Lecture 23 (Ch. 9)

In Ch. 8, we tested k specific proportions in 1 pop, and in r pops

1 pop	r pops
$H_0: \pi_1 = \pi_{01}, \pi_2 = \pi_{02}, \ldots, \pi_k = \pi_{0k}$	H_0: pops are homog. w.r.t. categ.
H_1: At least 1 of these is wrong.	H_1: --- not homog. -- --
chi-sqd dist. with $df = k-1$	chi-sqd dist. with $df = (k-1)(r-1)$
1 categorical/discrete var.	2 categorical/discrete vars.

Now, how about k population means?

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ (Not $\mu_1 = \mu_{01}, \mu_2 = \mu_{02}, \ldots$)

H_1: At least 2 μ's are different.

The dist. turns out to be the F-dist. with $(df_{num.}, df_{denom.})$.

The method is called 1-way (or single factor) ANOVA.
It deals with 1 continuous variable, y, whose mean is computed in k different levels of 1 categorical variable, x. ← "factor"

Example: Does knowledge of religion depend on religion?

THE PEW FORUM ON RELIGION & PUBLIC LIFE — A project of the PewResearchCenter

TOPICS

ISSUES
Abortion
Church-State Law
Death Penalty
Education
Gay Marriage & Homosexuality
Government
Politics & Elections
Science & Bioethics
Social Welfare

BELIEFS & PRACTICES
Belief in God
Frequency of Prayer
Importance of Religion
Religious Attendance
Other Beliefs & Practices

RELIGIOUS AFFILIATION
Christian
Jewish
Muslim
Other Affiliations
Unaffiliated

DEMOGRAPHICS
Age
Education & Income
Gender
Geography
Race
Other Demographics

REGIONS
Americas
Asia & the Pacific
Europe
Middle East & North Africa
Sub-Saharan Africa

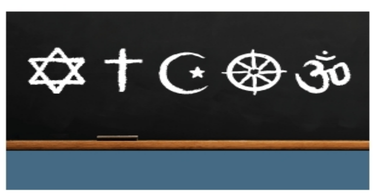
PUBLICATIONS
Analyses
Event Transcripts
Graphics

pewforum.org › Topics › Beliefs & Practices

U.S. Religious Knowledge Survey

POLL - September 28, 2010

Executive Summary



Print Email Share AAA Text Size

Atheists and agnostics, Jews and Mormons are among the highest-scoring groups on a new survey of religious knowledge, outperforming evangelical Protestants, mainline Protestants and Catholics on questions about the core teachings, history and leading figures of major world religions.

On average, Americans correctly answer 16 of the 32 religious knowledge questions on the survey by the Pew Research Center's Forum on Religion & Public Life. Atheists and agnostics average 20.9 correct answers. Jews and Mormons do about as well, averaging 20.5 and 20.3 correct answers, respectively. Protestants as a whole average 16 correct answers; Catholics as a whole, 14.7. Atheists and agnostics, Jews and Mormons perform better than other groups on the survey even after controlling for differing levels of education.

Atheists and Agnostics, Mormons and Jews Score Best on Religious Knowledge Survey

Average # of questions answered correctly out of 32

Total	**16.0**
Atheist/Agnostic	20.9
Jewish	20.5
Mormon	20.3
White evangelical Protestant	17.6
White Catholic	16.0
White mainline Protestant	15.8
Nothing in particular	15.2
Black Protestant	13.4
Hispanic Catholic	11.6

test scores (out of 32)

In This Report

Looks like Athiests know most!?

Even though we want to compare k means, it's not enough to look at sample means only.

SCIENTIFIC AMERICAN™

Winner of the 2011 National Magazine Award for General Excellence

Search ScientificAmerican.com

News & Features | Blogs | Multimedia | Education | Citizen Science | Topics

Home » Scientific American Magazine » December 2010

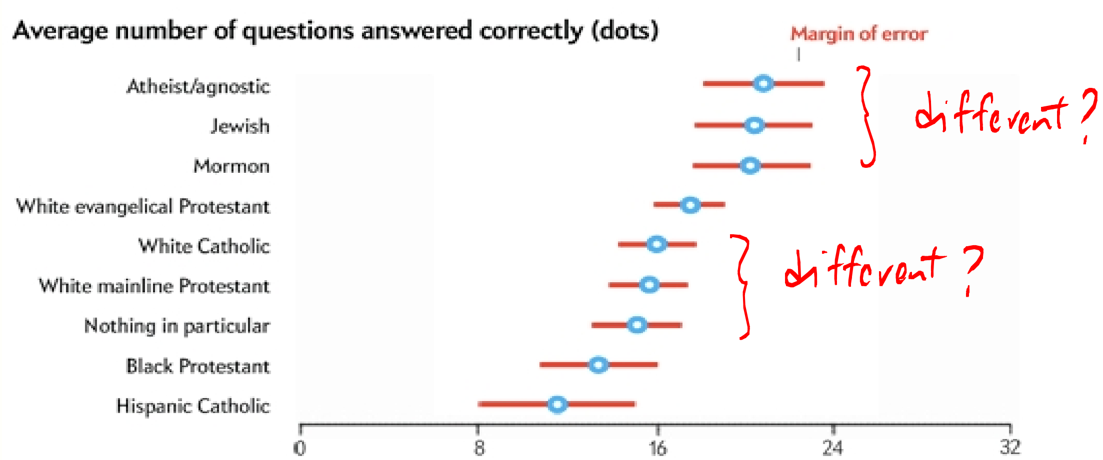
Advances | More Science

The Science of "Disestimation": The Shortcomings of Opinion Polls

Why we shouldn't put our faith in opinion polls

By Charles Seife | December 14, 2010 | 19

Share Email Print



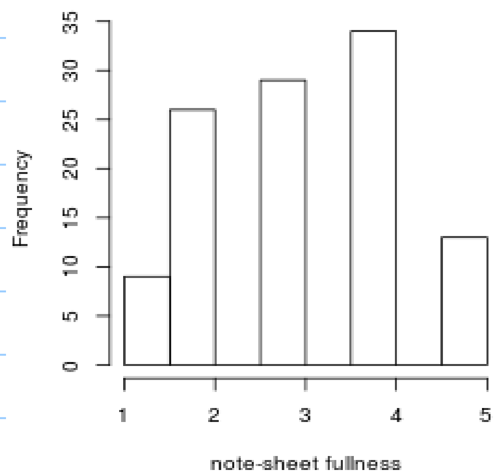
Moral: When testing means, std dev. is what's important

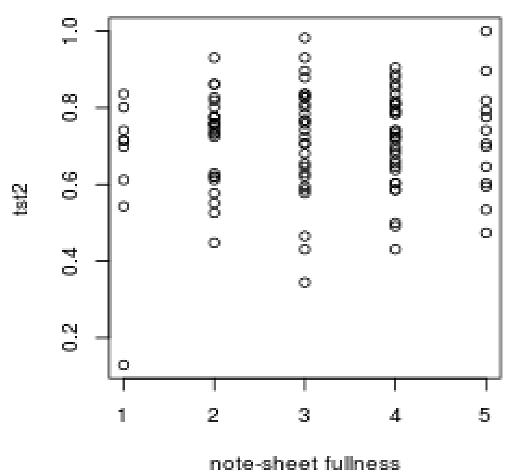
Also note That This is just a generalization of The 2-sample/pop test (for comparing μ_1, μ_2) to The case of k populations.

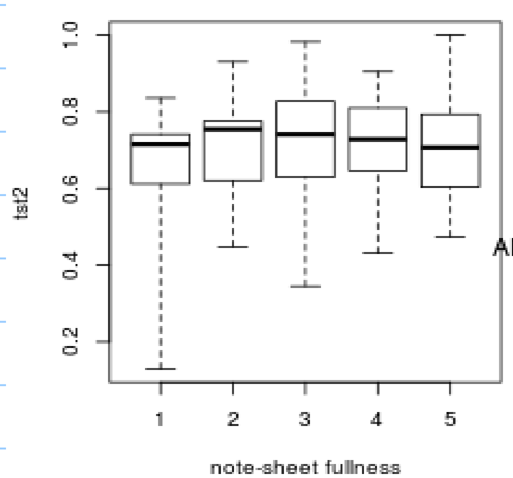
Does fullnes of note sheet have an effect on test2 scores?

	note-sheet fullness	mean test score
not-so-full	1	0.6437
	2	0.7205
	3	0.7179
	4	0.7201
very full	5	0.7142

Looking at means is not enough. Must also look at variance.







ANOVA F-test: F=0.5078 p-value= 0.7301

Example 9.1 (p.422-423)

Vibration (y) for 5 brands (x) of bearing:
(or "speed" for 5 brands (x) of computers.)

Data:

Brand 1	2	3	4	5
13.1	⋮	⋮	⋮	⋮
15.0				
14.0				
⋮				
11.6				
$\bar{Y}_1 = 13.68$	$\bar{Y}_2 = 15.97$	13.67	14.73	13.08
$S_1 = 1.194$	$S_2 = 1.167$			

We want to know if the data provide evidence that the 5 bearings have different means (of vibration), μ_i.

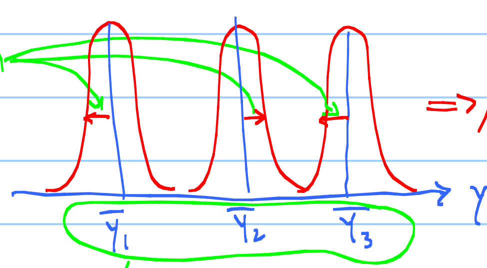
(ie. Are the 5 computers different in their speed?)

H_0: $\mu_1 = \mu_2 = \cdots = \mu_5$ ← Note we are talking about 5 populations,

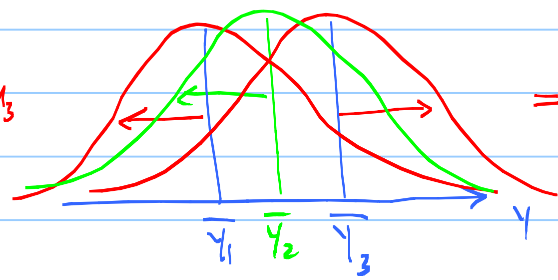
H_1: At least 2 μ's are diff.

The way ANOVA answers that question is by finding out how much of the total variation in y is "within" each category, and how much is "between" the categories.

variation within



$\Rightarrow \mu_1 \neq \mu_2 \neq \mu_3$



$\Rightarrow$ cannot Tell.

↳ variation between

⇒ See math, next page.

Recall the decomposition of SST from regression. Similarly,

$n = \sum_{i=1}^{k} n_i$

variability in the y_{ij} — k populations. — sample size in i^{th} pop.

$$S_{YY} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$$

Grand mean $= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$

Show! $= \sum_{i=1}^{k} \left(\frac{n_i}{n}\right) \bar{Y}_i$

j^{th} response in the i^{th} pop.

⋮

sample mean in i^{th} pop. — sample var. in i^{th} pop.

$$S_{YY} = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{\bar{y}})^2 + \sum_{i=1}^{k} \left[\overbrace{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}^{=(n_i-1)s_i^2}\right]$$

$$\begin{array}{ccccc} \| & & \| & & \| \\ SST & = & SS_{\text{between group}} & + & SS_{\text{within group}} \\ & & \| & & \| \\ & & SSTr \leftarrow \text{Treatment} & & SSE \\ & & \| & & \| \\ & & SS_{\text{explained}} & & SS_{\text{unexplained}} \end{array}$$

SS : Total = between + within

df : $n-1 = (k-1) + (n-k)$

k = # of levels in 1 factor (predictor) = # of pops.

FYI [Note: linear regression: $n-1 = k + [n-(k+1)]$

k = # of β's

different k

Q1: If all the sample means are equal, then SS_{within} is equal to

A) 0 B) $SS_{between}$ $\boxed{\text{C) } S_{YY}}$ D) None of the above.

$\bar{Y}_1 = \bar{Y}_2 = \cdots = \bar{\bar{y}} \Rightarrow SS_{between} = 0 \Rightarrow SS_{within} = \sum_i \sum_j (Y_{ij} - \bar{\bar{y}})^2 = S_{YY}$

Now, we can compare $SS_{between}$ and SS_{within}:

Note: If all the sample means are equal, then $F = 0$.

Theorem:

If H_0 = True, $F = \dfrac{SS_{between}/(k-1)}{SS_{within}/(n-k)} = \dfrac{MS_{between}}{MS_{within}}$

has an F-distribution with params. $df = (k-1, n-k)$

All we need is Table $\underline{VIII}$ to give us areas (p-values).

One assumption of this Theorem is that the y's in each of the k populations are normal, and that they all have the same variance, i.e. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$.

Use qq plots to test these assumptions. This assumption is called "homoscedasticity"!

$\boxed{\text{In lab}}$

Example 9.1 (p. 422-423)

H_0: $\mu_1 = \mu_2 = \cdots = \mu_5$

H_1: At least 2 μ's are diff.

y = Response = vibration

x = factor = brand type.

Data:

Brand 1	2	3	4	5
13.1	,	,	,	,
15.0	,	,	,	,
14.0	,	,	,	,
⋮	,	,	,	,
11.6	,	,	,	,
$n_1 = 6$	$n_2 = 6$			$n_5 = 6$
$\bar{Y}_1 = 13.68$	$\bar{Y}_2 = 15.97$	...		
$S_1 = 1.194$	$S_2 = 1.167$			

$$\bar{\bar{Y}} = \sum_{i=1}^{5} \left(\frac{n_i}{n}\right) \bar{Y}_i = \left(\frac{6}{30}\right)(13.68) + \cdots = 14.22$$

$$SS_{between} = \sum_{i=1}^{5} n_i \left(\bar{Y}_i - \bar{\bar{Y}}\right)^2 = 6\left(13.68 - 14.22\right)^2 + \cdots = 30.88$$

$$SS_{within} \quad \sum_{i=1}^{5} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots$$

$$= (6-1)(1.194)^2 + \cdots \approx 22.83$$

$$F_{obs} = \frac{30.88/(5-1)}{22.83/(30-5)} = 8.45 \qquad df = (5-1,\ 30-5)$$

p-value = (Table VIII) < .001

Conclusion:

∴ Reject H_0 ($\mu_1 = \mu_2 = \cdots$) in favor of H_1 (At least 2 μ's are different.) at $\alpha = .01$.

In "English":

∴ Sufficent evidence to reject "equal means" in favor of "at least 2 μ's are different." (which 2?)

∴ Bearing Type "has" an effect on vibration.
(At least one of the computers is faster than the others)

Section 9.3

We are skipping this, → Tukey's Test

This is the F-test that I showed you when testing whether the fullness of cheat sheet has an effect on test2 scores.

Most software produce an ANOVA Table for keeping Track of all The relevant numbers, similar to regression. The structure is :

Mean

Source	df	SS	MS	F_{obs}	p-value
Between Group (factor)	$k-1$	$SS_{between}$			table VIII
Within Group (error)	$n-k$	SS_{within}			
Total	$n-1$	SS_{Total}			

from formula

In Lab. you will produce The ANOVA Table for The above example. You will find:

Response = vibration

Factor = type of bearing (5 levels)

Source	df	SS	MS	F	P-value
factor	$5-1$	30.85	7.71	8.44	.00018
Error	$30-5$	22.84	0.91		
Total	$30-1$	53.7			

Summary:

σ unknown $\qquad$ σ known

z, t	$H_0: \mu = \mu_0$	$H_1: \mu \,\square\, \mu_0$	small / large sample
z	$H_0: \pi = \pi_0$	$H_1: \pi \,\square\, \pi_0$	large sample
z, t	$H_0: \mu_1 - \mu_2 = \Delta_0$	$H_1: \mu_1 - \mu_2 \,\square\, \Delta_0$	indep. or paired

Chisqd $H_0: \pi_1 = \pi_{01}, \pi_2 = \pi_{02}, \cdots \pi_k = \pi_{0k}$, H_1: Atleast 1 is wrong

chisqd H_0: homogeniety of r pops w.r.t. k categ. H_1: not.

Equivalently: H_0: 2 categ. variables are independent H_1: not.

F $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ $\quad H_1$: atleast 2 μ's are diff.

Note that the ANOVA F-test is a generalization of the 2-sample t-test to more than 2 populations.

hw_lect23_1

The following data refer to the melting temperature, y (in some unit), of a certain material at four different pressures, x (in some unit).

Pressure	Temperature
1.6	59.5, 53.3, 56.8, 63.1, 58.7
3.8	55.2, 59.1, 52.8, 54.4
6.0	51.7, 48.4, 53.9, 49.0
10.2	44.6, 48.5, 41.0, 47.3, 46.1

a) Make a comparative boxplot of y for the four pressure levels.

b) Based on the above boxplot, would you say there is a difference in the mean melting temperature for at least 2 of the pressure levels?

c) At alpha = 0.05, is there evidence that the mean melting temperature at the at least 2 of the four pressure levels are different? Report the p-value, and state the conclusion clearly.

d) Write code to compute the above p-value "by hand," i.e. without using aov() or lm(), but using the basic formulas for SS_between, SS_within, etc.

e) After (or before) a 1-way ANOVA test, one should check the two assumptions that the y's are normally distributed within each group, and with the same variance. To that end, make a plot that shows four qqplots (one for each pressure level) superimposed onto a single figure; make sure that the four qqplots have different colors. Are the 4 qqplots reasonably straight, and do they have approximately equal slopes? Hint: in the first call to plot(), use xlim=c(-2,2) and ylim=range(y).

hw_lect23_2 By R Optional

Do 1-way ANOVA on one of the 2 continous vars, and 1 of the categorical vars. in your data from hw_lect1. If you cannot, explain why not!