Psychology 205: Fall, 2015
Problem Set 1 - Solutions

William Revelle

Contents

1 Introduction to using R for statistics

Problem set 1 asked for a variety of analyses. Here I show the direct answers, but also do the analyses in a variety of ways. I use the statistical program R. For help on R, go to the short tutorial on using R for research methods `http://personality-project.org/r/r.205.tutorial.html`. In the following, I assume that you have downloaded R and installed the *psych* package.

2 Comparing two groups

2.1 A sample problem

An investigator believes that caffeine facilitates performance on a simple spelling test. Two groups of subjects are given either 200 mg of caffeine or a placebo. Although there are several ways of testing if these two groups differ, the most conventional would be a t-test. Apply a t-test to the data in Table 1:

Table 1: The effect of caffeine on spelling performance

placebo	caffeine
24	24
25	29
27	26
26	23
26	25
22	28
21	27
22	24
23	27
25	28
25	27
25	26

2.2 Review of variability of distributions of samples

Many statistical tests may be thought of as comparing a statistic to the error of the statistic. One of the most used tests, the t-test (developed by Gossett but published under the name of Student), compares the difference between two means to the expected error of the difference between to means. As we know, the standard error (se) of a single group with mean, $\bar{X}$ with standard deviation, s, and variance, s^2

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2/}{n-1} \tag{1}$$

is just

$$s.e. = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}} \tag{2}$$

and the standard error of the difference of two, uncorrelated groups is

$$se_{x_1-x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{3}$$

How best can we understand the notion of a standard error? One way is to draw repeated samples from a known population and examine their variability. Although this was the procedure used by Gossett, it is also possible to simulate this using random samples drawn by computer from a known or unknown distribution. Using R it is easy to simulate distributions, either the normal or resampled from our data. Consider 20 samples from a normal distribution of size 12 (Figure 1. For each sample we show the mean and the confidence interval of the mean. Note how some of the means are very far apart. That is, even though the mean for the population is known to be zero, the means of samples vary around that. The horizontal lines in the graph represent 1.96 * the standard error of the mean. Note how the confidence region around almost all sample means includes the population mean. But note how some do not. The confidence intervals are shown as "cats' eyes" to represent the point that most of the confidence is in the middle of the region.

```
> x <- matrix(rnorm(240),ncol=20)
> error.bars(x, xlab="sample", main="Means and Confidence Intervals")
> abline(h=0)
```

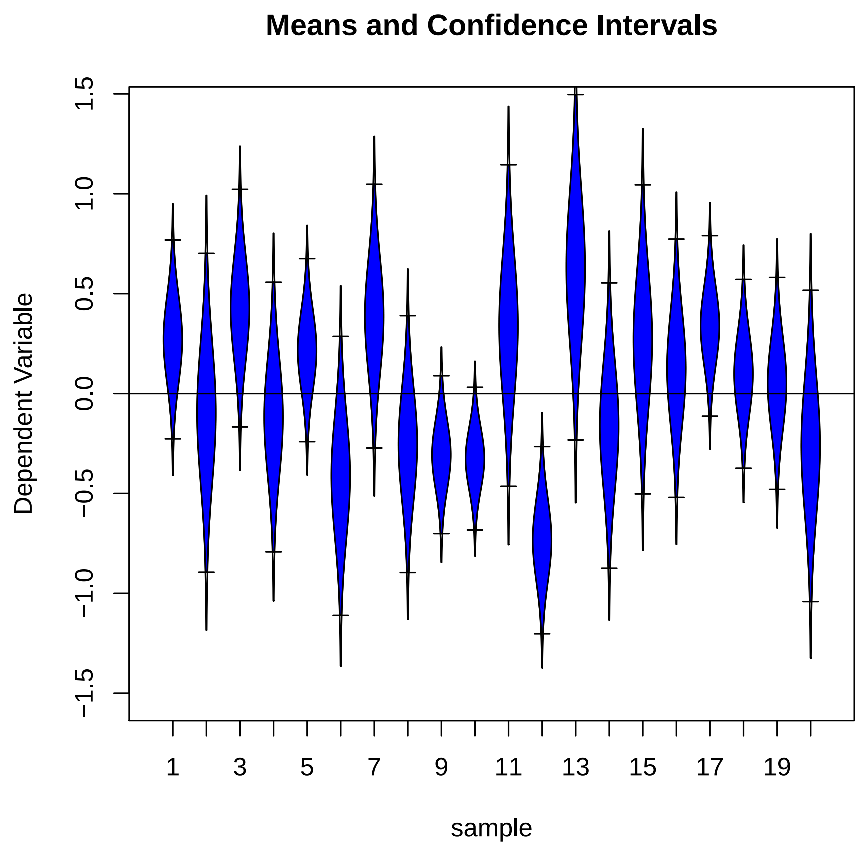

Figure 1: The mean and 95% confidence intervals for twenty randoms of size 12 from a normal distribution.

An alternative to sampling from the normal population is to resample from the actual data that we collect. Figure 2 shows the mean and confidence regions for 20 samples of size 12, where each sample was drawn with replacement from the original data. Once again, note how much variability there is from sample to sample, even though they come from the same population.

```
> x <- matrix(sample(spelling[,1],240,replace=TRUE),ncol=20)
> error.bars(x, xlab="sample", main="Means and Confidence Intervals")
> abline(h=24.25)
```

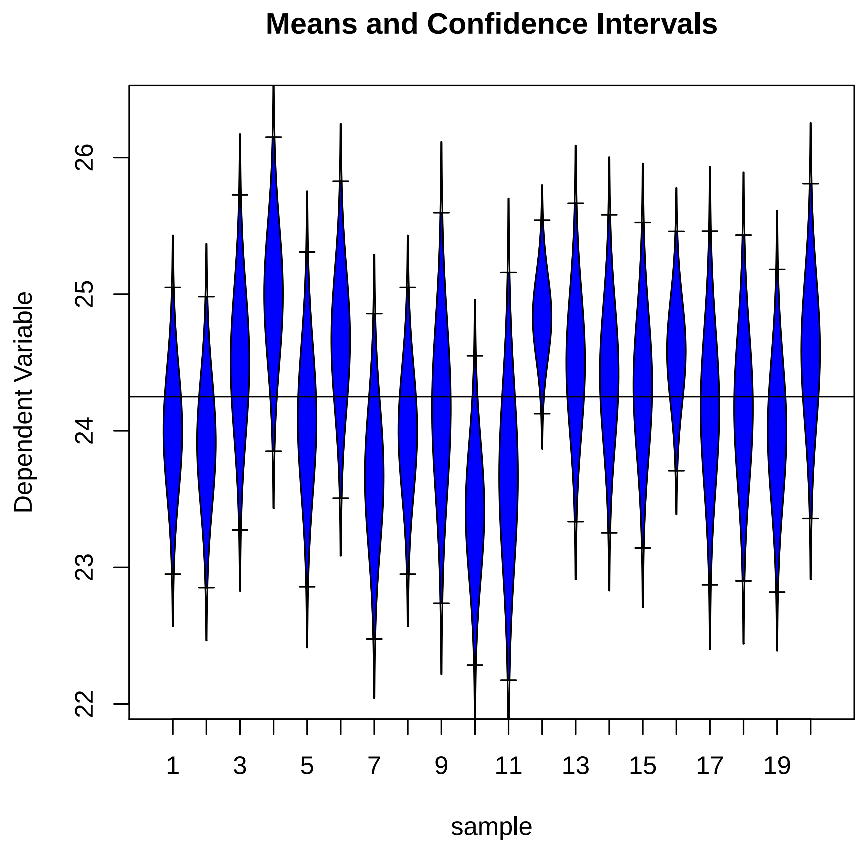

Figure 2: 20 random resamples (with replacement) of the spelling data. The horizontal line represents the mean of the original data.

Just as we can find the standard deviation of the data and standard error of the mean of a sample, so we can find the standard deviation and associated standard error of the mean for differences between two samples. The standard error of the difference of two, uncorrelated groups is two, uncorrelated groups is

$$se_{x_1-x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{4}$$

Given that samples from the same population differ a great deal, how much do the spelling scores of the placebo and caffeine groups differ? Do they differ more than would be expected by chance if in the population there was no effect of caffeine?

We can see this graphically by plotting 20 random samples from the *differences* between the two sets of data (Figure 3).

```
> x <- matrix(sample((spelling[,1]-spelling[,2]),240,replace=TRUE),ncol=20)
> error.bars(x, xlab="sample", main="Means and Confidence Intervals of the  difference between the t
> abline(h=0)
```

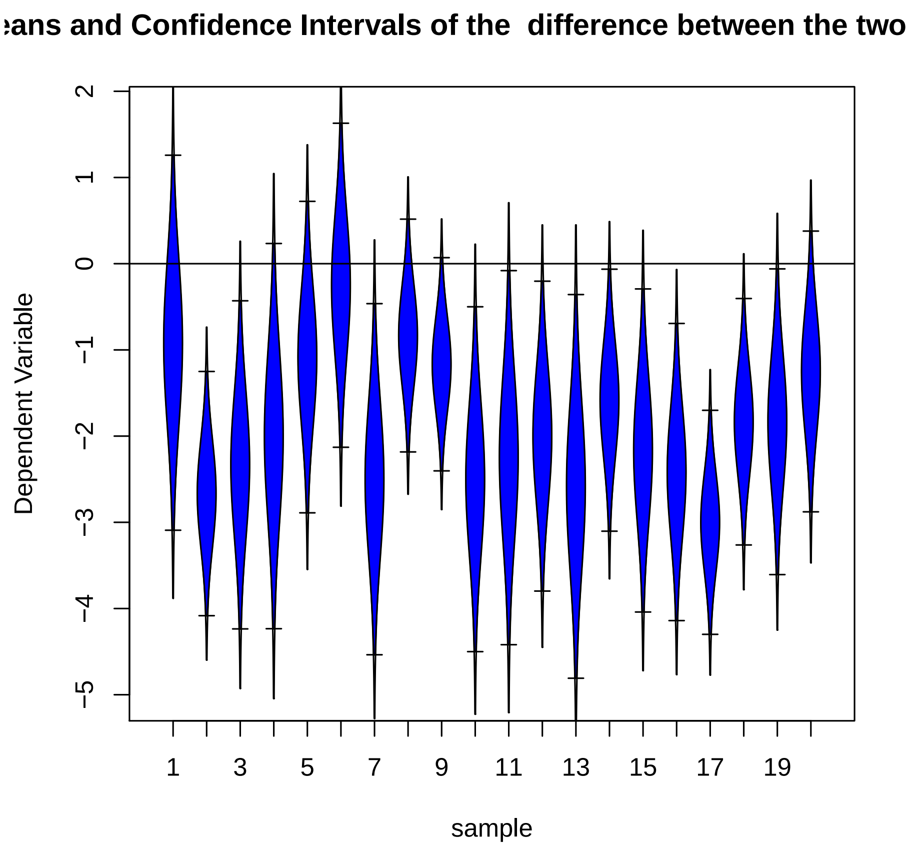

Figure 3: 20 random resamples (with replacement) of the spelling data. The horizontal line represents the mean of the original data.

2.3 The t-test

The t-test compares the differences between the means to the standard error of the differences between sample means.

That is,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{se_{x_1-x_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (5)$$

This looks somewhat complicated, but because it is such a common operation, the t-test is a basic function in R(as well as all major statistics programs).

2.4 Using R to do t-tests

From the point of view of most statistical programs, the data need to be rearranged to show the Independent Variable (IV) and the Dependent Variable (DV). Then we try to find how much the DV varies as a function of the IV.

In R, this is done by first loading in the *psych* package, then reading the clipboard using the `read.clipboard` and then using the `t.test` function

```
>library(psych)      #this loads the psych package into your active workspace
>spelling <- read.clipboard()    #copy into your clipboard and then read the clipboard into R
```

It is always useful to describe the data, both numerically and graphically. Numerically we can do this using the `describe` function.

```
> describe(spelling)

        vars  n  mean   sd median trimmed  mad min max range  skew kurtosis   se
Placebo    1 12 24.25 1.86   25.0    24.3 1.48  21  27     6 -0.33    -1.33 0.54
Drug       2 12 26.17 1.85   26.5    26.2 2.22  23  29     6 -0.22    -1.33 0.53
```

We can show this effect by plotting the two distributions back to back (Figure **??**). (This is a bit complicated and the code is included as an example.) But this figure does not reflect the standard error of the two measures.

Alternatively, (and probably better) we can do a boxplot and then add the standard errors to the data (Figure 5). This allows us to see how much we expect the groups to differ given their within group standard deviations and the sample size.

Now, we can do the t-test using the `t.test` function. The distribution of t depends upon the degrees of freedom. Figure 6 shows the .05 rejection region (.025 on the left tail, .025 on the right tail.))

```
> with(spelling, {t.test(Placebo,Drug)})

        Welch Two Sample t-test

data:  Placebo and Drug
t = -2.5273, df = 21.999, p-value = 0.01918
```

```
>  g1 <- spelling[,"Placebo"]
> g2 <- spelling[,"Drug"]
> t1 <- tabulate(g1-20)
> t2 <- tabulate(g2-20)
>  barplot(-t1,col = color[1],horiz=TRUE,xlim=c(-4,4),ylim=c(0,10),main="Counts from
+       Placebo and Drug conditions (-20)")
> barplot(t2,col = color[2],horiz=TRUE,add=TRUE)
>  axis(2)
```

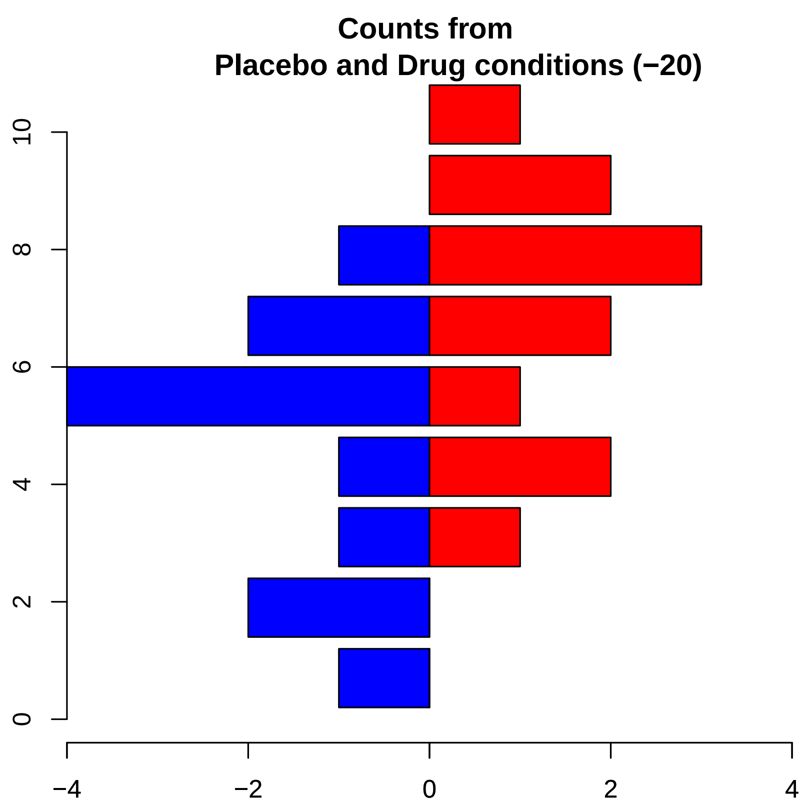

Figure 4: Compare the Placebo Condition (blue) with the Drug condition (red). At least to the eye, these appear different.

```
> boxplot(spelling,main="Spelling Performance as a function of drug")
> error.bars(spelling,add=TRUE)
```

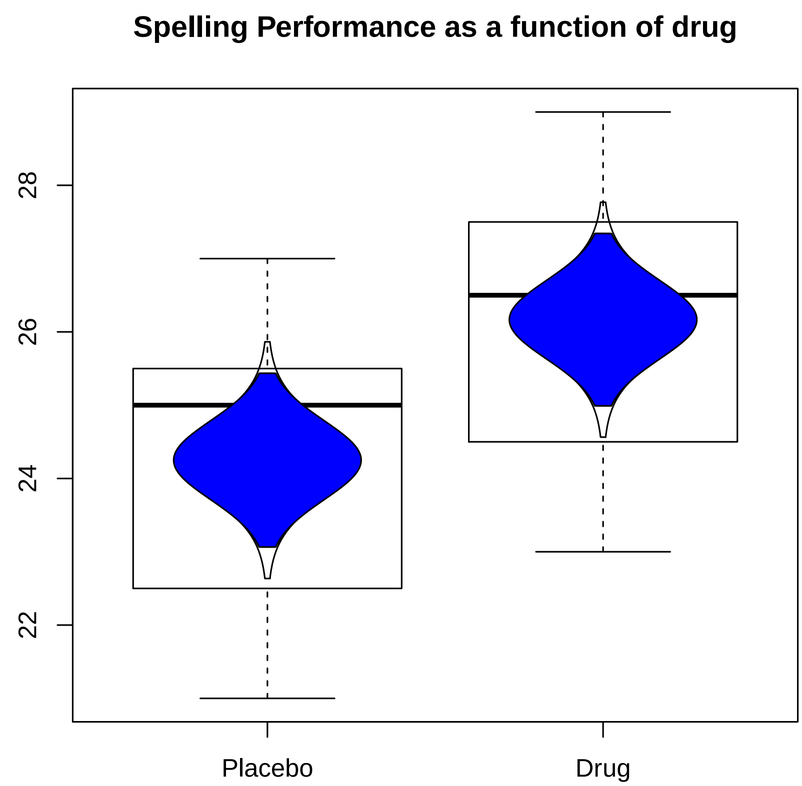

Figure 5: Spelling performance as a function of placebo and drug. Means and 95% confidence regions are shownin addition to the basic box plot. The boxplot shows the median, the upper and lower quartiles, and the "hinges" of the data.

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.4894368 -0.3438965
sample estimates:
mean of x mean of y
 24.25000  26.16667
```

2.4.1 ANOVA as a generalized t-test.

The t-test compares the difference between two means with respect to the standard error of the differences. Another test, developed by Ronald Fisher, is the Analysis of Variance (ANOVA). Here we are comparing an estimate of the population variance derived from the variance of the means to an estimate of the population variance derived from the variability within each group. For two groups, the variance estimate has 1 degree of freedom.

To do this, we need to reorganize the data so that we have one column of the dependent variable and another column showing the conditions. We do this with the `stack` function.

We use the `aov` function and then ask for the `summary` of the results. Compare the results of this analysis with the previous one. The F statistic for a 1 degree of freedom comparison (one between two groups) is the same as t^2. The probability of observing an F of this size or bigger is the same as observing the t of that size or larger (in absolute value.

```
> prob1 <- stack(spelling)
> summary(aov(values~ind,data=prob1))

            Df Sum Sq Mean Sq F value Pr(>F)
ind          1  22.04  22.042   6.387 0.0192 *
Residuals   22  75.92   3.451
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

2.4.2 Linear regression as a generalized ANOVA

Yet another way of thinking about this problem is to use linear regression. That is, if we estimate β in the linear regression equation:

$$\hat{y} = \beta x + e \tag{6}$$

and we use the `lm` (for linear model) function

```
> summary(lm(values~ind,data=prob1))

Call:
lm(formula = values ~ ind, data = prob1)

Residuals:
   Min     1Q Median     3Q    Max
-3.250 -1.479  0.750  1.062  2.833
```

```
> curve(dt(x,24),-3,3,xlab="t",ylab="probability of t",main="The t distribution")
> xvals <- seq(-2.07,2.07,length=50)
> dvals <- dnorm(xvals)
> polygon(c(xvals,rev(xvals)),c(rep(0,50),rev(dvals)),col="gray")
```

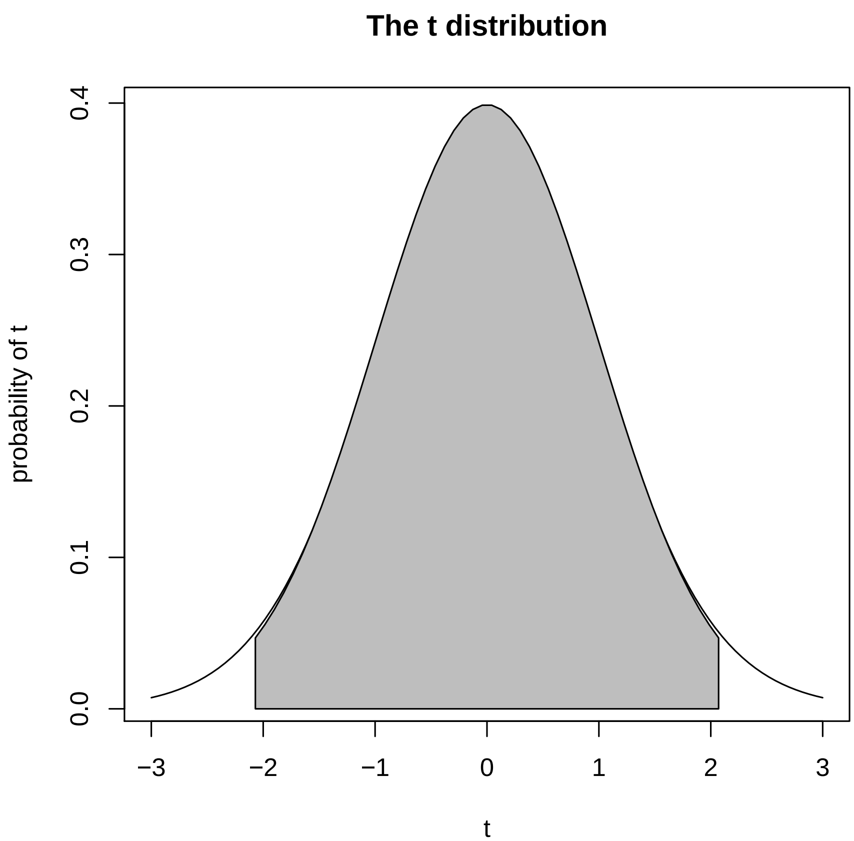

Figure 6: Finding area under the curve for for | t values | < may be done using the qnorm function. In this case, with df = 22, show the 5% rejection region. qt(.025,df=22) will yield the critical t value for the lower tail. qt(.975,df=22) for the upper region.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.1667     0.5362  48.796   <2e-16 ***
indPlacebo   -1.9167     0.7584  -2.527   0.0192 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 1.858 on 22 degrees of freedom
Multiple R-squared:  0.225,        Adjusted R-squared:  0.1898
F-statistic: 6.387 on 1 and 22 DF,  p-value: 0.01918
```

We find that the difference between the two IV conditions is 1.917 (this is the same as the difference between the means found in the t-test) and that the probability of this difference happening by chance if there were no difference is .0192. This is, of course, the same probability as that found by the t-test or the ANOVA.

3 Linear regression and correlation

Another investigator believes that introversion/extraversion has a linear relationship to spelling ability and reports the following data (Table 2). This can be solved by finding the linear regression of Spelling on Introversion or by finding the correlation between spelling and introversion. Do either one (or both).

Table 2: Does introversion predict spelling ability?

Introversion	Spelling
21	31
14	33
13	39
13	24
20	35
21	37
11	36
15	20
23	46
12	31
17	44
26	44

For this problem, we need to read in the data from the clipboard using the `read.clipboard` function and then can use the `cor` function to the find the correlation, or the `lm` function to find the linear regression, or use the `pairs.panels` function to find the correlation as well as to graph the data.

```
>int_spelling <- read.clipboard()
```

```
> round(cor(int_spelling),2)

             Introversion Spelling
Introversion         1.00     0.51
Spelling             0.51     1.00

> cor.test(int_spelling$Introversion,int_spelling$Spelling)

        Pearson's product-moment correlation

data:  int_spelling$Introversion and int_spelling$Spelling
t = 1.8761, df = 10, p-value = 0.0901
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.09002976  0.83857967
sample estimates:
      cor
0.5102348

> summary(lm(Spelling ~  Introversion,data=int_spelling))

Call:
lm(formula = Spelling ~ Introversion, data = int_spelling)

Residuals:
     Min       1Q   Median       3Q      Max
-13.2168  -3.5376   0.4292   6.1062   9.1372

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   20.8717     7.8064   2.674   0.0233 *
Introversion   0.8230     0.4387   1.876   0.0901 .
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 7.123 on 10 degrees of freedom
Multiple R-squared:  0.2603,       Adjusted R-squared:  0.1864
F-statistic:  3.52 on 1 and 10 DF,  p-value: 0.0901
```

4 Two way Analysis of Variance

Still another investigator believes that spelling performance is a function of the interaction of caffeine and time of day. She administors 0 or 200 mg of caffeine to subjects at 9 am and 9 pm. These data are typically examined using an Analysis of Variance (ANOVA), although a multiple regression using the general linear model would work as well. If the results are as below (Table 3), do the ANOVA.

We first read in the data (but without the labels for the columns) and then add colnames to the data

```
> pairs.panels(int_spelling)
```

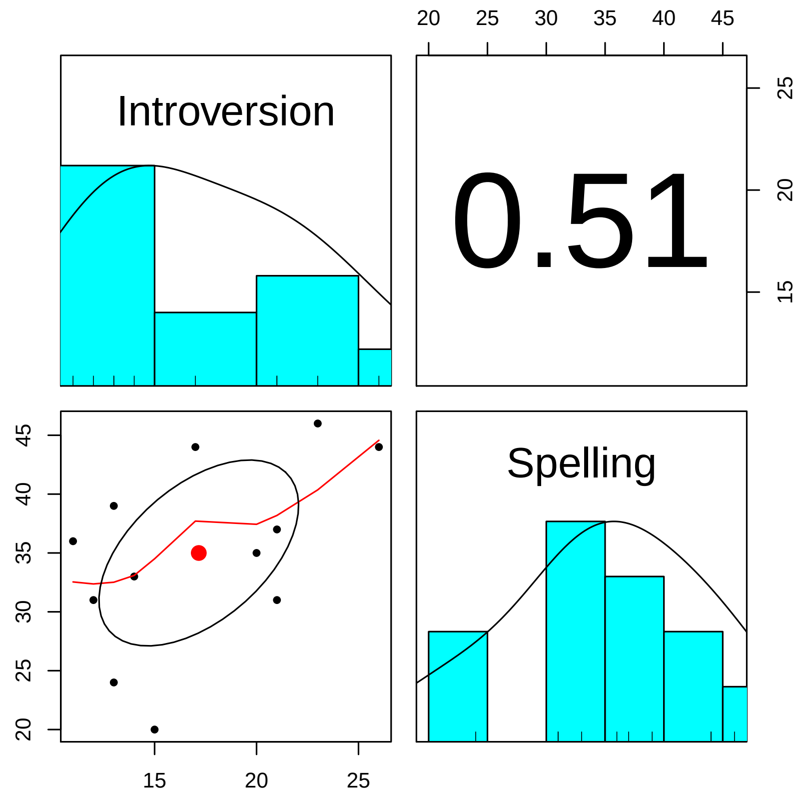

Figure 7: A Scatter Plot Matrix (splom) of the correlation between introversion and spelling

Table 3: Time of day, caffeine, and spelling performance

9am	9 am	9pm	9pm
0 mg	200 mg	0 mg	200 mg
26	27	28	24
27	30	27	23
25	28	25	25
22	32	25	21
27	25	31	23
23	29	32	21
21	31	25	25
28	28	32	21
21	28	26	26
23	26	25	22
20	29	27	23
23	31	26	26

```
>tod.data<- read.clipboard(header=FALSE)
```

Unfortunately, this analysis is a bit more complicated, because we need to string the data out and then add the conditions as additional variables. This will be discussed in more detail in subsequent handouts.

```
> colnames(tod.data) <- c("AP","AC","PP","PC")
> tod.stacked <- stack(tod.data)
> tod.df <- data.frame(spelling = tod.stacked$values,drug = rep(c(rep("P",12),rep("C",12)),2),time=c
> anova(lm(spelling~drug*time,data=tod.df))
```

```
Analysis of Variance Table

Response: spelling
          Df  Sum Sq Mean Sq F value    Pr(>F)
drug       1   1.688   1.688  0.2971    0.5885
time       1   9.187   9.187  1.6175    0.2101
drug:time  1 238.521 238.521 41.9937 6.633e-08 ***
Residuals 44 249.917   5.680
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

A more generic way of doing this analysis is as follows:

```
>raw.data <- read.clipboard(header=FALSE)
```

```
> nsub <- c(12,12)
> IV1.names <- c("Placebo","Caffeine")
```

```
> IV2.names <- c("AM","PM")
> nvar=2
> drug <- rep(rep(IV1.names,nsub),nvar)
> time <- rep(rep(IV2.names,nsub),nvar)
> data.df <- data.frame(stack(raw.data)$value,drug = drug,time=time)
> summary(aov(spelling~drug*time,data=tod.df))
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
drug         1   1.69    1.69   0.297    0.588    
time         1   9.19    9.19   1.618    0.210    
drug:time    1 238.52  238.52  41.994 6.63e-08 ***
Residuals   44 249.92    5.68                     
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

Table 4: A truncated version of the time of day data.frame.

```
   stack.raw.data..value     drug time
1                     26  Placebo   AM
2                     27  Placebo   AM
3                     25  Placebo   AM
4                     22  Placebo   AM
10                    23  Placebo   AM
11                    20  Placebo   AM
12                    23  Placebo   AM
13                    27 Caffeine   PM
14                    30 Caffeine   PM
34                    25  Placebo   AM
35                    27  Placebo   AM
36                    26  Placebo   AM
37                    24 Caffeine   PM
38                    23 Caffeine   PM
44                    21 Caffeine   PM
45                    26 Caffeine   PM
46                    22 Caffeine   PM
47                    23 Caffeine   PM
48                    26 Caffeine   PM
```

5 Chi Square tests of independence

Another experimenter wants to test the hypothesis that gender is related to interest in football. 100 subjects (50 male and 50 female) are asked whether or not they watched a recent football game. The results are in Table 5 The question of whether a relationship between two dichotomous variables is larger than chance is typically done by using a χ^2 test. Find the χ^2 to determine if there is a relationship between gender and watching the football game.

Table 5: Gender differences in football interest

	Watched	Did not watch
Male	30	20
Female	20	30

This is a question of the association between two categorical variables. We are given the counts and we can enter them into a matrix and run the χ^2 test directly. We have an option of correcting for continuity. We turn off this correction for consistency with the hand done version.

```
> football <- matrix(c(30,20,20,30),ncol=2)

     [,1] [,2]
[1,]   30   20
[2,]   20   30

> chisq.test(football,correct=FALSE)

        Pearson's Chi-squared test

data:  football
X-squared = 4, df = 1, p-value = 0.0455
```

6 Correlated and uncorrelated t-tests

A professor believes that taking statistics increases one's ability to reason analytically. To test this hypothesis, she develops a test of reasoning and gives it to two sets of students. Those who have just started a statistics course and those who have just finished a statistics course. The results are shown in Table 6

6.1 Uncorrelated t-tests

These data could be analyzed by using t-test (or by doing an ANOVA). Notice that this design is normally not as powerful as doing a pre-post within subjects design.

First, copy the data to a spreadsheet (with no extra lines) and then copy that to the clipboard. We then read the clipboard into R.

```
reasoning   <- read.clipboard()

>  t.test(reasoning$before,reasoning$after,equal.var=TRUE)

        Welch Two Sample t-test
```

Table 6: The effect of taking a statistics course on reasoning analytically.

before	after
12	15
11	23
15	17
14	22
11	18
10	17
11	21
12	21
18	16
17	17
13	23
16	18

```
data:  reasoning$before and reasoning$after
t = -5.0735, df = 21.896, p-value = 4.47e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.983620 -3.349713
sample estimates:
mean of x mean of y
 13.33333  19.00000
```

6.2 Correlated t-tests

Another professor has the same hypothesis, but decides to use a pre-post design. That is, each student takes the reasoning test twice, once before and once after the class. The data can now be analyzed by using a t-test for correlated scores, or a t-test comparing the difference scores to 0.

```
>  t.test(reasoning$before,reasoning$after,equal.var=TRUE,paired=TRUE)

        Paired t-test

data:  reasoning$before and reasoning$after
t = -4.363, df = 11, p-value = 0.001131
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.525295 -2.808038
sample estimates:
mean of the differences
              -5.666667
```

When examining these results, we notice that the assumption of independence between the pre and post scores yields a larger t value than when we allow them to be correlated. Examining this more closely, we discover that the correlation between the pre and post scores is actually negative!

```
> round(cor(reasoning),2)

       before after
before   1.00 -0.35
after   -0.35  1.00
```

If the numbers are the same as in problem 6, what test should be applied?

There are advantages and disadvantages of the designs used in questions 6 and 6.2. What are some of them?

7 Using the normal distribution

If a test is normally distributed and has a mean of 100 and a standard deviation of 15, then what percentage of students would you expect to have scores of 100 or greater?

Convert the observed score (in this 100) to a standard score by subtracting the mean and dividing the by the standard deviation:

$$z_x = (X - \bar{X})/s_x \tag{7}$$

Thus, z_x = (100-100) / 15 = 0.0. Then, using the `pnorm` function (probability of a normal) we find that
pnorm(0) = 0.5

This, of course, requires knowing how to think about the normal distribution. This one should be easy, the next one is also fairly easy.

With the same assumptions, what percentage of students would you expect to have scores greater than 115? That is to say, the number of people to the right of the value.

z_x = (115-100)/ 15 = 1
1- pnorm(1) = .16

8 The binomial distribution

If you flip a fair coin 10 times, how often would you expect to observe at least 8 heads?

This requires thinking about the binomial distribution and using the `dbinom` to help us. We create a vector, x, with 11 values, find the binomial probabilties of each value of x, and add them up for the cases of 8, 9, and 10. To better understand where these probabilities are coming from, we can multiply them by 1024 (2^{10}) to get the number of outcomes out of the 1024 different outcomes that match what we want:

```
> x  <-  0:10

 [1]  0  1  2  3  4  5  6  7  8  9 10
```

```
> curve(dnorm(x),-3,3,xlab="z = standard score units",ylab="probability of z",
+            main="The normal curve")
> xvals <- seq(-3,1,length=50)
> dvals <- dnorm(xvals)
> polygon(c(xvals,rev(xvals)),c(rep(0,50),rev(dvals)),col="gray")
```

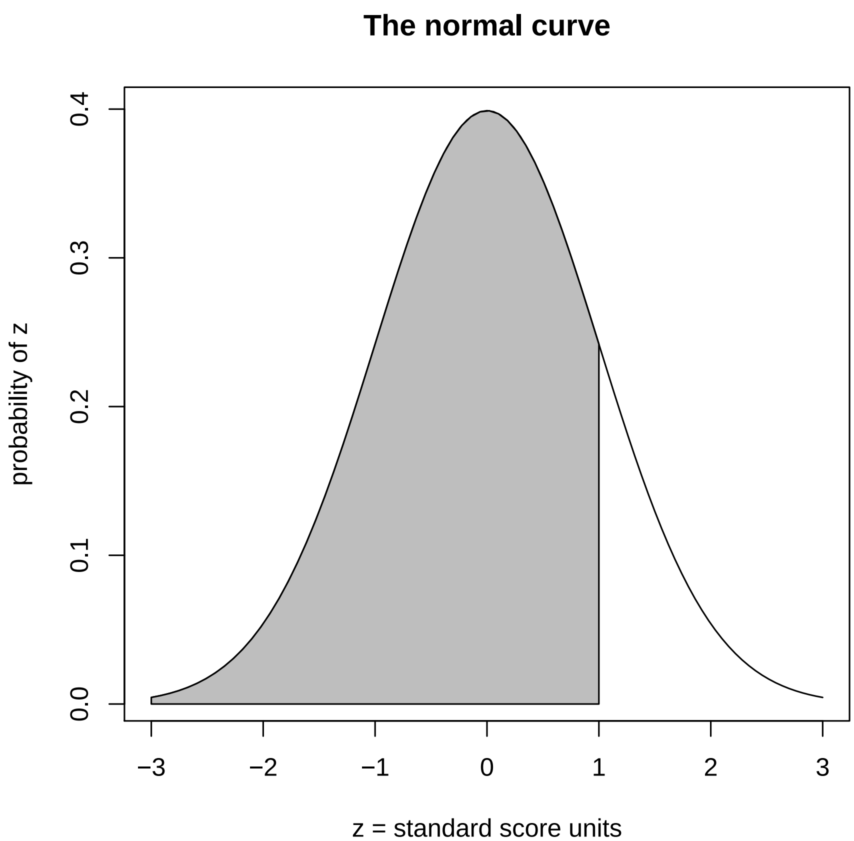

Figure 8: Finding area under the normal curve for for values < z may be done using the pnorm function. In this case, z = 1 and we want to find the shaded area. pnorm(1) will yield the area to the left of 1.

```
> round(dbinom(x,10,.5),3)

 [1] 0.001 0.010 0.044 0.117 0.205 0.246 0.205 0.117 0.044 0.010 0.001

> dbinom(x,10,.5) * 1024

 [1]   1  10  45 120 210 252 210 120  45  10   1

>  (1 + 10 + 45)/1024

[1] 0.0546875
```

Thus, the answer to our question of getting at least 8 is (1 + 10 + 45)/1024 or .0547
We can plot the binomial distribution using the `plot` function.

```
> barplot(dbinom(x,10,.5),main="The binomial distribution")
```

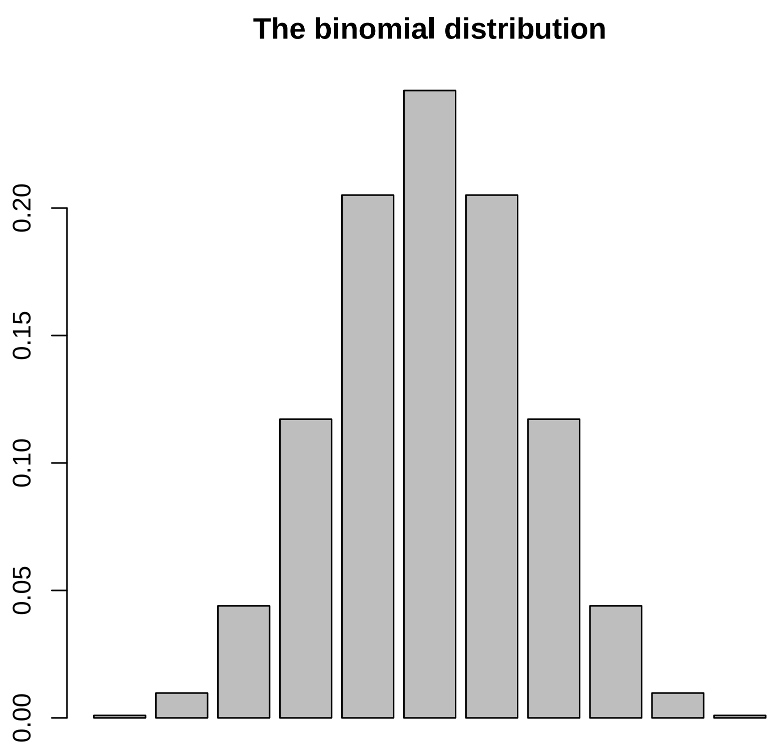

Figure 9: The binomial distribution for 10 coins (equal probability of a coin flip)