Chapter 6: Inferences Comparing Two Population Central Values

6.1 a. Reject H_o if $|t| \geq 2.064$

b. Reject H_o if $t \geq 2.624$

c. Reject H_o if $t \leq -1.860$

6.2 $H_o : \mu_1 - \mu_2 \geq 0$ versus $H_a : \mu_1 - \mu_2 < 0$; Reject H_o if $t \leq -1.703$

$t = \frac{71.5-79.8}{8.3186\sqrt{\frac{1}{16}+\frac{1}{13}}} = -2.6722 < -1.703 \Rightarrow$ Reject H_o and conclude there is significant evidence that μ_1 is less than μ_2.

6.3 p-value $= P(t \leq -2.6722) \Rightarrow 0.005 <$ p-value < 0.01

6.4 a. $H_o : \mu_{26} - \mu_5 \geq 0$ versus $H_a : \mu_{26} - \mu_5 < 0$; Reject H_o if $t \leq -1.812$

$t = \frac{165.8-378.5}{19.9\sqrt{\frac{1}{6}+\frac{1}{6}}} = -2.6722 < -1.812 \Rightarrow$ Reject H_o and conclude there is significant evidence that μ_{26} is less than μ_5, with p-value < 0.0005

b. The sample sizes are too small to evaluate the normality condition but the sample variances are fairly close considering the sample sizes. We would need to check with the experimenter to determine if the two random samples were independent.

c. A 95% C.I. on the mean difference is (-238.3, -187.1), which indicates that the average warm temperature rat blood pressure is between 187 and 239 units lower than the average $5°C$ rat blood pressure.

6.5 a. $H_o : \mu_A = \mu_B$ versus $H_a : \mu_A \neq \mu_B$; p-value $= 0.065 \Rightarrow$ The data do not provide sufficient evidence to conclude there is a difference in mean oxygen content.

b. The separate variance t'-test was used since it has df given by

$c = \frac{(.157)^2/15}{(.157)^2/15+(.320)^2/15} = 0.194; df = \frac{(15-1)(15-1)}{(1-.194)^2(15-1)+(.194)^2(15-1)} \approx 20.$

If the pooled t-test was used the df=15+15-2=28 but the separate variance test has $df \approx 20$, as is shown on printout.

c. The Above-town and Below-town data sets appear to be normally distributed, with the exception that the box-plot does display an outlier for the Below-town data. The Below-town data appears to be more variable than the Above-town data. The two data sets consist of independent random samples.

d. The 95% C.I. estimate for the difference in means is (-0.013, 0.378) ppm. The observed difference (0.183 ppm) is not significant.

6.6 a. $H_o : \mu_U \geq \mu_S$ versus $H_a : \mu_U < \mu_S$; p-value $< 0.0005 \Rightarrow$ The data provide sufficient evidence to conclude that successful companies have a lower percentage of returns than unsuccessful companies.

b. $n_1 + n_2 - 2 = 98 = df$ for pooled t-test. The printout shows df=86 which is the df for the separate variance test.

c. The box plots indicate that both data sets appear to be from normally distributed populations; however, the Successful data sets indicates a higher variability than the Unsuccessful.

d. A 95% C.I. on the difference in the mean percentages is (2.70%, 4.66%) $\Rightarrow$ We are 95% confident that successful businesses have roughly 3% to 5% fewer returns.

6.7 We want to test $H_o : \mu_{No} = \mu_{Sub}$ versus $H_a : \mu_{No} \neq \mu_{Sub}$; p-value = 0.0049 $\Rightarrow$ The data supports the contention that people that receive a daily newspaper have a greater knowledge of current events. The stem-and-leaf plots indicate that both data sets are from normally distributed populations but with different variances. The problem description indicates that the two random samples were independently selected. A 95% C.I. on the mean difference is (-15.5, -2.2) which reflects the lower values from the people who did not read a daily newspaper.

6.8 Probably not. If the data was normally distributed, then approximately 16% of the population values should be less than $\bar{y} - s$. However, $s > \bar{y}$ implies that $\bar{y} - s < 0$ and thus we would conclude that approximately 16% of the data is negative. Note that all the data values in this population are positive.

6.9 Since the sample sizes are very small, we would need to be assured that the samples were randomly selected from normally distributed populations.

6.10 a. $H_o : \mu_{96} \geq \mu_{82}$ versus $H_a : \mu_{96} < \mu_{82}$;

$t = \frac{7.00-54.30}{\sqrt{\frac{(3.89)^2}{13}+\frac{(15.7)^2}{13}}} = -10.58 \Rightarrow$ with $df = 13$, p-value $< 0.0005 \Rightarrow$

reject H_o and conclude the data provide sufficient evidence that there has been a significant decrease in mean PCB content.

b. A 95% C.I. on the difference in the mean PCB content of herring gull eggs is (-4.76, -2.24), which would indicate that the decrease in mean PCB content from 1982 to 1996 is between 2.24 and 4.76.

c. The box plots are given here:

The box plots of the PCB data from the two years both appear to support random samples from normal distributions, although the 1982 data is somewhat skewed to the left. The variances for the two years are substantially different hence the separate variance t-test was applied in a.

d. Since the data for 1982 and 1996 were collected at the same sites, there may be correlation between the two years. There may be also be spatial correlation depending on the distance between sites.

6.11 If the data were analyzed using the difference in PCB content (1996 data - 1982 data) at each site, the effect of between site variability could potentially be reduced. The data should be analyzed as a *before and after* study using paired data methodology.

6.12 $H_o : \mu_F = \mu_M$ versus $H_a : \mu_F \neq \mu_M$

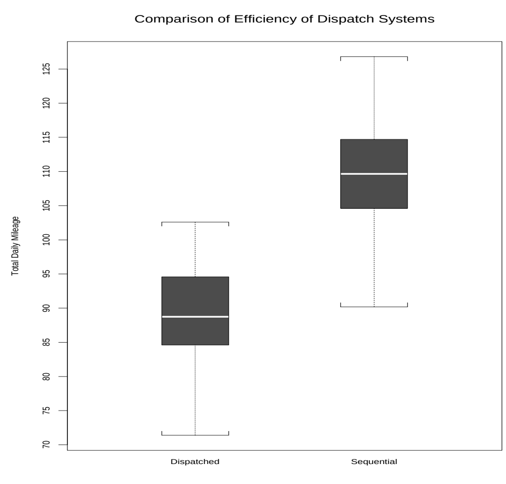
a. Pooled-variance t-test = -4.04, df = 58, p-value = 0.0001

b. Separate-variance t'-test = -3.90, df = 43, p-value = 0.0002

c. Since the p-value for both t-tests is very small, we would reject H_o using either of the two test statistics and conclude there is a significant difference in the average bonus percentage between males and females.

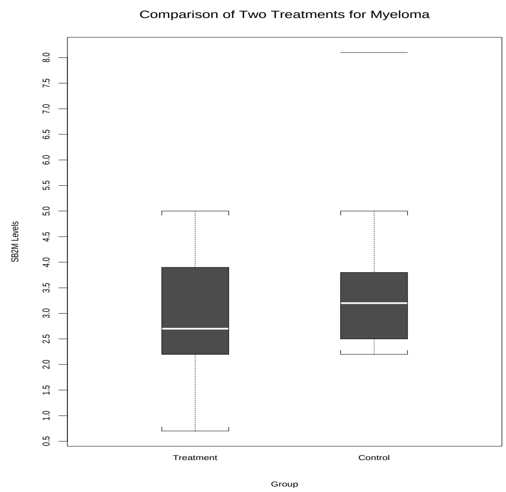
6.13 a. Plumber 1: $\bar{y}_1 = 88.81, s_1 = 7.89$
Plumber 2: $\bar{y}_2 = 108.93, s_2 = 8.73$

b. The box plots are given here:



Since both graphs show a roughly symmetrical distribution with no outliers, a t-test appears to be appropriate.

6.14 a. Box plots are given here:



Based on the plots the treatment data appears to be from a normally distributed population but the control data is right skewed with one large outlier.

b. The Wilcoxon rank sum statistic because of the small sample size and possible lack of normality.

c. H_o : The distributions are the same versus H_a : The distributions are different.
Using the Wilcoxon rank sum statistic with $\alpha = 0.05$,
reject H_o if $T < 79$ or $T > 131$, where T is the sum of the ranks for the treatment group.
$T = 93 \Rightarrow$ Fail to reject H_o. There is not significant evidence of a difference between the treatment and control groups.

d. The addition of alpha interferon (sumiferon) to the treatment regimen did not significantly change the effect on patient outcome.

6.15 a. The data from the treatment group appears to be a random sample from a normal distribution but the data from the control group is definitely not normally distributed. The variance from the control group is 13.4 times larger than the variance from the treatment group. The Wilcoxon rank sum test is probably the most appropriate test statistic since it is more robust to deviations from stated conditions than is the separate variance t test.

b. The p-value from the separate variance t test is 0.053 and the p-value from the Wilcoxon test is 0.0438. With $\alpha = 0.05$, the conclusions from the two tests differ with the t test failing to find that the mean daily mileage for the treatment group significantly smaller than the mean for the control group. The Wilcoxon did find a significant reduction in the mean of the treatment group.

c. The t test is probably not appropriate since the control group appears to have a nonnormal distribution.

d. The choice of test statistic would definitely have an effect on the final conclusions of the study. When the test statistics yield contradictory results, and the required conditions of the test statistics are not met, it is best to error on the side of the most conservative conclusion relative to which of the two Types of errors has the greatest consequence. Hence, reject H_o only if one of the tests is highly significant relative to specified α.

6.16 The researchers should pair the data based on soil pH. This would reduce or eliminate the effect of soil pH on corrosion, which would otherwise be confounded with the effectiveness of the coatings.

6.17 a. To conduct the study using independent samples, the 30 participants should be very similar relative to age, body fat percentage, diet, and general health prior to the beginning of the study. The 30 participants would then be randomly assigned to the two treatments.

b. The participants should be matched to the greatest extend possible based on age, body fat, diet, and general health before the treatment is applied. Once the 15 pairs are configured, the two treatments are randomly assigned within each pair of participants.

c. If there is a large difference in the participants with respect to age, body fat, diet and general health and if the pairing results in a strong positive correlation in the responses from paired participants, then the paired procedure would be more effective. If the participants are quite similar in the desired characteristics prior to the beginning of the study, then the independent samples procedure would yield a test statistic having twice as many df as the paired procedure and hence would be more powerful.

6.18 a. The paired t-test yields $t = \frac{2.58}{9.49/\sqrt{10}} = 0.86, df = 9, \Rightarrow$

p-value $= P(t \geq 0.86) \Rightarrow 0.10 <$ p-value < 0.25

There is not significant evidence that the mean SENS value decreased.

b. The 95% C.I. estimate of the change in the mean SENS value is (-4.21, 9.37).

c. The box plot indicates that results from patient number 9 are an outlier relative to the other patients. The remaining values appear to have a normal distribution but the results from patient number 9 should be carefully checked. The researchers should be interviewed to confirm that the results from the 9 patients are truely independent, i.e., the differences form a random sample from a normal distribution.

6.19 a. $H_o : \mu_d = 0$ versus $H_a : \mu_d \neq 0$.

$t = 4.95, df = 29, \Rightarrow$ p-value $= 2P(t \geq 4.91) < 0.001$

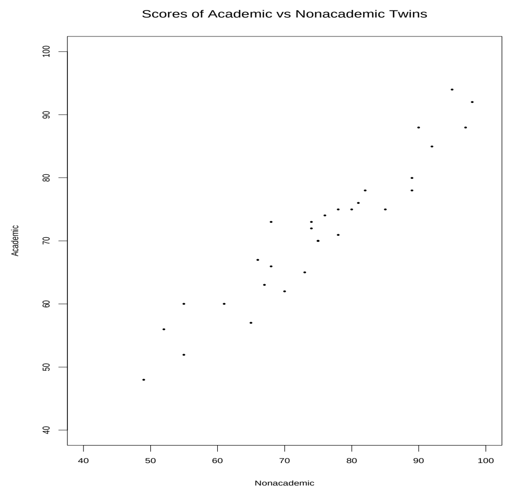
There is significant evidence of a difference in the mean final grades.

b. A 95% confidence interval estimate of the mean difference in mean final grades is (2.23, 5.37).

c. We would need to verify that the difference in the grades between the 30 twins are independent. The normal probability plot would indicate that the differences are a random sample from a normal distribution. Thus, the conditions for using a paired t-test appear to be valid.

d. Yes. The purpose of pairing is to reduce the subject to subject variability and there appears to be considerable differences in the students in the study. Also, a scatterplot of the data yields a strong positive correlation between the scores for the twins.

Scatterplot is given here:



6.20 a. Wilcoxon signed rank test with $\alpha = 0.05$:

Reject H_o if $T_- \leq 10$, since $T = T_- = 16 > 10 \Rightarrow$

Fail to reject H_o, with p-value$=P(T_- \leq 16) > 0.1$

b. The paired t-test in Exercise 6.18 also failed to reject H_o with a p-value > 0.10. If the differences have a normal distribution, then the paired t-test has greater power than the signed rank test.

6.21 H_o : The distribution of differences (Benzedrine minus placebo) is symmetric about 0 versus

H_a : The differences tend to be larger than 0

With $n = 14, \alpha = 0.05, T = T_-$, reject H_o if $T_- \leq 25$.

From the data we obtain $T_- = 16 < 25$, thus reject H_o and conclude that the distribution of heart rates for dogs receiving Benzedrine is shifted to the right of the dogs receiving the placebo.

6.22 a. The box plot and normal probability plots both indicate that the distribution of the data is somewhat skewed to the left. Hence, the Wilcoxon would be more appropriate, although the paired t-test would not be inappropriate since the differences are nearly normal in distribution.

b. H_o : The distribution of differences (female minus male) is symmetric about 0

H_a : The differences (female minus male) tend to be larger than 0

With $n = 20, \alpha = 0.05, T = T_-$, reject H_o if $T_- \leq 60$.

From the data we obtain $T_- = 18 < 60$, thus reject H_o and conclude that repair costs are generally higher for female customers.

6.23 Let d=After-Before

a. $H_o : \mu_d \leq 0$ versus $H_a : \mu_d > 0$;
$t = \frac{1.208-0}{1.077/\sqrt{12}} = 3.89$, with $df = 11 \Rightarrow .001 <$ p-value $< 0.005 \Rightarrow$
Reject H_o and conclude that the exposure has increased mean lung capacity.

b. 95% C.I. on $\mu_{After} - \mu_{Before}$: (0.52, 1.89)

c. If this was a well controlled experiment with all factors except ozone exposure controlled, the experimenter would be justified in making the claim concerning the population of rats from which the rats in the study were randomly selected. However, no statement can be made about the effects of ozone on humans.

6.24 a. $H_o : \mu_{Narrow} = \mu_{Wide}$ versus $H_a : \mu_{Narrow} \neq \mu_{Wide}$;
The separate variance t-test: $t' = \frac{118.37-110.20}{\sqrt{\frac{(7.87)^2}{12}+\frac{(4.71)^2}{15}}} = 3.17 \Rightarrow$
with $df \approx 17$, $0.002 <$ p-value $< 0.010 \Rightarrow$
reject H_o and conclude there is sufficient evidence in the data that the two types of jets have different average noise levels.

b. A 95% C.I. on $\mu_{Wide} - \mu_{Narrow}$ is (2.73, 13.60)

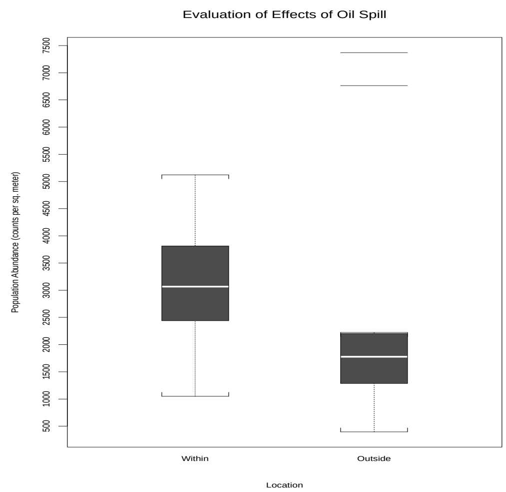
c. Because maintenance could affect noise levels, jets of both types from several different airlines and manufacturers should be selected. They should be of approximately the same age, etc. This study could possibly be improved by pairing Narrow and Wide body airplances based on factors that may affect noise level.

6.25 a. $H_o : \mu_{Within} = \mu_{Out}$ versus $H_a : \mu_{Within} \neq \mu_{Out}$;
The separate variance t-test: $t' = \frac{3092-2450}{\sqrt{\frac{(1191)^2}{14}+\frac{(2229)^2}{12}}} = 0.89 \Rightarrow$
with $df \approx 16$, p-value $\approx 0.384 \Rightarrow$
Fail to reject H_o and conclude the data does not provide sufficient evidence that there is a difference in average population abundance.

b. A 95% C.I. on $\mu_{Within} - \mu_{Out}$ is (-879, 2164)

c. The two samples are independently selected random samples from two normally distributed populations.

d. Box plots are given here:



The Within data set appears to be normally distributed but the Outside data may not be normally distributed since there were two outliers. The sample variance of the Outside is 3.5 times larger than the Within sample variance.

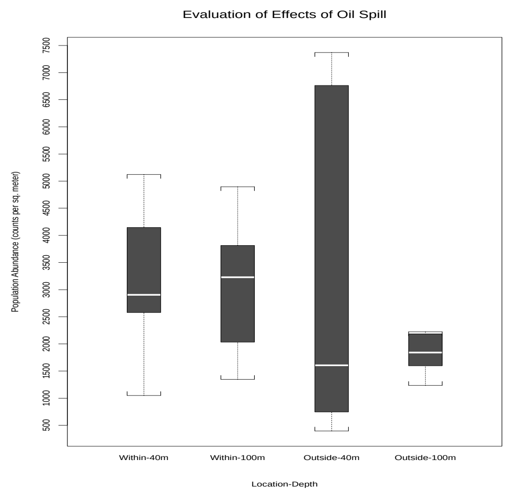
6.26 a. $H_o : \mu_{Within} = \mu_{Out}$ versus $H_a : \mu_{Within} \neq \mu_{Out}$;
Since both n_1, n_2 are greater than 10, the normal approximation can be used.
$T = 122, \mu_T = (12)(12 + 14 + 1)/2 = 162, \sigma = \sqrt{(12)(14)(12 + 14 + 1)/12} = 19.44$
$z = \frac{122-162}{19.44} = 2.06 \Rightarrow$ p-value $= 0.0394 \Rightarrow$
reject H_o and conclude the data provides sufficient evidence that there is a difference in average population abundance.

b. The Wilcoxon rank sum test requires independently selected random samples from two populations which have the same shape but may be shifted from one another.

c. The two population distributions may have different variances but the Wilcoxon rank sum test is very robust to departures from the required conditions.

d. The separate variance test failed to reject H_o with a p-value of 0.384. The Wilcoxon test rejected H_o with a p-value of 0.0394. The difference in the two procedures is probably due to the skewness observed in the Outside data set. This can result in inflated p-values for the t test which relies on a normal distribution when the sample sizes are small.

6.27 a. Box plot is given here:



b. 40M: $H_o : \mu_{Within} = \mu_{Out}$ versus $H_a : \mu_{Within} \neq \mu_{Out}$;
The separate variance t-test: $t' = \frac{3183-3080}{\sqrt{\frac{(1289)^2}{7}+\frac{(3135)^2}{6}}} = 0.08 \Rightarrow$
with $df \approx 6$, p-value $\approx 0.94 \Rightarrow$
fail to reject H_o and conclude the data does not provide sufficient evidence that there is a difference in average population abundance at 40 m.
100M: $H_o : \mu_{Within} = \mu_{Out}$ versus $H_a : \mu_{Within} \neq \mu_{Out}$;
The separate variance t-test: $t' = \frac{3002-1820}{\sqrt{\frac{(1181)^2}{7}+\frac{(385)^2}{6}}} = 2.50 \Rightarrow$
with $df \approx 7$, p-value $\approx 0.041 \Rightarrow$
reject H_o and conclude the data provides sufficient evidence that there is a difference in average population abundance at 100 m.

c. The conclusions are different at the two depths. The mean population abundances are fairly consistent at all but the 100 m depth outside the oil trajectory where the abundance is considerably smaller. However, the median at both within depths are nearly the same but at a higher level than the medians at two outside depths.

6.28 a. Such a statement cannot be made. In order to study the effect of the oil spill on the population, we would need some baseline data at these sites before the oil spill.

b. If baseline data were available, a paired t-test would be appropriate. The impact of the oil spill could be studies by a *before and after* analysis. If the pairing were effective, the paired data study would be more efficient than a two independent samples study, i.e., it would take fewer observations (sites) to achieve the same level of precision.

c. Spatial correlation might be a problem. Factors such as weather, food supply, etc., which are outside the control of the researcher might mask or exagerate the effect of the oil spill. These types of factors also prevent the researcher from making definitive *cause and effect* statements.

6.29 a. $H_o : \mu_{High} = \mu_{Con}$ versus $H_a : \mu_{High} \neq \mu_{Con}$;

Separate variance t-test: $t^{'} = 4.12$ with $df \approx 34$, p-value $= 0.0002$. $\Rightarrow$

Reject H_o and conclude there is significant evidence of a difference in the mean drop in blood pressure between the high-dose and control groups.

b. 95% C.I. on $\mu_{High} - \mu_{Con}$: (19.5, 57.6), i.e., the high dose groups mean drop in blood pressure was, with 95% confidence, 19.5 to 57.6 points greater than the mean drop observed in the control group.

c. Provided the researcher independently selected the two random samples of participants, the conditions for using a separate variance t-test were satisfied since the plots do not detect a departure from a normally distribution but the sample variances are somewhat different(1.9 to 1 ratio).

6.30 a. $H_o : \mu_{Low} = \mu_{Con}$ versus $H_a : \mu_{Low} \neq \mu_{Con}$;

Separate variance t-test: $t^{'} = -2.09$ with $df \approx 35$, p-value $= 0.044$. $\Rightarrow$

Reject H_o and conclude there is significant evidence of a difference in the mean drop in blood pressure between the low-dose and control groups.

b. 95% C.I. on $\mu_{Low} - \mu_{Con}$: (-51.3, -0.8), i.e., the low-dose groups mean drop in blood pressure was, with 95% confidence, 51.3 to 0.8 points less than the mean drop observed in the control group.

c. Provided the researcher independently selected the two random samples of participants, the conditions for using a pooled t-test were satisfied since the plots do not detect a departure from a normally distribution but the sample variances are somewhat different(1.7 to 1 ratio).

6.31 a. $H_o : \mu_{Low} = \mu_{High}$ versus $H_a : \mu_{Low} \neq \mu_{High}$;

Separate variance t-test: $t = -5.73$ with $df \approx 29$, p-value < 0.0005. $\Rightarrow$

Reject H_o and conclude there is significant evidence of a difference in the mean drop in blood pressure between the high-dose and low-dose groups.

b. 95% C.I. on $\mu_{Low} - \mu_{High}$: (-87.6, -41.5), i.e., the low-dose group mean drop in blood pressure was, with 95% confidence, 41.5 to 87.6 points lower than the mean drop observed in the high-dose group.

c. Provided the researcher independently selected the two random samples of participants, the conditions for using a pooled t-test were satisfied since the plots do not detect a departure from a normally distribution and the sample variances are somewhat different(3.2 to 1 ratio).

6.32 a. Let A_i be the event that a Type I error was made on the ith test, $i = 1, 2, 3$.

P(at least one Type I error in 3 tests) $= P(A_1 \cup A_2 \cup A_3)$

$= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$

$\leq P(A_1) + P(A_2) + P(A_3) = 0.05 + 0.05 + 0.05 = 3(0.05) = 0.15$

b. Set the value of α at $\frac{0.05}{3}$ for each of the 3 tests.

Then P(at least one Type I error in 3 tests) $\leq 3(\frac{0.05}{3}) = 0.05$.

Thus, we know that the chance of at least one Type I error in the 3 tests is at most 0.05.

6.33 a. $H_o : \mu_D \leq \mu_{RN}$ versus $H_a : \mu_D > \mu_{RN}$;

Separate variance t-test: $t = 9.04$ with $df \approx 76$, p-value $< 0.0001 \Rightarrow$

Reject H_o and conclude there is significant evidence that the mean score of the Degreed nurses is higher than the mean scores of the RN nurses.

b. p-value < 0.0001

c. 95% C.I. on $\mu_D - \mu_{RN}$: (35.3, 55.2) points

d. Since 40 is contained in the 95% C.I., there is a possibility that the difference in mean scores is less than 40. Hence, the differences observed may not be meaningful.

6.34 a. $H_o : \mu_F \geq \mu_M$ versus $H_a : \mu_F < \mu_M$;

b. 95% C.I. on $\mu_F - \mu_M$: (-142.30, -69.1) thousands of dollars

c. Since $s_1 \approx s_2$, use pooled t-test: $t = \frac{245.3-350.1}{57.2\sqrt{\frac{1}{20}+\frac{1}{20}}} = -5.85$ with $df = 38$

p-value $< 0.0001 \Rightarrow$

Reject H_o and conclude there is significant evidence that the mean campaign expenditures for females is less than mean candidates expenditures for males.

d. Yes, since the difference could be as much as \$142,300.

6.35 The required conditions are that the two samples are independently selected from populations having normal distributions with equal variances. The box plots do not reveal any indication that the population distributions were not normal. The sample variances have a ratio of 1.4 to 1.0, thus there is very little indication that the population variances were unequal.

6.36 a. The average potency after 1 year is different than the average potency right after production.

b. The two test statistics are equal since the sample sizes are equal.

c. The p-values are different since the test statistics have different degrees of freedom (df).

d. In this particular experiment, the test statistics reach the same conclusion, reject H_o.

e. Since $s_1 \approx s_2$ which yields a test of equal variances with p-value of 0.3917, the pooled t-test would be the more appropriate test statistic.

6.37 $H_o : \mu_{SUV} \leq \mu_{Mid}$ versus $H_a : \mu_{SUV} > \mu_{Mid}$

a. The box plots do not look particularly like samples from normal distributions, both indicate that the distributions are skewed to the right. The ratio of sample variances is 2.5 to 1, which is within acceptable limits for using the pooled t-test (provided both distributions are normally distributed). The lack of normality with such small sample sizes would invalidate the use of the t-test.

b. With t-test, we obtain p-value $= 0.11$. This would indicate that there is not significant evidence that mean damage is greater for the SUV's.

With the Wilcoxon rank sum test, we obtain p-value $= 0.0203$. This would indicate that there is significant evidence that mean damage is greater for the SUV's.

c. The Wilcoxon rank sum test is more appropriate because the populations appear to be nonnormally distributed. This would make the results of the t-test questionable, especially with the sample sizes being quite small.

d. The Wilcoxon and t-test tests yield contradictory results. The Wilcoxon rank sum test is insensitive to extreme values whereas the t-test can be invalidated when the distributions are highly skewed.

6.38 a. $n = \frac{2(700)^2(1.96)^2}{(500)^2} = 15.1 \Rightarrow n = 16$. The new study requires 16 vehicles of each type.

b. $n = \frac{2(700)^2(1.645+1.645)^2}{(500)^2} = 42.4 \Rightarrow n = 43$. The new study requires 43 vehicles of each type.

6.39 The concept of statistical significance involves inherent uncertainty since it is based on a probability. Because evidence in legal cases is admissible only if it can be asserted with near certainty, any statistical evidence will only be useful if its associated probability approaches 1. Common practice uses 95% significance as meaning near certainty and the legal profession accepts this figure as a scientific consensus.

6.40 a. The summary statistics for Exposed, Control, and Diff are given here.

```
Descriptive Statistics: Exposed, Control, Difference


Variable            N       Mean   Median   TrMean    StDev
Exposed            33      31.85    34.00    30.79    14.41
Control            33      15.88    16.00    15.83     4.54
Difference         33      15.97    15.00    15.00    15.86

Variable      SE Mean  Minimum    Maximum       Q1       Q3
Exposed          2.51    10.00      73.00    20.50    40.00
Control          0.79     7.00      25.00    12.50    19.00
Difference       2.76    -9.00      60.00     3.00    25.00
```

1. The median of the Exposed children is 34.00 whereas the maximum of the Control children is only 25.00. Thus, there appears to be a much higher level of lead in the Exposed children in comparison to the Control children.
2. The difference in the means for the Exposed and Control children is 31.85-15.88=15.97. Therefore, on the average the Exposed children have 15.97 units of lead more than the Control children.

b. The difference in the medians for the Exposed and Control children is 34.00-16.00=18.00, whereas, the median of the differences between paired Exposed and Control children is 15.00. The two differences are not the same because the median only considers the middle value of the data set and hence produces a different value for the difference of the medians in comparison to the median of the differences. The median of the differences is a more accurate view of the increase in lead exposure because many of the other factors that may have an effect on the amount of lead in the children have been controlled for by pairing the children and then taking the difference in their lead levels.

6.41 a. 1. The alternative hypothesis is that the mean lead level is greater in the Exposed children than in the Control children. Therefore, the null hypothesis is that the mean lead level in the Exposed children is less than or equal to the mean lead level in the Control children. In symbols, we have

$$H_o : \mu_D \leq 0 \text{ versus } H_a : \mu_D > 0,$$

where D =Exposed - Control. The p-value for the 1-sided test is given by p-value $= Pr[t_{32} \geq 5.78] = .00000103$. Thus, there is significant evidence that the mean difference in the lead levels between children in the Exposed and Control groups is greater than 0. The p-value for the 1-sided test is 1/2 of the p-value for the 2-sided test.

2. If the lead-using factory had implemented a safety program to warn its employees about the possible sourses of lead outside the factory, then the employees could educate their children how to avoid lead exposure and hence have lower levels than children of non-employees.

b. A Wilcoxon signed rank test and the sign test yield a p-value less than .000 and hence agree with the t-test.

```
Wilcoxon Signed Rank Test: Exposed-Control

Test of median = 0.000000 versus median  >  0.000000

                N for   Wilcoxon           Estimated
             N   Test  Statistic        P     Median
Exposed     33     33      561.0    0.000      30.50
Control     33                                 16.00

Sign Test for Median: Exposed-Control

Sign test of median = 0.00000 versus  >  0.00000

               N  Below  Equal  Above         P     Median
Exposed       33      0      0     33    0.0000      34.00
Control       33                                     16.00
```