

Announcements

- Homework
 - Homework 7 and Lab due tomorrow night, due to the late posting
- Test on Friday!
 - Over Chapters 2,3, and 4
 - Learning objectives have been posted
 - Should have sample problems and solutions out tonight
 - I'm also trying hard to have relevant grading complete
- Wednesday I think will be our last in-class lab, and will be on making fancy plots.
- Polling: `rembold-class.ddns.net`

Final Projects

- New location on the website for all final project related information
- Guidelines posted there, as well as on Campuswire
- Potential datasets have been uploaded to the website as well
 - All my datasets are comma-separated tables (csv) format. Can be read into R with `data <- read.csv("<Name of file>")`.
 - Not all datasets might work with different group numbers
 - You are always free to find your own as well. Just make sure they have enough variables to give you options for multiple regression or ANOVA (as needed).
- Please let me know by the end of the day on Wednesday who you are planning to work with (groups of 1, 2, or 3).

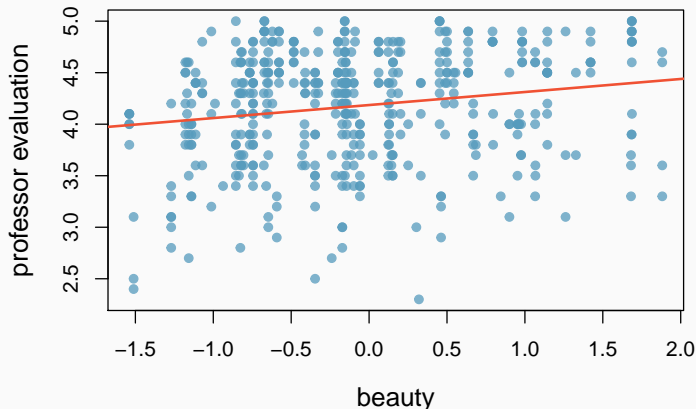
Beauty in the classroom

- Data: Student evaluations of instructors' beauty and teaching quality for 463 courses at the University of Texas.
- Evaluations conducted at the end of semester, and the beauty judgements were made later, by six students who had not attended the classes and were not aware of the course evaluations (2 upper level females, 2 upper level males, one lower level female, one lower level male).

Source: Hamermesh & Parker. (2004) "Beauty in the classroom: instructors pulchritude and putative pedagogical productivity Economics Education Review.

Professor rating vs. beauty

Professor evaluation score (higher score means better) vs. beauty score (a score of 0 means average, negative score means below average, and a positive score above average):



Review Question

Which of the below is correct based on the model output?

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.19	0.03	167.24	0.00
beauty	0.13	0.03	4.00	0.00

$R^2 = 0.0336$

- A) Model predicts 3.36% of professor ratings correctly.
- B) Beauty is not a significant predictor of professor evaluation.
- C) Professors who score 1 point above average in their beauty score are tend to also score 0.13 points higher in their evaluation.
- D) 3.36% of variability in beauty scores can be explained by professor evaluation.

Review Question

Which of the below is correct based on the model output?

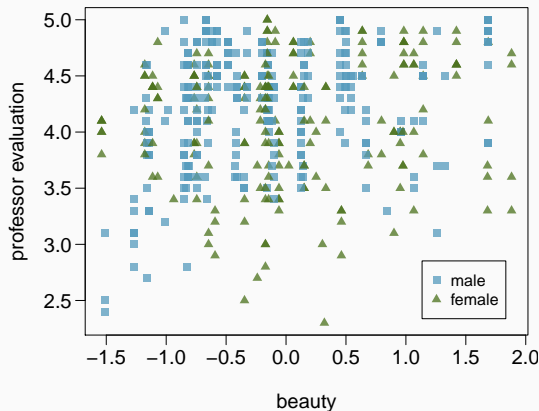
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.19	0.03	167.24	0.00
beauty	0.13	0.03	4.00	0.00
$R^2 = 0.0336$				

- A) Model predicts 3.36% of professor ratings correctly.
- B) Beauty is not a significant predictor of professor evaluation.
- C) Professors who score 1 point above average in their beauty score are tend to also score 0.13 points higher in their evaluation.
- D) 3.36% of variability in beauty scores can be explained by professor evaluation.

Exploratory analysis

Any interesting features?

For a given beauty score, are male professors evaluated higher, lower, or about the same as female professors?

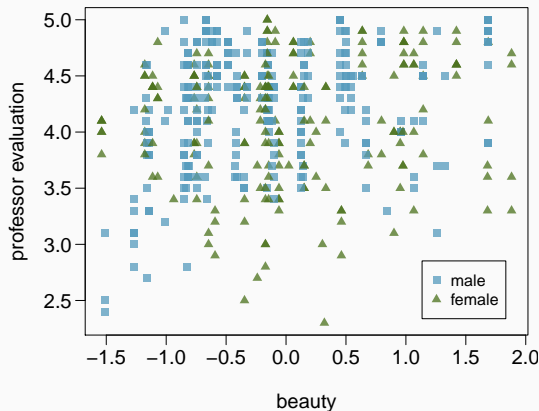


Exploratory analysis

Any interesting features?

- *Few females with very low beauty scores.*

For a given beauty score, are male professors evaluated higher, lower, or about the same as female professors?



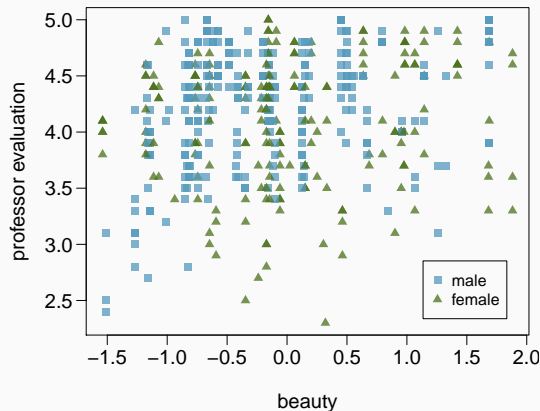
Exploratory analysis

Any interesting features?

- *Few females with very low beauty scores.*

For a given beauty score, are male professors evaluated higher, lower, or about the same as female professors?

- *Difficult to tell from this plot only.*



Professor rating vs. beauty + gender

For a given beauty score, are male professors evaluated higher, lower, or about the same as female professors?

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.09	0.04	107.85	0.00
beauty	0.14	0.03	4.44	0.00
gender.male	0.17	0.05	3.38	0.00
$R^2_{adj} = 0.057$				

- (a) higher
- (b) lower
- (c) about the same

Professor rating vs. beauty + gender

For a given beauty score, are male professors evaluated higher, lower, or about the same as female professors?

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.09	0.04	107.85	0.00
beauty	0.14	0.03	4.44	0.00
gender.male	0.17	0.05	3.38	0.00

$R^2_{adj} = 0.057$

- (a) higher → Beauty held constant, male professors are rated 0.17 points higher on average than female professors.
- (b) lower
- (c) about the same

Full model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.6282	0.1720	26.90	0.00
beauty	0.1080	0.0329	3.28	0.00
gender.male	0.2040	0.0528	3.87	0.00
age	-0.0089	0.0032	-2.75	0.01
formal.yes ¹	0.1511	0.0749	2.02	0.04
lower.yes ²	0.0582	0.0553	1.05	0.29
native.non english	-0.2158	0.1147	-1.88	0.06
minority.yes	-0.0707	0.0763	-0.93	0.35
students ³	-0.0004	0.0004	-1.03	0.30
tenure.tenure track ⁴	-0.1933	0.0847	-2.28	0.02
tenure.tenured	-0.1574	0.0656	-2.40	0.02

¹ formal: picture wearing tie&jacket/blouse, levels: yes, no

² lower: lower division course, levels: yes, no

³ students: number of students

⁴ tenure: tenure status, levels: non-tenure track, tenure track, tenured

Hypotheses

Just as the interpretation of the slope parameters take into account all other variables in the model, the hypotheses for testing for significance of a predictor also takes into account all other variables.

$H_0 : B_i = 0$ when other explanatory variables are included in the model.

$H_A : B_i \neq 0$ when other explanatory variables are included in the model.

Assessing significance: numerical variables

The p-value for age is 0.01. What does this indicate?

	Estimate	Std. Error	t value	Pr(> t)
...				
age	-0.0089	0.0032	-2.75	0.01
...				

- (a) Since p-value is positive, the higher the professor's age, the higher we would expect them to be rated.
- (b) If we keep all other variables in the model, there is strong evidence that professor's age is associated with their rating.
- (c) Probability that the true slope parameter for age is 0 is 0.01.
- (d) There is about 1% chance that the true slope parameter for age is not -0.0089.

Assessing significance: numerical variables

The p-value for age is 0.01. What does this indicate?

	Estimate	Std. Error	t value	Pr(> t)
...				
age	-0.0089	0.0032	-2.75	0.01
...				

- (a) Since p-value is positive, the higher the professor's age, the higher we would expect them to be rated.
- (b) If we keep all other variables in the model, there is strong evidence that professor's age is associated with their rating.
- (c) Probability that the true slope parameter for age is 0 is 0.01.
- (d) There is about 1% chance that the true slope parameter for age is not -0.0089.

Assessing significance: categorical variables

Tenure is a categorical variable with 3 levels: non tenure track, tenure track, tenured. Based on the model output given, which of the below is false?

	Estimate	Std. Error	t value	Pr(> t)
...				
tenure.tenure track	-0.1933	0.0847	-2.28	0.02
tenure.tenured	-0.1574	0.0656	-2.40	0.02

- (a) Reference level is non tenure track.
- (b) All else being equal, tenure track professors are rated, on average, about 0.19 points lower than non-tenure track professors.
- (c) All else being equal, tenured professors are rated, on average, about 0.16 points lower than non-tenure track professors.
- (d) All else being equal, there is a significant difference between the average ratings of tenure track and tenured professors.

Assessing significance: categorical variables

Tenure is a categorical variable with 3 levels: non tenure track, tenure track, tenured. Based on the model output given, which of the below is false?

	Estimate	Std. Error	t value	Pr(> t)
...				
tenure.tenure track	-0.1933	0.0847	-2.28	0.02
tenure.tenured	-0.1574	0.0656	-2.40	0.02

- (a) Reference level is non tenure track.
- (b) All else being equal, tenure track professors are rated, on average, about 0.19 points lower than non-tenure track professors.
- (c) All else being equal, tenured professors are rated, on average, about 0.16 points lower than non-tenure track professors.
- (d) All else being equal, there is a significant difference between the average ratings of tenure track and tenured professors.

Assessing significance

Which predictors do not seem to meaningfully contribute to the model, i.e. may not be significant predictors of professor's rating score?

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.6282	0.1720	26.90	0.00
beauty	0.1080	0.0329	3.28	0.00
gender.male	0.2040	0.0528	3.87	0.00
age	-0.0089	0.0032	-2.75	0.01
formal.yes	0.1511	0.0749	2.02	0.04
lower.yes	0.0582	0.0553	1.05	0.29
native.non english	-0.2158	0.1147	-1.88	0.06
minority.yes	-0.0707	0.0763	-0.93	0.35
students	-0.0004	0.0004	-1.03	0.30
tenure.tenure track	-0.1933	0.0847	-2.28	0.02
tenure.tenured	-0.1574	0.0656	-2.40	0.02

Approaches to Model Selection: Backward-elimination

1. R^2_{adj} approach:

- Start with the full model
- Drop one variable at a time and record R^2_{adj} of each smaller model
- Pick the model with the highest increase in R^2_{adj}
- Repeat until none of the models yield an increase in R^2_{adj}

2. p-value approach:

- Start with the full model
- Drop the variable with the highest p-value and refit a smaller model
- Repeat until all variables left in the model are significant

Backward-elimination: R^2_{adj} approach

Step	Variables included	R^2_{adj}
Full	beauty + gender + age + formal + lower + native + minority + students + tenure	0.0839

Backward-elimination: R^2_{adj} approach

Step	Variables included	R^2_{adj}
Full	beauty + gender + age + formal + lower + native + minority + students + tenure	0.0839
Step 1	gender + age + formal + lower + native + minority + students + tenure	0.0642
	beauty + age + formal + lower + native + minority + students + tenure	0.0557
	beauty + gender + formal + lower + native + minority + students + tenure	0.0706
	beauty + gender + age + lower + native + minority + students + tenure	0.0777
	beauty + gender + age + formal + native + minority + students + tenure	0.0837
	beauty + gender + age + formal + lower + minority + students + tenure	0.0788
	beauty + gender + age + formal + lower + native + students + tenure	0.0842
	beauty + gender + age + formal + lower + native + minority + tenure	0.0838
	beauty + gender + age + formal + lower + native + minority + students	0.0733

Backward-elimination: R^2_{adj} approach

Step	Variables included	R^2_{adj}
Full	beauty + gender + age + formal + lower + native + minority + students + tenure	0.0839
Step 1	gender + age + formal + lower + native + minority + students + tenure	0.0642
	beauty + age + formal + lower + native + minority + students + tenure	0.0557
	beauty + gender + formal + lower + native + minority + students + tenure	0.0706
	beauty + gender + age + lower + native + minority + students + tenure	0.0777
	beauty + gender + age + formal + native + minority + students + tenure	0.0837
	beauty + gender + age + formal + lower + minority + students + tenure	0.0788
	beauty + gender + age + formal + lower + native + students + tenure	0.0842
	beauty + gender + age + formal + lower + native + minority + tenure	0.0838
Step 2	beauty + gender + age + formal + lower + native + minority + students	0.0733
	gender + age + formal + lower + native + students + tenure	0.0647
	beauty + age + formal + lower + native + students + tenure	0.0543
	beauty + gender + formal + lower + native + students + tenure	0.0708
	beauty + gender + age + lower + native + students + tenure	0.0776
	beauty + gender + age + formal + native + students + tenure	0.0846
	beauty + gender + age + formal + lower + native + tenure	0.0844
	beauty + gender + age + formal + lower + native + students	0.0725

Backward-elimination: R^2_{adj} approach

Step	Variables included	R^2_{adj}
Full	beauty + gender + age + formal + lower + native + minority + students + tenure	0.0839
Step 2	gender + age + formal + lower + native + students + tenure	0.0647
	beauty + age + formal + lower + native + students + tenure	0.0543
	beauty + gender + formal + lower + native + students + tenure	0.0708
	beauty + gender + age + lower + native + students + tenure	0.0776
	beauty + gender + age + formal + native + students + tenure	0.0846
	beauty + gender + age + formal + lower + native + tenure	0.0844
	beauty + gender + age + formal + lower + native + students	0.0725
Step 3	gender + age + formal + native + students + tenure	0.0653
	beauty + age + formal + native + students + tenure	0.0534
	beauty + gender + formal + native + students + tenure	0.0707
	beauty + gender + age + native + students + tenure	0.0786
	beauty + gender + age + formal + students + tenure	0.0756
	beauty + gender + age + formal + native + tenure	0.0855
	beauty + gender + age + formal + native + students	0.0713
Step 4	gender + age + formal + native + tenure	0.0667
	beauty + age + formal + native + tenure	0.0553
	beauty + gender + formal + native + tenure	0.0723
	beauty + gender + age + native + tenure	0.0806
	beauty + gender + age + formal + tenure	0.0773
	beauty + gender + age + formal + native	0.0713

step function in R

Call:

```
lm(formula = profevaluation ~ beauty + gender + age + formal +  
    native + tenure, data = d)
```

Coefficients:

(Intercept)	beauty	gendermale	age
4.628435	0.105546	0.208079	-0.008844
formalyes	nativenon english	tenuretenure track	tenuretenured
0.132422	-0.243003	-0.206784	-0.175967

step function in R

Call:

```
lm(formula = profevaluation ~ beauty + gender + age + formal +  
    native + tenure, data = d)
```

Coefficients:

(Intercept)	beauty	gendermale	age
4.628435	0.105546	0.208079	-0.008844
formalyes	nativenon english	tenuretenure track	tenuretenured
0.132422	-0.243003	-0.206784	-0.175967

Best model: beauty + gender + age + formal + native + tenure

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender male	age	formal yes	lower yes	native non english	minority yes	students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender male	age	formal yes	lower yes	native non english	minority yes	students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02
Step 1	beauty	gender male	age	formal yes	lower yes	native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.38	0.03		0.34	0.02	0.01

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender	age	formal	lower	native	minority	students	tenure	tenure
		male		yes	yes	non english	yes		tenure track	tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02
Step 1	beauty	gender	age	formal	lower	native		students	tenure	tenure
		male		yes	yes	non english			tenure track	tenured
	0.00	0.00	0.01	0.04	0.38	0.03		0.34	0.02	0.01
Step 2	beauty	gender	age	formal		native		students	tenure	tenure
		male		yes		non english			tenure track	tenured
	0.00	0.00	0.01	0.05		0.02		0.44	0.01	0.01

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender male	age	formal yes	lower yes	native non english	minority yes	students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02
Step 1	beauty	gender male	age	formal yes	lower yes	native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.38	0.03		0.34	0.02	0.01
Step 2	beauty	gender male	age	formal yes		native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.05		0.02		0.44	0.01	0.01
Step 3	beauty	gender male	age	formal yes		native non english			tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.06		0.02			0.01	0.01

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender male	age	formal yes	lower yes	native non english	minority yes	students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02
Step 1	beauty	gender male	age	formal yes	lower yes	native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.38	0.03		0.34	0.02	0.01
Step 2	beauty	gender male	age	formal yes		native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.05		0.02		0.44	0.01	0.01
Step 3	beauty	gender male	age	formal yes		native non english			tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.06		0.02			0.01	0.01
Step 4	beauty	gender male	age			native non english			tenure tenure track	tenure tenured
	0.00	0.00	0.01			0.06			0.01	0.01

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender	age	formal	lower	native	minority	students	tenure	tenure
	0.00	male 0.00	0.01	yes 0.04	yes 0.29	non english 0.06	yes 0.35	0.30	tenure track 0.02	tenured 0.02
Step 1	beauty	gender	age	formal	lower	native		students	tenure	tenure
	0.00	male 0.00	0.01	yes 0.04	yes 0.38	non english 0.03		0.34	tenure track 0.02	tenured 0.01
Step 2	beauty	gender	age	formal		native		students	tenure	tenure
	0.00	male 0.00	0.01	yes 0.05		non english 0.02		0.44	tenure track 0.01	tenured 0.01
Step 3	beauty	gender	age	formal		native			tenure	tenure
	0.00	male 0.00	0.01	yes 0.06		non english 0.02			tenure track 0.01	tenured 0.01
Step 4	beauty	gender	age			native			tenure	tenure
	0.00	male 0.00	0.01			non english 0.06			tenure track 0.01	tenured 0.01
Step 5	beauty	gender	age						tenure	tenure
	0.00	male 0.00	0.01						tenure track 0.01	tenured 0.01

Backward-elimination: p – value approach

Step	Variables included & p-value									
Full	beauty	gender male	age	formal yes	lower yes	native non english	minority yes	students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.29	0.06	0.35	0.30	0.02	0.02
Step 1	beauty	gender male	age	formal yes	lower yes	native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.04	0.38	0.03		0.34	0.02	0.01
Step 2	beauty	gender male	age	formal yes		native non english		students	tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.05		0.02		0.44	0.01	0.01
Step 3	beauty	gender male	age	formal yes		native non english			tenure tenure track	tenure tenured
	0.00	0.00	0.01	0.06		0.02			0.01	0.01
Step 4	beauty	gender male	age			native non english			tenure tenure track	tenure tenured
	0.00	0.00	0.01			0.06			0.01	0.01
Step 5	beauty	gender male	age						tenure tenure track	tenure tenured
	0.00	0.00	0.01						0.01	0.01

Best model: beauty + gender + age + tenure

Forward-selection

1. R^2_{adj} approach:
 - Start with regressions of response vs. each explanatory variable
 - Pick the model with the highest R^2_{adj}
 - Add the remaining variables one at a time to the existing model, and once again pick the model with the highest R^2_{adj}
 - Repeat until the addition of any of the remaining variables does not result in a higher R^2_{adj}
2. p – value approach:
 - Start with regressions of response vs. each explanatory variable
 - Pick the variable with the lowest significant p-value
 - Add the remaining variables one at a time to the existing model, and pick the variable with the lowest significant p-value
 - Repeat until any of the remaining variables does not have a significant p-value

Note: There is no guarantee both forward and backward selection will result in the same model.

Selected model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.6284	0.1673	27.66	0.00
beauty	0.1055	0.0328	3.21	0.00
gender.male	0.2081	0.0519	4.01	0.00
age	-0.0088	0.0032	-2.75	0.01
formal.yes	0.1324	0.0714	1.85	0.06
native:non english	-0.2430	0.1080	-2.25	0.02
tenure:tenure track	-0.2068	0.0839	-2.46	0.01
tenure:tenured	-0.1760	0.0641	-2.74	0.01