Stat 301 – Case Studies

Example 2: The Biggest Loser
Dansinger, Griffith, Gleason, et al. (2005) report on a randomized, comparative experiment in which 160 subjects were randomly assigned to one of four diet plans: Atkins, Ornish, Weight Watchers, and Zone (40 subjects per diet). These subjects were recruited through newspaper and television advertisements in the greater Boston area; all were overweight or obese with body mass index values between 27 and 52. Among the variables measured were

- Which diet the subject was assigned to
- Whether or not the subject completed the 12-month study (0 = yes)
- The subject's weight loss after 2 months, 6 months, and 12-months (in kilograms, with a negative value indicating weight gain)
- The degree to which the subject adhered to the assigned diet, taken as the average of 12 monthly ratings, each on a 1-10 scale (with 1 indicating complete nonadherence and 10 indicating full adherence).

We will consider only the 80 subjects who were assigned to either the Atkins or Weight Watchers diets.

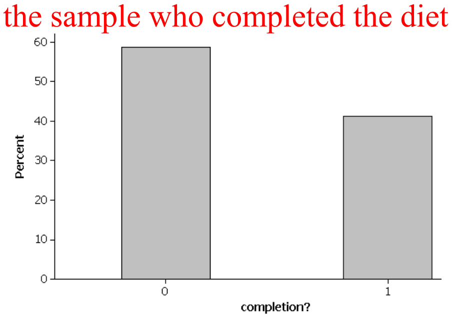
For each of the following research questions,

- Identify and classify the relevant variables
- Indicate which graphical display(s) would be appropriate
- Indicate which numerical summaries would be appropriate
- Specify an appropriate inference procedure to be used
- State the hypotheses to be tested (if appropriate)
- Comment on how to check the technical conditions of that procedure

(a) Did a statistically significant majority of subjects complete the 12 month study?

Variable: whether or not complete the 12 month study (categorical), so we are looking for one proportion

A bar graph such as the following would be appropriate and we would report $\hat{p}$, the proportion in the sample who completed the diet



$\hat{p}$ = 47/80 = .5875

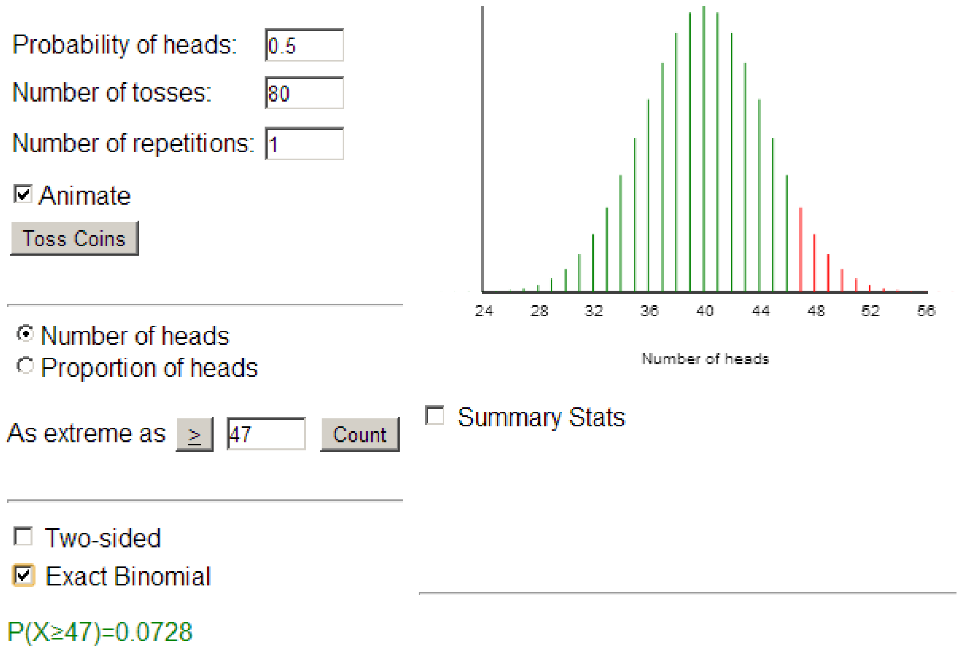
In this sample, a slight majority completed the study.

Let π represent the proportion of all such subjects who would complete the study in the population of dieters

H_0: π = .5   Ha: π > .5

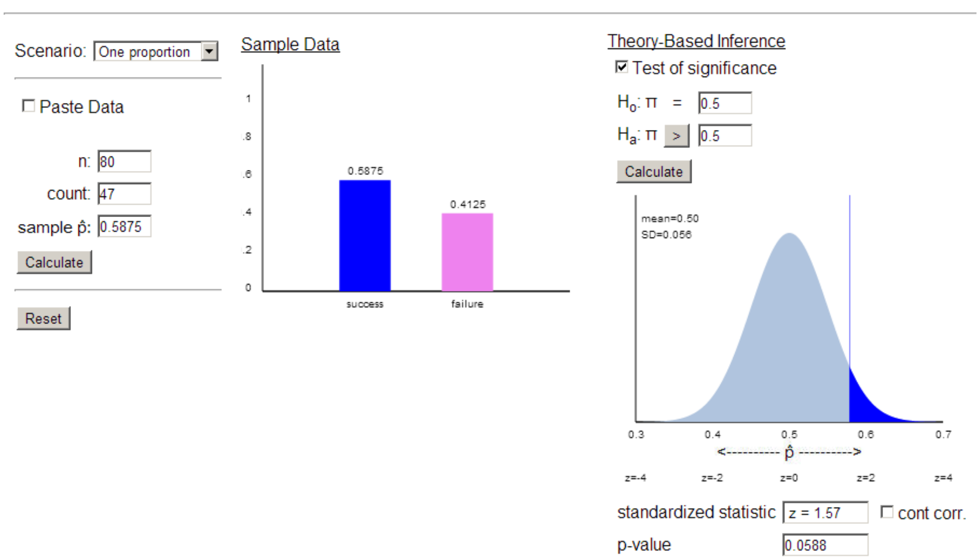
Binomial test

```
> iscambinomprob(47, 80, .5, lower.tail=FALSE)
Probability 47 and above = 0.07281773
```



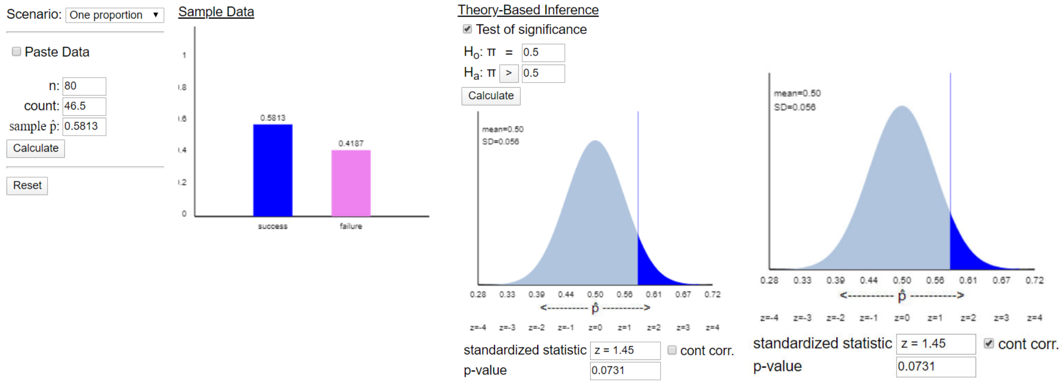
One sample z test (should be valid because 80(.5) = 40 >10 if we use the hypothesized probability here)

Theory-Based Inference




```
> iscamonepropztest(47, 80, .5, alternative="greater")

One Proportion z test

Data: observed successes = 47, sample size = 80, sample proportion = 0.5875

Null hypothesis        : pi = 0.5
Alternative hypothesis: pi > 0.5
z-statistic: 1.57
p-value: 0.05876
```

Better yet, use a continuity correction to improve the normal distribution's estimate of the binomial probability. This is equivalent to using 46.5 (and above).



This gives us a p-value of .0731, much closer to the exact binomial of .0728.

So there is weak evidence (p-value above .05 but less than .10) that most people will complete the study for whatever population this is representative of.

(b) Estimate the probability of a subject completing the 12-month study based on these data.
Even though this just says "estimate" – if you are giving an estimate for a population parameter, convey the precision and reliability of your estimate through a confidence interval!

Binomial interval

Confidence Interval

Confidence Level 0.95

Result	Value
Estimated Proportion	0.5875
Lower Limit	0.47185
Upper Limit	0.69648

```
> iscambinomtest(47, 80, conf.level=95)

Exact Binomial Test

Data: observed successes = 47, sample size = 80, sample proportion = 0.5875

95 % Confidence interval for pi: ( 0.47185 , 0.69648 )
```

One proportion z-interval for π: Technical conditions as in (a) (and have at least 10 successes and failures)

Confidence Interval

Confidence Level 0.95

Result	Value
Estimated Proportion	0.5875
Lower Limit	0.47963
Upper Limit	0.69537

```
> iscamonepropztest(47, 80, conf.level=95)

One Proportion z test

Data: observed successes = 47, sample size = 80, sample proportion = 0.5875

95 % Confidence interval for pi: ( 0.4796254 ,  0.6953746 )
```

Adjusted Wald (never a bad idea), using 49 successes out of 84 attempts

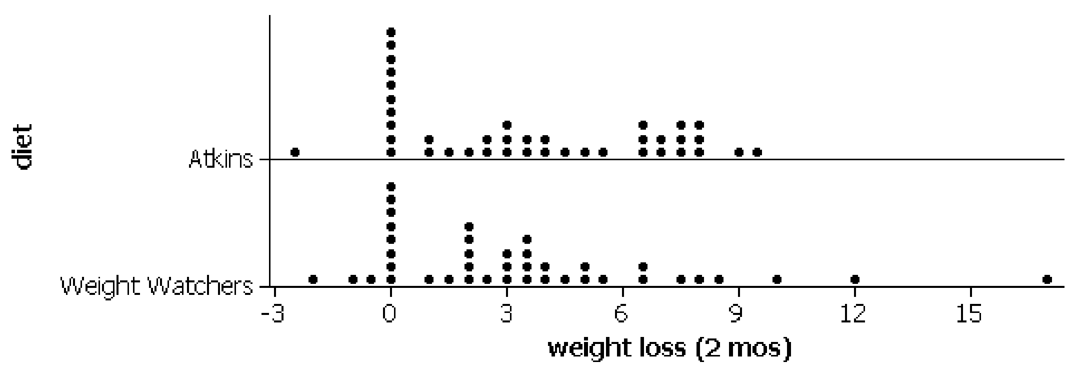
Result	Value
Estimated Proportion	0.58333
Lower Limit	0.4779
Upper Limit	0.68876

Notice this is narrower than the Binomial interval but pulled down a bit (closer to .5) from the Wald interval.

We are 95% confident that between 48% and 69% of the population will complete the program.

(c) Is there a statistically significant difference in the amount of weight lost between the two diets after 2 months?

Variables: diet (categorical) and weight loss after 2 months (quantitative), so we want to compare 2 means. Comparative graphs like dotplots or boxplots such as below would be appropriate, with means/medians and standard deviation/IQR as numerical summaries.



Both distributions are skewed to the right, including a large outlier (around 17 kilos) in the Weight Watchers group. The WW group averaged slightly more weight loss (3.465 kilos vs. 3.627 kilos) and had a bit more variability (std dev 3.83 kilos vs. 3.26 kilos). The difference between the groups does not appear substantial.

Let $\mu_A - \mu_{WW}$ represent the underlying difference in "population" means between Atkins diet population and Weight Watchers diet population.

H_0: $\mu_A - \mu_{WW} = 0$ (no treatment difference)

H_a: $\mu_A - \mu_{WW} \neq 0$ (one of the diets leads to more weight loss on average)

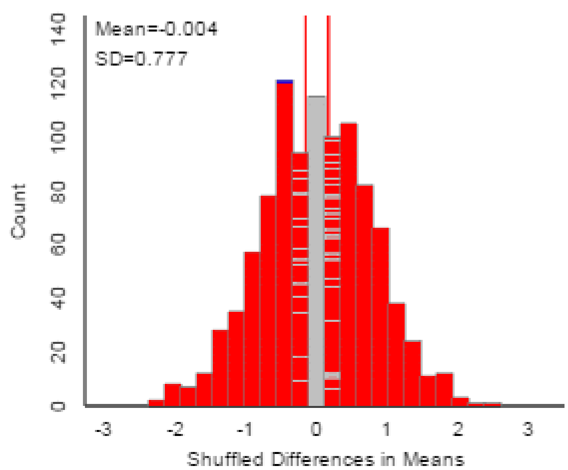
We have 40 people in each diet, so technically this passes the condition for a two-sample t-test. Subjects were randomly assigned to the diets so we will be willing to draw and effect conclusions at the end.

Test Results

Result	Value
Observed Difference (Mean 2 - Mean 1)	-0.1625
t-score	-0.2043
t critical values	+/- 1.99
Observed Significance (p-value)	0.8387

```
> t.test(WeightLoss2mos~diet, alternative="two.sided", conf.level=.95,
var.equal=FALSE)

 Welch Two Sample t-test

data:  WeightLoss2mos by diet
t = 0.2045, df = 76.019, p-value = 0.8385
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.420305  1.745305
sample estimates:
        mean in group Atkins mean in group WeightWatchers
                      3.6275                       3.4650
```

Total Shuffles = 1000



Count Samples Beyond .162 Count

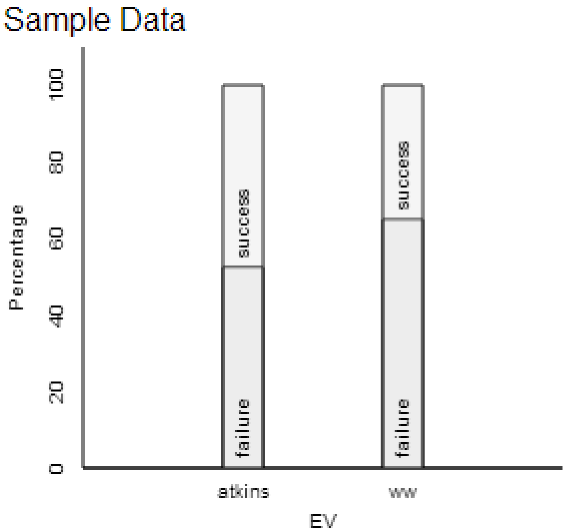
Count = 848/1000 (0.8480)

We have a large p-value (.839 > .05) and no reason to believe that after two months the diets differ with respect to average weight loss in the population.

(d) Is there a statistically significant difference in the completion rate between the two diets?

Variables: whether or not completed the study (categorical) and which diet (categorical), so want to compare two proportions.

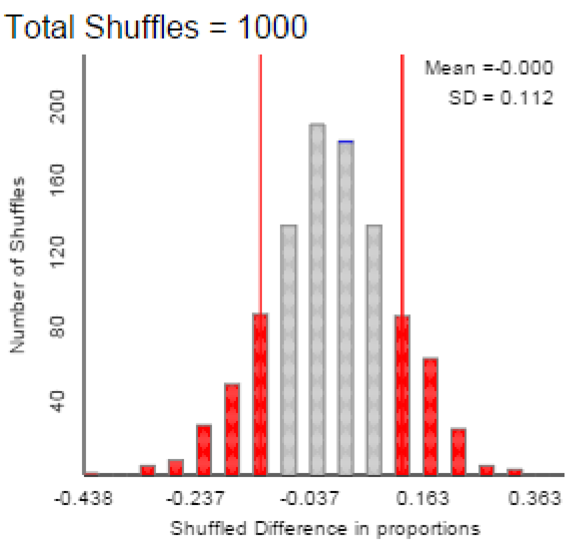
Appropriate graph is a segmented bar graph as below. Could look at conditional proportions (complete rates) or even relative risk and odds ratio.



	atkins	ww
success	19	14
failure	21	26
Prop	0.475	0.350

We have $\hat{p}_1 = 19/40 = .475$ and $\hat{p}_2 = 14/40 = .35$, indicating higher completion rates with the Atkins diet, but the difference does not seem large.

Simulation (Two-way table simulation)



Show previous

Count Samples Beyond .125

Count

Count= 359/1000 (0.3590)

Fisher's Exact Test

Show Fisher's Exact Test ☑

☑ two-sided

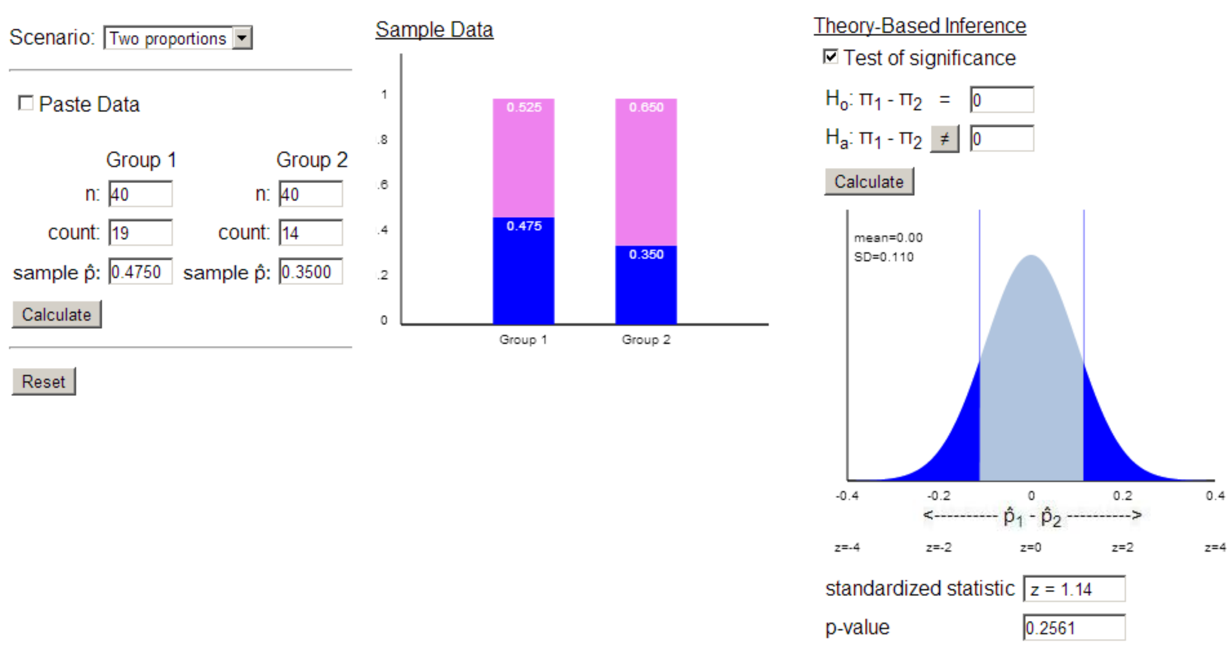
P(X ≤14 or X≥19) = 0.3638

```
> fisher.test(t(matrix(c(19,  21, 14, 26), nrow=2)))

 Fisher's Exact Test for Count Data

data:  t(matrix(c(19, 21, 14, 26), nrow = 2))
p-value = 0.3638
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.6245416 4.5526569
sample estimates:
odds ratio
  1.669333
```

Because we have at least 5 successes and at least 5 failures with each diet, and this was a randomized experiment, we can apply the two-sample z-test.

H_0: $\pi_{atkins} - \pi_{ww} = 0$

H_a: $\pi_{atkins} - \pi_{ww} \neq 0$ (one of the diets leads to a higher completion probability)

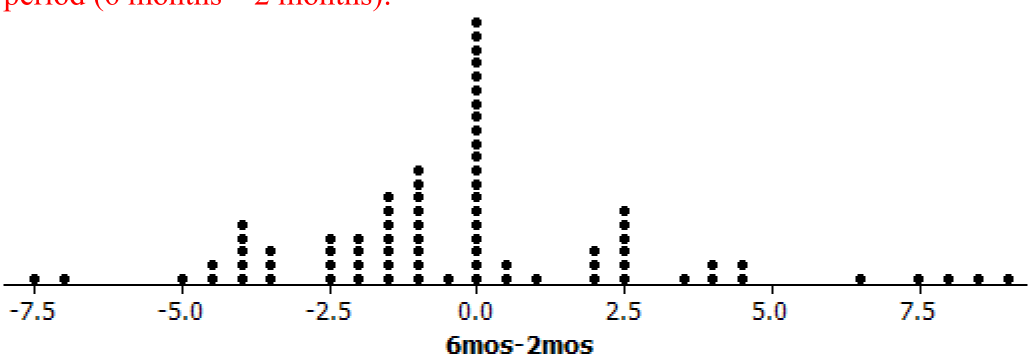


```
> iscamtwopropztest(19, 40, 14, 40, hypothesized=0, alternative = "two.sided",
conf.level=.95)

Two Proportion z test

Group1: observed successes = 19, sample size = 40, sample proportion = 0.475

Group2: observed successes = 14, sample size = 40, sample proportion = 0.35

Null hypothesis       : pi1-pi2 = 0
Alternative hypothesis: pi1-pi2 <> 0
z-statistic: 1.14
95 % Confidence interval for pi1-pi2: ( -0.08900325 ,  0.3390033 )
p-value: 0.2543
```

(The z-test p-value is noticeably lower than the simulation or Fisher's Exact Test p-value, so we will not use it. Or we could do a similar type of continuity correction.)

With the large p-value (.36 > .05) we fail to reject H_0. We do not have convincing evidence that one of the diets leads to a higher completion probability.

(e) Is there statistically significant evidence that the weight loss after 6 months tends to be smaller than the weight loss after 2 months?
Variable = difference in amount of weight loss
Because we have measured the same individuals at 2 months and at 6 months, we wanted to do a paired t test. We can create a new (quantitative) variable measure the additional weight lost in this four month period (6 months – 2 months):



There is some interesting clustering in this distribution. The mean (-.172 kilos) is a bit misleading and the standard deviation is large (3.179 kilos). With such a large sample size, we can still apply the one-sample t-test because we don't have severe skewness or outliers. The large spike at zero is probably due to the people who have already dropped out of the study.

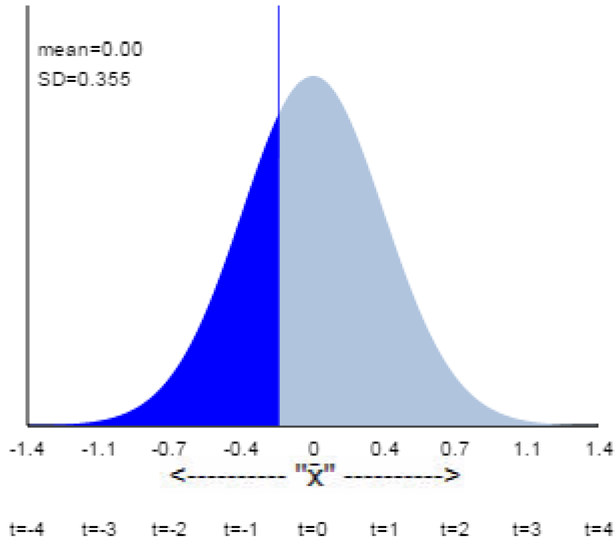
Let μ represent the average additional weight lost between 2 and 6 months by the dieter population.
H_0: $\mu = 0$ (on average, no change in weight change in this time period)
H_a: $\mu < 0$ (tend to lose more weight after 2 months compared to 6 months)

Paired t-test (sample size is above 30 so technically large enough)

H_0: μ = 0

H_a: μ < 0

Calculate



standardized statistic t = -0.48 df = 79

p-value 0.3149

weightloss(2mos)	3.54625	t-Ratio	0.485367
weightloss(6mos)	3.37375	DF	79
Mean Difference	0.1725	Prob > |t|	0.6288
Std Error	0.3554	Prob > t	0.3144
Upper 95%	0.87991	Prob < t	0.6856
Lower 95%	-0.5349		
N	80		

```
data:  Dataset$differences
t = -0.4854, df = 79, p-value = 0.3144
alternative hypothesis: true mean is less than 0
```

With the large p-value, we fail to reject H_0 and conclude that there is not a genuine decrease in weight loss, on average, between 2 and 6 months in this dieter population.

(f) What if the previous question had been: "Is there evidence that a majority of such dieters in the population would have lost less weight after 6 months than after 2 months?"

Now we could just count how many people had lost more weight after 2 months than after 6 months or how many are greater than or equal to zero – be sure to clarify how you define the variable – and then consider inference for one proportion as in (a).

Sign Test

Descriptive Statistics

Sample	N	Median
C14	80	-0.05

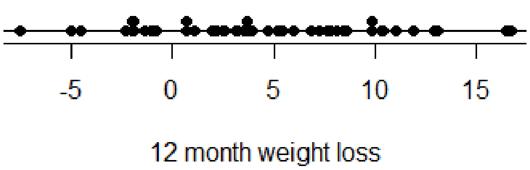
Test

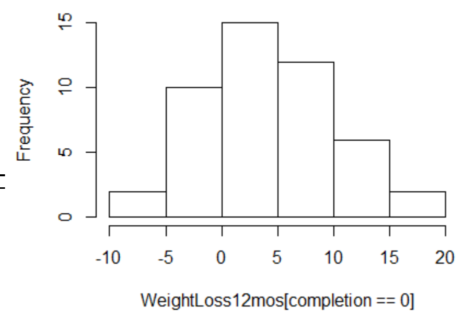
Null hypothesis H_0: $\eta = 0$
Alternative hypothesis H_1: $\eta < 0$

Sample	Number < 0	Number = 0	Number > 0	P-Value
C14	40	16	24	0.030

(g) Estimate the mean amount of weight loss <u>by all participants who complete</u> such a program after 12 months.

Variable = weight loss (after 12 months), quantitative, so we want to examine one mean. But I went ahead and deleted everyone who didn't complete the diet, leaving 47 subjects.






```
     n    Min      Q1 Median     Q3    Max   Mean     SD
47.000 -7.500 -0.050  3.700  8.250 16.700  4.291  5.640
```

On average, subjects the completed the program lost 4.291 kilos with standard deviation 5.64 kilos. The distribution is fairly symmetric, perhaps skewed to the right.

Because we have more than 30 subject, we can consider a one-sample t-interval. (Bootstrapping would be another option here.)

One sample t-interval

confidence level 95 %

(2.6350, 5.9470)

df = 46

We are 95% confident that dieters that stay on the program lose an average of 2.64 to 5.95 kilos.

(h) What can you say about generalizability and causation in this study?

For any of the research questions that involved comparing the two diets and we found a statistically significant difference, we can draw a cause and effect conclusion because the dieters were randomly assigned to the two diets. So question (c).

Generalizability might be limited because the subjects were volunteers and not a random sample from a larger population of dieters. *At most* we can say these results represent overweight or obese adults in the greater Boston area that are willing to respond to newspaper and television advertisements.

(i) Is there a statistically significance difference in the mean weight loss after 12 months among the four diets?

We don't know how to compare 4 diets at once! This is something you will learn about in Stat 302 ☺
But based on what you have learned in 301, you could design a randomization test that would randomly divide the observed weight loss values into 4 groups. Then you need some kind of statistic to measure the differences in the means across the four groups with a single number. Create a null distribution for this statistic and see how often the observed statistic or more extreme occurs by chance (re-randomizing the observations to the four groups) alone!