$

The parameters -3.6 and 0.8 are called *effect sizes*.

Model inference

Model inference involves asking two questions about each of the models inputs.

1. How large of an effect does this input have on the output?
2. How confident are we in our estimate of this effect?

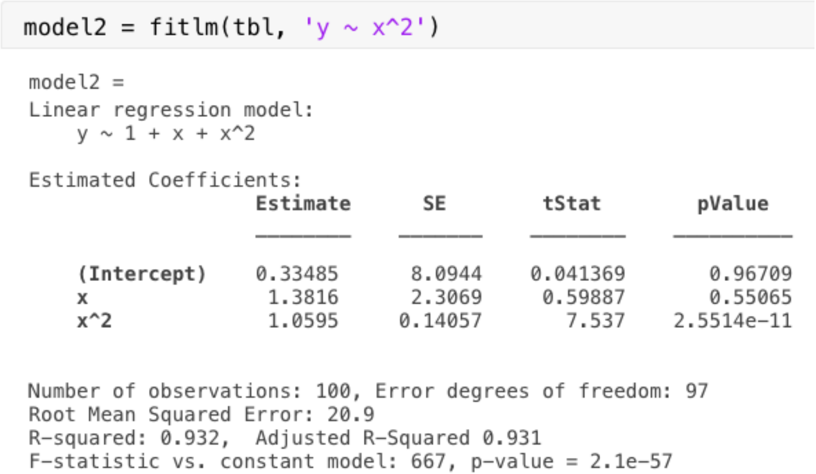
Consider the two input model

$$\mathbf{y} = 1.2 - 3.6\mathbf{x}_1 + 0.8\mathbf{x}_2 + \epsilon.$$

The parameters -3.6 and 0.8 are called *effect sizes*.

- A unit change in variable $\mathbf{x}_1$ would *decrease* the response by 3.6 units.
- A unit change in $\mathbf{x}_2$ would *increase* the response by 0.8 units.

How sure are we of the effect sizes?



```
model2 = fitlm(tbl, 'y ~ x^2')
```

```
model2 =
Linear regression model:
    y ~ 1 + x + x^2

Estimated Coefficients:
                   Estimate      SE        tStat       pValue
                   ________    _______    ________    __________

    (Intercept)    0.33485      8.0944    0.041369       0.96709
    x               1.3816      2.3069     0.59887       0.55065
    x^2             1.0595     0.14057       7.537    2.5514e-11


Number of observations: 100, Error degrees of freedom: 97
Root Mean Squared Error: 20.9
R-squared: 0.932,  Adjusted R-Squared 0.931
F-statistic vs. constant model: 667, p-value = 2.1e-57
```

What does this *p*-value mean?

- A low *p*-value indicates that an effect of this size was unlikely to occur randomly.
- It also means the confidence interval *excludes* zero, so we reject the hypothesis that the true effect is zero.

What does this *p*-value mean?

- A low *p*-value indicates that an effect of this size was unlikely to occur randomly.
- It also means the confidence interval *excludes* zero, so we reject the hypothesis that the true effect is zero.
- It does **not** mean that the effect is practically significant or important. (That's up to the effect size.)

RESEARCH ARTICLE

Marital satisfaction and break-ups differ across on-line and off-line meeting venues

John T. Cacioppo, Stephanie Cacioppo, Gian C. Gonzaga,
Elizabeth L. Ogburn, and Tyler J. VanderWeele

PNAS June 18, 2013 110 (25) 10135-10140; https://doi.org/10.1073/pnas.1222447110

For respondents categorized as currently married at the time of the survey, we examined marital satisfaction. Analyses indicated that currently married respondents who met their spouse on-line reported higher marital satisfaction (M = 5.64, SE = 0.02, n = 5,349) than currently married respondents who met their spouse off-line (M = 5.48, SE = 0.01, n = 12,253; mean difference = 0.18, $F_{(1,17,601)} = 46.67$, $P < 0.001$*).*

Interactions

Imagine we're modeling the response (y) from two input variables, $\mathbf{x}_1$ and $\mathbf{x}_2$. The simplest model is

$$y = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

Interactions

Imagine we're modeling the response (y) from two input variables, $\mathbf{x}_1$ and $\mathbf{x}_2$. The simplest model is

$$y = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

The coefficient β_1 measures the effect of $\mathbf{x}_1$ and β_2 measures the effect of $\mathbf{x}_2$. These effects are **independent**.

Interactions

Imagine we're modeling the response (y) from two input variables, $\mathbf{x}_1$ and $\mathbf{x}_2$. The simplest model is

$$y = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon$$

The coefficient β_1 measures the effect of $\mathbf{x}_1$ and β_2 measures the effect of $\mathbf{x}_2$. These effects are **independent**.

What is there is another effect that depends on both $\mathbf{x}_1$ and $\mathbf{x}_2$? This is an **interaction** between $\mathbf{x}_1$ and $\mathbf{x}_2$.

How do we model interactions?

We model the interaction of $\mathbf{x}_1$ and $\mathbf{x}_2$ using the product of these variables.

$$y = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_{12} \mathbf{x}_1 {:} \mathbf{x}_2 + \epsilon$$

The coefficient β_{12} is the effect size of the interaction.

How do we model interactions?

We model the interaction of $\mathbf{x}_1$ and $\mathbf{x}_2$ using the product of these variables.

$$y = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_{12} \mathbf{x}_1{:}\mathbf{x}_2 + \epsilon$$

The coefficient β_{12} is the effect size of the interaction.

Why do we multiply $\mathbf{x}_1$ and $\mathbf{x}_2$? There are at least two ways to interpret this term.

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

on/off → interaction when both "on"

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_1 : \mathbf{x}_2$
0	0	0
0	1	0
1	0	0
1	1	1

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

on/off → interaction when both "on"

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_1{:}\mathbf{x}_2$
0	0	0
0	1	0
1	0	0
1	1	1

high/low → interaction when both "high" or both "low"

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_1{:}\mathbf{x}_2$
-1	-1	$+1$
-1	$+1$	-1
$+1$	-1	-1
$+1$	$+1$	$+1$

The augmented slope interpretation

We can also interpret the interaction as one variable changing the effect of the other variable.

$$\begin{aligned} y &= \beta_1 \mathbf{x}_1 + \beta_2(\mathbf{x}_1){:}\mathbf{x}_2 + \epsilon \\ &= \beta_1 \mathbf{x}_1 + (\beta_2 + \beta_{12}\mathbf{x}_1){:}\mathbf{x}_2 + \epsilon \\ &= \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_{12}\mathbf{x}_1{:}\mathbf{x}_2 + \epsilon \end{aligned}$$

Things to remember about interactions

- Interaction are modeled as the product of variables.
- The interaction effect is “above and beyond" the independent effects (synergy/super-additivity, antagonism/sub-additivity).
- Higher-order interactions are possible (e.g. $\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3$), but these are rare.