Overview

X is…	Y is…	Use…
Discrete	Discrete	
Discrete	Continuous	
Continuous	Continuous	

Overview

X is…	Y is…	Use…
Discrete	Discrete	χ^2 and then Q or RR or OR
Discrete	Continuous	ANOVA
Continuous	Continuous	Regression , Correlation

Example #1

Do the populations of people who live in cities, in the suburbs, and in rural areas own different numbers of cars? (use α=0.05)

To find out, I randomly sampled 5 urban residents, 5 suburban residents, and 5 rural residents; I asked each person how many cars they own.

(*Note*: Treat "number of cars" as continuous)

Example #1

For urban residents:

0 1 1 0 0 (Avg = 0.4; s = 0.548)

For suburban residents:

2 1 1 2 1 (Avg = 1.4 ;s = 0.548)

For rural residents:

2 2 3 1 2 (Avg = 2.0; s = 0.707)

For the full sample: = 1.267

Example #1

Hypotheses:

H_0: $\mu_{urban} = \mu_{suburban} = \mu_{rural}$

H_a: Not all μ_{urban}, $\mu_{suburban}$, μ_{rural} are equal

Assumptions:

Temporarily waved for this example…

Confidence Level: α=0.05

Critical Value:

Critical value for $F_{2,12}$ when α=0.05 is 3.89

(Reject H_0 if F > 3.89)

Example #1

$$F = \frac{SS_{Between}/(J-1)}{SS_{Within}/(n-J)} = \frac{MS_{Between}}{MS_{Within}}$$

$$MS_{Between} = \frac{n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2}{J-1}$$

$$MS_{Between} = \frac{5(0.4-1.267)^2 + 5(1.4-1.267)^2 + 5(2.0-1.267)^2}{3-1}$$

$$MS_{Between} = \frac{6.533}{2} = 3.267$$

Example #1

$$F = \frac{SS_{Between}/(J-1)}{SS_{Within}/(n-J)} = \frac{MS_{Between}}{MS_{Within}}$$

$$SS_{Within} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_j)^2$$

$= (0-0.4)^2 + (1-0.4)^2 + (1-0.4)^2 + (0-0.4)^2 + (0-0.4)^2 +$
$(2-1.4)^2 + (1-1.4)^2 + (1-1.4)^2 + (2-1.4)^2 + (1-1.4)^2 +$
$(2-2)^2 + (2-2)^2 + (3-2)^2 + (1-2)^2 + (2-2)^2 =$ **4.40**

MS_{Within} = 4.40/12 = 0.367

$$F = \frac{SS_{Between}/(J-1)}{SS_{Within}/(n-J)} = \frac{MS_{Between}}{MS_{Within}} = \frac{3.267}{0.367} = 8.902$$

Reject H_0

Example #2

Political Party

Favor or Oppose CAPITAL PUNISHMENT for Murder?		Democrat	Republican	*Total*
	Favor	13,828	12,328	26,156
	Oppose	7,118	2,676	9,794
	Total	20,946	15,004	35,950

Source: 1972-2008 General Social Survey.
Note that "independents" and members of "other parties" are omitted.

1. Is there a statistically significant association between these two variables in the population?
2. Use relative risk and odds ratio to describe the strength and direction of the association

Example #2

Political Party

Favor or Oppose CAPITAL PUNISHMENT for Murder?		Democrat	Republican	*Total*
	Favor	13,828 E.V.=15,240	12,328 E.V.=10,916	26,156
	Oppose	7,118 E.V.=5,706	2,676 E.V.=4,088	9,794
	Total	20,946	15,004	35,950

Source: 1972-2008 General Social Survey.
Note that "independents" and members of "other parties" are omitted.

$$\chi^2 = \left(\frac{(13{,}828 - 15{,}240)^2}{15{,}240}\right) + \left(\frac{(12{,}328 - 10{,}916)^2}{10{,}916}\right)$$
$$+ \left(\frac{(7{,}118 - 5{,}706)^2}{5{,}706}\right) + \left(\frac{(2{,}676 - 4{,}088)^2}{4{,}088}\right) = 1{,}149.95$$

Example #2

		Political Party		
		Democrat	Republican	*Total*
Favor or Oppose CAPITAL PUNISHMENT for Murder?	Favor	13,828 **a**	12,328 **b**	26,156
	Oppose	7,118 **c**	2,676 **d**	9,794
	Total	20,946	15,004	35,950

Source: 1972-2008 General Social Survey.
Note that "independents" and members of "other parties" are omitted.

$$OR = \frac{b/d}{a/c} = \frac{12{,}328/2{,}676}{13{,}828/7{,}118} = 2.37$$

$$RR = \frac{b/(b+d)}{a/(a+c)} = \frac{12{,}328/(12{,}328+2{,}676)}{13{,}828/(13{,}828+7{,}118)} = 1.24$$

Example #3

Find the correlation between the number of murders in Minneapolis and St Paul between 2006 and 2010

	2006	2007	2008	2009	2010	Mean	SD
Minn.	60	47	41	20	40	41.6	14.47
St Paul	17	14	18	15	17	16.2	1.64

Example #3

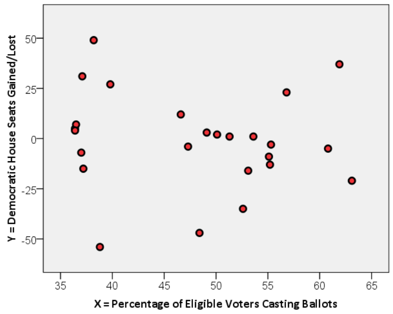
$$r_{yx} = \left(\frac{1}{N-1}\right)\sum_{i=1}^{N}\left(\frac{X_i-\bar{X}}{s_x}\right)\left(\frac{Y_i-\bar{Y}}{s_y}\right)$$

$$r_{yx} = \left(\frac{1}{5-1}\right)\left[\begin{array}{l}\left(\frac{60-41.6}{14.47}\right)\left(\frac{17-16.2}{1.64}\right)+\left(\frac{47-41.6}{14.47}\right)\left(\frac{14-16.2}{1.64}\right)+\\ \left(\frac{41-41.6}{14.47}\right)\left(\frac{18-16.2}{1.64}\right)+\left(\frac{20-41.6}{14.47}\right)\left(\frac{15-16.2}{1.64}\right)+\\ \left(\frac{40-41.6}{14.47}\right)\left(\frac{17-16.2}{1.64}\right)+\end{array}\right]$$

$$r_{yx} = (0.25)(0.6192-0.4997-0.0454+1.0904-0.0538)$$

$$r_{yx} = 0.2776$$

Example #4



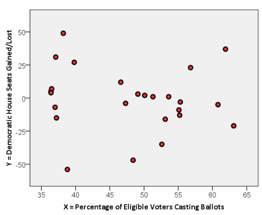
U.S. House Seats Gained or Lost (Y), by Voter Turnout (X), 1960-2008

	Mean	SD
X	47.908	8.947
Y	-1.080	24.005

r_{XY} = -0.115 n=25

1. Compute the least squares regression line relating Y to X
2. How much of the variation in Y is explained by X?

Example #4



U.S. House Seats Gained or Lost (Y), by Voter Turnout (X), 1960-2008

	Mean	SD
X	47.908	8.947
Y	-1.080	24.005

r_{XY} = -0.115 n=25

$$b_{YX} = r_{YX}\frac{s_Y}{s_X} = -0.115\frac{24.005}{8.947} = -0.31$$

$$a = \bar{Y} - b\bar{X} = -1.080 + (0.31)(47.908) = 13.77$$

$$R^2 = -0.115^2 = 0.013$$

Example #5

	Mean	SD	
X	2.5	1.0	N=200
Y	1.0	0.5	r_{YX}=0.20

1. Compute least squares regression line
2. Test hypothesis that $\rho^2=0$
3. Test hypothesis that $\rho=0$
4. Test hypothesis that $\beta=0$
5. Use α=0.05 throughout
6. Interpret!

Example #5

	Mean	SD	
X	2.5	1.0	N=200
Y	1.0	0.5	r_{YX}=0.20

$$b_{YX} = r_{YX}\frac{s_Y}{s_X} = 0.20\frac{0.5}{1.0} = 0.1$$

$$a = \overline{Y} - b\overline{X} = 1.0 - (0.1)(2.5) = 0.75$$

$$R^2 = 0.20^2 = 0.04$$

Example #5

	Mean	SD	
X	2.5	1.0	N=200
Y	1.0	0.5	r_{YX}=0.20

H_0: $\rho^2_{YX} = 0$; H_1: $\rho^2_{YX} > 0$; α=0.05 ; $F^*_{1,199}$=3.84

$$SS_{TOTAL} = (s_Y^2)(N-1) = (0.5^2)(199) = 49.75$$

$$SS_{REGRESSION} = (R^2_{YX})(SS_{TOTAL}) = (0.20^2)(49.75) = 1.99$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION} = 49.75 - 1.99 = 47.76$$

$$F_{1,N-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/N-2} = \frac{1.99/1}{47.76/198} = 8.25$$

Reject H_0

Example #5

	Mean	SD	
X	2.5	1.0	N=200
Y	1.0	0.5	r_{YX}=0.20

H_0: $\rho_{YX} = 0$; H_1: $\rho_{YX} \neq 0$; α=0.05 ; Z^*=1.96

$$Z_r = \left(\frac{1}{2}\right)\ln\left(\frac{1+r_{YX}}{1-r_{YX}}\right) = \left(\frac{1}{2}\right)\ln\left(\frac{1+0.20}{1-0.20}\right) = 0.202$$

$$Z = \frac{Z_r - 0}{\sqrt{1/N-3}} = \frac{0.202-0}{\sqrt{1/197}} = 2.84$$

Reject H_0

Example #5

	Mean	SD	
X	2.5	1.0	N=200
Y	1.0	0.5	r_{YX}=0.20

H_0: $\beta_{YX} = 0$; H_1: $\beta_{YX} \neq 0$; α=0.05 ; t^*_{198}=1.97

$$t_{N-2} = \frac{b_{YX} - 0}{s_b} = \frac{b_{YX} - 0}{\sqrt{\frac{MS_{ERROR}}{(s_X^2)(N-1)}}} = \frac{0.1}{\sqrt{\frac{47.76/198}{(1^2)(199)}}} = 2.87$$

Reject H_0