SOC 3811/5811: BASIC SOCIAL STATISTICS

Associations Between Continuous Variables

Please look at the graphic below.

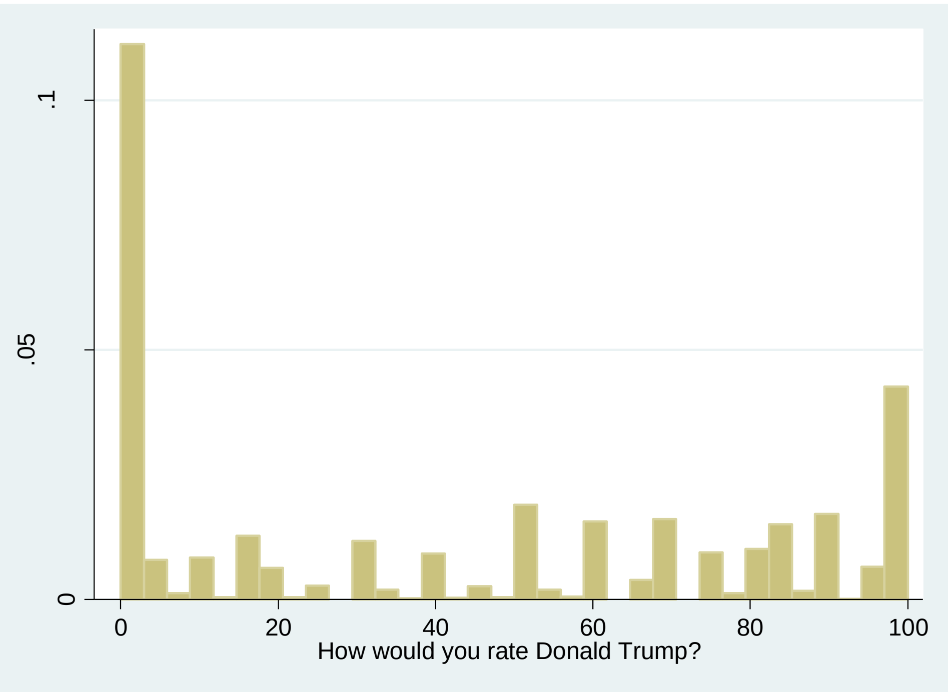
We'd like to get your feelings toward some people and groups who are in the news these days using something we call the feeling thermometer.

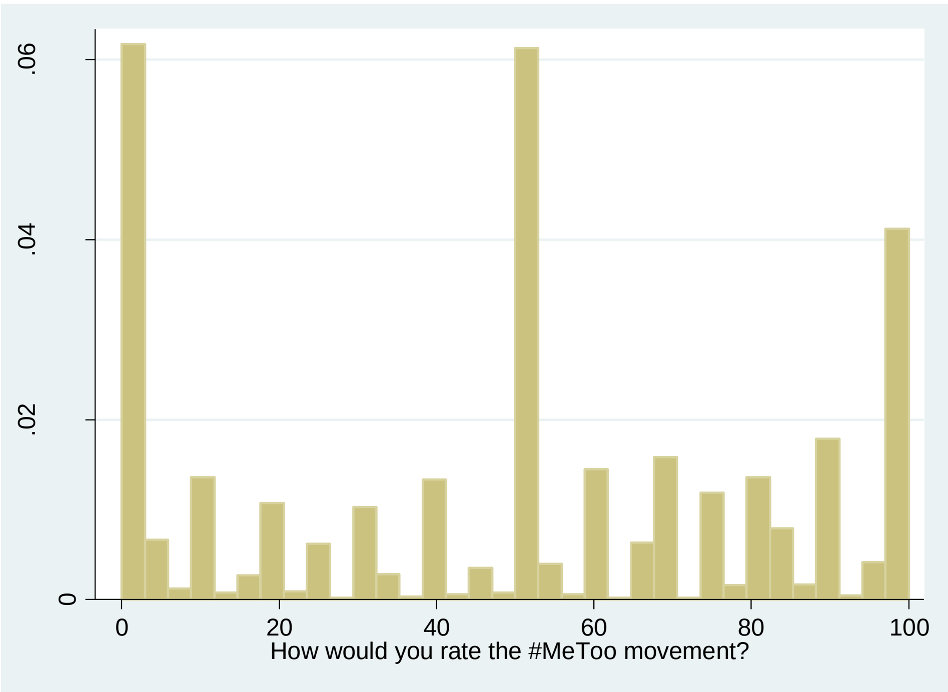
Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person or group. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable and warm toward the person or group. You would rate them at the 50 degree mark if you don't feel particularly warm or cold toward them.

If you come to a person or group that you don't recognize at all, you don't need to rate them. Just leave the box empty and go on to the next person or group.

“Feeling Thermometer” Data
2020 American National Election Survey (Pilot Test)

n = 2,999





$\overline{\mathrm{Y}} = 42.1 \quad s_{\mathrm{Y}} = 38.9$

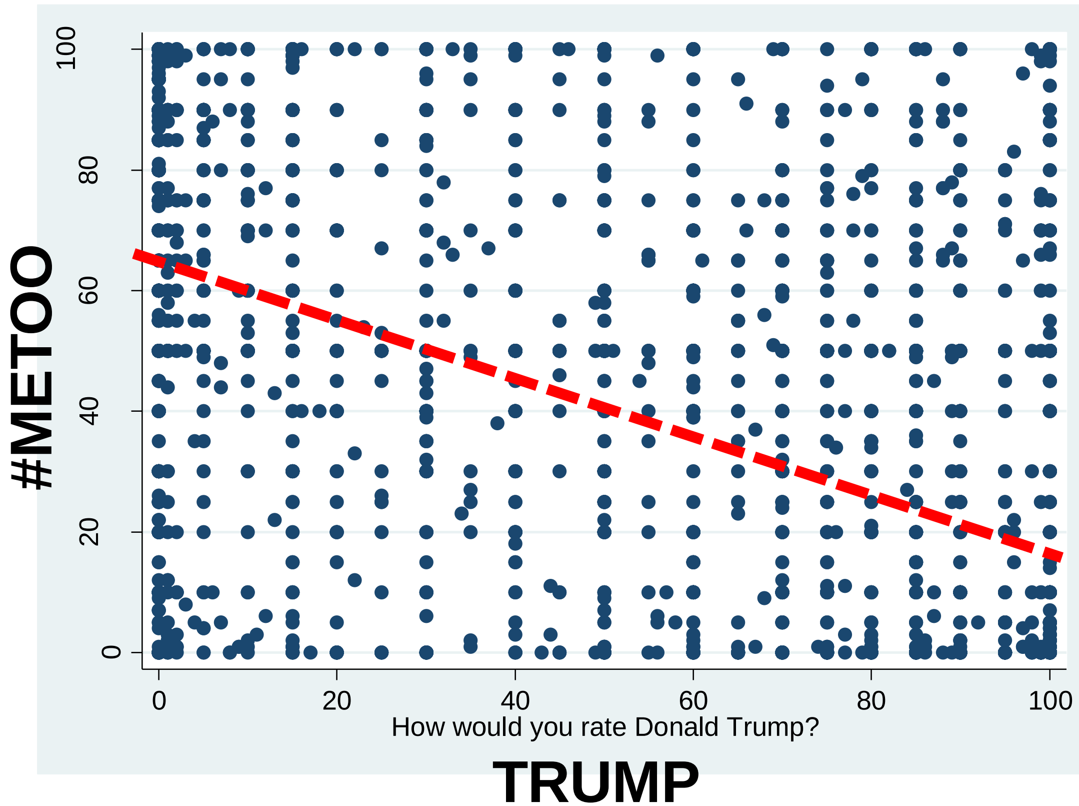
$\overline{\mathrm{X}} = 48.9 \quad s_{\mathrm{X}} = 34.3$

TRUMP

#METOO

“Feeling Thermometer” Data
2020 American National Election Survey (Pilot Test)

n = 2,999



$r_{yx} = -0.42$

$$b_{yx} = r_{yx}\frac{s_y}{s_x} = -0.42\,\frac{38.9}{34.3} = -\,0.48$$

$$a = \overline{Y} - b\overline{X} = 42.1 - (-0.48)(48.9) = 65.6$$

$R^2_{YX} = -0.42^2 = 0.18$

Inferences About Associations

Hypothesis Tests About ρ^2_{YX}

R^2_{YX} is a sample estimate of population parameter ρ^2_{YX}

If ρ^2_{YX} equals zero, then X does nothing to explain variability in Y

Hypothesis Tests About ρ_{YX}

r_{YX} is a sample estimate of population parameter ρ_{YX}

If ρ_{YX} equals zero, then there is no correlation between X and Y

Hypothesis Tests About Slope β_{YX}

b_{YX} is a sample estimate of population parameter β_{YX}

If β_{YX} equals zero, then the regression of Y on X has a zero slope

Inferences About Associations

Hypothesis Testing in 6 Steps

1. State the null (H_0) and alternative (H_1) hypotheses
2. Check that the sample data conform to basic assumptions; if they do not, then do not go any further
3. Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis
4. Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level
5. Calculate the test statistic … F or Z or t, depending
6. Compare the test statistic to the critical value

Inferences About ρ^2_{YX}

Inferences About ρ^2_{YX}

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\rho^2_{YX} = 0$

H_1: $\rho^2_{YX} > 0$

This is a one-sided test (with no $<$) because ρ^2_{YX} cannot possibly be less than zero

Failing to reject the null means failing to reject the hypothesis that X explains none of the variation in Y

Inferences About ρ^2_{YX}

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about ρ^2_{YX} to be valid

Inferences About ρ^2_{YX}

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's choose α=0.05

Inferences About ρ^2_{YX}

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

The hypothesis test for ρ^2_{YX} is (as described below) an F test with $df_{NUM}=1$ and $df_{DENOM}=n-2$

In our example, we want $F_{1,2997}$ for $\alpha=0.05$ which is 3.84

We will thus reject H_0 if our F statistic exceeds 3.84

Critical Values of F

(α=0.05)

DENOMINATOR Degrees of Freedom \ NUMERATOR Degrees of Freedom	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100	200	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01	250.10	251.14	251.77	253.04	253.68	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.47	19.48	19.49	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.58	8.55	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.70	5.66	5.65	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.44	4.41	4.39	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.75	3.71	3.69	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.32	3.27	3.25	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.02	2.97	2.95	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.80	2.76	2.73	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.64	2.59	2.56	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.57	2.53	2.51	2.46	2.43	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.47	2.43	2.40	2.35	2.32	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.38	2.34	2.31	2.26	2.23	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.31	2.27	2.24	2.19	2.16	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.18	2.12	2.10	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.19	2.15	2.12	2.07	2.04	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.15	2.10	2.08	2.02	1.99	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.11	2.06	2.04	1.98	1.95	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.07	2.03	2.00	1.94	1.91	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.97	1.91	1.88	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.18	2.10	2.01	1.96	1.94	1.88	1.84	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	1.98	1.94	1.91	1.85	1.82	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05	1.96	1.91	1.88	1.82	1.79	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.94	1.89	1.86	1.80	1.77	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.92	1.87	1.84	1.78	1.75	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.90	1.85	1.82	1.76	1.73	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.06	1.97	1.88	1.84	1.81	1.74	1.71	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.87	1.82	1.79	1.73	1.69	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.03	1.94	1.85	1.81	1.77	1.71	1.67	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.76	1.70	1.66	1.62
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.00	1.92	1.83	1.78	1.75	1.68	1.65	1.61
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	1.99	1.91	1.82	1.77	1.74	1.67	1.63	1.59
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	1.98	1.90	1.81	1.76	1.72	1.66	1.62	1.58
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	1.97	1.89	1.80	1.75	1.71	1.65	1.61	1.57
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.96	1.88	1.79	1.74	1.70	1.63	1.60	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.66	1.59	1.55	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.69	1.63	1.60	1.52	1.48	1.44
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96	1.80	1.71	1.61	1.55	1.52	1.44	1.39	1.34
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.77	1.68	1.57	1.52	1.48	1.39	1.34	1.28
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.52	1.46	1.41	1.32	1.26	1.19
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.35	1.24	1.17	1.00

Inferences About ρ^2_{YX}

Calculate the test statistic

$SS_{TOTAL} = (s_Y^2)(n-1) =$ (38.9²)(2999-1) = 4,536,603

$SS_{REGRESSION} = (R^2_{YX})(SS_{TOTAL}) =$ (0.18)(4,536,603) = 816,588

$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION} =$ 4,536,603 − 816,588 = 3,720,015

$$F_{1,n-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/n-2} = \frac{MS_{Regression}}{MS_{Error}} =$$

$$= \frac{816,588/1}{3,720,015/2,997} = 657$$

n = 2,999

$\bar{X} = 48.9$	$s_X = 34.3$
$\bar{Y} = 42.1$	$s_Y = 38.9$
$r_{yx} = -0.42$	$R^2_{YX} = 0.18$
$a = 65.6$	$b_{yx} = -0.48$

Inferences About ρ^2_{YX}

Calculate the test statistic

$$SS_{TOTAL} = (s_Y^2)(n-1) = (38.9^2)(2{,}999-1) = 4{,}536{,}603$$

$$SS_{REGRESSION} = (R_{YX}^2)(SS_{TOTAL}) = (0.18)(4{,}536{,}603) = 816{,}588$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION} = 3{,}720{,}015$$

$$F_{1,n-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/n-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{816{,}588/1}{3{,}720{,}015/2997} = 657$$

$\bar{X} = 48.9$	$s_X = 34.3$
$\bar{Y} = 42.1$	$s_Y = 38.9$
$r_{yx} = -0.42$	$R_{YX}^2 = 0.18$
$a = 65.6$	$b_{yx} = -0.48$

Inferences About ρ^2_{YX}

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do not reject H_0

We can restate the hypotheses:

H_0: $\rho^2_{YX} = 0$ ➔ Fail to reject H_0 if $F \leq 3.84$

H_1: $\rho^2_{YX} > 0$ ➔ Reject H_0 if $F > 3.84$

Since F=657, we reject H_0 … so it appears that in the population X and Y are associated, such that X accounts for some of the variability in Y

Inferences About Correlation (ρ_{YX})

Inferences About Correlation (ρ_{YX})

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\rho_{YX} = 0$

H_1: $\rho_{YX} \neq 0$

This is a two-sided test since ρ_{YX} can range from -1 to +1

Failing to reject the null means failing to reject the hypothesis that X and Y are uncorrelated in the population

Inferences About Correlation (ρ_{YX})

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about ρ_{YX} to be valid

Inferences About Correlation (ρ_{YX})

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's choose α=0.05

Inferences About Correlation (ρ_{YX})

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

Given α=0.05 for this two-sided test, the critical value Z* equals 1.96

Standard Normal Probabilities

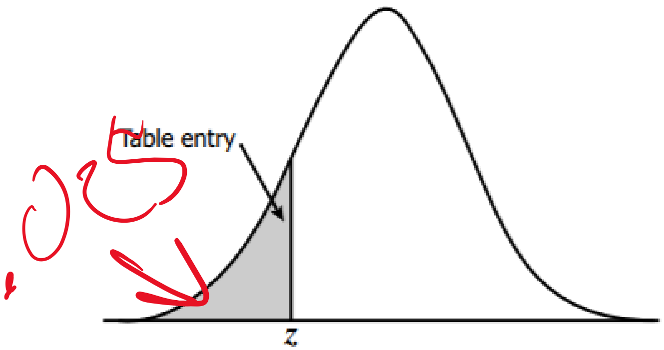


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Inferences About Correlation (ρ_{YX})

Calculate the test statistic

$$Z_r = \left(\frac{1}{2}\right) \ln\left(\frac{1 + r_{YX}}{1 - r_{YX}}\right) = \left(\frac{1}{2}\right) \ln\left(\frac{1 - .42}{1 + .42}\right) = -.447$$

$$Z = \frac{Z_r - 0}{\sqrt{1/n - 3}} = \frac{-.447 - 0}{\sqrt{1/2996}} = -2.446$$

$\overline{X} = 48.9$	$s_X = 34.3$
$\overline{Y} = 42.1$	$s_Y = 38.9$
$r_{yx} = -0.42$	$R^2_{YX} = 0.18$
$a = 65.6$	$b_{yx} = -0.48$

Inferences About Correlation (ρ_{YX})

Calculate the test statistic

$$Z_r = \left(\frac{1}{2}\right) \ln\left(\frac{1 + r_{YX}}{1 - r_{YX}}\right) = \left(\frac{1}{2}\right) ln\left(\frac{1 - 0.42}{1 + 0.42}\right) = -0.447$$

$$Z = \frac{Z_r - 0}{\sqrt{1/n - 3}} = \frac{-0.447}{\sqrt{1/2{,}996}} = -24.46$$

$\bar{X} = 48.9$ $s_X = 34.3$

$\bar{Y} = 42.1$ $s_Y = 38.9$

$r_{yx} = -0.42$ $R^2_{YX} = 0.18$

$a = 65.6$ $b_{yx} = -\ 0.48$

Inferences About Correlation (ρ_{YX})

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do no reject H_0

We can restate the hypotheses:

H_0: $\rho_{YX} = 0$ ➔ Fail to reject H_0 if $|Z| \leq 1.96$

H_1: $\rho_{YX} \neq 0$ ➔ Reject H_0 if $|Z| > 1.96$

Since Z=-24.46, we reject H_0 … so it appears that in the population X and Y are correlated

Inferences About Slope (β_{YX})

Inferences About Slope (β_{YX})

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\beta_{YX} = 0$

H_1: $\beta_{YX} \neq 0$

This is normally a two-sided test, although it needn't be

Inferences About Slope (β_{YX})

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about β_{YX} to be valid

Inferences About Slope (β_{YX})

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's go with α=0.05

Inferences About Slope (β_{YX})

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

Because we are using sample-based estimates of the variability in the sampling distribution of b_{YX}, we will conduct a t (instead of a Z) test

Because MS_{ERROR} has n-2 degrees of freedom, we will select a critical value of t with df=n-2

For a two-sided test with α=0.05 and n-2=2,997 degrees of freedom, the critical value t*=1.962

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

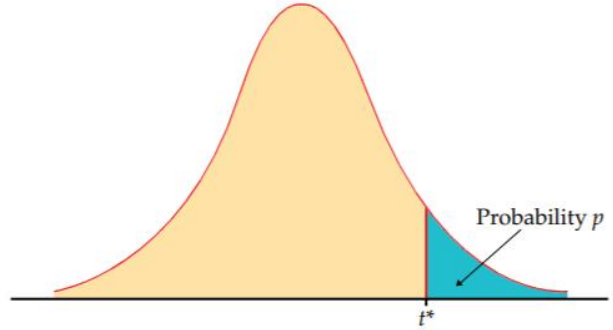


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291

Inferences About Slope (β_{YX})

Calculate the test statistic

$\bar{X} = 48.9$ $s_X = 34.3$

$\bar{Y} = 42.1$ $s_Y = 38.9$

$r_{yx} = -0.42$ $R^2_{YX} = 0.18$

$a = 65.6$ $b_{yx} = -0.48$

The test statistic t with df=N-2 equals:

$$t_{n-2} = \frac{b_{YX} - 0}{s_b} = \frac{b_{YX} - 0}{\sqrt{\dfrac{MS_{Error}}{(s_X^2)(n-1)}}}$$

In our example:

$$F_{1,n-2} = \frac{SS_{Regression}/1}{SS_{Error}/n-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{816{,}588/1}{3{,}720{,}015/2997} = 657$$

$$t_{n-2} = \frac{-.48 - 0}{\sqrt{\dfrac{3{,}720{,}015/2997}{(34.3^2)(2998)}}} = -25.6$$

Inferences About Slope (β_{YX})

$\bar{X} = 48.9 \quad s_X = 34.3$

$\bar{Y} = 42.1 \quad s_Y = 38.9$

$r_{yx} = -0.42 \quad R^2_{YX} = 0.18$

$a = 65.6 \quad b_{yx} = -0.48$

Calculate the test statistic

The test statistic t with df=N-2 equals:

$$t_{n-2} = \frac{b_{YX} - 0}{s_b} = \frac{b_{YX} - 0}{\sqrt{\frac{MS_{Error}}{(s_X^2)(n-1)}}}$$

In our example:

$$F_{1,n-2} = \frac{SS_{Regression}/1}{SS_{Error}/n-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{816{,}588/1}{3{,}720{,}015/2997} = 657$$

$$t_{n-2} = \frac{-0.48 - 0}{\sqrt{\frac{3{,}720{,}015/2997}{(34.3^2)(2{,}998)}}} = -25.6$$

Inferences About Slope (β_{YX})

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do no reject H_0

We can restate the hypotheses:

H_0: $\beta_{YX} = 0$ ➔ Fail to reject H_0 if $|t| \leq 1.96$

H_1: $\beta_{YX} \neq 0$ ➔ Reject H_0 if $|t| > 1.96$

Since $t = -25.6$, we reject H_0

Confidence Intervals for β_{YX}

Inferences About Slope (β_{YX})

Using the sample estimate (b_1) of β_{XY} and the standard error of the sampling distribution of β_{YX} (above), we can compute confidence intervals for β_{YX}

The standard error is:

$$s_b = \sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}}$$

So the confidence interval can be expressed as:

$$C.I. = b_1 \pm t* \sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}}$$

Inferences About Slope (β_{YX})

For our example, a 95% confidence interval would be:

$$C.I. = b_1 \pm 1.96\sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}} =$$

$\bar{X} = 48.9$ $s_X = 34.3$

$\bar{Y} = 42.1$ $s_Y = 38.9$

$r_{yx} = -0.42$ $R^2_{YX} = 0.18$

$a = 65.6$ $b_{yx} = -0.48$

$MS_{Error} = 1{,}241.24$

$$-.48 \pm 1.962\sqrt{\frac{3,770,015/2990}{(34.3^2)(2998)}}$$

$$-.48 \pm .037$$

$-.443$ and $-.516$

Inferences About Slope (β_{YX})

For our example, a 95% confidence interval would be:

$$\text{C.I.} = b_1 \pm 1.96\sqrt{\frac{\text{MS}_{\text{ERROR}}}{(s_X^2)(n-1)}} =$$

$$\text{C.I.} = -0.48 \pm 1.96\sqrt{\frac{1241.24}{(34.3^2)(2{,}999-1)}}$$

$$\text{C.I.} = -0.48 \pm 0.037$$

$$\text{C.I.} = -0.517 \text{ to } -0.443$$

$\bar{X} = 48.9$ $s_X = 34.3$

$\bar{Y} = 42.1$ $s_Y = 38.9$

$r_{yx} = -0.42$ $R^2_{YX} = 0.18$

$a = 65.6$ $b_{yx} = -0.48$

$MS_{Error} = 1{,}241.24$

Standardized Regression Coefficients

Standardized Coefficients (β^*_{YX})

For a variety of reasons researchers often like to express the slope of the regression line in standardized terms

This is useful when:

- The metric of X is in an arbitrary scale, or a scale that is not intrinsically meaningful
- We want to better understand the magnitude of the association between X and Y

Instead of asking…

"How many units does Y change as a result of a one unit change in X?"

we might ask,

"How many standard deviations does Y change as a result of a one standard deviation change in X?"

Standardized Coefficients (β^*_{YX})

The standardized slope, or beta coefficient (or beta weight) is expressed as

$$\beta^*_{YX} = (b_{YX})\left(\frac{s_X}{s_Y}\right)$$

In bivariate regression, the standardized slope thus equals the correlation, r_{YX}. In our example:

$$\beta^*_{YX} = (-0.48)\left(\frac{34.3}{38.9}\right) = -0.42$$

$\bar{X} = 48.9$ $s_X = 34.3$

$\bar{Y} = 42.1$ $s_Y = 38.9$

$r_{yx} = -0.42$ $R^2_{YX} = 0.18$

$a = 65.6$ $b_{yx} = -0.48$

$MS_{Error} = 1{,}241.24$