Answers

1. One obvious way to consider these data is by testing if the true average on time rate was different in the two years using a paired t–test. The hypotheses being tested are

$$H_0 : \mu_{1995} = \mu_{1994}$$

versus

$$H_a : \mu_{1995} \neq \mu_{1994}.$$

Here is output from the t–test, which indicates that there is no significant difference in average performance in the two years.

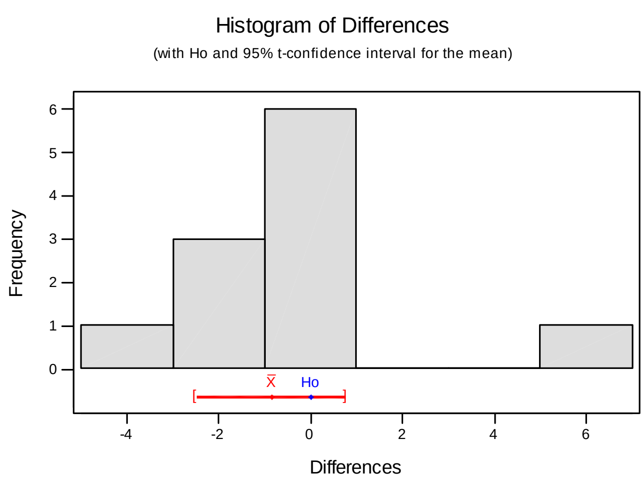
```
Paired T-Test and Confidence Interval

Paired T for May 1995 - May 1994

                  N      Mean     StDev   SE Mean
May 1995         11     91.44      3.08      0.93
May 1994         11     92.29      4.12      1.24
Difference       11    -0.855     2.393     0.722

95% CI for mean difference: (-2.462, 0.753)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.18  P-Value = 0.264
```

Unfortunately, the histogram of differences in on time performance given below shows that there is a problem. While the t–test assumes that there the differences in on time performance are a random sample from a normal distribution, there is a clear outlier in these data (Montauk, where the on time performance went up more than 5 percentage points from 1994 to 1995).



For this reason, we cannot trust the t–test. We could change our hypotheses to reflect a focus on medians:

$$H_0 : \text{Median}_{1995} = \text{Median}_{1994}$$

versus

$$H_a : \text{Median}_{1995} \neq \text{Median}_{1994}.$$

The sign test and signed rank test are then appropriate to test these hypotheses:

```
Sign Test for Median

Sign test of median = 0.00000 versus  not =  0.00000

                 N  Below  Equal  Above         P     Median
Change 9        11     10      0      1    0.0117    -0.5000



Wilcoxon Signed Rank Test

Test of median = 0.000000 versus median not = 0.000000

                 N for   Wilcoxon            Estimated
              N   Test  Statistic        P      Median
Change 9     11     11       11.0    0.056      -1.125
```

Both tests indicate moderate to strong evidence of a significant difference in performance from May 1994 to May 1995 (the signed rank test gives weaker evidence than the sign test, but since the distribution of 1994 to 1995 changes is not very symmetric, we might trust the stronger evidence of the sign test a little more). Unfortunately, it's in the wrong direction. The sign test is particularly important here. Not only was Montauk unusual in its large positive change from 1994 to 1995, it was the **only** branch where the performance didn't get worse from 1994 to 1995! You need to tell your boss that the "high tech" results say that LIRR performance has gotten worse — probably not what she wants to hear. Oh well — you didn't really want to work at the MTA anyway. Note, by the way, that we are treating these branch numbers as a random sample from a continuing process of on time figures, which might be reasonable for spring/summer months, but probably not for winter months (when on time performance is more dependent on the weather).

2. The prior beliefs are represented by the following hypotheses:

$$H_0 : \mu = 2$$

versus

$$H_a : \mu \neq 2,$$

where μ is the average serving time (an average serving time equaling 2 minutes corresponds to 30 customers per hour). We have that $\overline{X} = 3$ and $s = 1.5$, with $n = 20$. Thus, the t–test of interest is

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{3 - 2}{1.5/\sqrt{20}} = 2.98,$$

which is compared to a t–distribution on 19 degrees of freedom. This is highly statistically significant ($p = .0076$). Thus, the serving time is significantly higher than the hypothesized 2 minutes per person. We are assuming that the observations can be viewed as a random sample from a normal distribution.

3. In these data, $(.55)(.61) = 33.6\%$ of 500 respondents (i.e., 168) are having sex and using birth control every time. The question now is how to set the null and alternative hypotheses, since it's not clear what claim should "get the benefit of the doubt." I'll set them up this way:

$$H_0 : p \leq .3$$

versus

$$H_a : p > .3,$$

putting the burden of proof on the televangelist (note that even if you reversed the hypotheses, this is definitely a situation where a one–sided hypothesis is appropriate). The z–test is

$$z = \frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{.336 - .3}{\sqrt{(.3)(.7)/500}} = 1.757,$$

which has a (one–sided) tail probability of .0395, indicating weak support for the televangelist's claim. The exact test gives similar results:

```
Test and Confidence Interval for One Proportion

Test of p = 0.3 vs p > 0.3

                                                      Exact
Sample       X       N  Sample p         95.0 % CI       P-Value
1          168     500  0.336000  (0.294674, 0.379279)    0.045
```

4. One obvious way to consider these data is by testing if the true average number of attorneys was different in the two years using a paired t–test. The hypotheses being tested are

$$H_0 : \mu_{1988} = \mu_{1991}$$

versus

$$H_a : \mu_{1988} \neq \mu_{1991}.$$

Here is output from the t–test, which indicates that there is a significant increase in average number of attorneys from 1988 to 1991.

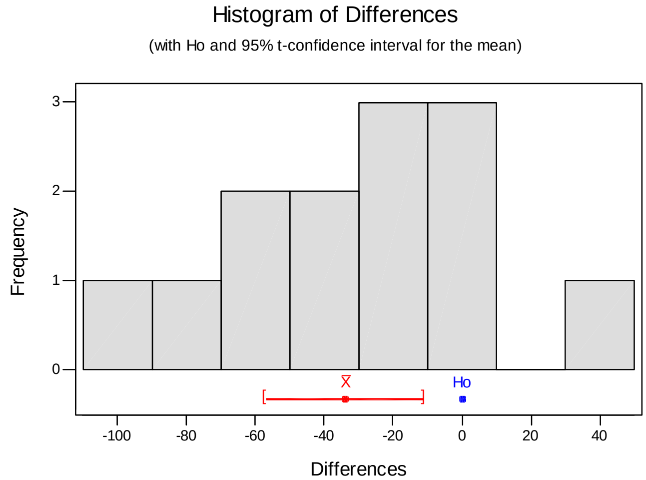
```
Paired T-Test and Confidence Interval

Paired T for Year 1988 - Year 1991

                  N       Mean     StDev   SE Mean
Year 198         13      282.0      92.4      25.6
Year 199         13      315.8     111.6      30.9
Difference       13      -33.8      37.8      10.5

95% CI for mean difference: (-56.7, -11.0)
T-Test of mean difference = 0 (vs not = 0): T-Value = -3.23  P-Value = 0.007
```

The histogram of differences in number of attorneys given below shows that the difference is somewhat right–tailed, casting some doubt on the normality assumption for this variable.



For this reason, we could consider changing our hypotheses to reflect a focus on medians:

$$H_0 : \text{Median}_{1988} = \text{Median}_{1991}$$

versus

$$H_a : \text{Median}_{1988} \neq \text{Median}_{1991}.$$

The sign test and signed rank test are then appropriate to test these hypotheses:

```
Sign Test for Median

Sign test of median = 0.00000 versus  not =  0.00000

               N  Below  Equal  Above         P     Median
Change 1      13     11      0      2    0.0225     -28.00



Wilcoxon Signed Rank Test

Test of median = 0.000000 versus median not = 0.000000

                 N for   Wilcoxon            Estimated
             N    Test  Statistic        P      Median
Change 1    13      13        9.5    0.013      -34.00
```

Fortunately, the implications are the same, as the median number of attorneys is significantly higher in 1991 than in 1988.

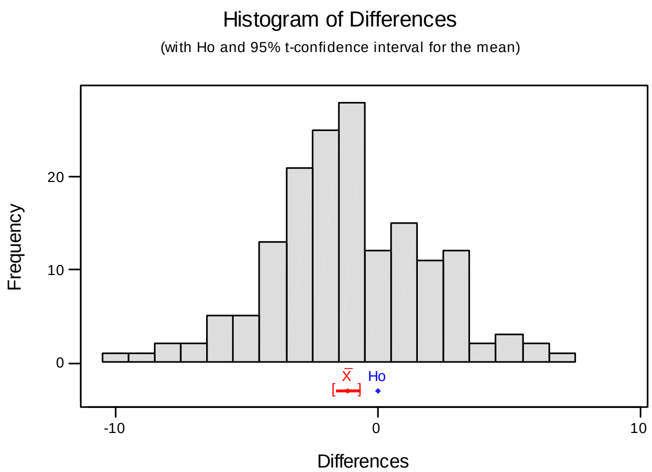
5.(a) Let μ_D and μ_S be the average Zagat ratings for décor and service, respectively. The hypotheses being tested here are

$$H_0 : \mu_D = \mu_S$$

versus

$$H_a : \mu_D \neq \mu_S.$$

The histogram below shows that the differences between décor and service rating are reasonably normally distributed, so a paired sample t–test is reasonable here.



Here is the output from the test:

```
Paired T-Test and Confidence Interval

Paired T for dcor rating - service rating

                  N      Mean     StDev   SE Mean
dcor ra         161    16.447    4.011     0.316
service          161    17.578    2.719     0.214
Difference       161    -1.130    2.986     0.235

95% CI for mean difference: (-1.595, -0.666)
T-Test of mean difference = 0 (vs not = 0): T-Value = -4.80  P-Value = 0.000
```

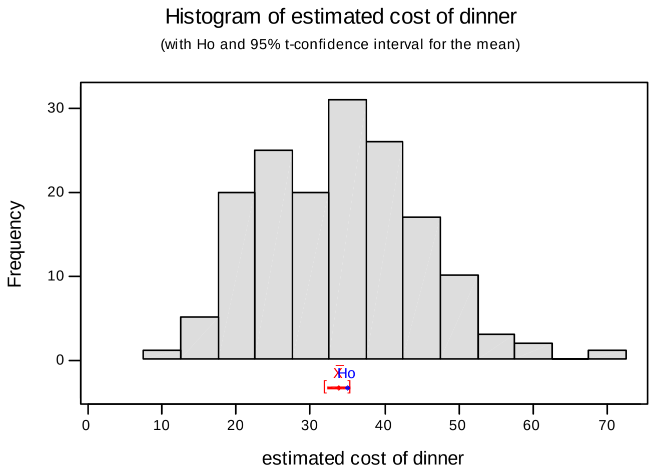
Obviously we strongly reject the null hypothesis here. Service ratings are, on average, 1.13 points higher than décor ratings, and this difference s statistically significant at any reasonable α level.

(b) Let μ_c be the average cost of a dinner in Manhattan. A histogram of the cost variable looks reasonably Gaussian, so we will use a t–test to test the hypotheses

$$H_0 : \mu_C = 35$$

versus

$$H_a : \mu_C \neq 35.$$



Here is output for the test:

```
T-Test of the Mean

Test of mu = 35.000 vs mu not = 35.000

Variable     N      Mean    StDev   SE Mean         T          P
estimate   161    33.845   10.349     0.816     -1.42       0.16
```

The observed sample mean cost is less than \$35 ($\overline{C} = \$33.85$), but is only marginally significantly lower, with a tail probability of .16.

(c) 76 of 161 restaurants have food rating at least 20. Let p be the probability that a randomly chosen restaurant achieves at least this rating. We are testing the hypotheses

$$H_0 : p = .5$$

versus

$$H_a : p \neq .5.$$

Here is output for the tests of these hypotheses:

```
EXACT TEST
----------

Test and Confidence Interval for One Proportion

Test of p = 0.5 vs p not = 0.5

                                                           Exact
Sample       X      N  Sample p        95.0 % CI         P-Value
1           76    161  0.472050  (0.392987, 0.552161)    0.529
```

```
NORMAL-BASED TEST
-----------------

Test and Confidence Interval for One Proportion

Test of p = 0.5 vs p not = 0.5

Sample      X      N  Sample p         95.0 % CI       Z-Value  P-Value
1          76    161  0.472050  (0.394937, 0.549162)    -0.71    0.478
```

In this case the two tests agree that the observed proportion of restaurants rated over 20 is not significantly different from .5.

6. Let p be the true proportion of Internet users who have a problem with companies doing this. We are testing the hypotheses

$$H_0 : p = .2$$

versus

$$H_a : p < .2$$

(note the one–sided alternative hypothesis). Here is output for the tests of these hypotheses:

```
EXACT TEST
----------

Test and Confidence Interval for One Proportion

Test of p = 0.2 vs p < 0.2

                                                         Exact
Sample      X      N  Sample p         95.0 % CI       P-Value
1          64    457  0.140044  (0.109549, 0.175286)    0.001

NORMAL-BASED TEST
-----------------

Test and Confidence Interval for One Proportion

Test of p = 0.2 vs p < 0.2

Sample      X      N  Sample p         95.0 % CI       Z-Value  P-Value
1          64    457  0.140044  (0.108227, 0.171861)    -3.20    0.001
```

We clearly reject the null hypothesis here; the observed sample proportion of people who have a problem with companies doing this is significantly smaller than 20%.

7. Let p be the true probability that the higher–seeded team wins the game (you could just as well have defined as the probability that the lower–seeded team wins). We are testing the hypotheses

$$H_0 : p = .5$$

versus

$$H_a : p \neq .5.$$

Here is output for the tests of these hypotheses:

```
EXACT TEST
----------

Test and Confidence Interval for One Proportion

Test of p = 0.5 vs p not = 0.5

                                                            Exact
Sample      X      N  Sample p         95.0 % CI          P-Value
1          63    116  0.543103  (0.448058, 0.635895)       0.403

NORMAL-BASED TEST
-----------------

Test and Confidence Interval for One Proportion

Test of p = 0.5 vs p not = 0.5

Sample      X      N  Sample p         95.0 % CI          Z-Value  P-Value
1          63    116  0.543103  (0.452453, 0.633754)         0.93    0.353
```

Clearly there is no evidence here to suggest that either the higher– or lower–seeded teams have an advantage. Apparently we can view a difference of one level in the seeding as mostly reflecting random noise.

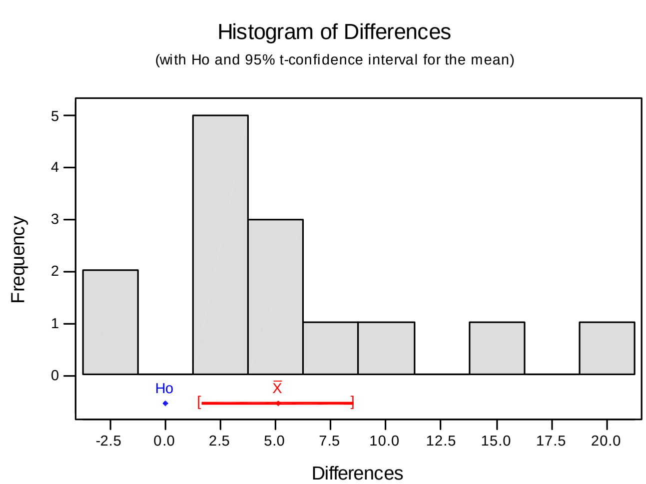
8.(a) Let μ_{SC} and μ_{SP} be the average small cap and S&P returns, respectively. The hypotheses being tested here are

$$H_0 : \mu_{SC} = \mu_{SP}$$

versus

$$H_a : \mu_{SC} \neq \mu_{SP}.$$

The histogram below shows that the differences between small cap and S&P returns are somewhat long right–tailed, so a paired sample t–test might not be reasonable here. Here is the output for the t–test, which indicates that the S&P Composite does significantly better than small caps, on average during market corrections:



```
Paired T-Test and Confidence Interval

Paired T for Correction S&P - Correction small caps

                     N      Mean     StDev   SE Mean
Correcti            14    -16.71     10.60      2.83
Correcti            14    -21.81     13.16      3.52
Difference          14      5.09      5.94      1.59

95% CI for mean difference: (1.66, 8.52)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.21  P-Value = 0.007
```

Since we're not sure about the validity of the t–test here, we could look at nonparametric tests. Let m_{SC} and m_{SP} be the median small cap and S&P returns, respectively. The hypotheses being tested here are now

$$H_0 : m_{SC} = m_{SP}$$

versus

$$H_a : m_{SC} \neq m_{SP}.$$

Here is output for the sign test and signed rank tests, respectively:

```
Sign Test for Median

Sign test of median = 0.00000 versus  not =  0.00000

                N  Below  Equal  Above         P     Median
S&P-smal      14      2      0     12    0.0129      3.750
```

```
Wilcoxon Signed Rank Test

Test of median = 0.000000 versus median not = 0.000000

                 N for   Wilcoxon            Estimated
             N    Test  Statistic         P     Median
S&P-smal    14      14       95.0     0.008      4.050
```

The nonparametric tests agree that the true median S&P return is larger than the true median small cap return during correction periods (the sample values are −13.35% and −18.3%, respectively). Thus, all three tests agree that it's better to be out of the small caps when the market drops.

(b) The histogram below shows that the differences between small cap and mid cap returns are extremely long left–tailed, with one very negative value (during the 10/76–2/78 correction, small caps dropped 25.6%, while mid caps actually went up 4.9%). A paired sample t–test is not be reasonable here. Instead, let m_{MC} be the median mid cap return. The hypotheses being tested here are

$$H_0 : m_{SC} = m_{MC}$$

versus

$$H_a : m_{SC} \neq m_{MC}.$$

Here is output for the sign test and signed rank tests, respectively:

```
Sign Test for Median

Sign test of median = 0.00000 versus  not =  0.00000

               N  Below  Equal  Above         P     Median
Small ca      14     12      0      2    0.0129     -2.350



Wilcoxon Signed Rank Test

Test of median = 0.000000 versus median not = 0.000000

                N for   Wilcoxon           Estimated
             N   Test  Statistic        P     Median
Small ca    14     14       11.0    0.010     -2.600
```

The nonparametric tests agree that the true median mid cap return is larger than the true median small cap return during correction periods (the sample median mid cap return is -15.2%). Thus, the two tests again agree that it's better to be out of the small caps when the market drops.