

How to run T-tests using R

Presenters: Kyle Ward and Jon Wayland

Spring 2013

R is a free software downloadable at http://www.r-project.org/

1. R Console Setup: > prompts you for formula or function. The result appears on the next line(5). 2. Comments begin with # Anything in the line following a # is a comment. 3. Installing a Package Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) We will be using a function called leveneTest() which is the package called "car". 1) Select "Install Packages" in the dropdown menu "Packages" at the top of the screen. 2) Select the country, and state that is nearest you. 3) Select the package "car" and press "ok". 4) Activate the package strains for the scores on a memory test. 5. Alternate Form of the Data > status scores i) One vector with all scores (smokers and nonsmokers) ii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iiii) One vector with all scores (smokers and nonsmokers) iiii) One vector with all scores (smokers and nonsmokers) iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	Notes:	Code and Output:
The result appears on the next line(s). 2. Comments begin with # Anything in the line following a # is a comment. 3. Installing a Package Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) We will be using a function called leveneTest() which is the package called "car". 1) Select function called leveneTest() which is the package "car" and press "ok". 2) Select the country, and state that is nearest you. 3) Select the package "car" and press "ok". 4) Activate the package "car" and press "ok". 4) Activate the package "car" and press "ok". 4) Activate the package "car" and press "ok". 5. Alternate Form of the Data > Scores = c(1onsmokers, smokers) > status scores i) One vector with all scores (smokers and nonsmokers) ii) One vector identifying which group the individual belongs in. iii) One vector identifying which group the individual belongs in. iii) One vector identifying which group the individual belongs in. iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	1. R Console Setup:	
2. Comments begin with # Anything in the line following a # is a comment. 3. Installing a Package Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA (Larnegie Mellon) We will be using a function called leveneTest() which is the package called "car". 1. Select th "Install Packages" in the dropdown menu "Packages" at the top of the screen. 2. Select the package "car" and press "ok". 4. Creating a Dataset We will look at the differences between smokers and nonsmokers in terms of their scores on a memory test. 5. Alternate Form of the Data * Internate Form of the Data * One vector with all scores (smokers and nonsmokers) i) One vector with all scores (smokers and nonsmokers) ii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) ii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iii) One vector with all scores (smokers and nonsmokers) iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	> prompts you for formula or function.	
Anything in the line following a # is a comment. # This is a comment! 3. Installing a Package Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) Ketk memory too package too package called "car". 10. Select "Install Packages" in the dropdown menu "Packages" at the too pot the screen. # In step 1) you can also use the command 2) Select the country, and state that is nearest you. 3. Select the package "car" and press "ok". > library(car) 4. Creating a Dataset > nonsmokers = c(18,22,21,17,20,17,23,20,22,21) > soncers = c(16,20,14,21,20,18,13,15,17,21) 5. Alternate Form of the Data > scores = c (nonsmokers, smokers) onsmokers) i) One vector with all scores (smokers and nonsmokers) > no 18 ii) One vector videntifying which group the individual belongs in. 1 on 18 2 no 22 3. Be core identifying which group the individual belongs in. 1 no 18 2 no 22 3. Ro 20 9 no 21 11 yes 16 11 yes 16 12 yes 20 11 yes 16 12 yes 20 13 yes 10 11 yes 16		
3. Installing a Package Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) We will be using a function called levenetest() which is the packages in the dropdown menu "Packages" at the top of the screen. # In step 1) you can also use the command build (A) USA NO USA N	2. Comments begin with #	
Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) We will be using a function called leveneTest() which is the package called "car".If use p1) you can also use the command USA (CA) USA RAD USA RA		# This is a comment!
We will look at the differences between smokers and nonsmokers in terms of their scores on a memory test. > smokers = c(16,20,14,21,20,18,13,15,17,21) S. Alternate Form of the Data > scores = c(nonsmokers, smokers) S. Alternate Form of the Data > scores = c(nonsmokers, smokers) Rather than the scores being in separate vectors, data frame(status, scores) > status=c(rep("no", length(nonsmokers)), rep("yes", length(smokers))) Point vector with all scores (smokers and nonsmokers) 1 no 18 ii) One vector identifying which group the individual belongs in. 1 no 18 iii) One vector identifying which group the individual belongs in. 1 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 2 yes 20 13 yes 14 14 yes 21 15 yes 20 15 yes 20 15 yes 18 14 yes 21 15 yes 18	 Many functions and data sets are available in packages that be downloaded from a CRAN site. We generally use PA 1 (Carnegie Mellon) We will be using a function called leveneTest() which is the package called "car". Select "Install Packages" in the dropdown menu "Packages" at the top of the screen. Select the country, and state that is nearest you. Select the package "car" and press "ok". Activate the package using the library command. 	UK (St Andrews) USA (CA 1) USA (CA 2) USA (IA) USA (IN) USA (MD) USA (M
nonsmokers in terms of their scores on a memory test.5. Alternate Form of the Data>scores = c(nonsmokers, smokers)Rather than the scores being in separate vectors, data for t-tests is sometimes in this format:i)One vector with all scores (smokers and nonsmokers)ii)One vector identifying which group the individual belongs in.iii)One vector identifying which group the individual belongs in.iiii)One vector identifying which group the individual belongs in.iiiiiOne vector identifying which group the individual belongs in.iiiiOne vector identifying which group the individual belongs in.iiiiOne vector identifying which group the individual belong in.iiiiOne vector identifying which group the individual belong in.iiiiOne vector identifying which group the individual belong in.iiiiOne vector ide	-	
5. Alternate Form of the Data >scores = c(nonsmokers, smokers) 7. Alternate Form of the Data >scores = c(nonsmokers, smokers) 8. Alternate Form of the Data >scores = c(nonsmokers, smokers) 9. Construction >status=c(rep("no", length(nonsmokers)), rep("yes", length(smokers))) 9. Construction >data.frame(status, scores) 10. One vector with all scores (smokers and nonsmokers) >data.frame(status, scores) 11. One vector identifying which group the individual belongs in. 1 no 18 2 no 21 4 no 17 5 no 20 6 no 17 7 no 23 8 no 21 11. yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18		> Smokers = c(16,20,14,21,20,18,13,15,17,21)
Rather than the scores being in separate vectors, data for t-tests is sometimes in this format: >status=c(rep("no", length(nonsmokers)), rep("yes", length(smokers))) i) One vector with all scores (smokers and nonsmokers) >data.frame(status, scores) ii) One vector identifying which group the individual belongs in. 1 no 18 iii) One vector identifying which group the individual belongs in. 1 no 12 no 21 iiii One vector identifying which group the individual belongs in. 1 no 17 5 no 20 iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii		Scores = c(nonsmellers _ smellers)
Rather than the scores being in separate vectors, data for t-tests is sometimes in this format: i) One vector with all scores (smokers and nonsmokers) ii) One vector identifying which group the individual belongs in. 1 no 18 2 no 22 3 no 21 4 no 17 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 4 yes 21 15 yes 20 16 yes 18 	5. Alternate Form of the Data	
Rather than the scores being in separate vectors, data for t-tests is sometimes in this format:>data.frame(status, scores)i)One vector with all scores (smokers and nonsmokers)status scoresii)One vector identifying which group the individual belongs in.1iii)One vector identifying which group the individual belongs in.1iiii)One vector identifying which group the individual belongs in.1iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii		
for t-tests is sometimes in this format: i) One vector with all scores (smokers and nonsmokers) ii) One vector identifying which group the individual belongs in. 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18	Bather than the scores being in separate vectors, data	
i) One vector with all scores (smokers and nonsmokers) 1 no 18 ii) One vector identifying which group the individual belongs in. 3 no 21 4 no 17 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 11 yes 10 12 yes 20 13 yes 14 14 yes 10 15 yes 10 16 yes 12 15 yes 14 14 yes 14 15 yes 14 14 yes 14 15 yes 14 15 yes 14 16 yes 18		
nonsmokers) 2 no 22 ii) One vector identifying which group the individual belongs in. 17 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 16 12 yes 20 13 yes 14 14 yes 20 15 yes 20 16 12 yes 15 yes 14 14 yes 21 15 yes 20 16 yes 14 14 yes 21 15 yes 20 16 yes 14		1 no 18
 ii) One vector identifying which group the individual belongs in. 3 no 21 4 no 17 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18 		2 no 22
individual belongs in. 4 no 17 5 no 20 6 no 17 7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 15 yes 20 16 yes 18 18 18		
5 NO 20 6 NO 17 7 NO 23 8 NO 20 9 NO 22 10 NO 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18		
7 no 23 8 no 20 9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18		
8no209no2210no2111yes1612yes2013yes1414yes2115yes2016yes18		
9 no 22 10 no 21 11 yes 16 12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18		
10no2111yes1612yes2013yes1414yes2115yes2016yes18		
12 yes 20 13 yes 14 14 yes 21 15 yes 20 16 yes 18		
13 yes 14 14 yes 21 15 yes 20 16 yes 18		11 yes 16
14 yes 21 15 yes 20 16 yes 18		12 yes 20
15 yes 20 16 yes 18		
16 yes 18		
18 yes 15		
19 yes 17		
20 yes 21		

6. Boxplots

Boxplots are a useful graphical method for comparing multiple groups. It is important to keep in mind that boxplots are median oriented graphics, while the t-test is comparing means.

ylab is the label given to the y axis

> boxplot(nonsmokers, smokers, ylab="Scores on Memory Test", range=1.5, names=c("nonsmokers","smokers"), main="Memory Performance by\n Smoking Status")

# the \n indicates that you want the main label to split onto a second line	Memory Performance by Smoking Status
	Scores on Memory Test Scores on Memory Test 16 16 17 18 20 22 18 20 22 22 22 23 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 18 20 23 20 18 20 23 20 23 20 23 20 23 20 23 20 23 20 23 20 23 20 23 20 20 20 20 20 20 20 20 20 20 20 20 20
7. Descriptive Statistics Running the Mean and Standard Deviation of each group's scores gives you information to use during your interpretation of the t-test.	<pre>> mean(nonsmokers) [1] 20.1 > mean(smokers) [1] 17.5 > sd(nonsmokers) [1] 2.131770 > sd(smokers) [1] 2.953341</pre>
8. Independent-Samples T-test We will be running an independent sample t-test comparing mean scores between two independent groups.	
9. t-test() command	# This is the general formula for a t-test
To the right is the general function for the t-test It shows the default settings for a t-test run in R if you were to simply type in t.test(x,y)	<pre>> t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)</pre>
 10. Checking Assumptions: Homogeneity of Variance To check the assumption of equal variances, we will run a Levene's test in R Levene's Test 	 > # Be sure to do step 3. before using the levelTest() command. > leveneTest(scores, status)
In running a Levene's test in R, place the outcome variable first and the grouping variable second.	Levene's Test for Homogeneity of Variance (center = median)
 A p-value less than .05 violates the assumption of homogeneity of variance We do not want a p-value less than .05 	Df Fvalue Pr(>F) group 1 1.9459 0.18 18

11. Checking Assumptions: Normality	
Shapiro-Wilk Test	>shapiro.test(scores)
•	Shapiro-Wilk normality test
The Shapiro-Wilk test compares the scores in your	
sample to a normally distributed set of scores with the	
same mean and standard deviation.	data: scores
• If the test is not significant (p > .05) it tells us	W = 0.9369, p-value = 0.2098
that our distribution is not statistically	
different from the normal distribution.	
 Like the Levene's test, we do not want this test 	
to be statistically significant	
12. Checking Assumptions: Normality	> qqnorm(scores)
Q-Q plots	>qqline(scores, col= "red ")
The Q-Q chart plots the values you would expect to get if the distribution were normal (theoretical values) against the values actually seen in our dataset (sample	Normal Q-Q Plot
 values). If the data were normally distributed, the data would fall along a straight line. 	Sample Quantiles 0 0 0 0 0 0 0 22 0 0 0 0 0 0 0 22 0 0 0 0 0 0 0 22 0 0 0 0 0 0 0 0 23 0 0 0 0 0 0 0 0 0 23 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 Any deviation from a straight line represents a deviation from normality. 	24 Sample 24 Sam
	-2 -1 0 1 2
	Theoretical Quantiles
13. Independent Samples T-Test	>t.test(nonsmokers,smokers, var.equal = TRUE)
To run an independent samples, two-tailed t-test,	Two Sample t-test
simply input the variable names in the t-test command.	
This command assumes the default that it is a two-	data: nonsmokers and smokers
sample test, it is a two-tailed test, equal variances are	t = 2.2573, df = 18, p-value = 0.03665
not assumed, and the confidence level is set at .95.	
not assumed, and the confidence level is set at .95.	alternative hypothesis: true difference in means is not equal to 0
	alternative hypothesis: true difference in means is not equal to 0
Note: We are using the default assumptions:	
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE,	95 percent confidence interval:
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE,	
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE,	95 percent confidence interval: 0.1801366 5.0198634 sample estimates:
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE,	95 percent confidence interval: 0.1801366 5.0198634
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages"
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female	<pre>95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign"</pre>
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign"
14. Paired Sample T-Test Do Male employees tend to earn more than female employees?	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" R RGui (32-bit) File Edit View Misc Packages Windows Help
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set:	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" Elie Edit View Misc Packages Windows Help Elie Edit View Misc Packages Windows Help Elie Edit View Misc Packages Undows Help
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set:	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages Windows Help
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees?	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" [Regui (32-bit)] File Edit View Misc Packages Windows Help [Deal Packages] Set CRAN mirror
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages. Set CRAN mirror Select repositories
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages. Select repositories Select repositori
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # fore t
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were matched by an indicator of salary potential and	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages Windows Help East CRAN mirror Select one Select repositories Select repositories.
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were matched by an indicator of salary potential and represent the entire spectrum of earnings. The data is	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages Select repositories Select repositories Show to cite R of Install package(s) > library(foreign) >salaries=read.spss (file.choose(), to.data.frame=TRUE)
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were matched by an indicator of salary potential and represent the entire spectrum of earnings. The data is not real but was generated to reflect actual 2011	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" File Edit View Misc Packages Windows Help File Edit View Misc Packages (File choose (), select repositories Install package(s) > library(foreign) > salaries=read.spss (file.choose(), to.data.frame=TRUE) to.data.frame=TRUE) # colorspace
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were matched by an indicator of salary potential and represent the entire spectrum of earnings. The data is	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" R RGui (32-bit) File Edit View Misc Packages Windows Help Cod package Select repositories Select file from the folder it has been saved Settach(salaries)
Note: We are using the default assumptions: alternative =c("two.sided"), mu = 0, paired = FALSE, var.equal = FALSE, conflevel = .095 14. Paired Sample T-Test Do Male employees tend to earn more than female employees? The following analyses uses the SPSS data set: Salaries.sav The data set contains weekly salaries (in \$) for pairs of 100 male and female employees. The individuals were matched by an indicator of salary potential and represent the entire spectrum of earnings. The data is not real but was generated to reflect actual 2011 earning distributions published by the Bureau of Labor	95 percent confidence interval: 0.1801366 5.0198634 sample estimates: mean of x mean of y 20.1 17.5 # to load spss file, select "packages" in the top menu and choose "Load Packages" # from the package list, choose "foreign" # from the package list, choose "foreign" # RGui (32-bit) File Edit View Misc Packages Windows Help Coad package Select repositories Select reposit

15. Looking at your data	>head(salaries)
Using the head() function in R allows you to view 6	
respondents. This is an easy way to see what the	PairNo SalaryF SalaryM
variables are named and how the data has been	1 1 287 256
recorded.	2 2 291 314
	3 3 180 257
	4 4 177 243
	5 5 304 276
	6 6 322 268
16. Descriptive Statistics	> mean(SalaryM)
	[1] 1004.98
	> sd(SalaryM)
	[1] 604.1701
	> mean(SalaryF)
	[1] 840.38
	> sd(SalaryF)
	[1] 520.951
17. Boxplots	<pre>> boxplot(SalaryM,SalaryF,ylab="Weekly Salary",</pre>
Like the example above comparing smokers and	names=c("Males","Females"), main="Weekly Salaries by Gender
nonsmokers on a memory task, we want to look at the	range=1.5)
distribution of the data using a boxplot.	
	Weekly Salaries by Geneder
	°
Note: The range=1.5 tells R to mark any points that are	5 5000
further than 1.5*IQR from the box as outliers.	ary 22000
where	weekly Salary
$IQR=Interquartile range = Q_3-Q_1$	See
	Males Females
18. Paired Sample T-Test	<pre>> t.test(SalaryM,SalaryF, alternative = c("greater"), paired = TRUE</pre>
	var.equal = FALSE, conf.level = 0.95)
• Because the equation runs left to right, the	· · · · · ·
original measurement should be placed first.	
 We will make the alternative hypothesis 	Paired t-test
"greater" because we are testing if males have	
a larger weekly income compared to a paired	data: SalaryM and SalaryF
group of females	t = 9.2209, df = 99, p-value = 2.738e-15
 We change "paired" to true to change it to a 	
paired sample t-test	alternative hypothesis: true difference in means is greater than 0
• Based on the results, the p-value is less than	
.05, meaning that men have a significantly	95 percent confidence interval:
higher weekly income when compared to	134.9607 Inf
women.	
women.	sample estimates:
	mean of the differences

