it's all **politics**

Poll: Support For Obama Among Young Americans Eroding

December 4, 2013 · 12:53 PM ET

ADAM WOLLNER

President Obama speaks at a town hall meeting at Binghamton University in Vestal, N.Y., in August.

Mike Groll/AP

After voting for him in large numbers in 2008 and 2012, young Americans are souring on President Obama.

According to a new Harvard University Institute of Politics poll, just 41 percent of millennials — adults ages 18-29 — approve of Obama's job performance, his lowest-ever standing among the group and an 11-point drop from April.

Obama's signature health care law is also unpopular among millennials. Fifty-seven percent of those surveyed said they disapprove of Obamacare, compared with 38 percent who said they approve.

A majority of respondents also said they disapprove of the way Obama is handling the economy, Syria, Iran and the budget deficit.

The results reflect a similar downward trend among the public at large. Recent polls ranging from Gallup to CNN show Obama's approval rating hovering around 40 percent, while disapproval of the health care law is in the mid-to-high 50s.

"Millennials are starting to look a lot more like their older brothers and sisters, parents and grandparents," IOP polling director John Della Volpe said in a conference call with reporters Wednesday.

The online survey of 2,089 adults was conducted from Oct. 30 to Nov. 11, just weeks after the federal government shutdown ended and the problems surrounding the implementation of the Affordable Care Act began to take center stage. The poll's margin of error was plus or minus 2.1 percentage points.

Fifty-five percent of the survey's respondents said they voted for Obama in the last presidential election, while 33 percent said they voted for Republican Mitt Romney. If the election were held again, Obama would still come out on top, but by a tighter 46 to 35 percent vote; 13 percent said they would vote for someone else.

According to the Pew Research Center, 66 percent of 18- to 29-year-olds voted for Obama in 2008, and 60 percent voted for his re-election in 2012.

Harvard's poll found millennials, like the rest of the public, aren't happy with Congress either. Just 19 percent of respondents said they approve of congressional Republicans, while 35 percent approve of their Democratic counterparts. Both figures are single-digit drops from April. Forty-five percent also said they would "recall and replace" their member of Congress if they had the option.

millennials president obama

1 Polling

(Preface: The use of this example should not be interpreted as an endorsement of President Obama). The previous two pages are an NPR article reporting the results of a Harvard University Institute of Politics poll conducted in between October 30 and November 11, 2013 of millennials' (adults aged 18-29) approval of President Obama's job performance. **Please start writing all your responses where indicated below.**

a) Who is the study population?

b) In this study, what does the value of true population proportion p correspond to?

c) Say we had access to infinite resources, what would be the best way to measure p?

d) What condition must be met for the sample of n young Americans to be representative of the population?

e) In this example, what is the sample proportion $\widehat{p}$, the point estimate of the population proportion p? Give the numerical value.

f) Recall that the standard error of $\widehat{p}$ based on samples of size n is

$$\text{SE}_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

and that when we don't know p as is typically the case, we use the *plug-in* estimate $\widehat{p}$ of p, resulting in

$$\text{SE}_{\widehat{p}} \approx \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Compute the standard error for this poll.

g) What is the margin of error for this poll? Show all your work.

h) Construct a 95% confidence interval for the proportion of all young Americans who approved of Obama's job performance. Assume all conditions for inference are met.

Please start writing all your responses here:

2 Teaching evaluations

Recall the teaching evaluation data from class. Let's look at a random sample of 5 out of the 463 rows of this dataset:

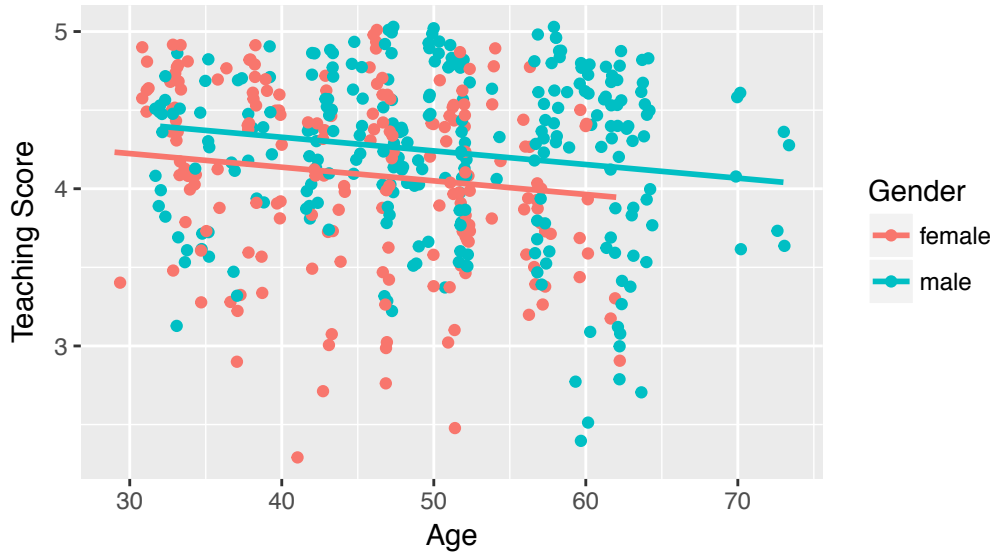
	score	gender	age
341	4.9	male	43
108	5.0	female	46
187	4.3	female	47
206	4.1	male	62
176	4.7	male	39

We are interested in modeling the outcome variable y = teacher evaluation score as a function of two explanatory variables:

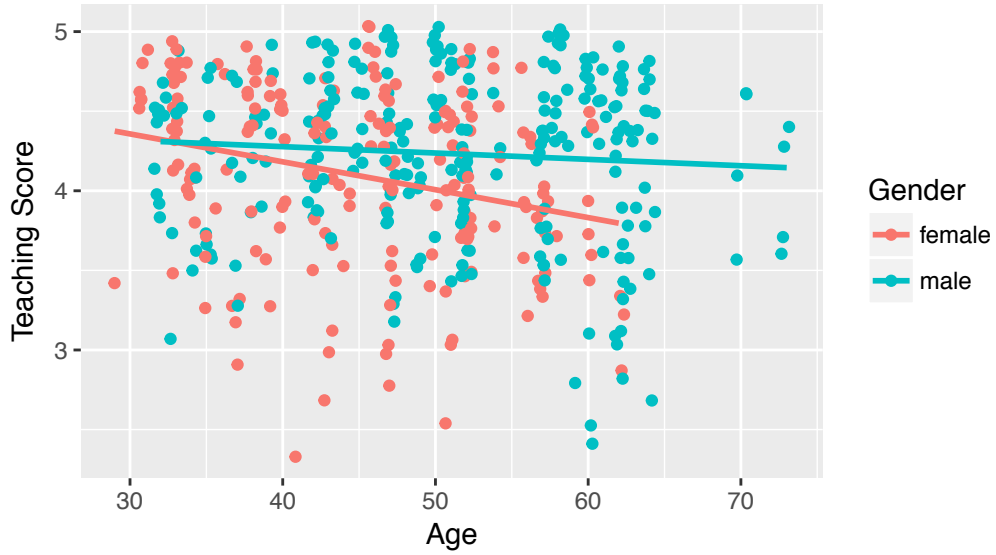
1. x_1: numerical explanatory/predictor variable of age
2. x_2: categorical explanatory/predictor variable of gender

You decide two fit regression models, one without an interaction term and the other with an interaction term, and compare the results.

a) Does this plot represent a visualization of the regression model with or without the interaction term? Circle either "with" or "without."



b) Does this plot represent a visualization of the regression model with or without the interaction term? Circle either "with" or "without."



c) You fit the regression model which includes the interaction term and obtain the following regression table output:

term	estimate	std.error	statistic	p.value
(Intercept)	4.88	0.21	23.8	0.00
age	-0.02	0.00	-3.9	0.00
gendermale	-0.45	0.27	-1.7	0.09
age:gendermale	0.01	0.01	2.5	0.01

Rewrite the equation for the line $\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$, but specifically for this data. In other words, put appropriate subscripts on the b's and write down what each of the three's x and the $\widehat{y}$ correspond to.

d) Interpret the 4 resulting fitted coefficients of the line (the intercept b_0 and the three slopes b_1, b_1, and b_3) not in abstract mathematical terms, but in the context of our outcome variable: teaching score.

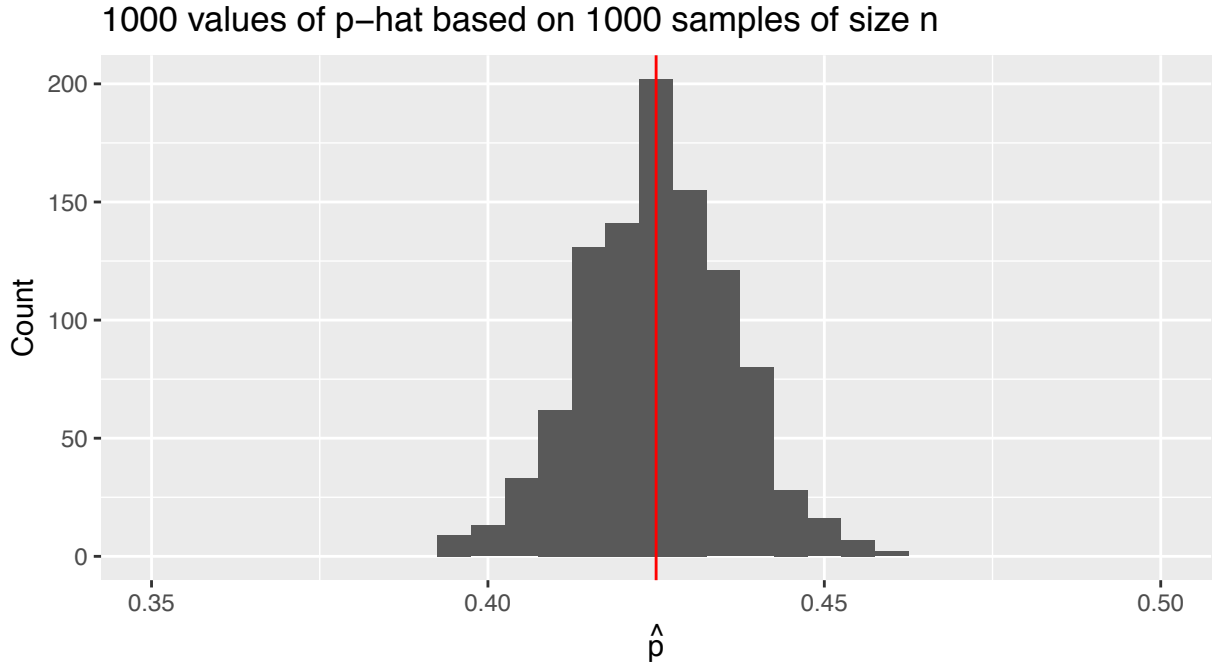
e) If a faculty member at UT Austin is female and aged 50, what do you predict their teaching score to be? Do not perform any arithmetic, but write down what you would enter into your calculator if you had one.

2 Sampling

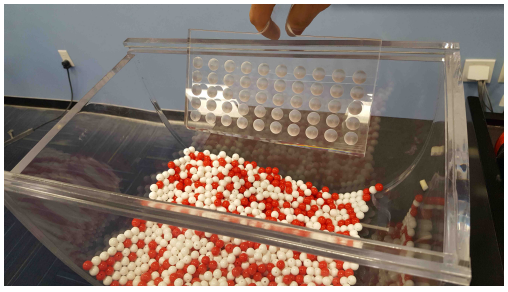
Consider the following hypothetical extension of the previous question on the poll of Obama's approval ratings among young Americans in 2013. First, say that you are an all knowing "higher power" and you know that $p = 0.425$. Furthermore, the following 1000 polling companies conduct polls on the same dates as the Harvard poll, using the same sampling methodology and the same sample size n. You observe the following results:

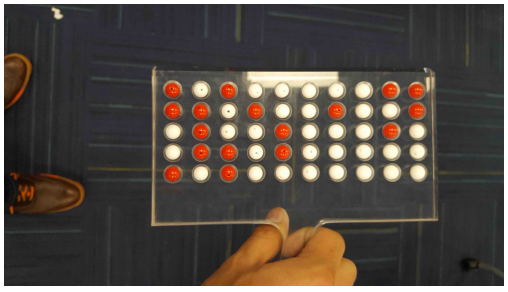
- Poll 1: Gallup finds that out of their sample of size n, 844 young Americans approve of Obama.
- Poll 2: Ipsos finds that out of their sample of size n, 857 young Americans approve of Obama.
- ...
- Poll 1000: Monmouth University finds that out of their sample of size n, 871 young Americans approve of Obama.

Based on these 1000 polls based on samples of size n, you compute 1000 values of $\widehat{p}$. You plot these in a histogram:



Hint: Think back to:





Please start writing all your responses where indicated below.

a) The above normal-shaped distribution is called the X distribution of the sample proportion $\widehat{p}$. It describes how different values of $\widehat{p}$ vary from sample to sample due to Y. What are X and Y?

b) What is the numerical value of the red line on the x-axis i.e. the center of this distribution?

c) The standard deviation of the above distribution of 1000 values of $\widehat{p}$ is called what?

d) Between what two values will 95% of sample proportions based on samples of size n occur? Show your work.

e) Mark these two values on the plot on the previous page.

f) What would happen to this histogram if the sample size were $n = 5000$?

g) What does this change in sample size to $n = 5000$ mean in terms of the quality of the polling results?

h) Bringing things back to real life and focusing on what we would do in practice, we would *not* take 1000 different samples of size $n = 2089$ and compute 1000 different values of $\widehat{p}$, but instead take only a single sample of $n = 2089$ and compute a single value of $\widehat{p}$. What was the point of the above exercise then? Respond in two sentences or less.

i) Say instead of estimating population proportions p with sample proportions, we want to estimate population means μ with sample means $\overline{x}$. The Central Limit Theorem guarantees that as we consider samples of size n that get larger and larger, the distribution of different values of $\overline{x}$ based on different samples of size n behaves more and more X and gets Y. What are X and Y?

j) Which is the correct form of the standard error of $\overline{x}$: A) $\sqrt{s^2}$ or B) $\sqrt{\frac{s^2}{n}}$? How do you know this?

Please start writing all your responses here:

4 Pennies

Recall the "sack" of $N = 800$ pennies in the dataframe `pennies` from which we virtually sampled $n = 50$ pennies from in Problem Set 10.

```
library(ggplot2)
library(dplyr)
library(moderndive)

mean(pennies$year)

## [1] 1989.848

sd(pennies$year)

## [1] 12.43956

ggplot(pennies, aes(x = year)) +
  geom_histogram(binwidth = 5) +
  labs(x = "year", title = "Fig 1: Population distribution of year of 800 pennies") +
  geom_vline(xintercept = mean(pennies$year), col = "red", size = 1)
```

Fig 1: Population distribution of year of 800 pennies

START WRITING YOUR RESPONSES WHERE INDICATED BELOW.

a) What are the (study) population, the name of the population parameter, the numerical value of the population parameter, the name of the point estimate, the formula for the true standard error, and the numerical value of the true standard error.

b) Say you (virtually) sample $n = 50$ pennies from the above population, compute the sample mean, and replace the pennies. Then everyone on your floor does the same. In what range of values do you expect 95%

of the resulting sample means to lie?

c) Say at the end of the day, you have 679 such sample means and you plotted a histogram of it. What is the name (not shape) of this distribution?

d) Why is the distribution of sample means in c) normal even though the population distribution of the $N = 800$ pennies is left-skewed as seen above?

e) In practice we would never perform such a simulation; rather we would just take a single sample of size n. What is the point of this simulation then?

f) Say all your floormates are germophobes and instead of randomly sampling pennies, they selectively only select the shinier and cleaner pennies. What impact will this have on the distribution in part c)?

g) Say instead the population of interest is not the sack of 800 pennies, but all pennies in circulation in the US. List the steps needed to construct a 99% confidence interval for the population parameter of interest based on $n = 50$.

START WRITING YOUR RESPONSES HERE: