R code

Students also viewed

ecture 7 - Repeated-Measures AN...

20 terms

nishka1509 Preview

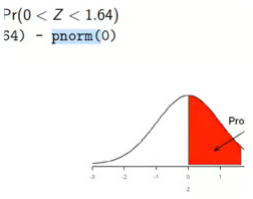
Study Designs and Models

14 terms

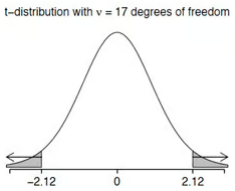
A andrecamerino_ Preview

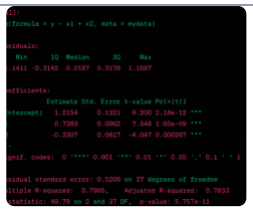
Terms in this set (13)

Usage of pnorm()	**pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)** **Probability in a specific area under a normal distribution curve.** *q is the "value" from which you want the area above/below* Example questions: *Find the proportion of students with a height between 180-190cm* **pnorm(q, mean = 1, sd = 2)** = pnorm(q, mean = 1, sd = 2, lower.tail = **TRUE**) **pnorm(q)** = pnorm(q, **mean = 0, sd = 1, lower.tail = TRUE**) **1 - pnorm(q)** = pnorm(q, mean = 0, sd = 1, lower.tail = **FALSE**)

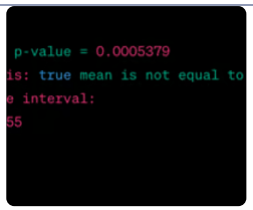


Usage of *qnorm()*	**qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)** To find the value "q" in pnorm() Example questions: *Find the height which is exceeded by 10% of students.* *Find the values between which 95% of the population cholesterol levels lie:* **qnorm(p, mean = 1, sd = 2)** = qnorm(p, mean = 1, sd = 2, lower.tail = **TRUE**) **qnorm(p)** = qnorm(p, **mean = 0, sd = 1, lower.tail = TRUE**) **1 - qnorm(p)** = qnorm(p, mean = 0, sd = 1, lower.tail = **FALSE**)
Usage of dbinom()	**dbinom(x=15, size=20, prob=0.75)** *(to get the value of Pr(x = 15) for X~B(20, 0.75))* *The command dbinom() will provide the individual binomial probabilities associated with a given outcome, provided the number of trials (size) and the probability of 'success' (prob).*
Usage of pbinom()	**pbinom(q=10,size=20,prob=0.75)** (to get the total probability of *Pr (X ≤ 10) = Pr (X = 0, 1, 2, . . . , 10)*) Example question: *Find the probability that 10 or fewer live with both parents.* *The command pbinom will provide the **sum** of all individual binomial probabilities less than or equal to a given outcome q, provided the number of trials (size) and the probability of 'success' (prob).*

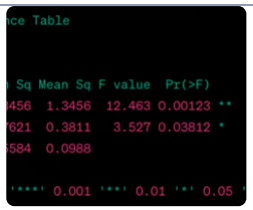
Usage of pt()	**2*pt(q=2.12, df=17, lower.tail=FALSE)** To get the p-value. Because there are two tails, so we need to *2. Alternative commands: 2*(1-pt(2.12, 17)) **2*pt(q=-2.12, df=17)** **2*(1-pt(q=-2.12, df=17))** 
Usage of qt()	**qt(p, df)** To find the multiplier. e.g., *from R, qt(0.975,4) = 2.776, so the confidence interval becomes: 0.492 +/- 2.776 * 0.08513*
Usage of pchisq()	**1-pchisq(q = 9.70, df = 1)** **pchisq(9.70, 1, lower.tail = FALSE)** *q is the t - stat value (chi^2 = 9.7)* To find the p-value of a chi-square
Usage of qchisq()	**qchisq(p = 0.95, df = 1)** To find the critical value of chi-square
Usage of pf()	**pf(1.0242, df1=3, df2=28, lower.tail=F)** To find the p-value of the F-distribution **1.0242 is the F test statistic**
Usage of qf()	**qf(0.05, 3, 15, lower.tail=FALSE)** To find the critical value of F statistic

Output for regression model fit	The **"Residual standard error"** represents the **standard deviation** of the residuals, which is an estimate of the average distance between the observed and predicted values. The **"Multiple R-squared"** and "Adjusted R-squared" values indicate the goodness of fit of the model. They represent the proportion of variance in the response variable explained by the predictors. **Adjusted R-squared** *takes into account the number of predictors and the sample size*. The **"F-statistic"** is a measure of overall significance of the model. It assesses whether the regression model as a whole is statistically significant. The associated **p-value** indicates the probability of obtaining such an F-statistic by chance.	

Output for a t test	The **"data"** line indicates the **name of the variable** or group being tested. In this example, it's denoted as "x." The **"t"** value represents the **calculated t-statistic** for the test. It measures the difference between the sample mean and the hypothesised mean relative to the variability in the data. The "**df"** value stands for **degrees of freedom,** which is a measure of the amount of information available for the test. It represents the sample size minus one. The "**p-value"** is the probability of obtaining the observed test statistic (t-value) or a more extreme value under the null hypothesis. It indicates the level of statistical significance. The **"alternative hypothesis"** states the alternative to the null hypothesis. In this example, it states that the t**rue mean is not equal to 0.** The **"95 percent confidence interval"** provides a range of values within which we can be 95% confident that the true population mean lies. It is calculated based on the sample data and reflects the



	precision of the estimate. The **"sample estimates"** section presents the estimated mean of the variable or group being tested. In this example, it shows the estimated mean of "x" as 0.1271137.

Output for an ANOVA table	The **"Response"** line shows the name of the **dependent variable,** which in this example is denoted as "y". The **"Df"** column represents the **degrees of freedom** associated with each factor or source of variation. In this example, there are **two predictor variables (x1 and x2),** each with their respective degrees of freedom, and the **"Residuals"** row represents the **degrees of freedom associated with the error or unexplained variation.** The **"Sum Sq"** column shows the sum of squares associated with each factor or source of variation. It represents the **total variability** explained by each factor. The **"Mean Sq"** column represents the mean square, which is calculated by dividing the sum of squares by the degrees of freedom. It represents the **average variability explained by each factor.** The **"F value"** column displays the F-statistic, which is calculated by dividing the mean square of each factor by the mean square of the residuals. It **measures the ratio** 

of explained variation to unexplained variation and is used to **test the significance of each factor.**

The **"Pr(>F)"** column shows the **p-value associated with each factor.** It indicates the **probability of obtaining the observed F-statistic** or a more extreme value under the null hypothesis.