Question 1):

I) A t-test was run to test the ***null-hypothesis (H_0)*** that: the **mean weight** of the bumblebee bat population being examined is 1.8g.

$(\mu_{Bumble-bat} = 1.8)$.

The ***alternative-hypothesis (H_A)*** would be that: **the mean weight** of the bumblebee bat population being examined is not equal to 1.8g.

$(\mu_{Bumble-bat} \neq 1.8)$.

II) For a sample size of *n* = 15, the **critical value** at the $\propto = .05$ level with *df* = 14 would be: $\mathbf{t_{crit}} = \pm \mathbf{2.145}$.

III) For this given sample, the **test statistic** would be: $\mathbf{t_{stat}} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \approx \frac{1.71\bar{3} - 1.8}{\frac{.2588}{\sqrt{15}}} \approx \mathbf{-1.297}$,

which is rather far from the critical value.

IV) Assuming the null-hypothesis, statistical software determined that the **p-value** for such a sample would be: $\mathbf{p_{val} = .2155}$, which is much larger than the $\propto = .05$ threshold.

The **95% confidence interval** for the mean would be: $1.5700 \leq \mu \leq 1.8566$, or ≈ **[1.57, 1.86]**

V) On this basis, one would **fail to reject** (FTR) **the null-hypothesis** (H_0), determining that:

VI) There is **not statistically significant evidence** to suggest that the mean weight of the bumblebee bat population being examined is not equal to 1.8g $(\mu_{Bumble-bat} \neq 1.8)$ → [$\mathbf{p_{val} = .2155}$].

Question 2):

Part A)

To test for age discrimination, one-thousand permutations were run to test the ***null-hypothesis (H_0)*** that: there is **no difference in the mean** age of workers who were fired vs. those not fired:

$(\mu_F - \mu_{NF} = 0)$.

The ***alternative-hypothesis (H_A)*** would be that: there is, in fact, a statistically significant **difference in the mean**: $(\mu_F - \mu_{NF} \neq 0)$.

According to Figure 1, two-hundred-ninety-six (296) of one-thousand (1000) values were more extreme than the observed difference in sample means (1.9238 → see Figure 2), creating a permutation test p-value of **p_{val} = .2960** , which is greater than the $\propto = .05$ threshold for statistical significance.

On this basis, one would **fail to reject** (FTR) **the null-hypothesis** (H_0), determining that there is **not statistically significant evidence** to suggest that: $(\mu_F - \mu_{NF} \neq 0)$ → **[p_{val} = .296].**

Given the evidence, inferences regarding this difference **cannot** be drawn to the population being examined, and as assignment to groups was not randomized, inferences regarding causal relations also **cannot** be drawn.

Figure 1:

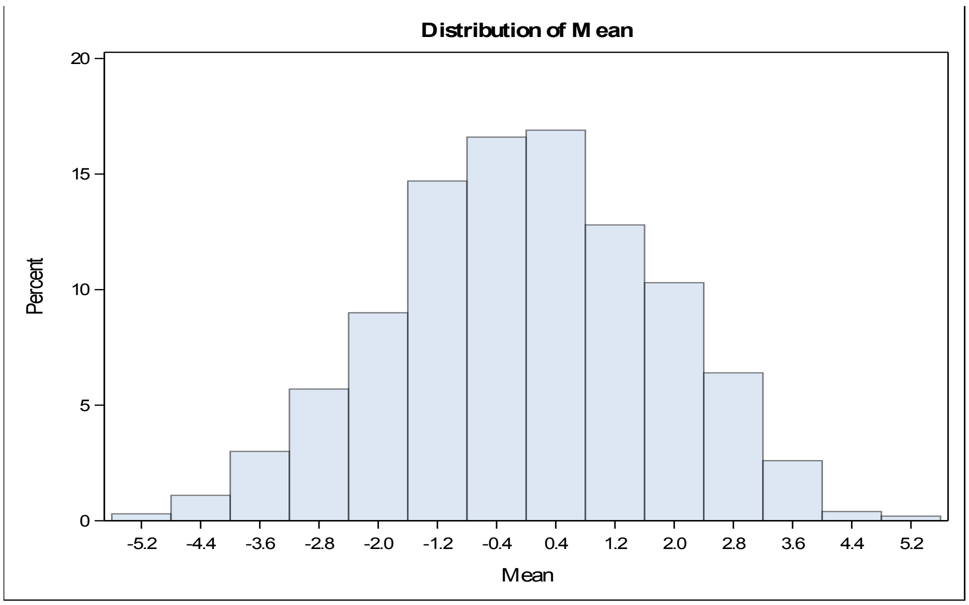


Figure 2:

Obs	Variable	Class	Method	Variances	Mean	LowerCLMean	UpperCLMean	StdDev
1	COL3	Diff (1-2)	Pooled	Equal	-2.5286	-6.0143	0.9572	6.0965
2	COL4	Diff (1-2)	Pooled	Equal	2.4905	-0.9975	5.9785	6.1004
3	COL6	Diff (1-2)	Pooled	Equal	2.8143	-0.6534	6.2819	6.0648

…

294	COL995	Diff (1-2)	Pooled	Equal	-2.4476	-5.9381	1.0429	6.1048
295	COL996	Diff (1-2)	Pooled	Equal	-4.2286	-7.5758	-0.8814	5.8541
296	COL998	Diff (1-2)	Pooled	Equal	3.4619	0.0429	6.8809	5.9797

Part B)

I) To test for age discrimination against the elderly, a one-sided, two sample t-test was run to test the ***null-hypothesis (H_0)*** that: there is **no difference in the mean** age of workers who were fired vs. those not fired:

$$(\mu_F - \mu_{NF} = 0).$$

This time, the ***alternative-hypothesis (H_A)*** would be that: the **mean age of those fired is greater** than that of those who were not: $(\mu_F - \mu_{NF} > 0)$.

II) For a sample size of *n* = 51, the **critical value** at the $\propto = .05$ level with *df* = 49 would be: $\mathbf{t_{crit}} = \mathbf{1.678}$

III) For this given sample, the **test statistic** would be: $\mathbf{t_{stat}} = \frac{\bar{x}_F - \bar{x}_{NF}}{s_p(\sqrt{\frac{1}{n_F} + \frac{1}{n_{NF}}})} \approx \frac{\mathbf{45.8571 - 43.9333}}{\mathbf{6.1519(.28452)}} \approx$

$\mathbf{1.0991}$, which is somewhat far from the critical value.

IV) Assuming the null-hypothesis and unequal variance **(Satterthwaite)** for this one-sided test, statistical software determined that the **p-value** for such a sample would be **roughly half of the double-sided** p-value of .2870, or: $\mathbf{p_{val} = .1435}$, which is larger than the $\propto = .05$ threshold.

V) On this basis, one would **fail to reject** (FTR) **the null-hypothesis** (H_0), determining that:

VI) There is **not statistically significant evidence** to suggest that the mean age of those fired is greater than that of those who were not: $(\mu_F - \mu_{NF} > 0)$. →[$\mathbf{p_{val} = .1435}$].

As with the permutation test, given the evidence, inferences regarding this difference **cannot** be drawn to the population being examined, and as assignment to groups was not randomized, inferences regarding causal relations also **cannot** be drawn.

Part C)

[**NB**: An error in this question's phrasing was identified by Wen Tao in office hours (1/21/17)!]

Assuming that one is being asked to compare the **p-values** of the one-sided, two sample t-test (Part B) and the above two-sided permutation test (Part A), the former value (**.296**) is slightly larger than twice the latter value (2 * **.1435**), which is not unreasonable, factoring in the randomization mechanisms of the permutation test. Presumably, these mechanisms would produce the relatively more reliable statistical figure of the two.

Part D)

According to the results of the one-sided t-test used in Part B), the **95% confidence interval** for the difference in means, assuming equal variance **(Pooled)**, would be:

$-1.0107 \leq \mu_F - \mu_{NF} \leq \infty$, or: **[-1.01**, ∞**]**.

According to the results of the two-sided t-test used in Part A), the **95% confidence interval** for the difference in means, assuming equal variance **(Pooled)**, would be:

$-1.5936 \leq \mu_F - \mu_{NF} \leq 5.4413$, or: **[-1.594**, **5.441]**.

Noting that the interval contains zero, this is consistent with the above data indicating there is **no statistically significant evidence of a greater mean age** for those who were fired.

Part E)

i) Given the sample standard deviations, the pooled standard deviation would be:

$$\boldsymbol{s_p} = \sqrt{\frac{(n_F-1)(s_F{}^2)+(n_{NF}-1)(s_{NF}{}^2)}{(n_F+n_{NF})-2}} \approx \sqrt{\frac{(20)(6.5214^2)+(29)(5.8835^2)}{(21+30)-2}} \approx \mathbf{6.1519}$$

ii) The standard error of the difference in sample means would be:

E = $\boldsymbol{t_{crit}} * \boldsymbol{s_p}(\sqrt{\frac{\mathbf{1}}{\boldsymbol{n_F}}+\frac{\mathbf{1}}{\boldsymbol{n_{NF}}}}) \approx (1.678)(6.1519)(.28452) \approx \mathbf{2.9371}$,

which is roughly consistent with the above one-sided confidence interval:

$(\bar{x}_F - \bar{x}_{NF)-E} < \mu_F - \mu_{NF} < (\bar{x}_F - \bar{x}_{NF)+E}$ → or,

(1.9238 – 2.9371) < $\mu_F - \mu_{NF}$ < (1.9238 + 2.9371) → or,

(-1.0133) < $\mu_F - \mu_{NF}$ < (4.8609)

Question 3)

Part A)

I) A two-sided, two sample t-test was run to test the ***null-hypothesis (H_0)*** that: there is **no difference in the mean** amount of pocket cash carried by students at SMU and Seattle U: $(\mu_{SMU} - \mu_{Seattle} = 0)$.

The ***alternative-hypothesis (H_A)*** would be that: there is, in fact, a statistically significant **difference in the mean**: $(\mu_{SMU} - \mu_{Seattle} \neq 0)$.

II) For a sample size of *n* = 30, the **critical value** at the $\propto = .05$ level with *df* = 28 would be: $\mathbf{t_{crit}} = \mathbf{2.048}$

III) For this given sample, the **test statistic** would be:

$\mathbf{t_{stat}} = \frac{\bar{x}_{SMU} - \bar{x}_{Seattle}}{s_p\left(\sqrt{\frac{1}{n_{SMU}} + \frac{1}{n_{Seattle}}}\right)} \approx \frac{141.6 - 27.0}{224.1(.36596)} \approx \mathbf{1.397}$, which is rather far from the critical value.

IV) Assuming the null-hypothesis and unequal variance **(Satterthwaite)** for this two-sided test, statistical software determined that the **p-value** for such a sample would be $\mathbf{p_{val} = .1551}$, which is substantially larger than the $\propto = .05$ threshold.

Presuming unequal variance **(Satterthwaite)**, according to statistical software, the **95% confidence interval** for the differences in means would be:
$-48.395 \leq \mu_{SMU} - \mu_{Seattle} \leq 277.6$, or ≈ **[-48.4**, **277.6]**

V) On this basis, one would **fail to reject** (FTR) **the null-hypothesis** (H_0), determining that:

VI) There is **not statistically significant evidence** to suggest that the mean amount of pocket cash carried by students at SMU and Seattle U is not equal $(\mu_{SMU} - \mu_{Seattle} \neq 0)$ →
[$\mathbf{p_{val} = .1551}$].

Given the evidence, inferences regarding this difference **cannot** be drawn to the population being examined, and as assignment to groups was not randomized, inferences regarding causal relations also **cannot** be drawn.

Part B)

A comparison of the **p-value** of this two-sided, two sample t-test (**.1551**) with that of the similar permutation test run for last week's homework (**.164**) is consistent with the kind of random variation one encounters running permutation tests, whose values fluctuate with each trial. The fact that the permutation test p-value is between that of the values for the **Pooled (.1732)** and **Satterthwaite (.1551)** t-test suggests an unequal variance between groups to which the randomizing mechanisms of the permutation test is more sensitive.

Question 4)

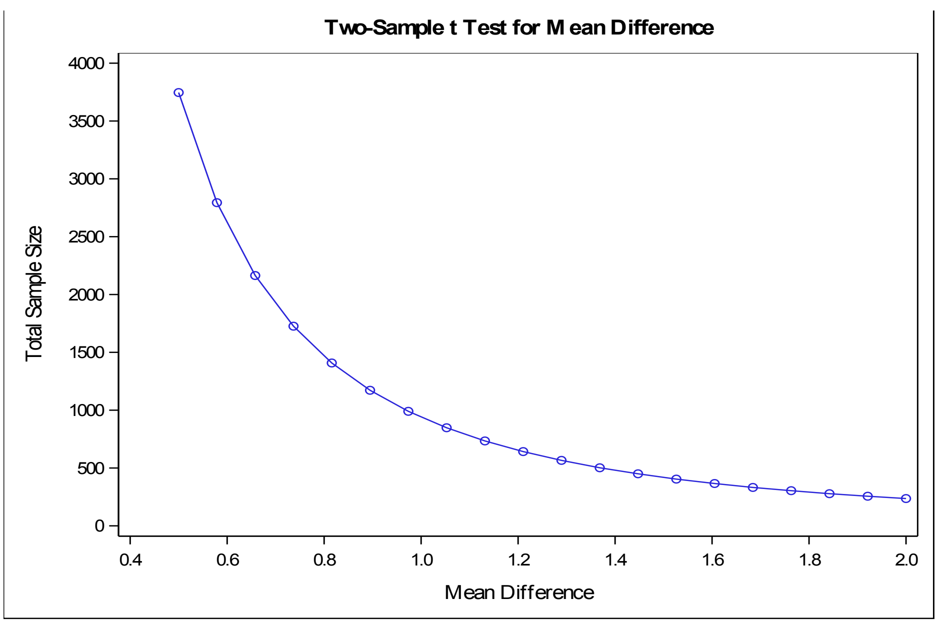
Part A)

Given the Pooled standard deviation from Question 2 **(6.1519)**, a power curve was built to detect effect sizes between (0.5 and 2), calculating the sample size required to have a test with power of (.8). The curve below (Figure 4) was generated with the following SAS code (Figure 3), and produced an estimated **required sample size** of **(n = 256)**:

Figure 3:

```
proc power;
twosamplemeans test = diff
meandiff = 1.9238
stddev = 6.1519
power = .8
alpha = .05
sides = 1
ntotal = . ;
plot x = effect min = .5 max = 2;
run;
```

Figure 4:



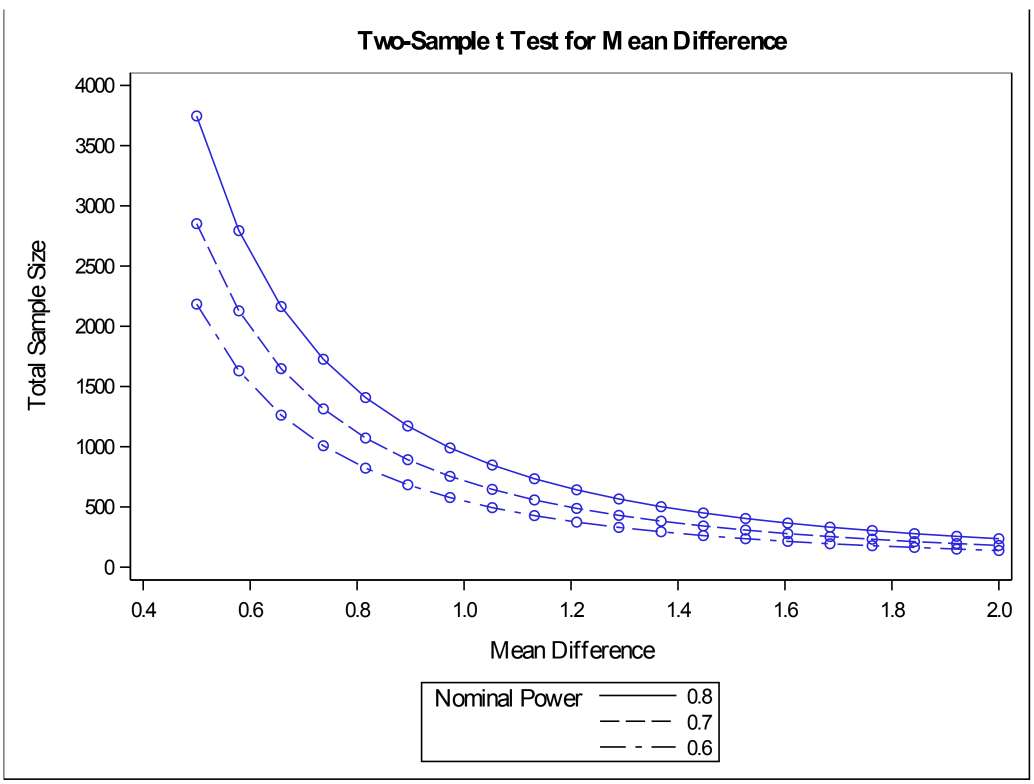
Part B)

The following SAS Code (Figure 5) and power plot (Figure 6) calculates curves of sample size (Y-axis) vs. effect size (X-axis) for a power of (.8), (.7), and (.6):

Figure 5:

```
proc power;
twosamplemeans test = diff
meandiff = 1.9238
stddev = 6.1519
power = .8 .7 .6
alpha = .05
sides = 1
ntotal = . ;
plot x = effect min = .5 max = 2;
run;
```

Figure 6:



Part C)

Given the above plot of differences in required sample size, if one wants to detect an effect size of (.8), the estimated savings in sample size between tests with powers of 80% and 60% would be roughly: (1400 samples at .8 power) – (800 samples at .6 power), or: ≈ **600 samples.**