Science Dynamics Masterclass
© 2025 Science Dynamics
Research Labs, Proverbs

14 - Correlation and Simple Regression as Inferential Techniques

Bivariate Normal Distribution

- Taken separately - X and Y are normally distributed
- When X is held constant all Y values have normal distribution
- Y constant = dtto for X
- All conditional distributions of Y given X have the same standard deviation and vice versa
- the means of all the conditional distributions of Y given X fall along straight line and vice versa

Normality

Homoscedasticity

Linearity

Testing Whether the Pearson equals to zero

Create a scatterplot

- linearity
- spread of values must be "reasonable"
- no severe outliers

$$t = \frac{r - 0}{\sqrt{(1 - r^2)/(N - 2)}}$$

$$df = N - 2$$

Calculate the t statistics, find the p-value

Calculate the t statistics, find the...

r is the sample Pearson Product Moment CC and N-2 is the number of the degrees of freedom (two less than the number of pairs in the sample)
p-value in SPSS: Analyze, Correlate,

...

$$Z_r = 0.5 ln \frac{1 + r}{1 - r}$$

$$Upper\ limit = Z_r + (z_{critical})\left(\frac{1}{\sqrt{N - 3}}\right)$$

$$Lower\ limit = Z_r - (z_{critical})\left(\frac{1}{\sqrt{N - 3}}\right)$$

$$z_{critical} = 1.645(90\%CI); 1.960(95\%CI)$$

$$z_{critical} = 2.576(99\%CI)$$

Using CI to estimate size of the population correlation coef

Using CI to estimate size of...

to transform r to Z using SPSS: Transform, Compute, Z in Target Variable, Numeric Expression: 0.5* LN((1+r)/(1-r))

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

Then transform obtained Z values back to r

Then transform obtained...

in SPSS: Transform, Compute, espression: (Exp(2*Z)-1)/(Exp(2*Z)+1)

Exploring goodness of Fit: Regression Diagnostics

$$\rho_{Y|X} = standard\ \ error\ \ of\ \ prediction\ \ (estimate\ )$$

estimate of the standard error based on sample data:

$$\hat{\rho}_{X|Y} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{N - 2}}$$

Population standard Error of Prediction

Population standard...

where Yi is the actual Y score of individual i in the sample ^Yi is the predicted Y score of individual i in the sample using linear regression equation, N is the number of pairs

$$t = \frac{b - 0}{\hat{\sigma}_b}$$

$$df = N - 2$$

Testing the b-weight for statistical significance

$$\hat{\sigma}_b = \sqrt{\frac{1}{N - 2}\frac{\sum(Y_i - \hat{Y}_i)^2}{(N - 1)\hat{\sigma}_X^2}}$$

standard error of b

SPSS - Constructing linear Regression Equation for Prediction

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	75,779	8,775		8,635	,000	57,803	93,755
	Temperature in degrees Fahrenheit	1,180	,116	,887	10,149	,000	,942	1,419

a. Dependent Variable: Number of ice cream bars sold

Example - Coefficients

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3232,599	1	3232,599	103,000	,000[a]
	Residual	878,767	28	31,385		
	Total	4111,367	29			

a. Predictors: (Constant), Temperature in degrees Fahrenheit
b. Dependent Variable: Number of ice cream bars sold

Using ANOVA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,887[a]	,786	,779	5,602

a. Predictors: (Constant), Temperature in degrees Fahrenheit

Measuring the fit of the overall regression equation

Measuring the fit of the overall...

measere of the fit of th eoverall regressio equation is given by R, the correlatoin between te actual values of the dependent variable and te values of the predictd dependent variable Y and Y^.

Relating R^2 and Ro^2yx

$$F_{P,N-P-1} = \frac{R^2/P}{(1 - R^2)/(N - P - 1)}$$

Testing R^2 for statistical significance

Testing R^2 for statistical significance

There are four ways to test stat. significance - t-test of the significance of r, t-test of the significance of b, ANOVA F-test of the overall regresion model and R^2 (equation above)
where P is the number of predictors in the regression equation, N is sample

...

$$R^2_{ADJ} = 1 - (1 - R^2)\frac{(N - 1)}{N - P - 1}$$

Estimating the true population R^2

Estimating the true population R^2

Sample R^2 overestimates the population R^2. For ratios P:N 1:30 the difference is minimal, for 1:10 it could be great.